《经济科学译丛》

编辑委员会

《经济科学译丛》总序

中国是一个文明古国，有着几千年的辉煌历史。近百年来，中国由盛而衰，一度成为世界上最贫穷、落后的国家之一。1949年中国共产党领导的革命，把中国从饥饿、贫困、被欺侮、被奴役的境地中解放出来。1978年以来的改革开放，使中国真正走上了通向繁荣富强的道路。

中国改革开放的目标是建立一个有效的社会主义市场经济体制，加速发展经济，提高人民生活水平。但是，要完成这一历史使命决非易事，我们不仅需要从自己的实践中总结教训，也要从别人的实践中获取经验，还要用理论来指导我们的改革。市场经济虽然对我们这个共和国来说是全新的，但市场经济的运行在发达国家已有几百年的历史，市场经济的理论亦在不断发展完善，并形成了一个现代经济学理论体系。虽然许多经济学名著出自西方学者之手，研究的是西方国家的经济问题，但他们归纳出来的许多经济学理论反映的是人类社会的普遍行为，这些理论是全人类的共同财富。要想迅速稳定地改革和发展我国的经济，我们必须学习和借鉴世界各国包括西方国家在内的先进经济学理论与知识。

本着这一目的，我们组织翻译了这套经济学教科书系列。这套译丛的特点是：第一，全面系统。除了经济学、宏观经济学、微观经济学等基本原理之外，这套译丛还包括了产业组织理论、国际经济学、发展经济学、货币金融学、公共财政、劳动经济学、计量经济学等重要领域。第二，简明通俗。与经济学的经典名著不同，这套丛书都是国外大学通用的经济学教科书，大部分都已发行了几版或十几版。作者尽可能地用简明通俗的语言来阐述深奥的经济学原理，并附有案例与习题，对于初学者来说，更容易理解与掌握。

经济学是一门社会科学，许多基本原理的应用受各种不同的社会、政治或经济体制的影响，许多经济学理论是建立在一定的假设条件上的，假设条件不同，结论也就不一定成立。因此，正确理解掌握经济分析的方法而不是生搬硬套某些不同条件下产生的结论，才是我们学习当代经济学的正确方法。

本套译丛于1995年春由中国人民大学出版社发起筹备并成立了由许多经济学专家学者组织的编辑委员会。中国留美经济学会的许多学者参与了原著的推荐工作。中国人民大学出版社向所有原著的出版社购买了翻译版权。北京大学、中国人民大学、复旦大学以及中国社会科学院的许多专家教授参与了翻译工作。在中国经济体制转轨的历史时期，我们把这套译丛献给读者，希望为中国经济的深入改革与发展做出贡献。

《经济科学译丛》编辑委员会

1996年12月

献给我的妻子普西芭（Pushpa）及
两个女儿琼（Joan）和黛安娜（Diane）

作者简介

iii 达摩达尔·N·古扎拉蒂（Damodar N.Gujarati）在执教于纽约市立大学28年之后，现在是纽约州西点美国军事学院社会科学系的经济学教授。古扎拉蒂博士于1960年获孟买大学商学硕士学位，1963年获芝加哥大学工商行政硕士学位，并于1965年获芝加哥大学博士学位。古扎拉蒂博士曾在知名的国内和国际期刊诸如 *Review of Economics and Statistics*，*Economic Journal*，*Journal of Financial and Quantitative Analysis*，*Journal of Business*，*American Statistician* 和 *Journal of Industrial and Labor Relations* 发表论文多篇。古扎拉蒂博士现任多种期刊和图书出版社的编辑评判人，并且是印度官方刊物 *Journal of Quantitative Economics* 的编委会成员。古扎拉蒂博士还是 *Pensions and the New York City Fiscal Crisis*（the American Enterprise Institute，1978），*Government and Business*（McGraw-Hill，1984）和 *Essentials of Econometrics*（McGraw-Hill，2d ed.，1999）的作者。古扎拉蒂博士在计量经济学方面的书已被译成多种文字出版。

古扎拉蒂博士曾是英国 Sheffield 大学访问教授（1970—1971年），是访问印度的 Fulbright 教授（1981—1982年），新加坡国立大学管理学院访问教授（1985—1986年），以及澳大利亚 New South Wales 大学计量经济学教授（1988年夏）。作为美国新闻署赴海外讲学的一位参与者，古扎拉蒂博士曾

在澳大利亚、孟加拉国、德国、印度、以色列、毛里求斯、韩国等广泛讲授微观和宏观经济学专题。古扎拉蒂博士还在加拿大和墨西哥举办过学术研讨会并讲演。

前 言

写作的背景和宗旨

xxv 写作《计量经济学》第四版的主要目的和前三版一样，是用不超过初等水平的矩阵代数、微积分或统计学，对计量经济学做一个初等而又全面的介绍。

在本版中，我力图把1995年第三版问世以来计量经济学理论和实践中所出现的某些进展包括进来。有了高级而又被使用者所钟爱的统计软件，如Eviews，Limdep，Microfit，Minitab，PcGive，SAS，Shazam和Stata等，现在就能对本书前三版尚不能包括的几个计量方法展开讨论。我在本版的一些例子和习题中充分利用了这些统计软件。

我很愉悦而又吃惊地发现，不仅经济学和商学院的学生在使用我的书，而且政治学、国际关系学、农业学和健康科学等其他学科的学生也在用我这本书。这些学科的学生会发现，本书对几个专题所展开的讨论是很有用的。

第四版

本版的主要变化如下：

1. 在引言一章中，在讨论过传统计量方法论的步骤之后，我将讨论如何从几个不相上下的计量模型中选择一个合适模型的问题，这是一个很重要的问题。

xxvi 2. 在第 1 章中，我很简要地讨论了经济变量的度量尺度。其重要性在于，知道变量是比率尺度、区间尺度、序数尺度还是名义尺度，就能决定在给定情形中用哪种计量方法比较适当。

3. 第 3 章的附录包含有 OLS 估计量的大样本性质，特别是一致性。

4. 第 5 章的附录引入了本书中广泛使用的四个重要概率分布（即正态分布、t 分布、χ^2 分布和 F 分布）的性质及其相互关系。

5. 在关于回归模型函数形式的第 6 章中，现在包含有对标准化变量回归的讨论。

6. 为了让非专家能更容易理解，我将对线性回归的矩阵表述方法从原来的第 9 章移到附录 C。并将附录 C 略加扩充，以包括某些较喜欢数学的学生感兴趣的高深内容。新的第 9 章就讨论虚拟变量回归模型。

7. 关于多重共线性的第 10 章，包含有对著名的朗利（Longley）数据的广泛讨论，朗利数据非常有助于阐明多重共线性的性质和范围。

8. 关于异方差性的第 11 章，现在在附录中包含有对怀特（White）稳健标准误的直观讨论。

9. 关于自相关的第 12 章，现在包含有对纠正 OLS 标准误的尼威-怀斯特（Newey-West）方法，以考虑误差项中可能出现自相关的问题。纠正后的标准误被称为 HAC 标准误。本章还将简要地讨论含有自相关误差项的预测这一专题。

10. 关于计量建模的第 13 章，取代了原来的第 13、14 章。应用研究者将发现，本章包含有一些特别有用的新专题。这些专题包括对模型选择准则的简明讨论，如赤池信息准则（Akaike information criterion）、施瓦茨信息准则（Schwarz information criterion）、马娄斯的 C_p（Mallows's C_p criterion）准则和预测 χ^2（forecast chi squre）准则。本章还讨论诸如异常数据（outliers）、杠杆数据（leverage）、有影响力数据（influence）、递归最小二乘法（recursive least squares）和邹至庄的预测失灵检验（Chow's prediction failure test）等专题。本章结束时，对从事计量理论和计量实践的专家提出了警戒性的忠告。

11. 关于非线性回归模型的第 14 章是全新的。由于可轻松自如地使用统计软件，所以估计非线性于参数的回归模型就不再困难。有些计量模型天生就不是参数的线性函数，并需要用迭代方法来估计。本章讨论并解释某些估计非线性参数回归模型的相对简单的方法。

xxvii 12. 关于定性响应回归模型的第 15 章取代了原书关于虚拟因变量回归模型的第 16 章，对涉及具有定性性质因变量的回归模型展开了广泛的讨论。主要集中讨论 logit 和 probit 模型及其变异模型。本章还讨论了用于对计数数据（如一个企业一年内得到的专利数，5 分钟内接听的电话数等）建模的泊松回归模型（Poisson regression model）。本章也对多项式 logit 和 probit 模型

及持续期间模型进行简单讨论。

13. 关于综列数据回归模型的第 16 章也是新的一章。综列数据兼具时间序列和横截面数据的特点。由于在社会科学中可以利用的综列数据越来越多，所以综列数据回归模型也被许多领域的研究者越来越多地采用。针对估计基于综列数据的回归模型时通常使用的固定效应（fixed effects）模型和随机效应（random effects）模型，本章进行了非技术性的讨论。

14. 关于动态计量模型的第 17 章，现在包含有对格兰杰（Granger）因果性检验相当广泛的讨论，在应用研究中都会例行使用到（或误用到）这个检验。格兰杰因果性检验对模型中所用到的滞后项的个数是很敏感的。还要假定模型背后的时间序列是平稳的。

15. 除了增加一些新的解答题和对现有估计方法略微引申之外，关于联立方程模型的第 18、19 和 20 章基本上没有变化。这些都反映了如下事实，由于一系列原因，包括它们在 20 世纪 70 年代 OPEC 石油冲击预测中的不良表现，人们对这种模型的兴趣逐渐淡漠。

16. 第 21 章对原 21 章进行了重大修订。在这一章中，提出并说明了时间序列计量经济学中的几个概念。本章的要点是平稳时间序列的性质和重要性。本章给出了几种查明某给定时间序列是否平稳的方法。一个时间序列的平稳性，对应用本书中所讨论的各种计量方法而言至关重要。

17. 第 22 章也对原 22 章进行了明显的修订。它讨论了基于博克斯-詹金斯（Box-Jenkins）（ARIMA）和向量自回归（vector autoregression，VAR）方法论而进行的经济预测专题，还讨论了用条件异方差自回归（autoregressive conditional heteroscedasticity，ARCH）和广义条件异方差自回归（generalized autoregressive conditional heteroscedasticity，GARCH）方法度量金融时间序列中的易变性这个专题。

18. 关于统计概念的附录 A 略加丰富。附录 C 利用矩阵代数讨论了线性回归模型。这些有益于更高年级的学生。

和前面的版本一样，本书所讨论的所有计量方法都由例子加以说明，其中有些例子是基于各个学科坚实的数据基础。章末的问答题和解答题增加了一些新的例子和数据集。对于更高层次的读者，每章都有一些技术性的附录，给出书中提出的各个定理和/或公式的证明。

课程安排和内容选择

xxviii 此版的变化大大地扩展了本书的论述范围。我希望教师因此在选择适合其听众的专题方面，有充分的灵活性。这里给出如何使用本书的一些建议。

非专业人员一学期的课程：附录 A、第 19 章及对第 10～12 章做一浏览（略去全部证明）。

经济学专业一学期的课程：附录 A 和第 1～13 章。

经济学专业两学期的课程：附录 A、B、C 和第 1～22 章。第 14 和 16

章可以有选择性地学习。某些技术性的附录可以不讲。

毕业生、研究生和研究者：将本书作为计量经济学主题方面必备的参考书。

配套材料*

数据 CD

每本书都配有一张含有文中数据的 CD，CD 是 ASCII 或文本格式的，用绝大多数软件包都能读。

《学生习题解答手册》

对教师免费、对学生出售的《学生习题解答手册》(ISBN: 0072427922)，包含了书中 475 道问答题和解答题的答案。

EViews

我们很高兴本书的第四版在一张 CD 上与书中所有数据一起提供 EViews 的学生版 3.1。这个软件和教材（ISBN：0072565705）一起可以从出版商那里获得。EViews 学生版也可以单独从 QMS 获得。详细信息在如下网址查阅：http://www.eviews.com。

网址

有一个综合性的网址提供了支持计量经济学学习所需要的辅助材料：www.mhhe.com/econometrics/gujarati4。

* 中译本出版者未购买 CD 和《学生习题解答手册》的版权。——译者注

引 言

§1.1 什么是计量经济学?

1 从字面上解释，计量经济学（econometrics）意味着“经济测量”。虽然测量是计量经济学的一个重要部分，但计量经济学涉及的范围要广泛得多，这可以从下面的一些文献摘录中看出：

> 计量经济学，是对经济学的作用存在有某种期待的结果，它把数理统计学应用于经济数据，以使数理经济学构造出来的模型得到经验上的支持，并获得数值结果。[1]
>
> 计量经济学可定义为实际经济现象的数量分析。这种分析乃是基于理论与观测的并行发展，而理论与观测又通过适当的推断方法而得以联系。[2]
>
> 计量经济学可定义为这样的社会科学：它把经济理论、数学和统计推断作为工具，应用于经济现象的分析。[3]
>
> 计量经济学研究经济定律的经验判定。[4]

2 计量经济学家的功夫，就在于找出一组足够具体且足够现实的假定，使他尽可能最好地利用他所获得的数据。[5]

> 计量经济学有助于在积极意义上驱散公众对经济学科（数量的或非数量的）的不良印象：这门学科犹如一个空箱子，即使有打开它的钥匙，对于其空洞的内容，十位经济学家会做出十一种解释。[6]

> 本质上，计量经济学的研究方法是，利用统计推断的理论和技术作为桥头堡，以达到经济理论和实际测算相衔接的目的。[7]

§1.2 为什么是一门单独的学科?

如上述诸定义所示意，计量经济学是经济理论，数理经济，经济统计与数理统计的混合物。但是，这门学科值得作为一门独立的学科来研究。其理由如下：

经济理论所作的陈述或假说大多数是定性分析的。例如，微观经济理论声称，在其他条件不变的情况下，一种商品的价格下降可望增加对该商品的需求量，即经济理论设想（postulates）商品价格与其需求量之间有一负的或逆向关系。但此理论并没有对这两者的关系提供任何数值度量，也就是说，它没有说出随着商品价格的某一变化，需求量将会上升或下降多少。计量经济学家的工作就是要提供这一数值估计。换言之，计量经济学对大多数的经济理论赋予经验内容。

数量经济学的主要问题，是要用数学形式（方程式）表述经济理论而不去问理论的可度量性或其经验方面的可论证性。如前所示，计量经济学的主要兴趣在于经济理论的经验论证。我们将看到，计量经济学家常常使用数理经济学家所提出的数学方程式，但要把这些方程式改造成适合于经验检验的形式。这种从数学方程到计量经济方程的转换需要有许多的创造性和实际技巧。

经济统计学的问题，主要是收集、加工并通过图或表的形式以展现经济
3 数据。这也是经济统计学家的工作。他们是收集国民总产值(GNP)、就业、失业、价格等数据的主要负责人。这些数据从此构成了计量经济工作的原始资料。但是，经济统计学家的工作却到此为止。他们不考虑怎样用所收集来的数据去检验经济理论。当然，如果他们考虑的话，他们就变成计量经济学家了。

虽然数理统计学提供了这一行业中使用的许多工具，但由于大多数经济数据的独特性，即数据并非受控下的实验结果，计量经济学家常常需要有特殊的方法。好比气象学家那样，计量经济学家通常依赖不能由他们来直接控制的数据。如斯班诺斯（Spanos）所正确观察到的那样：

> 在计量经济学中，建模者通常面对的是观测数据而非实验数据。这

对计量经济学中的经验建模有两方面的重要含义。首先，要求建模者掌握与分析实验数据极为不同的技巧……其次，数据搜集者和分析者的分离要求建模者十分熟悉所用数据的性质和结构。[8]

§1.3 计量经济学方法论

对于一个经济问题，计量经济学家是怎样进行分析的？他们的方法论是什么？尽管关于计量经济学的思想方法存在若干学派，我们这里讲述的则是至今在经济学及有关领域的经验研究中仍占统治地位的**传统或经典方法论**。[9]

大致地说，传统的计量经济学方法论按下列路线进行：

（1）理论或假说的陈述；
（2）理论的数学模型的设定；
（3）理论的计量经济模型的设定；
（4）获取数据；
（5）计量经济模型的参数估计；
（6）假设检验；
（7）预报或预测；
（8）利用模型进行控制或制定政策。

为了说明以上步骤，让我们考虑一下著名的凯恩斯消费理论。

（1）理论或假说的陈述

4 凯恩斯说：

> 基本的心理定律……是，通常或平均而言，人们倾向于随着他们收入的增加而增加其消费，但比不上收入增加的那么多。[10]

简言之，凯恩斯设想，**边际消费倾向**（marginal propensity to consume, MPC），即收入每变化一个单位的消费变化率，大于 0 而小于 1。

（2）消费的数学模型的设定

虽然凯恩斯假设消费与收入之间有一正向的相关关系，但他并没有明确

指出这两者之间准确的函数关系。为简单起见，数理经济学家也许建议采用如下的凯恩斯消费函数形式：

$$Y=\beta_1+\beta_2X \qquad 0<\beta_2<1 \tag{I.3.1}$$

其中 Y – 消费支出，X = 收入，而被称为模型**参数**的 β_1 和 β_2 分别代表**截距**和**斜率**系数。

斜率系数 β_2 就是 MPC 的度量，为说明其几何意义，将方程（I.3.1）表示如图 I.1。该方程表明消费与收入存在线性关系。这种关系仅是展示消
5 费与收入之间关系（即经济学中所称的消费函数）的数学模型之一。所谓数学模型不外是一组数学方程而已。如果模型只有一个方程，像上例那样，就称为单一方程模型；如果模型有多于一个方程，就称多方程模型（后者将在以后讨论）。

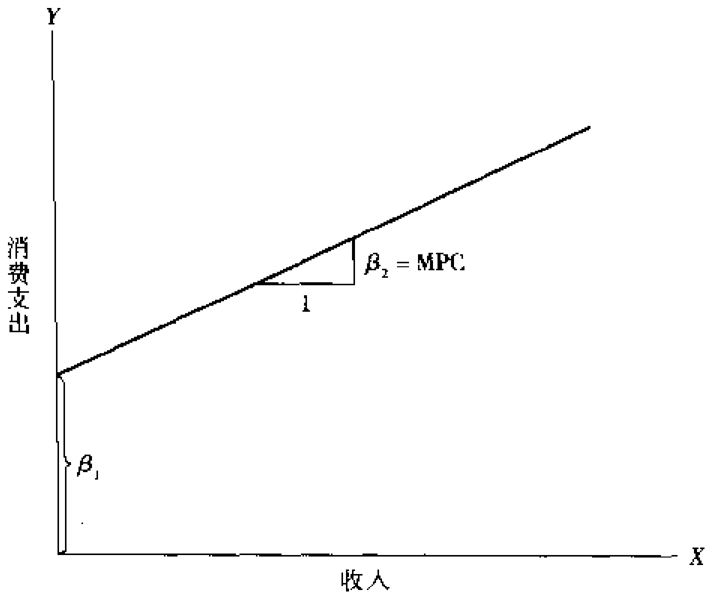


图 I.1 凯恩斯消费函数

出现在方程（I.3.1）等号左边的变量称为**因变量**，而出现在右边的变量（一个或多个）则称**自变量**或**解释变量**。这样，在代表凯恩斯消费函数的方程（I.3.1）中，消费（支出）是因变量而收入是解释变量。

（3）消费的计量经济模型的设定

由方程（I.3.1）给出的消费函数的纯数学模型，它假定消费与收入之间有一个准确的或确定性的关系，因而它对计量经济学家的用处是有限的。一般地说，经济变量之间的关系是非准确的。例如，我们获得了（比如说）500 个美国家庭消费支出和可支配收入的一个样本数据，并把这些数据画在以消费支出为纵坐标，以可支配收入为横坐标的图纸上。我们不能指望所有

的观测值都恰好落在方程（I.3.1）这条直线上，因为除了收入外，还有其他变量也在影响着消费支出。比如说，家庭大小、家庭成员的年龄、家庭的宗教信仰等等，都会对消费产生一定的影响。

考虑到经济变量之间的非准确关系，计量经济学家会把确定性的消费函数（I.3.1）修改如下：

$$Y = \beta_1 + \beta_2 X + u \tag{I.3.2}$$

其中 u 被称为**干扰或误差项**，是一个**随机变量**，它有定义良好的概率性质。干扰项 u 可用来代表所有未经指明的对消费有所影响的因素。

方程（I.3.2）是**计量经济模型**的一个例子。更专门地说，它是本书所主要论述的**线性回归模型**之一例。该计量经济消费函数假设了因变量 Y（消费）对解释变量 X（收入）存在线性关系。然而两者的关系不是准确的，它因家庭而异。

6 可以把消费函数的计量经济模型描绘成像图 I.2 那样。

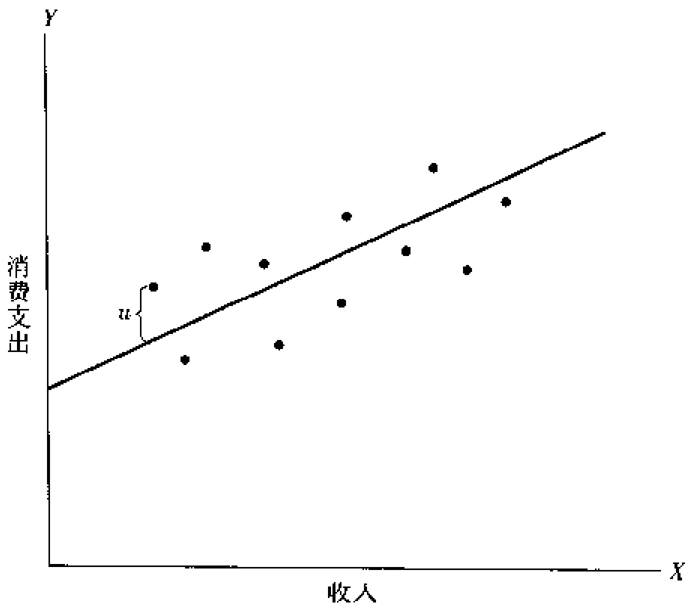


图 I.2 凯恩斯消费函数的计量经济模型

（4）获得数据

为了估计方程（I.3.2）所给出的计量经济模型，也就是为了得到 β_1 和 β_2 的数值，需要有数据。虽然我们在第 1 章里将要更详细地谈论数据对经
7 济分析的根本重要性，但现在不妨先看一下由表 I.1 给出的美国经济 1982—1996 年间的数据。该表中的 Y 变量是对个人加总（aggregate）的消费支出，而 X 变量是国内生产总值（GDP），代表加总收入的一个度量，X 和 Y 均以 1992 年 10 亿美元为单位计算。因此，所列数据代表以 1992 年不变价格计

算的“实际”消费和“实际”收入。现将这些数据描绘在图 I.3 上（比较图 I.2)。图中忽略了时间因素。

(5) 计量经济模型的估计

既然有了数据，下一步的工作就是估计消费函数中的参数。参数的数值估计将对消费函数赋予经验内容。估计参数的具体步骤将在第 3 章中说明。这里仅指出，**回归分析**的统计技术是获得估计值的主要手段。利用这种技术以及表 I.1 所给的数据，我们获得 β_1 和 β_2 的估计值为 -184.08 和 0.706 4。于是所估计的消费函数是

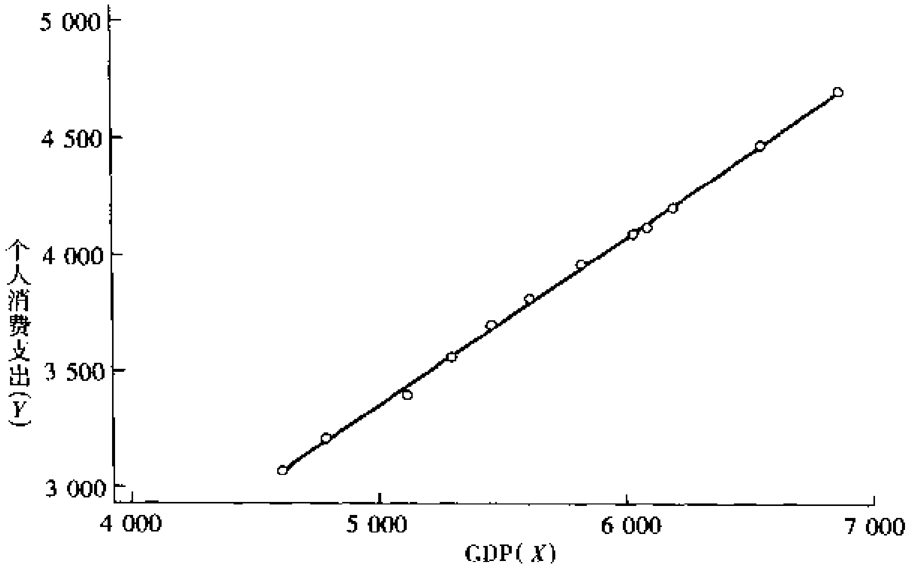


图 I.3 1982—1996 年个人消费支出（Y）与 GDP（X）的关系
（均以 1992 年 10 亿美元为单位）

表 I.1 Y（个人消费支出）和 X（国内生产总值）数据，1982—1996 年，均以 1992 年 10 亿美元为单位

年份	Y	X
1982	3 081.5	4 620.3
1983	3 240.6	4 803.7
1984	3 407.6	5 140.1
1985	3 566.5	5 323.5
1986	3 708.7	5 487.7
1987	3 822.3	5 649.5

1988	3 972.7	5 865.2
1989	4 064.6	6 062.0
1990	4 132.2	6 136.3
1991	4 105.8	6 079.4
1992	4 219.8	6 244.4
1993	4 343.6	6 389.6
1994	4 486.0	6 610.7
1995	4 595.3	6 742.1
1996	4 714.1	6 928.4

资料来源：*Economic Report of the President*，1998，Table B－2，p.282.

$$\hat{Y} = -184.08 + 0.7064 X_i \qquad (I.3.3)$$

Y 顶上的帽符表示一种估计值。[11] 图 I.3 给出了估计的消费函数（即回归线）。

8 如图 I.3 所示，因为数据点很靠近回归线，所以回归线对数据拟合得相当好。从以上方程我们发现，在 1982—1996 年期间，斜率系数（即 MPC）约为 0.70，表明在此样本期间，实际收入每增加 1 美元，平均而言，实际消费支出将增加约 70 美分。[12] 我们说平均而言，是因为消费和收入之间没有准确的关系。这点可以从图 I.3 所展示的、得自（I.3.3）的回归线看出；但并非所有数据点都恰好位于回归线上。我们可以简单地说，据我们的数据，真实收入每提高 1 美元，平均消费支出或消费支出均值会增加约 70 美分。

(6) 假设检验

假定所拟合的模型是实际情况的一个较好的近似，还必须制定适当的准则，借以判断如同方程（I.3.3）中的估计值是否与待检验的理论预期值相一致。根据像米尔顿·弗里德曼（Milton Friedman）这样的"实证"经济学家的意见，凡是不能通过经验证据来证实的理论或假设，都不可作为科学探索的一个部分。[13]

如前所说，凯恩斯曾预期 MPC 是正的，但小于 1。在我们的例子中，我们求得 MPC 约为 0.70。但在把这一发现看做是对凯恩斯消费理论的认可之前，还要追问这一估计值是否充分地低于 1，以使我们不再怀疑这个估计值仅是一次偶然的机会得来，或者怀疑我们用的数据太特殊了。换言之，0.70 是不是在统计意义上（statistically）小于1？如果是，就可用来支持凯恩斯理论。

以样本证据为依据去确认或否定经济理论，是以统计理论的一个分支**统计推断（假设检验）**作为其理论基础的。在本书中，我们会处处看到这种推断过程实际上是如何操作的。

（7）预报或预测

如果所选的模型确认了我们所考虑的假说或理论，就可以根据解释变量或**预测元**（predictor）X 的已知或预期未来值来预测因变量或**预报变量**（forecast variable）Y 的未来值。

为了说明起见，假设我们想预测 1997 年的平均消费支出。1997 年 GDP
9 的值为72 698亿美元。[14]将 GDP 的这个数字代入（I.3.3）的右边，我们得到：

$$\begin{aligned}\hat{Y}_{1997} &= -184.077\,9+0.706\,4\ (7\,269.8)\\ &=4\,951.316\,7\end{aligned} \tag{I.3.4}$$

即约 49 510 亿美元。

因此，给定 GDP 值，预测的平均消费支出或消费支出的均值约为49 510亿美元。1997 年报告的实际消费支出值为49 135亿美元。于是估计模型（I.3.3）约高估了实际消费支出 378.2 亿美元。我们可以说预测误差约为 378.2 亿美元，占 1997 年实际 GDP 值的 0.76%。在我们充分讨论以后章节的线性回归模型之后，我们才会知道这样的误差是“小”还是“大”。但目前重要的是要注意到，在给定这种分析的统计性质下，这种预测误差是无法避免的。

估计模型（I.3.3）还有另外一个用处。假设总统决定减少收入税。这种政策对收入及消费支出和最终的就业会有什么影响？

假若政策改变的结果使投资有所下降，其对经济的影响将如何？宏观经济理论告诉我们，投资支出每改变 1 元，收入的改变由**收入乘数**（$\boldsymbol{M}$）：

$$M=\frac{1}{1-\mathrm{MPC}} \tag{I.3.5}$$

给出。如利用由（I.3.3）得到的 $\mathrm{MPC}=0.70$，此乘数就变成 $M=1/(1-0.70)=3.33$。就是说，投资减少（增加）1 美元，将最终导致收入减少（增加）3 倍之多；注意，乘数的实现需要时间。

在这一计算中 MPC 是个关键值。因为，M 依赖于它。但 MPC 的估计来自诸如（I.3.3）的回归模型。所以，MPC 的数量估计为政策的制定提供了有价值的信息。一旦获知 MPC，即可跟踪政府财政政策的改变，预测收入和消费支出的未来变化过程。

(8) 利用模型进行控制或制定政策

假若我们已估计出由（I.3.3）给出的凯恩斯消费函数，而且政府认为
10 4.9万亿美元（以1992年的美元计）的（消费）支出水平即可维持当前约4.2%的失业率（2000年初），那么什么收入水平将保证消费支出的这一目标值？

如果消费函数（I.3.3）是合理的，简单的算术将表明：

$$4\,900 = -184.077\,9 + 0.706\,4X \tag{I.3.6}$$

解得 $X = 7\,197$。就是说，给定约为0.70的一个MPC，71 970亿美元的收入水平将产生4.9万亿美元的消费支出。

上述计算提示我们，一个已估计出来的模型可服务于控制或政策的目的。通过适当的财政与货币政策的配合，政府可操纵**控制度量**（control variable）X 以产生**目标变量**（target variable）Y 的期望水平。

图 I.4 剖析了经典计量经济学的建模方法。

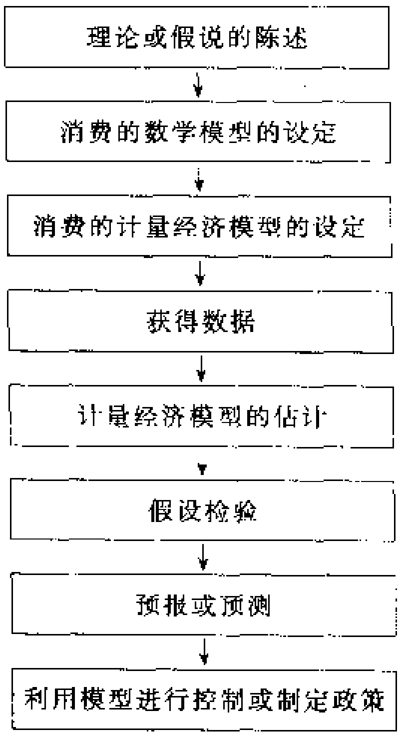


图 I.4　计量经济建模剖析

在竞争模型之间进行选择

当一家政府机构（如美国商业部）搜集经济数据时（如表 I.1 中所示），

其脑海中不一定有什么经济理论。那人们如何知道这些数据实际上是支持凯恩斯消费理论的呢？是因为图I.3所示的凯恩斯消费函数（即回归线）与实
11 际数据点极为接近吗？还有其他的消费模型（理论）能同样好地拟合这些数据吗？比如，米尔顿·弗里德曼提出了一个被称为**持久收入假说**（permanent income hypothesis）[15]的消费模型。罗伯特·霍尔（Robert Hall）也提出了一个被称为**生命周期持久收入假说**（life-cycle permanent income hypothesis）[16]的消费模型。这些模型中有没有一两个也能拟合表 I.1 中的数据呢？

简言之，实践中的研究者面临的问题是，给定一种现象，如消费—收入关系，如何在不相上下的假设或模型之间做出选择。如米勒（Miller）所据理力争的那样：

> 除非假设比某些自然的对手在处理数据上有更好的表现，否则，不针对具体数据便可得出真正确定性的结论……这里，正是原假设的胜利，同时也是貌似可信的对立假设的失败强化了假设。[17]

那么，人们如何在竞争的模型或假设中进行选择呢？这里值得记住克莱夫·格兰杰（Clive Granger）的建议：[18]

> 我想给的建议是，在你提出一种新理论或新的经验模型时，你要考虑这些问题：
>
> (i) 它的用途是什么？它有助于什么样的经济决策？
>
> (ii) 在已经提出的证据中，有没有某个证据让我能将这种新理论或新模型与其他理论或模型做比较？
>
> 我认为关注这样的问题将会加强经济研究或讨论。

通览全书，在解释各种经济现象时，我们将遇到几个竞争的假设。比如，经济学的学生都熟悉生产函数的概念，它基本上就是指产出与投入（如资本和劳动）之间的关系。在文献中，最有名的两个就是柯布－道格拉斯（Cobb-Douglas）生产函数和常替代弹性（constant elasticity of substitution）生产函数。给定投入和产出的数据，我们将不得不找出这两个生产函数中的哪一个能很好地拟合数据（如果有一个能很好地拟合数据的话）。

在能用于检验这些竞争假设中的任何一个时，前面讨论的八步骤经典计量经济方法论在意义上是中性的。

有没有可能提出一种足以包含这些竞争假设的综合方法论呢？这是一个复杂而又有争议的问题。我们将在掌握了必要的计量经济理论之后，在第13 章来讨论这个问题。

§1.4 计量经济学的类型

12 如图1.5中的分类框架所提示，计量经济学可划分为两大类：**理论计量经济学**和**应用计量经济学**。在每一大类中均可按经典或贝叶斯（Bayesian）的传统方法去探讨这一学科。本书的重点在于经典方法。至于贝叶斯方法读者可参阅本章末所附参考文献。

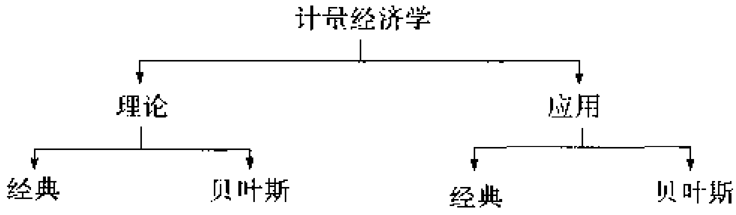


图1.5 计量经济学分类

理论计量经济学是要找出适当的方法，去测度由计量经济模型设定的经济关系式。为此，计量经济学家大量地取材于数理统计学。例如，本书中广泛使用的方法之一是**最小二乘法**（least squares）。理论计量经济学家必须知道此方法所涉及的假定、方法的性质以及当某一或某些假定不成立时，这些性质将会受到什么影响。

在应用计量经济学中，我们利用理论计量经济学作为工具，去研究经济学或商业中的某些特定领域，诸如生产函数，投资函数，供求函数，组合证券理论，等等。

本书主要讨论计量经济学方法的发展、假定、用途及其局限性，并引用经济学和商业中各个方面的例子来进行说明。然而，这不是一本应用计量经济学的书，它并不深入研究任何一个特殊的经济应用领域。关于这方面的某些参考书，见于本书的末尾。

§1.5 数学与统计学的预备知识

虽然本书是在一个初等水平的基础上写作的，但作者仍假定读者熟悉一些统计估计和假设检验的基本概念。不过，为了便于读者重新复习有关知识，
13 在附录A中提供了本书中所使用的基本统计概念的主要而非技术性的概述。至于数学方面，希望读者对微积分的概念还不陌生，虽然这不是必要的。在许多研究生用的计量经济学书中大量使用矩阵代数，作者却愿意声明，阅读本书并无此必要。作者坚信，为了表达计量经济学的基本意念，并不需要用矩阵代数。然而，为了顾及喜欢数学的学生，附录C仍然摘要地

给出了基本回归理论的矩阵表述。对于这些学生，附录B还提供了关于矩阵代数主要结果的一个简明摘要。

§1.6 计算机的作用

当今，回归分析已是计量经济学的家常便饭，似乎没有计算机和某些统计软件将是不可思议的。（但请相信我，我本人是在计算尺时代成长起来的！）幸亏，现在已有好些优等的商用回归软件包，兼容于主机和微机，并且与日俱增。诸如ET、LIMDEP、SHAZAM、MICRO TSP、MINITAB、E-VIEWS、SAS、SPSS、STATA、Microfit、PcGive和BMD等回归软件包，容纳了本书所讲的大多数计量经济学方法和检验讨论的内容。

本书有时要求读者用一种或多种统计软件包做**蒙特卡罗**（Monte Carlo）模拟实验。蒙特卡罗实验是能使读者很好地体会本书所讲的几种统计方法性质的一种“有趣的”练习。在适当的地方，我们将讨论蒙特卡罗实验的细节。

§1.7 文献阅读

计量经济学方法论是一个非常广泛且富有争议的论题。对有兴趣的读者，我建议读以下几本书：

Neil de Marchi and Christopher Gilbert, eds., *History and Methodology of Econometrics*, Oxford University Press, New York, 1989. 本书收集了早期的一些关于计量经济学方法论的读物，对关于时间序列数据（即对一个时期收集的数据）的英国计量经济学方法有广泛的讨论。

Wojciech W. Charemza and Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autogression*, 2d ed., Edward Elgar Publishing Ltd., Hants, England, 1997. 作者们批判了传统计量经济学方法，并详细注释了计量经济学方法论的新动向。

Adrian C. Darnell and J. Lynne Evans, *The Limits of Econometrics*, Edword Elgar Publishers Ltd., Hants, England, 1990. 该书对计量经济学的各
14 种方法论作了一个较为中立的论述，并表示要重新加盟到传统的计量经济学方法论中去。

Mary S. Morgan, *The History of Econometric Ideas*, Cambridge University Press, New York, 1990. 作者对计量经济学的理论和实践做了出色的历史剖析，还对哈默维（Haavelmo, 1990年诺贝尔经济学奖得主）对计量经济学的早期贡献做了深入的讨论。出于同样想法，David F. Hendry and Mary

S.Morgan, *The Foundation of Econometric Analysis*, Cambridge University Press, U.K., 1995. 也搜集了计量经济学研讨会上的作品，以说明计量经济思想随着时间的演进过程。

David Colander and Reuven Brenner, eds. *Educating Economists*, University of Michigan Press, Ann Arbor, Michigan, 1992. 该书对经济学教学和实践提出了尖锐的、有时是不可知论的看法。

关于贝叶斯统计学和计量经济学，如下书目很有用处：John H. Dey, *Data in Doubt*, Basic Blackwell Ltd., Oxford University Press, England, 1985. Peter M.Lee, *Bayesian Statistics: An Introduction*, Oxford University Press, England, 1989. Dale J. Porier, *Intermediate Statisitcs and Econometrics: A Comparative Approach*, MIT Press, Cambridge, Massachusetts, 1995. 一本高深的参考书是 Arnold Zeller, *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York, 1971.

【注释】

[1] Gerhard Tintner, *Methodology of Mathematical Economics and Econometrics*, The University of Chicago Press, Chicago, 1968, P.74.

[2] P.A.Samuelson, T.C.Koopmans and J.R.N.Stone, "Report of the Evaluative Committee for Econometrica", *Econometrica*, Vol.22, No.2, April 1954, pp. 141–146.

[3] Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, New York, 1964, p. 1.

[4] H.Theil, *Principles of Econometrics*, John Wiley & Sons, New York, 1971, P.1.

[5] E.Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1966, P.514.

[6] Adrian C.Darnell and J. Lynne Evans, *The Limits of Econometrics*, Edward Elgar Publishing, Hants, England, 1990, P.54.

[7] T.Haavelmo, "The Probability Approach in Econometrics", Supplement to *Econometrica*, Vol.12, 1944, preface P.iii.

[8] Aris Spanos, *Probability Theory and Statistical Inference: Econometric, Modeling with Observational Data*, Cambridge University Press, United Kingdom, 1999, P.21.

[9] 关于对计量经济学方法论更完善但可能属于高级的讨论，可参见 David F.Hendry, *Dynamic Econometrics*, Oxford University Press, New York, 1995。也可参见 Aris Spanos 的前引文献。

[10] John Maynard Keynes, *The General Theory of Employment, Interest and Money*, Harcourt Brace Jovanovich, New York, 1936, p.96.

[11] 在一个变量或参量的上方画一个帽符以表示它是一个估计值，已成为惯例。

[12] 第 3 章将表明，最小二乘法给出了这些估计值。这里暂且不去管它们是怎样得来的，而且也不必问为什么截距是个负值。

[13] 参看 Milton Friedman, "The Methodology of Positive Economics", *Essays in Positive Economics*," University of Chicago Press, Chicago, 1953。

[14] 有 1997 年 PCE 和 GDP 的数据可用，但我们在说明本节所讨论的专题时故意不用。如我们在以后章节中将讨论的那样，留下一部分数据用来检查拟合模型对样本外观测的预测力如何，是一个很好的主意。

[15] Milton Friedman, *A Theory of Consumption Function*, Princeton University Press, Princeton, N.J., 1957.

[16] R.Hall, "Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 1978, vol.86, pp.971－987.

[17] R.W.Miller, *Fact and Method*: *Explanation*, *Confirmation*, *and Reality in the Natural and Social Sciences*, Princeton University Press, Princeton, N.J., 1978, p.176.

[18] Clive W.J.Granger, *Empirical Modeling in Economics*, Cambridge University Press, U.K., 1999, p.58.

简要目录

目　录

目 录

第1篇

单一方程回归模型

15 本书的第 1 篇介绍各种单一方程回归模型。在这些模型中，一个叫做因变量（dependent variable）的变量被表达为一个或多个叫做解释变量（explanatory variables）的其他变量的线性函数。在这样的模型中，有一个隐含的假定：如果在因变量与解释变量之间存在因果关系的话，这个关系只有一个流向，就是从解释变量到因变量。

在第 1 章中，我们讨论回归（regression）的历史和现代释义，并用几个取自经济学和其他领域的例子来说明这两种释义的区别。

在第 2 章中，我们借助于双变量线性回归模型来介绍回归分析的一些基本概念。双变量线性模型，是指其中的因变量被表达成仅仅一个解释变量的线性函数。

在第 3 章中，我们继续同双变量模型打交道，并引进以经典线性回归模型（classical linear regression model）为名的一种涉及若干简化假定的模型。有了这些假定，我们随之介绍普通最小二乘法（ordinary least squares, OLS），用以估计双变量回归模型中的参数。OLS 法易于应用，且有一些非常良好的统计性质。

在第 4 章中，我们介绍（双变量）经典正态线性回归模型。它假定随机因变量遵循正态概率分布。有了这一假定，在第 3 章中得到的 OLS 估计量就有了一些比非正态经典线性回归模型更强的统计性质，使我们能从事于统计推断——假设检验。

16 第 5 章专门讨论假设检验。这里，我们试图分辨回归系数的估计值是否与它们的假设值（hypothesized values）相符（无矛盾），假设值是指从理论或先前的经验工作中得到的值。

第 6 章考虑由双变量回归模型延伸出来的一些枝节问题。具体地说，本章讨论：（1）过原点回归；（2）尺度与测量单位；（3）回归模型的函数形式，如双对数，半对数和倒数等模型的问题。

在第 7 章中，我们考虑含有多于 1 个解释变量的复回归或多变量回归模型（multiple regression），并说明怎样能把 OLS 法推广应用到这些模型的参数估计上。

第 8 章把第 5 章介绍的概念推广到复回归模型上，并指出由多个解释变量的引入而诱发的若干复杂性。

第 1 篇以讨论虚拟或定性解释变量的第 9 章作为结束。本章强调，并不是所有的解释变量都必须是定量的（即比率尺度）。虽然诸如性别、种族、宗教、国籍和居住地等变量都不容易量化，但它们在解释许多经济现象时却能够起到重要作用。

第 1 章　回归分析的性质

17 如引言中所说，回归是计量经济学的主要工具。本章中，我们将扼要地阐释这一工具的性质。

§1.1　“回归”一词的历史渊源

回归一词最先由弗朗西斯·高尔顿（Francis Galton）引入。在一篇著名的论文中，高尔顿发现，虽然有一个趋势——父母高，儿女也高；父母矮，儿女也矮，但给定父母的身高，儿女辈的平均身高却趋向于或者“回归”到全体人口的平均身高。[1]换言之，尽管父母都异常高或异常矮，但儿女的身高却有走向人口总体平均身高的趋势。高尔顿的普遍回归定律（law of universal regression）还被他的朋友卡尔·皮尔逊（Karl Pearson）证实。皮尔逊曾收集过一些家庭群体的1 000多名成员的身高记录。[2]他发现，对于一个父亲高的群体，儿辈的平均身高低于他们父辈的身高，而对于一个父亲矮的群体，儿辈的平均身高则高于其父辈的身高。这样就把高的和矮的儿辈身高一同“回归”到所有男子的平均身高。用高尔顿的话说，这是“回归到中等”（regression to mediocrity）。

§1.2 回归的现代释义

18 然而，对回归的现代解释却是非常不同的，大致上，我们可以这样说：

> 回归分析是关于研究一个叫做**因变量**的变量对另一个或多个叫做**解释变量**的变量的依赖关系，其用意在于通过后者（在重复抽样中）的已知或设定值，去估计和（或）预测前者的（总体）均值。

对回归分析的这种看法的全部含义，将随着本书的进程而渐明。但是用少数几个简单的例子，就能把回归的基本概念弄得一清二楚。

例 子

1. 再考虑高尔顿的普遍回归定律。高尔顿的兴趣在于发现为什么人口的身高分布有一种稳定性。但从现代的观点考虑，我们并不关心这种解释。我们关心的，却是给定父辈身高的情形下找出儿辈平均身高的变化。换言之，我们关心一旦知道了父辈的身高，怎样预测儿辈的平均身高。为了看清楚怎样能做到这一点，考虑图 1.1 这个**散点图**（scatter diagram 或 scattergram）。

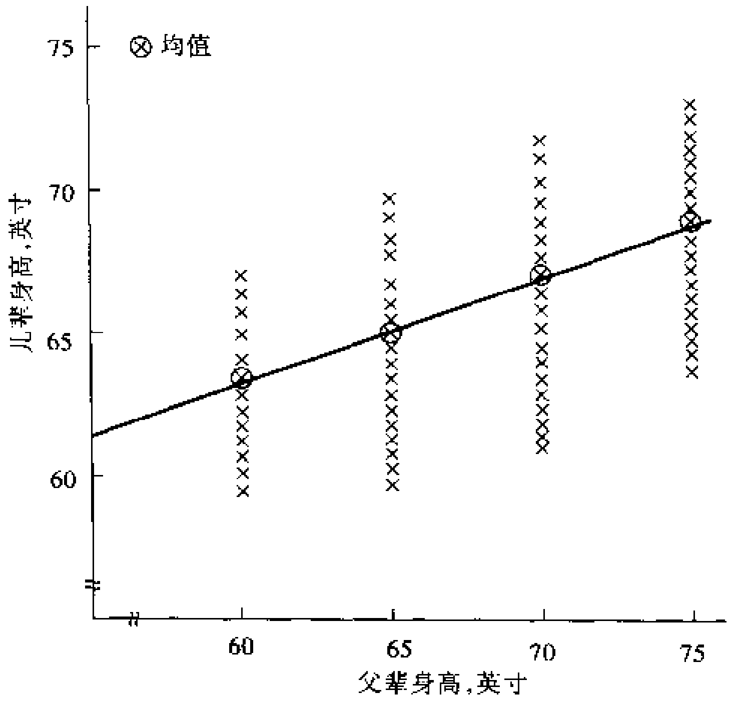


图 1.1 对应于给定父亲身高的儿子身高的假想分布

19 该图展示了对应于设定的父亲身高，儿子在一个假想人口总体中的身高分布。注意，对应于任一给定的父亲身高，都有着儿子身高的一个分布范围。然而，值得注意的是，随着父亲身高的增加，儿子的平均身高也增加。为了看得明白无误，我们勾画了一条通过这些散点的直线，以表明儿子的平均身高是怎样随父亲身高的增加而增加的。我们以后知道，这条线就叫做**回归线**(regression line)。[3]

2. 考虑图 1.2 中的散布图形。这是在不同的固定年龄处测度的一个假想的男孩身高的总体分布。注意，对应于每一给定年龄都有一个高度的范围。显然，同一个给定年龄的男孩们不会完全一样高。但身高随年龄增加而增加（当然，到一定的年龄为止），如果我们通过表示给定年龄下平均身高

20 的圆点画一条线（回归线），就可以清楚地看出这一点。于是，知道了男孩的年龄，就能预测相当于这个年龄的平均身高。

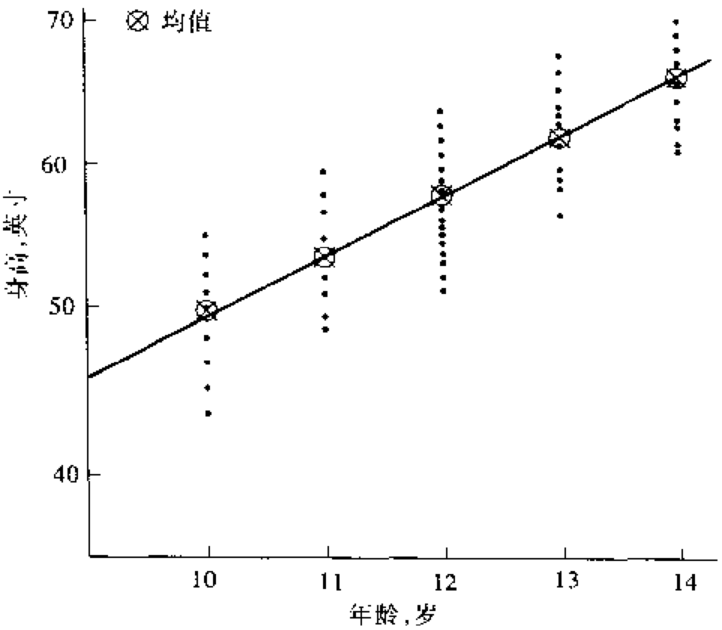


图 1.2　对应于选定年龄的假想身高分布

3. 转到经济学中的例子。经济学家也许想研究个人消费支出对税后或可支配实际个人收入的依赖关系。这种分析会有助于估计边际消费倾向(MPC)，就是实际收入每美元价值的变化所引起的消费支出的平均变化(见图 1.3)。

4. 一位能设定价格或产出（但不能同时设定两者）的垄断商，也许想知道产品需求对价格变化的实际反应，通过这种定价实验，也许能估计出产品需求的**价格弹性**（price elasticity，即产品需求对价格变化的灵敏程度），从而有助于确定最有利可图的价格。

5. 一位劳动经济学家也许要研究货币工资变化率对失业率的关系。图 1.3 给出了历史数据所表现的散点图。图中的曲线是把货币工资变化率同失业率联系起来的著名的**菲利普斯曲线**（Phillips curve）之一例。这样的散点

图能使劳动经济学家预测在给定的某个失业率下货币工资的平均变化。这种知识也许有助于发表一些关于经济扩张过程的见解。因为，工资的增长不免要反映到物价的增长上来。

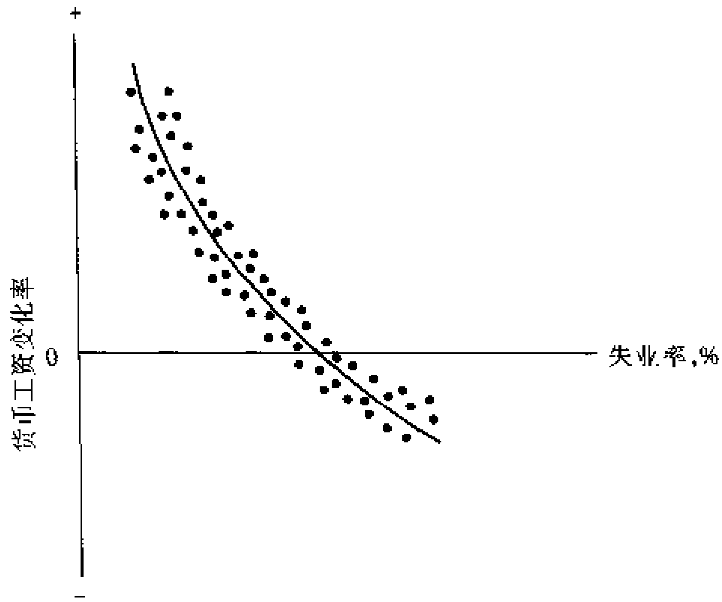


图 1.3　假想菲利普斯曲线

21 6. 由货币经济学中得知，其他条件不变，通货膨胀率 π 愈高，人们愿意以货币形式保存的收入比例 k 愈低。图 1.4 描述了这种关系。对这种关系进行数量分析，将使货币经济学家能够预测在各种通货膨胀率下人们愿意以货币形式保存的收入比例。

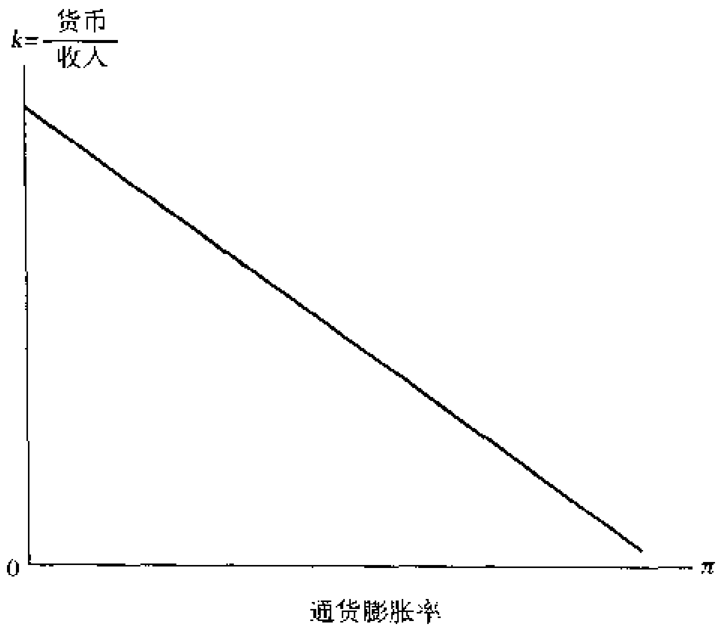


图 1.4　货币持有与通货膨胀率 π 的关系

7. 公司的销售部主任想知道人们对公司产品的需求与（比如说）广告费开支的关系。这种研究在很大程度上有助于计算出相对于广告费支出的**需求弹性**，即广告费预算每变化百分之一时需求变化的百分比。这种知识有助于制定“最优”广告费预算。

8. 最后，也许农业经济学家想研究作物（比如说小麦）收成对气温、降雨量、阳光量和施肥量的依赖关系。这种依赖性分析能使他对给定的解释变量进行信息预测或预报作物的平均收成。

读者也能提供关于一个变量依赖于另一个或多个变量的大量事例。本书讨论的回归分析技巧，就是用来研究这种变量之间的依从关系的。

§1.3 统计关系与确定性关系

22 读者能从 1.2 节所举的例子看到，不像经典物理学中考虑的那种变量之间的函数或确定性依赖关系，在回归分析中，我们考虑的是一种统计依赖关系。在变量之间的统计关系式中，我们主要处理的是随机（random 或 stochastic）变量[4]，也就是有着概率分布的变量。但是在函数或确定性依赖关系中，我们要处理的变量不是随机的。

例如，作物收成对气温、降雨、阳光以及施肥的依赖关系是统计性质的。这个性质的意义在于：这些解释变量固然重要，但并不能使农业经济学家准确地预测作物的收成。因为一则对这些变量的测量有误差，二则还有一大堆集体影响收成的因素（变量），但难以一一辨认出来。因此，无论我们考虑了多少个解释变量，都无法完全地解释作物收成这个因变量。它的一些“内在的”或随机的变异是注定存在的。

另一方面，在确定性现象中，我们处理这样一类关系式，比如说，牛顿的引力定律所表示的关系式：宇宙间的每个粒子吸引着另一个粒子，其引力与它们的质量乘积成正比，而与它们之间距离的平方成反比，用符号表示就是$F=k(m_1m_2/r^2)$，其中 F = 引力，m_1 和 m_2 为两个粒子的质量，r = 距离，而 k = 比例常数。另一个例子是欧姆定律：对于金属导体在有限的温度范围内，电流 C 正比于电压 V，即 $C=\left(\frac{1}{k}\right)V$，其中$\frac{1}{k}$是比例常数。这类确定性现象的其他例子包括玻意耳（Boyle）的气体定律，基尔霍夫（Kirchhoff）的电流定律和牛顿的运动定律。

在本书中，我们不去研究这类确定性现象。当然，如果，比方说，在牛顿的引力定律中，k 的测量有误差，则原来的确定性关系就变成了一个统计关系式。这时，引力只能按给定的 k 值（还有 m_1、m_2 和 r）近似地加以预测。于是，变量 F 变成了一个随机变量。

§1.4 回归与因果关系

虽然回归分析研究一个变量对另一（些）变量的依赖关系，但它并不一定意味着因果关系。用肯德尔（Kendall）和斯图尔特（Stuart）的话说：“一个统计关系式，不管多强也不管多么有启发性，却永远不能确立因果方面的
23 联系：对因果关系的理念，必须来自于统计学以外，最终来自这种或那种理论。”[5]

在前面所引的农作物收成一例中，没有任何统计上的理由可以认为降雨量不依赖于作物收成。我们把作物收成看做依赖于降雨量等因变量，是出于非统计上的考虑。普通常识提示了我们不能把这种关系倒转过来，因为我们不能用改变作物收成的方法来控制降雨。

所有 1.2 节引用的例子都指出一个要点：**从逻辑上说，统计关系式本身不可能意味着任何因果关系**。要谈因果律，必须诉诸先验的或理论上的思考。例如，在上面所引的第三个例子中，我们说消费支出依赖于实际收入，是引用了经济理论的。[6]

§1.5 回归与相关

与回归分析密切相关而在概念上则迥异的，是以测度两个变量之间的线性关联力度为其主要目的的**相关分析**（correlation analysis）。第 3 章中我们将要详细讨论的**相关系数**（correlation coefficient）就是用来测度这种（线性）关联强度的。例如，我们也许有兴趣去求吸烟与肺癌、统计学考分与数学考分、中学成绩与大学成绩等等之间的相关（系数）。而在回归分析中，如前所述，我们并不主要对这种度量感兴趣。而感兴趣的是试图根据其他变量的设定值来估计或预测某一变量的平均值。例如，我们也许想知道能否从一个学生的已知数学考分，来预测他的统计学平均考分。

回归和相关有一些值得提出的基本不同点。在回归分析中，对因变量和解释变量的处理方法存在着不对称性。因变量被当作是统计的，随机的，也就是它有一个概率分布。而解释变量则被看做是（在重复抽样中）取有固定值的。[7] 这点在第 1.2 节所给的回归定义中已作说明。因此，在图 1.2 中，我们假定年龄变量被固定在给定的水平上，而身高则是在这些水平上加以度量的。但在相关分析中，我们对称地对待任何（两个）变量；对因变量和解
24 释变量不加以区别。毕竟，数学考分与统计学考分之间的相关就是统计学考分与数学考分之间的相关。此外，两个变量都被看做是随机的，如我们将会看到的，大部分相关理论都建立在变量是随机性的假定之上。但是，本书要

阐述的回归理论大部分都以下述假定为条件：因变量是随机的，而解释变量是固定的或非随机的。[8]

§1.6 术语与符号

在我们进入正式的回归理论分析之前，先来斟酌一下有关术语与符号的问题。因变量和解释变量两名词在文献中都有过种种其他描述。一个有代表性的清单如下：

因变量（Dependent variable）	解释变量（Explanatory variable）
⇕	⇕
被解释变量（Explained variable）	自变量（Independent variable）
⇕	⇕
预测子（Predictand）	预测元（Predictor）
⇕	⇕
回归子（Regressand）	回归元（Regressor）
⇕	⇕
响应（Response）	刺激变量（Stimulus）
⇕	⇕
内生（Endogenous）	外生（Exogenous）
⇕	⇕
结果（Outcome）	共变（Covariate）
⇕	⇕
被控变量（Controlled variable）	控制变量（Control variable）

虽然采用什么名词术语是一个个人爱好和传统习惯的问题，但本书中我们采用因变量—解释变量术语。

如果我们在研究一个变量对一个解释变量的依从关系，如消费支出对实际收入的依赖，则称这种研究为简单（simple）或**双变量回归分析**（two-variable regression analysis）。但是，如果我们在研究一个变量对多于一个解释
25 变量的依赖性，有如农作物收成依赖于降雨、气温、阳光和施肥一例，则称它为（多元或多变量）**复回归分析**（multiple regression analysis）。换句话说，在双变量回归中只有一个解释变量，而在复回归中则有多于一个解释变量。

Random 一词和 **Stochastic** 是同义语，都是随机的意思。如前所述，一个随机变量是指这样的一个变量：它以给定的概率取任一数值集合，正数或负数。

除非另作声明，字母 Y 一律指因变量，而诸 $X's(X_1, X_2, \cdots, X_k)$ 一律指解释变量。其中 X_k 代表第 k 个解释变量。下标 i 或 t 则指第 i 次或第 t 次观测值。这样，X_{ki}(或 X_{kt})就指对变量 X_k 的第 i(或 t)次观测值。N(或 T)指总体中的观测值的总个(次)数，而 n(或 t)则指样本中的观测值总个数。作为一种惯例，观测值下标 i 将用于**横截面数据**（cross-sectionaldata）（即在一个时间点上收集的数据），而下标 t 将用于**时间序列数据**（time series data）(即在一段时期收集的数据)。关于横截面和时间序列的性质以及经验分析所用数据的性质与来源等重要议题，将在下节讨论。

§1.7 计量经济分析所用数据的性质与来源[10]

任何计量经济分析的成功最终都有赖于适当数据的获得。因此，我们花点时间，来讨论经验分析中所遇到的数据的性质、来源和局限性是非常必要的。

数据类型

用于经验分析的数据有三类：**时间序列**，**横截面**以及**混合**（时间序列与横截面合并）数据。

时间序列数据。引言中的表 I.1 展示的数据就是时间序列数据之一例。一个时间序列是对一个变量在不同时间取值的一组观测结果。这些数据可以是在有规则的时间间隔收集的。譬如每日（如股票价格），每周（如联邦储备委员会提供的货币供给数字），每月（如失业率，消费者价格指数 CPI），每季（如 GNP），每年（如政府预算），每 5 年（如制造业普查资料），每10年（如人口普查资料），有些数据每季和每年都有公布，如 GDP 和消费者支
26
出数据。随着高速计算机的出现，可以搜集极短的时间区间内的数据，如股票价格数据，几乎可以得到其连续数据（即所谓的实时牌价）。

虽然许多计量经济研究都使用时间序列数据，但它们的使用给计量经济学家提出了特殊的问题。如我们在以**时间序列计量经济学**为名的篇章中所表明的，大多数基于时间序列数据的分析之作，大多假定所依据的时间序列是**平稳的**（stationary）。虽然要介绍平稳性的准确技术含义为时尚早，但粗略地说，如果一个时间序列的均值和方差不随时间而系统地变化，那它就是平稳的。为了看出其含义，考虑图 1.5，它描绘的是美国从 1959 年 1 月 1 日至 1999 年 7 月 31 日期间 M1 货币供给的行为。(实际数据在习题 1.4 中给出。)如你从图中所见，随着时间推移，M1 货币供给表现出稳定上升的**趋势**（trend）和易变动性，这就表明 M1 的时间序列不是平稳的。[11]我们在第 21

章将详尽讨论这一点。

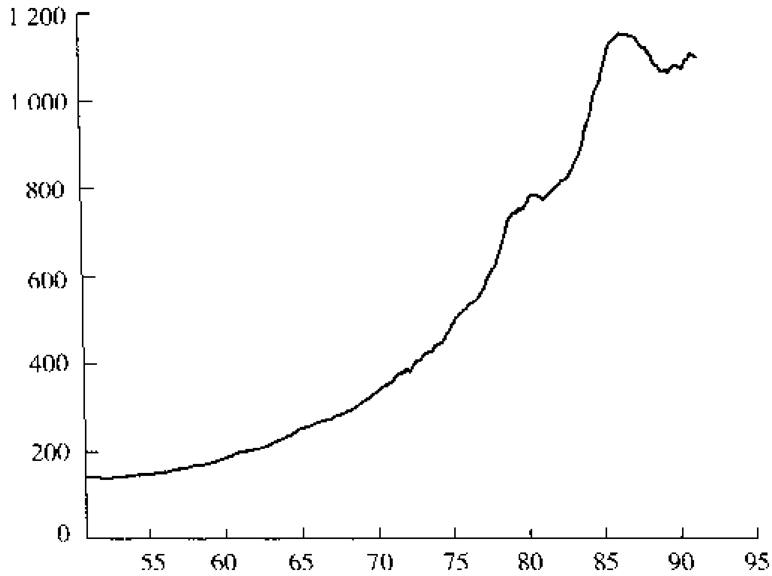


图 1.5　美国 1951 年 1 月—1999 年 9 月的 M1 货币供给

横截面数据。横截面数据指对一个或多个变量在同一时间点上收集的数
27 据。诸如人口普查局每 10 年进行一次的人口普查（最近一次是在 2000 年），密歇根大学举办的消费者支出普查。当然，由盖洛普（Gallup）和其他机构主导的一类民意调查更是属于这类数据。表 1.1 给出了横截面数据的一个实例。该表给出 1990 和 1991 年美国 50 个州的劳工会蛋产量和蛋价格。对每一年份 50 个州的数据构成一个横截面数据样本。这样，表 1.1 中就有两个横截面样本。

正如时间序列数据由于平稳性的争论产生了它独有的特殊问题，横截面数据也有其自身的问题，特别是异类性（heterogeneity）的问题。我们有些州生产大量的蛋类（如宾夕法尼亚州），而有些州则生产甚少（如阿拉斯加
28 州）。当我们的统计分析包含有相异的单元时，我们必须考虑**尺度**或**规模**效应以避免把苹果和橙子混同起来。为了清楚地看出这一点，我们将美国 1990 年 50 个州的蛋产量及其价格数据描绘在图 1.6 上。这幅图表明观测值散布得多么分散。在第 11 章中，我们将看到，在评价经济变量之间的关系时，规模效应如何会成为一个重要的因素。

混合数据。在混合或组合数据中兼有时间序列和横截面数据的成分。表 1.1 中的数据即混合数据之一例。对每一个年份我们有 50 个横截面观测值，而对每一个州我们有蛋价和蛋产量的两个时间的观测序列，总共 100 个混合（或组合）观测值。类似地，习题 1.1 给出的数据是混合数据。因为1973—1997 年每个国家的通货膨胀率构成一个时间序列。而对某一年说，7 个国家的通货膨胀率又构成一个横截面。在此混合数据中，我们有 175 个观测值——对 7 个国家中的每一个，有 25 个年观测值。

表 1.1 美国蛋类生产

州	Y_1	Y_2	X_1	X_2	州	Y_1	Y_2	X_1	X_2
AL	2 206	2 186	92.7	91.4	MT	172	164	68.0	66.0
AK	0.7	0.7	151.0	149.0	NE	1 202	1 400	50.3	48.9
AZ	73	74	61.0	56.0	NV	2.2	1.8	53.9	52.7
AR	3 620	3 737	86.3	91.8	NH	43	49	109.0	104.0
CA	7 472	7 444	63.4	58.4	NJ	442	491	85.0	83.0
CO	788	873	77.8	73.0	NM	283	302	74.0	70.0
CT	1 029	948	106.0	104.0	NY	975	987	68.1	64.0
DE	168	164	117.0	113.0	NC	3 033	3 045	82.8	78.7
FL	2 586	2 537	62.0	57.2	ND	51	45	55.2	48.0
GA	4 302	4 301	80.6	80.8	OH	4 667	4 637	59.1	54.7
HI	227.5	224.5	85.0	85.5	OK	869	830	101.0	100.0
ID	187	203	79.1	72.9	OR	652	686	77.0	74.6
IL	793	809	65.0	70.5	PA	4 976	5 130	61.0	52.0
IN	5 445	5 290	62.7	60.1	RI	53	50	102.0	99.0
IA	2 151	2 247	56.5	53.0	SC	1 422	1 420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48.0	45.8
KY	412	483	67.7	73.5	TN	277	279	71.0	80.7
LA	273	254	115.0	115.0	TX	3 317	3 356	76.7	72.6
ME	1 069	1 070	101.0	97.0	UT	456	486	64.0	59.0
MD	885	898	76.6	75.4	VT	31	30	106.0	102.0
MA	235	237	105.0	102.0	VA	943	988	86.3	81.2
MI	1 406	1 396	58.0	53.8	WA	1 287	1 313	74.1	71.5
MN	2 499	2 697	57.7	54.0	WV	136	174	104.0	109.0
MS	1 434	1 468	87.8	86.7	WI	910	873	60.1	54.0
MO	1 580	1 622	55.4	51.5	WY	1.7	1.7	83.0	83.0

注：Y_1 = 1990 年蛋产量（百万个）　　X_1 = 1990 年每打价格（美分）

Y_2 = 1991 年蛋产量（百万个）　　X_2 = 1991 年每打价格（美分）

资料来源：*World Almanac*, 1993, p.119. The data are from the Economic Research Service, U.S. Department of Agriculture.

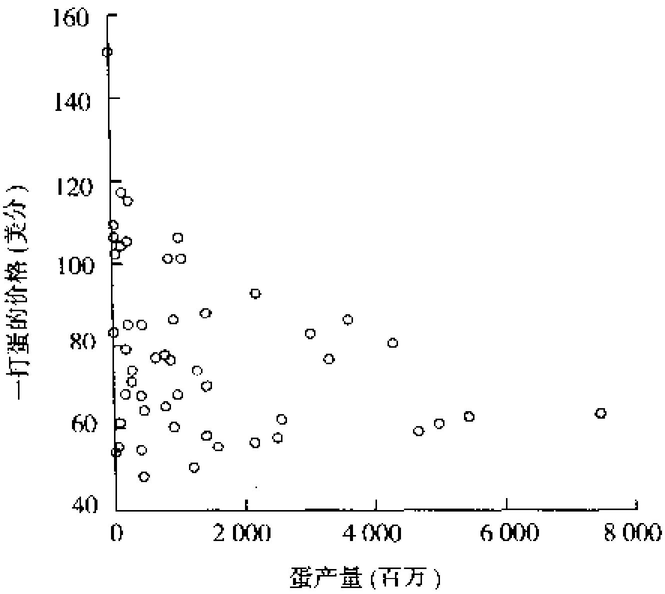


图 1.6　1990 年蛋产量与价格的关系

综列、纵列或微观综列数据。这是混合数据的一种特殊类型。指对相同的横截面单元（比如家庭或厂家）在时间轴上进行跟踪调查的数据。例如，美国商业部定期进行的一种住房调查，每一次调查都对同样的住户（或对住在同样地址的人们）进行采访，以便发现自从上次调查以来该户的住房和财务条件是否发生变化。通过对相同住户的定期采访，综列数据对住户行为的动态特性提供了非常有用的信息，我们将在第 16 章看到这一点。

数据来源[12]

29 用于经验分析的数据可以由一个政府机构（如商业部），一个国际机构（如国际货币基金组织或世界银行），一个私有组织（如标准普尔公司）或某一个人来收集。从表面上看，有成千的这种机构在收集着各种用途的数据。

互联网　互联网真的使数据搜集发生了革命性变化。如果只是用键盘"在网上冲浪"（比如汇率），那你将淹没在各种各样的数据来源之中。我们在附录 E 中提供了一些被频繁访问的网址，它们能提供各类经济和金融数据。许多数据都无须太多费用就能下载。你可能想把那些能为你提供有用经济数据的各种网址制成书签。

这些机构所收集的数据可以是实验（experimental）或非实验（nonexperimental）性质的。在自然科学中，经常收集的是实验数据。这时，研究者希望在保持一些因素不变的情况下收集数据，以便评价另一些因素对某一现象的影响。例如，在评价肥胖对血压的影响时，研究者要在人们饮食、烟酒习惯都不变的情况下收集数据，以便尽可能减少这些变量的变化对血压的影响。

在社会科学中，人们通常获得的数据是非实验性质的，就是说，这些数据不受研究者的控制。[13]例如，GNP、失业、股票价格等数据并不受研究者的直接控制。如我们将要看到的，对数据缺乏控制，常常给研究者在寻觅某种事态的准确原因时造成特别的难题。例如，究竟是货币供给决定（名义）GDP呢，还是（名义）GDP决定货币供给呢？

数据的准确性[14]

虽然有大量的数据可供经济研究之用，但是数据的质量常常不那么好。对此有几点理由。第一，如上所说，大部分社会科学数据是非实验性质的。因此就有观测误差的可能；或出于疏漏，或出于差错。第二，即使是实验得来的数据，测量误差也可由近似计算或进位而产生。第三，在问卷型调查 30 中，非应答问题可能十分严重；问卷能有40%的应答者就算幸运。根据这样的部分答卷做的分析未必能够真正反映60%非应答者的行为。由此导致所谓（样本）**选择性偏误**（selectivity bias）。不但如此，回答问卷的人不一定回答所有的问题，特别是那些财务上敏感的问题，从而导致更多的选择性偏误。第四，获取数据的抽样方法可能变化很大，要比较不同样本得来的结果通常是困难的。第五，通常获得的经济数据都是高度加总的。例如，大多数的宏观数据（如GNP、就业、通货膨胀、失业）都是对整个经济或者至少是对一些广大地区给出的。这种高度加总数据未必能告诉我们多少有关个人或微观单位的情况，而后者才是研究的最终目标。第六，由于保密性质，某些数据只能以高度加总的形式公布。例如，法律不允许IRS公开个人税收回执数据；它只能透露一些概括的数据。因此，尽管你想知道某一收入水平的个人在卫生保健方面花了多少钱，但你是做不到这种分析的，除非是在非常高度的加总水平上，然而，这样的宏观分析往往揭示不了微观单位的行为动态。类似地，商业部每5年进行一次企业普查，但法律却不允许它公布关于任何厂家的生产、人员雇佣、能源消耗、研究与开发费用等方面的信息，因此要研究在这些项目上的厂际差异是困难的。

因为有这些和许多其他的问题，研究者应时刻记住：**研究结果不能比数据的质量更好。**所以，如果在一定情况下，研究者发现研究结果“不能令人满意”的话，原因不一定是误用模型，而是数据的质量不好。不幸的是，由于大多数社会科学研究所用的数据是非实验性质的，研究者常常别无选择，惟有依赖其所能获得的数据。但他还应时刻记住：所用的数据未必是最好的。因此不要过于教条地对待研究的结果，尤其当数据的质量受到怀疑时。

对变量测量尺度的注解[15]

我们通常遇到的变量分为如下四大类：比率尺度（ratio scale）、区间尺

度（interval scale）、序数尺度（ordinal scale）和名义尺度（nominal scale）。理解其中的每一类对我们都很重要。

31 **比率尺度** 对于一个变量 X，取其两个值 X_1 和 X_2，比率 X_1/X_2 和距离（X_2-X_1）都是有意义的量。此外，这些值在这种尺度下存在着一种自然顺序（上升或下降）。因此，诸如 $X_2 \leqslant X_1$ 或 $X_2 \geqslant X_1$ 之类的比较也是有意义的。大多数经济变量都属于这一类。于是，问今年的 GDP 与去年的 GDP 相差多少是有意义的。

区间尺度 一个区间尺度变量满足比率尺度变量的后面两个性质，但不满足第一个性质。因此，两个时期之内的距离（如 2000—1995）是有意义的，但两个时期的比率（2000/1995）就没有意义。

序数尺度 只要一个变量满足比率尺度的第三个性质（即自然顺序），那它就属于这一类变量。例子有考试分数体系（A、B、C）或收入阶层（高、中、低）等。对于这些变量，自然顺序存在，但不同类别之间的差别不能量化。经济学的学生可能会想起两种商品之间的无差异曲线（indifference curves），虽然每条更高的无差异曲线标志着更高的效用水平，但你不能量化一条无差异曲线比另一条到底高多少。

名义尺度 此类变量不具备比率尺度变量的任何一个特征。诸如性别（男，女）和婚姻状况（已婚、未婚、离婚、分居）之类的变量只表示了不同的类别。**问题**：这类变量不能用比率尺度、区间尺度或名义尺度表示的原因是什么？

以后将看到，适合于比率尺度变量的计量经济方法可能不适合于名义尺度变量。因此，记住上面讨论的四类测量尺度之间的区别就很重要。

§1.8 要点与结论

1. 回归分析的主要用意，是分析一个叫做因变量的变量，对另一个或多个叫做解释变量的变量的统计依赖性。

2. 这种分析的目的，是要在解释变量的已知或固定值的基础上，估计和（或）预测因变量的均值。

3. 实际上，回归分析的成功有赖于适用数据的获得。本章讨论了研究者通常能获得的，特别是在社会科学中数据的性质，其来源和局限性。

4. 在任何一项研究中，研究者都应清楚地说明分析中所用数据的来源、定义、收集方法，以及数据中的任何差错或疏漏以及数据中的任何改动。须知，政府公布的宏观经济数据通常是有修改的。

32 5. 因为读者未必有这种时间、精力和资源去跟踪数据，所以他有权假定研究者所用的数据是适当地采集的，并且计算和分析也都是正确的。

习 题

1.1 表 1.2 给出了 7 个工业化国家的消费者价格指数数据，以 1982—1984 年为指数的基期并令 1982－1984＝100。

表 1.2　7 个工业化国家 1973—1997 年间的 CPI（1982－1984＝100）

年份	加拿大	法国	德国	意大利	日本	英国	美国
1973	40.800 00	34.600 00	62.800 00	20.600 00	47.900 00	27.900 00	44.400 00
1974	45.200 00	39.300 00	67.100 00	24.600 00	59.000 00	32.300 00	49.300 00
1975	50.100 00	43.900 00	71.100 00	28.800 00	65.900 00	40.200 00	53.800 00
1976	53.900 00	48.100 00	74.200 00	33.600 00	72.200 00	46.800 00	56.900 00
1977	58.100 00	52.700 00	76.900 00	40.100 00	78.100 00	54.200 00	60.600 00
1978	63.300 00	57.500 00	79.000 00	45.100 00	81.400 00	58.700 00	65.200 00
1979	69.200 00	63.600 00	82.200 00	52.100 00	84.400 00	66.600 00	72.600 00
1980	76.100 00	72.300 00	86.700 00	63.200 00	90.900 00	78.500 00	82.400 00
1981	85.600 00	81.900 00	92.200 00	75.400 00	95.300 00	87.900 00	90.900 00
1982	94.900 00	91.700 00	97.100 00	87.700 00	98.100 00	95.400 00	96.500 00
1983	100.400 0	100.400 0	100.300 0	100.800 0	99.800 00	99.800 00	99.600 00
1984	104.700 0	108.100 0	102.700 0	111.500 0	102.100 0	104.800 0	103.900 0
1985	109.000 0	114.400 0	104.800 0	121.100 0	104.100 0	111.100 0	107.600 0
1986	113.500 0	117.300 0	104.700 0	128.500 0	104.800 0	114.900 0	109.600 0
1987	118.400 0	121.100 0	104.900 0	134.400 0	104.800 0	119.700 0	113.600 0
1988	123.200 0	124.400 0	106.300 0	141.100 0	105.600 0	125.600 0	118.300 0
1989	129.300 0	128.700 0	109.200 0	150.400 0	108.100 0	135.300 0	124.000 0
1990	135.500 0	133.000 0	112.200 0	159.600 0	111.400 0	148.200 0	130.700 0
1991	143.100 0	137.200 0	116.300 0	169.800 0	115.000 0	156.900 0	136.200 0
1992	145.300 0	140.500 0	122.100 0	178.800 0	116.900 0	162.700 0	140.300 0
1993	147.900 0	143.500 0	127.600 0	186.400 0	118.400 0	165.300 0	144.500 0
1994	148.200 0	145.800 0	131.100 0	193.700 0	119.300 0	169.400 0	148.200 0
1995	151.400 0	148.400 0	133.500 0	204.100 0	119.100 0	175.100 0	152.400 0
1996	153.800 0	151.400 0	135.500 0	212.000 0	119.300 0	179.400 0	156.900 0
1997	156.300 0	153.200 0	137.800 0	215.700 0	121.300 0	185.000 0	160.500 0

a. 利用所给数据计算每个国家的通货膨胀率。[1]

b. 对每个国家按时间顺序描出其通货膨胀率（即以时间为横轴，以通货膨胀率为纵轴）。

c. 你从这 7 个国家的通货膨胀经历中能得出什么宽泛的结论?

d. 哪个国家的通货膨胀率变动性最大? 你能给出什么样的解释呢?

33 1.2 a. 将加拿大、法国、德国、意大利、日本和英国的通货膨胀率相对美国的通货膨胀率画出散点图。

b. 一般性地评论这六个国家的通货膨胀相对美国通货膨胀的表现。

c. 如果你发现这六个国家的通货膨胀率与美国的通货膨胀率同向变化，那是否表明美国的通货膨胀"导致"了其他国家的通货膨胀? 为什么是或为什么不是?

1.3 表 1.3 给出了7个工业化国家1977—1998 年间的外汇汇率数据。除英国外，汇率都定义为一美元兑换外币的数量；而英国的汇率定义为一英镑兑换美元的数量。

a. 画出这些汇率相对于时间的散点图，并评论汇率在给定期间内的一般表现。

b. 如果一美元能买到更多单位的外币，则称之为美元**升值** (appreciate)。相反，如果一美元只能买更少的外币，则称之为美元**贬值** (depreciate)。在 1977—1998 年间，美元的一般表现如何? 顺便查阅一本宏观经济学或国际经济学教科书，以探明是哪些因素决定了货币的升值或贬值。

表 1.3 **七国汇率：1977—1998 年**

年份	加拿大	法国	德国	日本	瑞典	瑞士	英国
1977	1.063 300	4.916 100	2.323 600	268.620 0	4.480 200	2.406 500	1.744 900
1978	1.140 500	4.509 100	2.009 700	210.390 0	4.520 700	1.790 700	1.918 400
1979	1.171 300	4.256 700	1.834 300	219.020 0	4.289 300	1.664 400	2.122 400
1980	1.169 300	4.225 100	1.817 500	226.630 0	4.231 000	1.677 200	2.324 600
1981	1.199 000	5.439 700	2.263 200	220.630 0	5.066 000	1.967 500	2.024 300
1982	1.234 400	6.579 400	2.428 100	249.060 0	6.283 900	2.032 700	1.748 000
1983	1.232 500	7.620 400	2.553 900	237.550 0	7.671 800	2.100 700	1.515 900
1984	1.295 200	8.735 600	2.845 500	237.460 0	8.270 800	2.350 000	1.336 800
1985	1.365 900	8.980 000	2.942 000	238.470 0	8.603 200	2.455 200	1.297 400
1986	1.389 600	6.925 700	2.170 500	168.350 0	7.127 300	1.797 900	1.467 700
1987	1.325 900	6.012 200	1.798 100	144.600 0	6.346 900	1.491 800	1.639 800
1988	1.230 600	5.959 500	1.757 000	128.170 0	6.137 000	1.464 300	1.781 300
1989	1.184 200	6.380 200	1.880 800	138.070 0	6.455 900	1.636 900	1.638 200
1990	1.166 800	5.446 700	1.616 600	145.000 0	5.923 100	1.390 100	1.784 100

1991	1.146 000	5.646 800	1.661 000	134.590 0	6.052 100	1.435 600	1.767 400
1992	1.208 500	5.293 500	1.561 800	126.780 0	5.825 800	1.406 400	1.766 300
1993	1.290 200	5.666 900	1.654 500	111.080 0	7.795 600	1.478 100	1.501 600
1994	1.366 400	5.545 900	1.621 600	102.180 0	7.716 100	1.366 700	1.531 900
1995	1.372 500	4.986 400	1.432 100	93.960 00	7.140 600	1.181 200	1.578 500
1996	1.363 800	5.115 800	1.504 900	108.780 0	6.708 200	1.236 100	1.560 700
1997	1.384 900	5.839 300	1.734 800	121.060 0	7.644 600	1.451 400	1.637 600
1998	1.483 600	5.899 500	1.759 700	130.990 0	7.952 200	1.450 600	1.657 300

资料来源：*Economic Report of the President*, January 2000 and January 2001.

34 1.4 图 1.5 背后的 M1 货币供给数据由表 1.4 给出。你能给出货币供给在表中所示时期上升的原因吗？

表 1.4　季节调整的 M1 供给：1959 年 1 月—1999 年 9 月（10 亿美元）

1959：01	138.890 0	139.390 0	139.740 0	139.690 0	140.680 0	141.170 0
1959：07	141.700 0	141.900 0	141.010 0	140.470 0	140.380 0	139.950 0
1960：01	139.980 0	139.870 0	139.750 0	139.560 0	139.610 0	139.580 0
1960：07	140.180 0	141.310 0	141.180 0	140.920 0	140.860 0	140.690 0
1961：01	141.060 0	141.600 0	141.870 0	142.130 0	142.660 0	142.880 0
1961：07	142.920 0	143.490 0	143.780 0	144.140 0	144.760 0	145.200 0
1962：01	145.240 0	145.660 0	145.960 0	146.400 0	146.840 0	146.580 0
1962：07	146.460 0	146.570 0	146.300 0	146.710 0	147.290 0	147.820 0
1963：01	148.260 0	148.900 0	149.170 0	149.700 0	150.390 0	150.430 0
1963：07	151.340 0	151.780 0	151.980 0	152.550 0	153.650 0	153.290 0
1964：01	153.740 0	154.310 0	154.480 0	154.770 0	155.330 0	155.620 0
1964：07	156.800 0	157.820 0	158.750 0	159.240 0	159.960 0	160.300 0
1965：01	160.710 0	160.940 0	161.470 0	162.030 0	161.700 0	162.190 0
1965：07	163.050 0	163.680 0	164.850 0	165.970 0	166.710 0	167.850 0
1966：01	169.080 0	169.620 0	170.510 0	171.810 0	171.330 0	171.570 0
1966：07	170.310 0	170.810 0	171.970 0	171.160 0	171.380 0	172.030 0
1967：01	171.860 0	172.990 0	174.810 0	174.170 0	175.680 0	177.020 0
1967：07	178.130 0	179.710 0	180.680 0	181.640 0	182.380 0	183.260 0
1968：01	184.330 0	184.710 0	185.470 0	186.600 0	187.990 0	189.420 0
1968：07	190.490 0	191.840 0	192.740 0	194.020 0	196.020 0	197.410 0
1969：01	198.690 0	199.350 0	200.020 0	200.710 0	200.810 0	201.270 0

1969: 07	201.660 0	201.730 0	202.100 0	202.900 0	203.570 0	203.880 0
1970: 01	206.220 0	205.000 0	205.750 0	206.720 0	207.220 0	207.540 0
1970: 07	207.980 0	209.930 0	211.800 0	212.880 0	213.660 0	214.410 0
1971: 01	215.540 0	217.420 0	218.770 0	220.000 0	222.020 0	223.450 0
1971: 07	224.850 0	225.580 0	226.470 0	227.160 0	227.760 0	228.320 0
1972: 01	230.090 0	232.320 0	234.300 0	235.580 0	235.890 0	236.620 0
1972: 07	238.790 0	240.930 0	243.180 0	245.020 0	246.410 0	249.250 0
1973: 01	251.470 0	252.150 0	251.670 0	252.740 0	254.890 0	256.690 0
1973: 07	257.540 0	257.760 0	257.860 0	259.040 0	260.980 0	262.880 0
1974: 01	263.760 0	265.310 0	266.680 0	267.200 0	267.560 0	268.440 0
1974: 07	269.270 0	270.120 0	271.050 0	272.350 0	273.710 0	274.200 0
1975: 01	273.900 0	275.000 0	276.420 0	276.170 0	279.200 0	282.430 0
1975: 07	283.680 0	284.150 0	285.690 0	285.390 0	286.830 0	287.070 0
1976: 01	288.420 0	290.760 0	292.700 0	294.660 0	295.930 0	296.160 0
1976: 07	297.200 0	299.050 0	299.670 0	302.040 0	303.590 0	306.250 0
1977: 01	308.260 0	311.540 0	313.940 0	316.020 0	317.190 0	318.710 0
1977: 07	320.190 0	322.270 0	324.480 0	326.400 0	328.640 0	330.870 0
1978: 01	334.400 0	335.300 0	336.960 0	339.920 0	344.860 0	346.800 0
1978: 07	347.630 0	349.660 0	352.260 0	353.350 0	355.410 0	357.280 0
1979: 01	358.600 0	359.910 0	362.450 0	368.050 0	369.590 0	373.340 0
1979: 07	377.210 0	378.820 0	379.280 0	380.870 0	380.810 0	381.770 0
1980: 01	385.850 0	389.700 0	388.130 0	383.440 0	384.600 0	389.460 0
1980: 07	394.910 0	400.060 0	405.360 0	409.060 0	410.370 0	408.060 0
1981: 01	410.830 0	414.380 0	418.690 0	427.060 0	424.430 0	425.500 0
1981: 07	427.900 0	427.850 0	427.460 0	428.450 0	430.880 0	436.170 0
1982: 01	442.130 0	441.490 0	442.370 0	446.780 0	446.530 0	447.890 0
1982: 07	449.090 0	452.490 0	457.500 0	464.570 0	471.120 0	474.300 0
1983: 01	476.680 0	483.850 0	490.180 0	492.770 0	499.780 0	504.350 0
1983: 07	508.960 0	511.600 0	513.410 0	517.210 0	518.530 0	520.790 0
1984: 01	524.400 0	526.990 0	530.780 0	534.030 0	536.590 0	540.540 0
1984: 07	542.130 0	542.390 0	543.860 0	543.870 0	547.320 0	551.190 0
1985: 01	555.660 0	562.480 0	565.740 0	569.550 0	575.070 0	583.170 0
1985: 07	590.820 0	598.060 0	604.470 0	607.910 0	611.830 0	619.360 0
1986: 01	620.400 0	624.140 0	632.810 0	640.350 0	652.010 0	661.520 0
1986: 07	672.200 0	680.770 0	688.510 0	695.260 0	705.240 0	724.280 0
1987: 01	729.340 0	729.840 0	733.010 0	743.390 0	746.000 0	743.720 0
1987: 07	744.960 0	746.960 0	748.660 0	756.500 0	752.830 0	749.680 0

1988：01	755.550 0	757.070 0	761.180 0	767.570 0	771.680 0	779.100 0
1988：07	783.400 0	785.080 0	784.820 0	783.630 0	784.460 0	786.260 0
1989：01	784.920 0	783.400 0	782.740 0	778.820 0	774.790 0	774.220 0
1989：07	779.710 0	781.140 0	782.200 0	787.050 0	787.950 0	792.570 0
1990：01	794.930 0	797.650 0	801.250 0	806.240 0	804.360 0	810.330 0
1990：07	811.800 0	817.850 0	821.830 0	820.300 0	822.060 0	824.560 0
1991：01	826.730 0	832.400 0	838.620 0	842.730 0	848.960 0	858.330 0
1991：07	862.950 0	868.650 0	871.560 0	878.400 0	887.950 0	896.700 0
1992：01	910.490 0	925.130 0	936.000 0	943.890 0	950.780 0	954.710 0
1992：07	964.600 0	975.710 0	988.840 0	1 004.340	1 016.040	1 024.450
1993：01	1 030.900	1 033.150	1 037.990	1 047.470	1 066.220	1 075.610
1993：07	1 085.880	1 095.560	1 105.430	1 113.800	1 123.900	1 129.310
1994：01	1 132.200	1 136.130	1 139.910	1 141.420	1 142.850	1 145.650
1994：07	1 151.490	1 151.390	1 152.440	1 150.410	1 150.440	1 149.750
1995：01	1 150.640	1 146.740	1 146.520	1 149.480	1 144.650	1 144.240
1995：07	1 146.500	1 146.100	1 142.270	1 136.430	1 133.550	1 126.730
1996：01	1 122.580	1 117.530	1 122.590	1 124.520	1 116.300	1 115.470
1996：07	1 112.340	1 102.180	1 095.610	1 082.560	1 080.490	1 081.340
1997：01	1 080.520	1 076.200	1 072.420	1 067.450	1 063.370	1 065.990
1997：07	1 067.570	1 072.080	1 064.820	1 062.060	1 067.530	1 074.870
1998：01	1 073.810	1 076.020	1 080.650	1 082.090	1 078.170	1 077.780
1998：07	1 075.370	1 072.210	1 074.650	1 080.400	1 088.960	1 093.350
1999：01	1 091.000	1 092.650	1 102.010	1 108.400	1 104.750	1 101.110
1999：07	1 099.530	1 102.400	1 093.460			

资料来源：Board of Governors，Federal Reserve Bank，USA.

35 **1.5** 假设你要做一个犯罪活动的经济模型，比方说花在犯罪活动（如卖毒品）上的小时数。在做这样的一个模型时你要考虑哪些变量？看一下你的模型能否与诺贝尔经济学奖得主加里·贝克尔（Gary Becker）的模型相媲美。[17]

1.6 经济学中的控制试验：2000 年 4 月 7 日，克林顿总统签署了一项参众两院同时通过的法案，取消对社会保障金领取者的收入限制。此前，年龄介于 65 岁和 69 岁之间的受济者，年收入超过 1.7 万美元者，超出的部分按每 3 美元减少 1 美元的社会保障救济金。你如何设计一个研究方案来分析这项法律修改的影响？注：原有法律对 70 岁以上的受济者没有设定收入限制。

36 **1.7** 表1.5 中的数据发表在 1984 年 3 月 1 日的《华尔街日报》上。它将 1983 年 21 家企业的广告预算（以百万美元计）与这些企业产

品的参观者每周保留的印象次数（以百万次计）相联系。这些数据基于对4 000个成人的调查，在调查中要求产品使用者列出一条在过去一周里见过的该类产品的商业广告。

a. 以印象数为纵轴、以广告支出为横轴画散点图。

b. 你认为这两个变量之间的关系具有什么样的性质？

c. 看一下你的图，你认为值得做广告吗？想想那些出现在星期天的超级杯赛上和世界职业棒球锦标赛期间的商业广告。

注：我们在以后的章节中将进一步探讨表 1.5 中给出的数据。

表 1.5　广告支出的影响

企业	印象（百万次）	支出（以 1983 年的百万美元计）
1. 美乐	32.1	50.1
2. 百事	99.6	74.1
3. 金鹰	11.7	19.3
4. 联邦快递	21.9	22.9
5. 汉堡王	60.8	82.4
6. 可口可乐	78.6	40.1
7. 麦当劳	92.4	185.9
8. 前世通公司	50.7	26.9
9. 健怡可乐	21.4	20.4
10. 福特	40.1	166.2
11. 利维	40.8	27.0
12. 百威	10.4	45.6
13. 贝尔	88.9	154.9
14.CK	12.0	5.0
15. 温迪快餐	29.2	49.7
16. 宝丽来	38.0	26.9
17.Shasta	10.0	5.7
18.Meow Mix	12.3	7.6
19. 卡夫食品	23.4	9.2
20. 佳洁士	71.1	32.4
21.Kibbles' N Bits	4.4	6.1

资料来源：http://lib.stat.cmu.edu/DASL/Datafiles/tvadsdat.html.

【习题注释】

[1] 将当年的 CPI 减去上一年度的 CPI 后，再除以上一年度的 CPI，然后乘以 100 即可得到通货膨胀。因此，加拿大 1974 年的通货膨胀率就是[(45.2－40.8)/40.8]×100＝10.78%(近似)。

[2] G.S.Becker, "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, vol.76, 1968, pp.169－217.

【注释】

[1] Francis Galton, "Family Likeness in Stature", *Proceedings of Royal Society, London*, vol.40, 1886, pp.42－72.

[2] K.Pearson and A.Lee, "On the Laws of Inheritance", *Biometrika*, vol.2, Nov.1903, pp.357－462.

[3] 在此主题讨论的现阶段，我们干脆把这根回归线叫做因变量（儿高）与解释变量（父高）之间的平均关系线。注意，此线有一正的斜率；但此斜率小于 1，这和加尔顿的"回归到中等"相一致。(为什么?)

[4] "stochastic"（随机）一词来自希腊语 stokhos，意为"公牛的眼睛"。把矛抛向一块盾牌的结果是一个随机过程，即一个带有多少击不中的过程。

[5] M.G.Kendall and A.Stuart, *The Advanced Theory of Statistics*, Charles Griffin Publishers, New York, 1961.vol.2, chap.26, p.279.

[6] 但在第 3 章中我们将看到，经典回归分析中假定用于分析的模型是正确的。因此，在所设的模型中，隐含着因果的流向。

[7] 须知，解释变量可能本来就是随机的。但为了回归分析的目的，我们假定它们的值在重复抽样中固定不变（即 X 在不同的多个样本中取同样的一组值），从而把它们化成实效上是非随机的。对此第 3 章 3.2 节有更多的讨论。

[8] 在高级的计量经济学教材中，解释变量非随机的这一假定可以去掉（见第 2 篇篇首语）。

[9] 正式定义和更多的细节见附录 A。

[10] 一个富于信息的叙述，见于 Michael D.Intriligator, *Econometric Models, Techniques, and Applications*, Prentice Hall, Englewood Cliffs, N.J., 1978, chap.3。

[11] 为了更清楚地看出这一点，我们把数据分为四个时期：1951 年 1 月—1962 年 12 月；1963 年 1 月—1974 年 12 月；1975 年 1 月—1986 年 12 月和 1987 年 1 月—1999 年 9 月。对这些子期间，货币供给的均值（及括号中相应的标准差）分别是 165.88 (23.27)，323.20 (72.66)，788.12 (195.43) 和 1 099 (27.84)，所有数字都以 10 亿美元为单位。这大致表明了货币供给在整个期间非平稳的事实。

[12] 为看到一种明确的叙述，可参考 Albert T.Somers, *The U.S.Economy Demystified: What the Major Economic Statistics Mean and their Significance for Business*, D.C.Heath, Lexington, Mass., 1985。

[13] 人们在社会科学中有时也能做控制试验，习题 1.6 就给出了一个例子。

[14] O.Morgenstern, *The Accuracy of Economic Observations*, 2nd ed., Princeton University Press, Princeton, N.J., 1963，该书提有尖锐的意见。

[15] 以下讨论很大程度上依据于 Aris Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, New York, 1999, p.24。

第 2 章　双变量回归分析：一些基本概念

37 在第 1 章中我们概括地讨论了回归的概念。本章中我们将正式地继续探讨这一主题。具体地说，本章和随后的两章，我们将向读者介绍最为简单的双变量回归分析所依据的理论，其中因变量（回归子）仅与惟一的一个解释变量（回归元）相关。我们首先考虑双变量情形，并不是因为它有实用上的适宜性，而是因为它能使回归分析的基本概念表述得尽可能简单，而且，某些概念还能借助于二维图形来说明。不仅如此，我们还将看到，更为一般的多重回归分析，在许多方面都是双变量情形的逻辑推广。

§2.1　一个人为的例子[1]

如第 1.2 节所指出的，回归分析大体上说，是要根据解释变量的已知或给定值，去估计和（或）预测因变量的（总体）均值。[2]为了弄懂怎样才能
38 做到这一点，观察表 2.1 中的数据。

表中数据指的是在一个假想的经济社会中，构成总体的 60 个家庭及其周收入（X）和周消费支出（Y）的美元数量。这 60 个家庭被分成 10 个收入组（从 80 美元到 260 美元），各组中每个家庭的周支出都列在表中。因

此，我们就有 10 个固定的 X 值和与每个 X 相对应的 Y 值；可以说，有 10 个 Y 的子总体。

表 2.1 **周家庭收入 X，美元**

$Y\downarrow$ $X\rightarrow$	80	100	120	140	160	180	200	220	240	260
	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
周家庭	65	74	90	95	110	120	140	140	155	175
消费支出	70	80	94	103	116	130	144	152	165	178
Y 美元	75	85	98	108	118	135	145	157	175	180
		88	—	113	125	140	—	160	189	185
	—	—	—	115	—		—	162	—	191
共计	325	462	445	707	678	750	685	1 043	966	1 211
Y 的条件均值 $E(Y/X)$	65	77	89	101	113	125	137	149	161	173

39 从图 2.1 可以清楚地看出，每个收入组的周消费支出都有相当大的变化。但人们一般得到的图是，尽管每个收入组中的周消费支出可以变化，但是平均来讲，周消费支出随着收入的上升而增加。为了清楚地看出这一点，我们在表 2.1 中已经给出了与 10 个收入水平中的每一个相对应的平均周消费支出或周消费支出的均值。于是，对应于 80 美元的周收入水平，平均消费支出是 65 美元，而对应于 200 美元的收入水平的支出则是 137 美元。对 Y 的 10 个子总体，我们共有 10 个均值。我们称这些均值为**条件期望值**(conditional expected values)，因为它们取决于(条件)变量 X 的给定值。我们用符号表示为 $E(Y|X)$，读作"给定 X 值下 Y 的期望值"(也可参见表 2.2)。

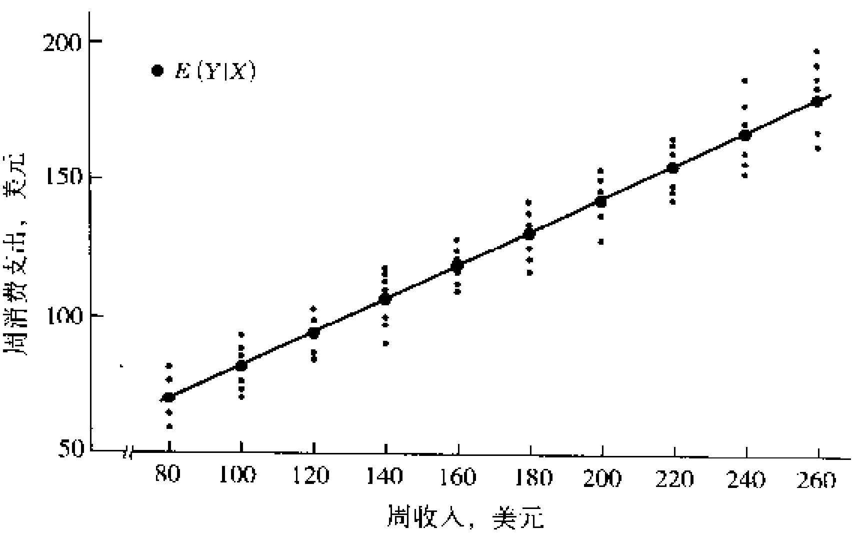


图 2.1 不同收入水平下支出的条件分布（表 2.1 的数据）

表 2.2　　　与表 2.1 的数据相对应的条件概率 $P(Y|X_i)$

$P(Y\|X)$ ↓ \ X →	80	100	120	140	160	180	200	220	240	260
条件概率 $P(Y\|X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	–	$\frac{1}{6}$	–	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	–	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	–	–	–	$\frac{1}{7}$	–	–	–	$\frac{1}{7}$		$\frac{1}{7}$
Y 的条件均值	65	77	89	101	113	125	137	149	161	173

于是，在几何意义上，总体回归曲线就是解释变量取给定值时因变量的条件均值或期望值的轨迹。将周消费支出的这些条件期望值与无条件期望值 $E(Y)$ 区别开来至关重要。如果我们将总体中所有 60 个家庭的周消费支出都加起来，再将得到的总和除以 60，则得到的数字 121.20（7 272/60）美元就是周消费支出的无条件均值或期望值 $E(Y)$；我们在得到这个数字时并不关心各个家庭的收入水平，从这个意义上讲，它是无条件的。[3] 显然，表 2.1 中给出的 Y 的各个条件期望值都不同于 Y 的无条件期望值 121.20 美元。当我们问"一个家庭周消费支出的期望值是多少"时，我们得到的回答是 121.20 美元（无条件均值）。但如果我们问"一个月收入为 140 美元的家
40 庭的周消费支出的期望值是多少"时，我们得到的回答是 101 美元（条件均值）。换言之，如果我们问"对一个周收入 140 美元的家庭的周消费的最佳（均值）预测是多少"，回答将是 101 美元。因此，对收入水平的了解能使我们相对于在不了解这些时更好地预测消费支出的均值。[4] 如我们将在本书中讨论的那样，这可能正是回归分析的本质。

图 2.1 中的黑圆点表示了不同 X 值下 Y 的条件均值。将这些条件均值连起来，就得到所谓的**总体回归线**（population regression line, PRL）或更一般地称为**总体回归曲线**（population regression curve）。[5] 更简单地说，就是 **Y 对 X 的回归**（regression of Y on X）。形容词"总体"源于以下事实：我们一直在用 60 个家庭的整个总体来讨论这个例子。当然，在现实中，一个总体

可能有许多个家庭。更简单地说，对应于回归元 X 的给定值都有 Y 的一个子总体，连接这些子总体的均值就得到了总体回归曲线。它可以画成图 2.2 的形状。

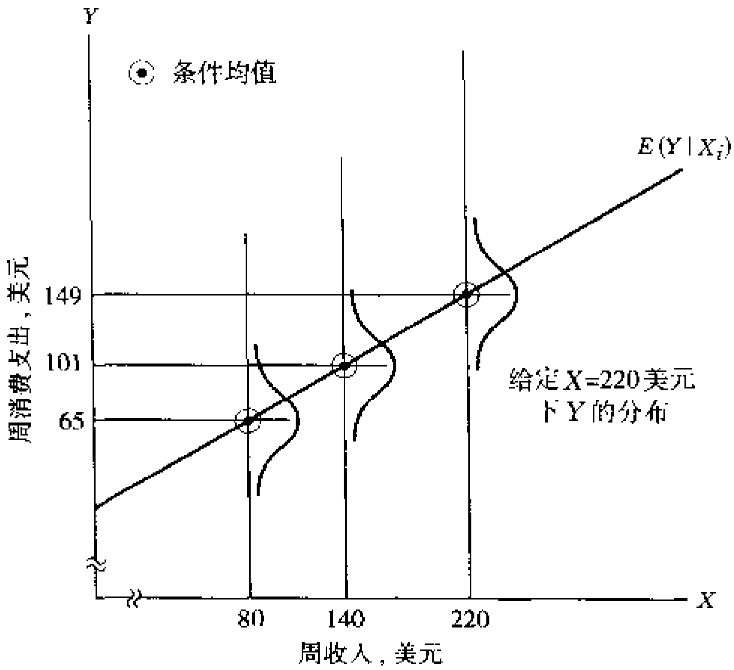


图 2.2　总体回归线（表 2.1 的数据）

41 此图表明，对于每个 X（即收入水平），都有周消费支出 Y 值的一个总体，这些 Y 值分布在其（条件）均值的左右。为简便起见，我们假定这些 Y 值对称地分布在其相应（条件）均值周围，并且回归线（或曲线）穿过这些（条件）均值。

以此为背景，读者可能发现重温第 1.2 节中对回归的定义会有所裨益。

§2.2　总体回归函数的概念

从前面的讨论中，特别是从图 2.1 和图 2.2，我们清楚地看到，每一条件均值 $E(Y|X_i)$ 都是 X_i 的一个函数，用符号表示：

$$E(Y|X_i)=f(X_i) \tag{2.2.1}$$

其中 $f(X_i)$ 表示解释变量 X_i 的某个函数。在我们的人为的例子中，$E(Y|X_i)$ 是 X_i 的一个线性函数。方程（2.2.1）被称为**条件期望函数**（conditional expectation function, CEF）或**总体回归函数**（PRF）或简称**总体回归**（PR）。它仅仅表明在给定 X_i 下 Y 的分布的（总体）均值与 X_i 存在着函数关系。简言之，它说出了 Y 的均值或平均对应值是怎样地随 X 而变化的。

函数 $f(X_i)$ 采取什么形式？这是一个重要的问题，因为在实际情况中

我们不会得到全部总体值用来做分析研究。因此，PRF 的函数形式是一个经验方面的问题。对一些特殊情形，理论也许能告诉我们一点什么。例如，一位经济学家可能提出消费支出与收入有线性关系。作为一个初次逼近或一个暂行的假设，他假定，比如说，PRF $E(Y \mid X_i)$ 是 X_i 的线性函数，其形式是：

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i \tag{2.2.2}$$

其中 β_1 和 β_2 为未知但却固定的参数，称为**回归系数**（regression coefficients）；β_1 和 β_2 也分别称为**截距**（intercept）和**斜率系数**（slope coefficients）。方程（2.2.1）本身则称为**线性总体回归函数**或简称线性总体回归。一些文献中曾用过的一些其他术语还有线性总体回归模型或线性总体回归方程。在本教材中，名词**回归**、**回归方程**和**回归模型**将不加区别地同义地使用。

在回归分析中，我们的兴趣在于估计像方程式（2.2.2）那样的 PRF。就是说，根据对 Y 和 X 的观测值估计未知数 β_1 和 β_2 的值。这个问题将在第 3 章中详细研究。

§2.3 "线性"一词的含义

42 由于本教材主要讨论像方程式（2.2.2）那样的线性模型，因而我们必须知道线性一词的真正含义，因为对它可以有两种解释。

对变量为线性

对线性的第一种并且也许是更"自然"的一种解释是，Y 的条件期望值是 X_i 的线性函数，比如说，如同方程式（2.2.2）那样。[6] 从几何意义上说，这时回归曲线是一条直线。同样按照这种解释，诸如 $E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$ 的回归函数，由于变量 X 以幂或指数 2 出现，就不是线性的。

对参数为线性

对线性的第二种解释是，Y 的条件期望 $E(Y \mid X_i)$ 是诸参数如 β_1，β_2 的一个线性函数；它可以是或不是变量 X 的线性函数。[7] 对于这种解释，$E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$ 就算是一个线性（对于参数）回归模型。为了看出这一点，让我们假设 X 取值为 3。因此 $E(Y \mid X) = \beta_1 + 9\beta_2$，显然它是 β_1 和 β_2 的线性函数。图 2.3 中所示的所有模型也都是线性回归模型，即对于参数是线性的模型。

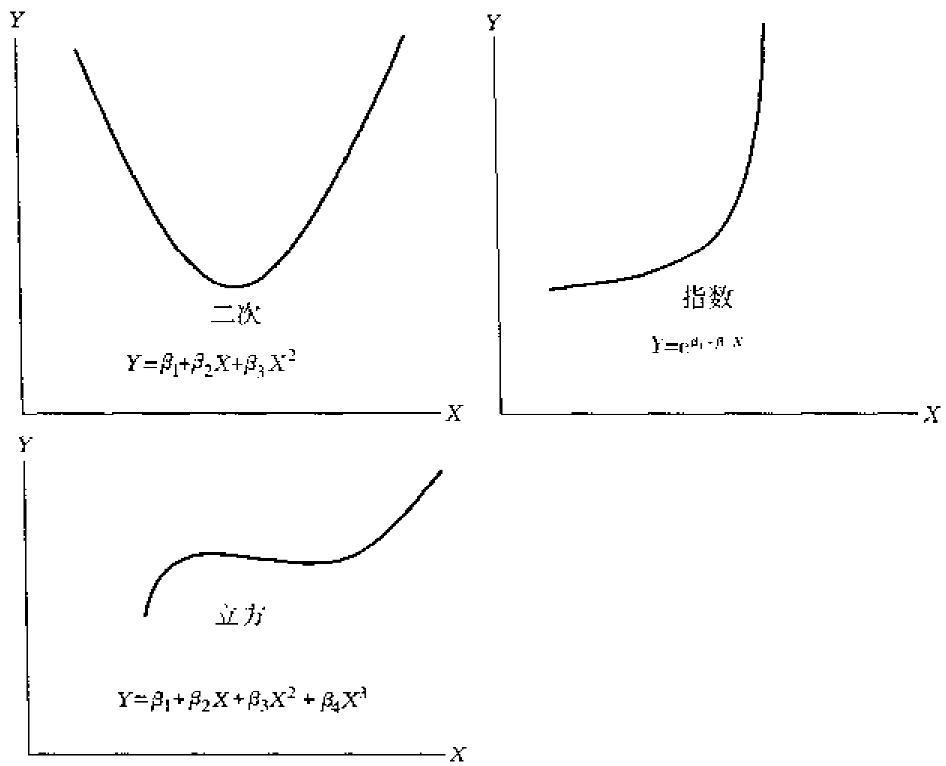


图 2.3　对于参数为线性的函数

现在考虑模型 $E(Y|X_i)=\beta_1+\beta_2^2X_i$。假设 $X=3$，我们则得到 $E(Y|X_i)=\beta_1+3\beta_2^2$，它显然不是 β_2 的线性函数。上述模型正是**非线性（对于参数）回归模型**（nonlinear regression model，in the parameter）。我们将在第 14 章讨论这种模型。

在两种线性的解释中，对于下面即将展开讨论的回归理论说，主要考虑的是参数为线性的情形。因此，从现在起，“线性”回归一词总是指对参数 β 为线性的一种回归（即参数只以它的 1 次方出现）；对解释变量 X 则可以是或不是线性的。把上面的讨论排成表格的形式，我们得到表 2.3。这样，$E(Y|X_i)=\beta_1+\beta_2X_i$ 兼对参数和变量为线性，是一个线性回归模型（LRM），而对参数为线性但对变量 X 为非线性的 $E(Y|X_i)=\beta_1+\beta_2X_i^2$，也是一个 LRM。

43 **表 2.3**　　**线性回归模型**

模型对参数为线性？	模型对变量为线性？	
	是	不是
是	LRM	LRM
不是	NLRM	NLRM

注：LRM = 线性回归模型，　NLRM = 非线性回归模型。

§2.4 PRF 的随机设定

从图 2.1 清楚地看到，随着家庭收入的增加，家庭消费支出平均地说也增加。但是，对某一个别家庭来说，消费支出和它的（固定）收入水平的关系会是怎样？从表 2.1 和图 2.1 中明显看出，某一个别家庭的消费支出不一定随收入水平的增加而增加。例如，从表 2.1 中我们观察到，对应于每周 100 美元的收入水平，有一户人家的消费支出是 65 美元，少于每周收入仅为 80 美元的两户人家的消费支出（70 美元和 75 美元）。但应看到，家庭每
44 周收入为 100 美元的平均消费支出比家庭每周收入为 80 美元的平均消费支出大（77 美元比 65 美元大）。

那么，个别家庭的消费支出与给定收入水平之间能有什么关系呢？我们从图 2.1 看到，给定收入水平 X_i 的个别家庭的消费支出聚集在收入为 X_i 的所有家庭的平均消费支出的周围，也就是围绕着它的条件均值。因此，我们可以把个别的 Y_i 围绕它的期望值的离差（deviation）表述如下：

$$u_i = Y_i - E(Y \mid X_i)$$

或

$$Y_i = E(Y \mid X_i) + u_i \tag{2.4.1}$$

其中离差 u_i 是一个不可观测的可正可负的随机变量，在专业术语中，把 u_i 称为**随机干扰**（stochastic disturbance）或**随机误差**（stochastic error）项。

怎样解释方程式（2.4.1）呢？我们可以说，给定 X 水平，个别家庭的支出可表示为两个成分之和：(1) $E(Y \mid X_i)$ 代表相同收入水平的所有家庭的平均消费支出。这一成分称之为**系统性**或**确定性**成分；(2) u_i 为随机或**非系统性**成分。我们很快就要分析这个随机误差项的性质。但现在假定它是所有可能影响 Y，但又未能包括到回归模型中来的被忽略变量的替代（surrogate）或代理（proxy）变量。

假定 $E(Y \mid X_i)$ 是对 X_i 为线性的，好比方程式（2.2.2）那样，方程（2.4.1）就可写为：

$$\begin{aligned} Y_i &= E(Y \mid X_i) + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned} \tag{2.4.2}$$

方程（2.4.2）拟定，一个家庭的消费支出，线性地依赖于它的收入和干扰项。例如，给定 $X = 80$，各家庭的消费支出可表达为：

$$Y_1 = 55 = \beta_1 + \beta_2(80) + u_1$$
$$Y_2 = 60 = \beta_1 + \beta_2(80) + u_2$$
$$Y_3 = 65 = \beta_1 + \beta_2(80) + u_3$$

$$
\begin{aligned}
Y_4 &= 70 = \beta_1 + \beta_2(80) + u_4 \qquad (2.4.3)\\
Y_5 &= 75 = \beta_1 + \beta_2(80) + u_5
\end{aligned}
$$

45 现在，如果在方程式（2.4.1）的两边取期望值，就得到

$$
\begin{aligned}
E(Y_i \mid X_i) &= E[E(Y \mid X_i)] + E(u_i \mid X_i)\\
&= E(Y \mid X_i) + E(u_i \mid X_i) \qquad (2.4.4)
\end{aligned}
$$

这里我们利用了常数的期望值就是它本身这一事实。[8]要仔细看清，在方程式（2.4.4）中我们取的是以给定的 X 值为条件的条件期望。

因为$E(Y_i \mid X_i)$就是 $E(Y \mid X_i)$，故可从方程（2.4.4）推出：

$$E(u_i \mid X_i) = 0 \qquad (2.4.5)$$

因此，假定回归线通过 Y 的条件均值（见图 2.2），就意味着 u_i 的条件均值（以给定的诸 X_i 为条件）为零。

从前面的讨论可以看出，如果$E(u_i \mid X_i) = 0$，则方程式（2.2.2）和方程式（2.4.2）是等价的。[9]但是方程式（2.4.2）有它的优点，因为它清楚地表明，除收入外，还有影响着消费支出的其他变量，因此并不能单凭回归模型中含有的（1 个或多个）变量就能完全解释个别家庭的消费支出。

§2.5　随机干扰项的意义

如 2.4 节所指出，干扰项是从模型中省略下来的而又集体地影响着 Y 的全部变量的替代物。显然的问题是：为什么不把这些变量明显地引进到模型中来？换句话说，为什么不构造一个含有尽可能多个变量的复回归模型？理由是多方面的。

1. 理论的模糊性。即使有决定 Y 的行为的理论，也许是，而且常常是不完全的。我们可以肯定每周收入 X 影响每周消费支出 Y。但还有什么其他影响 Y 的变量则不是无所知就是知道而不确定。因此不妨用 u_i 作为模型所排除或忽略的全部变量的替代变量。

2. 数据的欠缺。即使我们明知被忽略变量中的一些变量，并因而考虑用一个复回归而不是一个简单回归，我们却不一定能得到关于这些变量的数
46 量信息。在经验研究中，人们得不到他们最想要的数据是司空见惯的事。例如，在原理上，除收入外，我们还可引进财富作为家庭消费支出的解释变量。但不幸的是，一般是得不到关于家庭财富的信息的。因此，我们不得不把财富变量从我们的模型中割舍掉。哪怕它在解释消费支出方面有很强的理论重要性。

3. 核心变量与周边变量。假定在我们的消费—收入例子中，除了收入 X_1 外，家庭的孩子数 X_2、性别 X_3、宗教 X_4、教育 X_5 和地区 X_6 也会影响到消费支出。但很可能这些变量的全部或其中的一些，合起来的影响也是非

常之小，充其量是一种非系统的或随机的影响。从实际出发以及从成本上计算，把它们一一引入模型是划不来的。所以人们希望把它们的联合效应当作一个随机变量来看待。[10]

4. 人类行为的内在随机性。即使我们成功地把所有有关的变量都引进到模型中来，在个别的 Y 中仍不免有一些“内在”的随机性，是无论我们花了多少力气都解释不了的。干扰项 u_i 也许能很好地反映这种随机性。

5. 糟糕的替代变量。虽然经典回归模型（将在第 3 章中讨论）假定变量 Y 和 X 能准确地观测，但实际上数据会受到测量误差的干扰。试看弗里德曼的著名的消费函数理论。[11]他把**永久消费**（Y^p）看做**永久收入**（X^p）的函数。但由于这些变量不可直接观测，故实际上我们利用替代变量，诸如可观测的当前消费（Y）和当前收入（X）。而由于所观测的 Y 和 X 未必等于 Y^p 和 X^p，这里就有一个测量误差的问题。这时干扰项 u 又可用来代表测量误差。如在以后的一章中我们将看到，如果有这种误差，回归系数 β 的估计会受到严重的影响。

6. 节省原则。仿效简单性原则（Occam's razor）[12]，我们想保持一个尽可能简单的回归模型。如果我们能用两个或三个变量就“基本上”解释了
47 Y 的行为，并且如果我们的理论完善或扎实的程度还没有达到足以提出可包含进来的其他变量，那么为什么要引进更多的变量？让 u_i 去代表所有的其他变量好了。当然，我们不应该只是为了保持回归模型简单而排除有关的和重要的变量。

7. 错误的函数形式。即使我们有了解释一种现象的在理论上正确的变量，并且我们能获得这些变量的数据，我们却常常不知道回归子和回归元之间的函数关系式是什么形式。消费支出是收入的线性（对变量而言）函数抑或非线性（对变量而言）函数？如果属于前者，$Y_i = \beta_1 + \beta_2 X_i + u_i$ 就是 Y 和 X 之间的适当函数关系式；但如果属于后者，$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$ 也许才是正确的函数形式。在双变量模型中，人们往往能通过散点图来判断关系式的函数形式。而在多变量回归模型中，由于无法从图形上想像一个多维的散点图，因此要决定适当的函数形式就不那么容易了。

因为所有的这些理由，我们随后将看到，随机干扰项在回归分析中扮演着极为重要的角色。

§2.6 样本回归函数

至今我们有意把讨论局限于与固定 X 值相对应的 Y 值总体，以避免考虑抽样的问题。（注：表 2.1 的数据代表总体，而不是一个样本。）但在大多数的实际情况中，我们仅有对应于某些固定 X 的 Y 值的一个样本。所以现在是面对抽样问题的时候了。我们现在的任务是要在样本信息的基础上估计 PRF。

作为一个说明，假定我们不知道表 2.1 的总体数据，我们仅有的信息是表 2.4 给出的对应于固定 X 值的 Y 值的一个随机（抽取的）样本。它和表 2.1 不同，对应于给定的每个 X_i 只有一个 Y 值。表 2.4 中的每个 Y 都是从表 2.1 的总体中对应于同一 X_i 的同组 Y 值随机抽取的。

表 2.4　　表 2.1 总体的一个随机样本

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

问题是：我们能通过表 2.4 的样本预测整个总体中对应于选定 X 的每周平均消费支出 Y 吗？换句话说，我们能通过这些样本数据估计 PRF 吗？读者一定怀疑，由于抽样波动，我们未必能“准确”估计 PRF。为说明这点，设想我们从表 2.1 的总体中抽取另一个随机样本。如表 2.5。

表 2.5　　表 2.1 总体的另一个随机样本

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

将表 2.4 和表 2.5 的数据描点，得到图 2.4 中的散点图。在这个散点图

中画两根样本回归线以尽可能地拟合这些散点：SRF_1 是根据第一个样本画的；而 SRF_2 是根据第二个样本画的。那么，两条回归线中的哪一条代表"真实"的总体回归线呢？如果我们避免偷看展现着 PR 的图 2.1 的诱惑，我们就不可能有绝对的把握知道图 2.4 中的两条回归线中哪一条代表真实的
49 总体回归线（或曲线）。图 2.4 中的回归线称为**样本回归线**（sample regression lines）。姑且假定它们都代表总体回归线，但因抽样波动，它们最多也不过是真实 PR 的一个逼近而已。一般地说，从 N 个不同的样本会得到 N 个不同的 SRF，并且这些 SRF 不大可能是一样的。

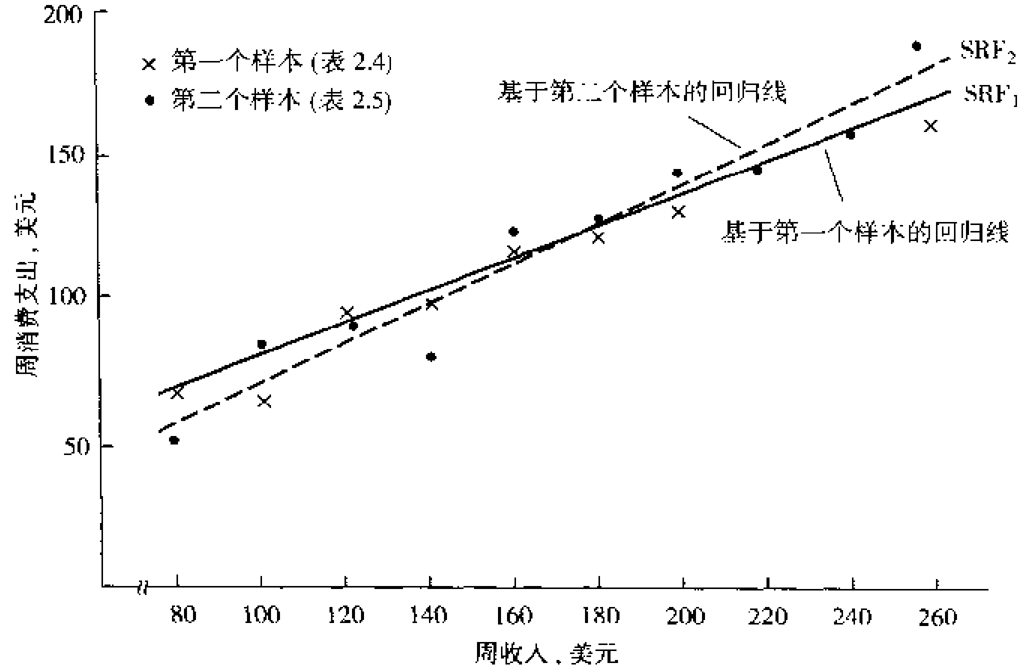


图 2.4　根据两个不同样本的回归线

类比于总体回归线有一个 PRF 作为其基础，现在我们能够写出一个代表样本回归线的**样本回归函数**（sample regression function，SRF）概念。相当于方程式（2.2.2）的样本关系式可写为：

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \qquad (2.6.1)$$

其中，$\hat{Y}_i$ 读为 "Y-帽"（"Y-hat"或"Y-cap"）

$\hat{Y}_i = E(Y \mid X_i)$ 的估计量

$\hat{\beta}_1 = \beta_1$ 的估计量

$\hat{\beta}_2 = \beta_2$ 的估计量

注意，一个**估计量**，又称（样本）**统计量**，是指一个规则或公式或方法，它告诉人们怎样用手中样本所提供的信息去估计总体参数。在一项应用中，由估计量算出的一个具体的数值，称为**估计值**。[13]

正如同我们可以把 PRF 表达成两种等价形式（2.2.2）和（2.4.2）那

样，我们还能把 SRF（2.6.1）表达成它的随机形式如下：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{2.6.2}$$

其中，除已定义过的记号外，$\hat{u}_i$ 表示（样本）**残差**(或**剩余**)项。概念上，$\hat{u}_i$ 类似于 u_i，并可把它当作 u_i 的估计量，把它引进到 SRF 中来和把 u_i 引进到 PRF 中来，是同样道理。

至此，总的说来，由于我们的分析仅仅依据某总体的单一样本的时候要比不是这样的时候多，我们看到，我们在回归分析中的主要目的是根据 SRF

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i \tag{2.6.2}$$

来估计 PRF

$$Y_i = \beta_1 + \beta_2 x_i + u_i \tag{2.4.2}$$

50 然而，由于抽样的波动，我们根据 SRF 估计出来的 PRF 充其量也不过是一个近似的结果。图 2.5 对这种近似作了解析。

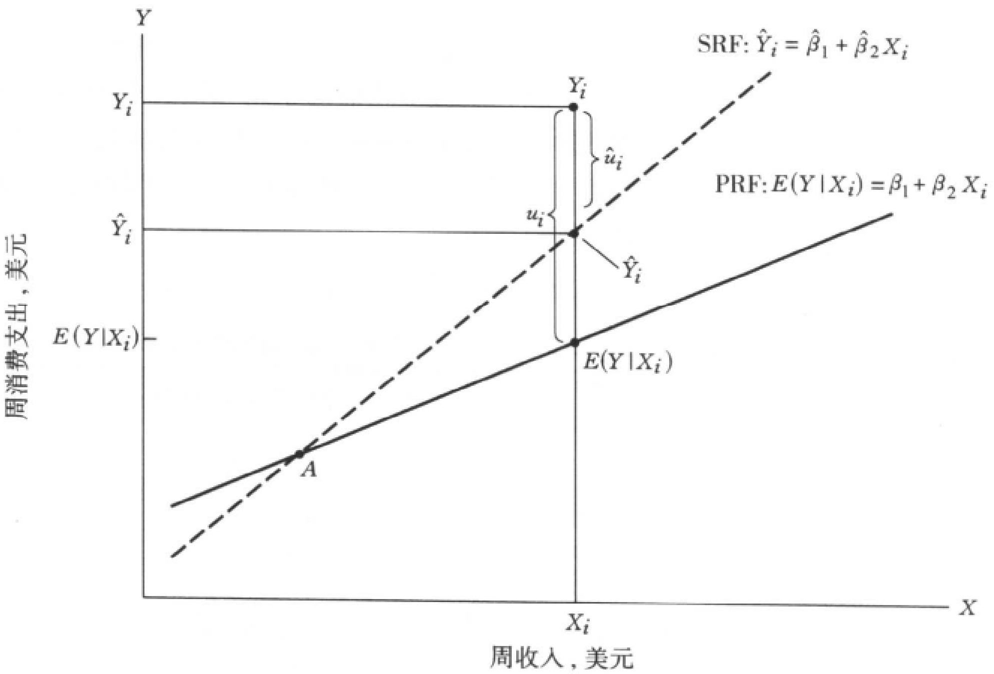


图 2.5　样本和总体回归线

对 $X = X_i$ 我们有一个观测值 $Y = Y_i$。利用 SRF 的表达式，可将所观测的 Y_i 表达为：

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{2.6.3}$$

而通过 PRF，又可把它表达为：

$$Y_i = E(Y \mid X_i) + u_i \tag{2.6.4}$$

现在，对于图 2.5 中示意的 X_i，$\hat{Y}_i$ 明显过高地估计了在那里的真实的

$E(Y|X_i)$。类似的图形分析，对 A 点以左的任何 X_i，SRF 过低估计了真实的 PRF。但读者能容易看到，由于抽样的波动，这种过高或过低的估计是不可避免的。

现在，重要的问题是：既然认识到 SRF 只不过是 PRF 的一个近似，能不能设计一种规则或方法，使得这种近似是一种尽可能"接近"的近似？或者说，怎样构造 SRF 能使 $\hat{\beta}_1$ 尽可能"接近"真实的 β_1，$\hat{\beta}_2$ 尽可能地"接近"真实的 β_2？尽管真实的 β_1 和 β_2 永远不得而知。

51 对这些问题的回答，需要我们在第 3 章中倾注大量的注意力。这里仅仅指出，我们能够给出这样一种程序，它告诉我们怎样构造 SRF，以尽可能忠实地反映 PRF。试想一想，虽然我们从来没有真正地确定过 PRF，但也能做到这一点，这是多么的不可思议。

§2.7 一个说明性的例子

我们以一个例子来结束本章。表 2.6 给出了受教育水平（以读书的年数度量）、每个受教育层次中人们的平均小时工资和所述受教育水平的人数三类数据。伯恩特（Ernst Berndt）最早获得表中列出的数据，他是从 1985 年 5 月的人口普查中推导出这些数据的。[14] 我们在下一章还将解释这些数据（以及其他的解释变量）。

表 2.6　　小时工资均值与受教育水平

读书年数	工资均值，美元	人数
6	4.456 7	3
7	5.770 0	5
8	5.978 7	15
9	7.331 7	12
10	7.318 2	17
11	6.584 4	27
12	7.818 2	218
13	7.835 1	37
14	11.022 3	56
15	10.673 8	13
16	10.836 1	70
17	13.615 0	24
18	13.531 0	31
		总计 528

资料来源：Arthur S. Goldberger, *Introductory Econometrics*, Harvard University Press, Cambridge, Mass., 1998, Table 1.1, p.5 (adapted).

以受教育水平为横轴、以（条件）平均工资为纵轴画图，可得到图2.6。图中的回归曲线表明了平均工资如何随受教育水平而变化；它们通常随着受教育水平的提高而增加，这是一个无足为奇的结论。我们在下一章将研究，其他变量如何影响平均工资。

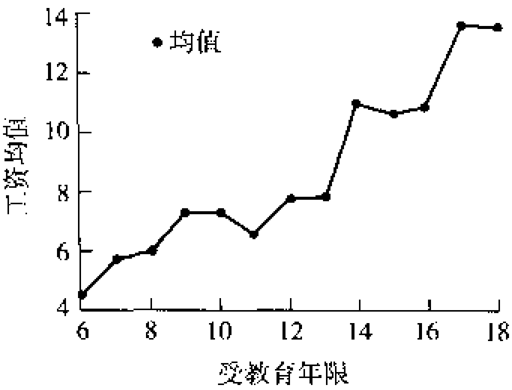


图 2.6　工资均值与受教育水平之间的关系

§2.8　要点与结论

52 1. 作为回归分析基础的主要概念是**总体回归函数**。我们做回归分析的目标就是要发现，因变量的均值如何随给定解释变量的变化而变化。

2. 本书研究线性 PRF，也就是对未知参数为线性的回归。这些回归对因变量和自变量（一或多个）来说，可以是或不是线性的。

3. 为了经验研究的目的，重要的是随机的 PRF。在 PRF 的估计中，随机干扰项 u_i 起着关键性的作用。

4. PRF 是一个理想化的概念。实际上，人们很少得知他们所研究的整个总体。通常他们只拥有对这个总体的观测值的一个样本，因此要用随机样本回归函数去估计 PRF。第 3 章将讨论如何做到这点。

习　题

问答题

2.1　什么是条件期望函数或总体回归函数？

2.2　总体和样本回归函数之间的差别是什么？这是不是人为的区别？

2.3　回归分析中的随机误差项 u_i 有什么作用？它与残差 $\hat{u}_i$ 有何区别？

2.4　为什么需要回归分析？为什么不简单地用回归子的均值作为最优值？

2.5 线性回归模型的含义是什么？

2.6 判别如下模型是线性对于参数、线性对于变量、或者是同时线性对于参数和变量，哪些模型是线性回归模型？

模型	描述性名称
a. $Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i$	倒数
b. $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$	半对数
c. $\ln Y_i = \beta_1 + \beta_2 X_i - u_i$	反半对数
d. $\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + u_i$	对数或双对数
e. $\ln Y_i = \beta_1 - \beta_2\left(\frac{1}{X_i}\right) + u_i$	对数倒数

注：ln 表示自然对数（即对数的底为 e）；u_i 为随机干扰项。我们将在第 6 章研究这些模型。

2.7 如下模型是线性回归模型吗？为什么是或为什么不是？

a. $Y_i = e^{\beta_1 + \beta_2 X_i + u_i}$　　b. $Y_i = \dfrac{1}{1 + e^{\beta_1 + \beta_2 X_i + u_i}}$

55 c. $\ln Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i$　　d. $Y_i = \beta_1 + (0.75 - \beta_1)e^{-\beta_2(X_i - 2)} + u_i$

e. $Y_i = \beta_1 - \beta_2^3 X_i + u_i$

2.8 内在线性回归模型的含义是什么？如果习题 2.7d 中的 β_2 为 0.8，那它是一个线性回归模型、还是非线性回归模型？

***2.9** 考虑如下非随机模型（即不含随机误差项的模型）。它们是线性回归模型吗？若不是，可能通过适当的代数变换使之转化成线性模型吗？

a. $Y_i = \dfrac{1}{\beta_1 + \beta_2 X_i}$

b. $Y_i = \dfrac{X_i}{\beta_1 + \beta_2 X_i}$

c. $Y_i = \dfrac{1}{1 + \exp(-\beta_1 - \beta_2 X_i)}$

2.10 图 2.7 是一个散点图及其回归线。你从此图中能得到什么一般性的结论？图中勾画出的回归线是总体回归线还是样本回归线？

54 **2.11** 你能从图2.8 的散点图中得出什么一般性结论？其背后有什么经济理论？［提示：查一本国际经济学教科书并阅读贸易的赫克谢尔－俄林（Heckscher-Ohlin）模型。］

2.12 图2.9 揭示了什么关系？基于此图，你是否认为最低工资法有利于经济福利？

2.13 引言中图 I.3 所示的回归线是 PRF 还是 SRF？为什么？你如何解释回归线周围的散点？除 GDP 外，还会有什么别的因素或变量在决定着个人消费支出？

* 选做题

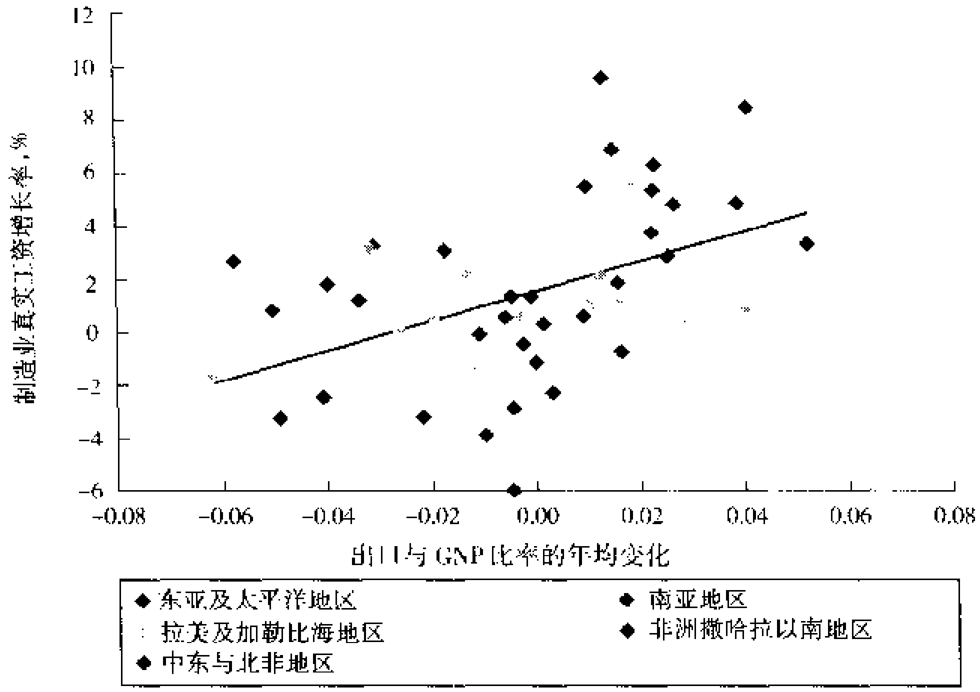


图 2.7　制造业真实工资的增长率%，出口与 GNP 比率的年均变化

注：数据为 50 个发展中国家 1970—1990 年的数据。

资料来源：The World Bank, *World Development Report* 1995, p.55. The original source is UNIDO data, World Bank data.

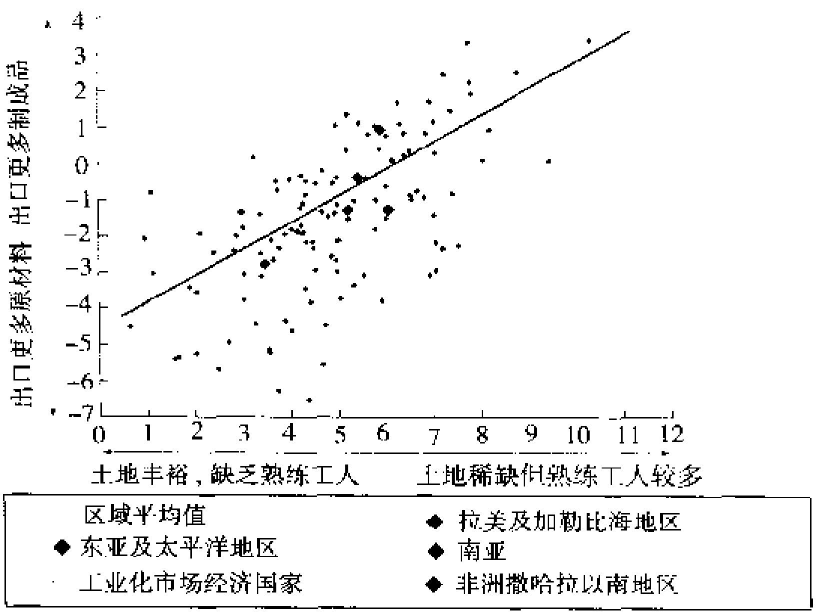


图 2.8　出口的技术密集性与人力资本

注：采集了 1985 年 126 个工业化和发展中国家的数据。横轴数据为一个国家平均受教育年龄与其土地面积之比的对数；纵轴为制造品出口与初级产品出口之比的对数。

资料来源：World Bank, *World Development Report* 1995, p.59. Original sources: Export data from United Nations Statistical Office COMTRADE data base; education data from UNDP 1990; land data from the World Bank.

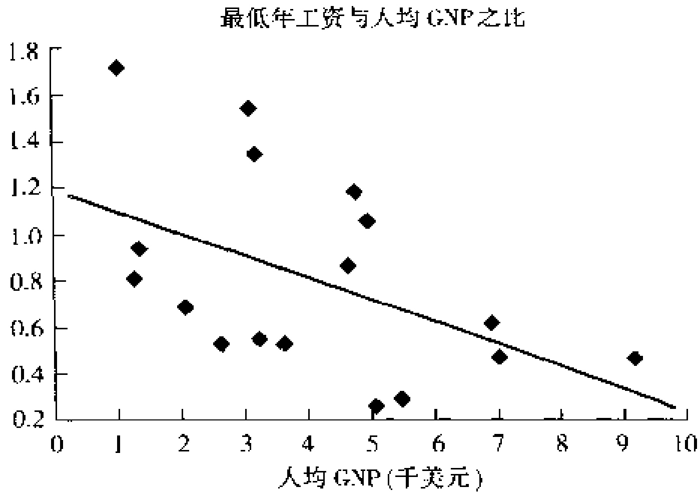


图 2.9 最低工资与人均GNP

资料来源：World Bank, *World Development Report* 1995, p.75.

解答题

2.14 表2.7给出了美国1980—1996年间的数据。

样本由17个发展中国家构成。年份从1988年到1992年因国家而有所变化。数据以国际价格度量。

表 2.7 **劳动力参与数据**

年份	CLFPRM[1]	CLFPRF[2]	UNRM[3]	UNRF[4]	AHE82[5]	AHE[6]
1980	77.4	51.5	6.9	7.4	7.78	6.66
1981	77.0	52.1	7.4	7.9	7.69	7.25
1982	76.6	52.6	9.9	9.4	7.68	7.68
1983	76.4	53.9	9.9	9.2	7.79	8.02
1984	76.4	53.6	7.4	7.6	7.80	8.32
1985	76.3	54.5	7.0	7.4	7.77	8.57
1986	76.3	55.3	6.9	7.1	7.81	8.76
1987	76.2	56.0	6.2	6.2	7.73	8.98
1988	76.2	56.6	5.5	5.6	7.69	9.28
1989	76.4	57.4	5.2	5.4	7.64	9.66
1990	76.4	57.5	5.7	5.5	7.52	10.01
1991	75.8	57.4	7.2	6.4	7.45	10.32
1992	75.8	57.8	7.9	7.0	7.41	10.57
1993	75.4	57.9	7.2	6.6	7.39	10.83
1994	75.1	58.8	6.2	6.0	7.40	11.12

1995	75.0	58.9	5.6	5.6	7.40	11.44
1996	74.9	59.3	5.4	5.4	7.43	11.82

资料来源：*Economic Report of the President*，1997. 如下标引指的是源文件。

[1]CLFPRM，Civilian labor force participation rate，male（%），Table B－37，p.343.

[2]CLFPRF，Civilian labor force participation rate，female（%），Table B－37，P.343.

[3]UNRM，Civilian unemployment rate，male（%），Table B－40，P.346.

[4]UNRF，Civilian unemployment rate，female（%），Table B－40，P.346.

[5]AHE82，Average hourly earnings（1982 dollars），Table B－45，P.352.

[6]AHE，Average hourly earnings（current dollars），Table B－45，P.352.

56

a. 将城市男性劳动力参与率相对城市男性失业率描点。目测一条穿过散点的回归线。推测二者之间的关系，其背后的经济理论是什么？这个散点图支持该理论吗？

b. 对女性也做 a 部分的问题。

c. 现在同时将男性和女性的劳动力参与率相对平均小时工资（以 1982 年美元计）描点（你可以分开画图）。现在又有何发现？又将如何使你的发现理性化？

d. 你可以将劳动力参与率同时对失业率和平均小时工资描点吗？若不能，你如何口头说明这三个变量之间的关系。

2.15 表2.8 给出的是以卢比度量的食物支出和总支出数据，样本是印度的 55 个农户。（在 2000 年初，1 美元约兑换 40 印度卢比。）

表 2.8 **食物与总支出（卢比）**

观测	食物支出	总支出	观测	食物支出	总支出
1	217.000 0	382.000 0	29	390.000 0	655.000 0
2	196.000 0	388.000 0	30	385.000 0	662.000 0
3	303.000 0	391.000 0	31	470.000 0	663.000 0
4	270.000 0	415.000 0	32	322.000 0	677.000 0
5	325.000 0	456.000 0	33	540.000 0	680.000 0
6	260.000 0	460.000 0	34	433.000 0	690.000 0
7	300.000 0	472.000 0	35	295.000 0	695.000 0
8	325.000 0	478.000 0	36	340.000 0	695.000 0
9	336.000 0	494.000 0	37	500.000 0	695.000 0
10	345.000 0	516.000 0	38	450.000 0	720.000 0
11	325.000 0	525.000 0	39	415.000 0	721.000 0
12	362.000 0	554.000 0	40	540.000 0	730.000 0

13	315.000 0	575.000 0	41	360.000 0	731.000 0
14	355.000 0	579.000 0	42	450.000 0	733.000 0
15	325.000 0	585.000 0	43	395.000 0	745.000 0
16	370.000 0	586.000 0	44	430.000 0	751.000 0
17	390.000 0	590.000 0	45	332.000 0	752.000 0
18	420.000 0	608.000 0	46	397.000 0	752.000 0
19	410.000 0	610.000 0	47	446.000 0	769.000 0
20	383.000 0	616.000 0	48	480.000 0	773.000 0
21	315.000 0	618.000 0	49	352.000 0	773.000 0
22	267.000 0	623.000 0	50	410.000 0	775.000 0
23	420.000 0	627.000 0	51	380.000 0	785.000 0
24	300.000 0	630.000 0	52	610.000 0	788.000 0
25	410.000 0	635.000 0	53	530.000 0	790.000 0
26	220.000 0	640.000 0	54	360.000 0	795.000 0
27	403.000 0	648.000 0	55	305.000 0	801.000 0
28	350.000 0	650.000 0			

资料来源：Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, New York, 1998, p.457.

57 a. 以总支出为横轴，食物支出为纵轴将数据描点，并画出一条穿过散点的回归线。

b. 你从此例中能得出什么一般性结论？

c. 据经验，你会预测无论总支出水平如何，食物支出点总是随总支出线性增加吗？你可以用总支出作为总收入的一个代理变量。

2.16 表 2.9 给出 1967—1990 年拟读大学的应届高中毕业生的平均学力测验（SAT）得分。

表 2.9　1967—1990 年* 即将进入大学的应届生的平均学力测验得分

	语文			数学		
年度	男	女	总平均	男	女	总平均
1967	463	468	466	514	467	492
1968	464	466	466	512	470	492
1969	459	466	463	513	470	493

1970	459	461	460	509	465	488
1971	454	457	455	507	466	488
1972	454	452	453	505	461	484
1973	446	443	445	502	460	481
1974	447	442	444	501	459	480
1975	437	431	434	495	449	472
1976	433	430	431	497	446	472
1977	431	427	429	497	445	470
1978	433	425	429	494	444	468
1979	431	423	427	493	443	467
1980	428	420	424	491	443	466
1981	430	418	424	492	443	466
1982	431	421	426	493	443	467
1983	430	420	425	493	445	468
1984	433	420	426	495	449	471
1985	437	425	431	499	452	475
1986	437	426	431	501	451	475
1987	435	425	430	500	453	476
1988	435	422	428	498	455	476
1989	434	421	427	500	454	476
1990	429	419	424	499	455	476

* 1967—1971 年的数据是估计值。

资料来源：The College Board. The *New York Times*, Aug.28, 1990, p.B-5.

a. 用横轴代表年度，纵轴代表 SAT 得分，分别描绘男生和女生的语文和数学得分。

b. 你能得出一些什么一般性结论？

c. 知道了男（女）生的语文分数，你会怎样预测他（她）们的数学分数？

d. 将女生总（平均）SAT 得分对男生总（平均）SAT 得分描点。通过这些散点，勾画一条回归线。你看到了什么？

【注释】

[1] 对统计学知识多少有些生疏的读者可能愿意在阅读本章之前，先阅读统计学附录——附录 A 以温故知新。

[2] 一个随机变量 Y 的期望值或期望或总体平均可记为$E(Y)$。另一方面，从 Y 的总体的一个样本值中计算出来的均值记为 $\bar{Y}$，读作“Y-横”（Y bar）。

[3] 如附录 A 中所示，条件均值和无条件均值通常是不同的。

[4] 十分感激戴维森（James Davidson）的这一见解。参见 James Davidson, *Econometric Theory*, Blackwell Publishers, Oxford, U.K., 2000, p.11。

[5] 虽然在本例中 PRL 是一条直线，但它可以是一条曲线（见图 2.3）。

[6] 我们说函数 $Y=f(X)$ 对 X 而言是线性的，如果 X 仅以幂或指数 1 出现（即不包括有 X^2、$\sqrt{X}$等项）并且不与其他变量相乘或相除（诸如 $X\cdot Z$ 或 X/Z，其中 Z 为另一变量）。如果 Y 仅依赖于 X，那么，Y 与 X 有线性关系的另一说法是，Y 对 X 的变化率（即 Y 对 X 的斜率）或导数（dY/dX）与 X 值无关。例如，若 $Y=4X$，则 $dY/dX=4$，这就与 X 值无关。但若 $Y=4X^2$，则 $dY/dX=8X$，这就不是与 X 值无关的了，从而它不是对 X 线性的。

[7] 我们说一个函数是对参数（比方说 β_1）为线性的，如果 β_1 仅以 1 次方出现，而不乘以或除以任何其他参数（例如，$\beta_1\beta_2$，β_2/β_1，等等）。

[8] 关于期望运算子 E 的性质的一个简要的讨论，见附录 A。注意一旦 X_i 值被固定，$E(Y|X_i)$ 就是常数。

[9] 事实上，在第 3 章所讲的最小二乘法中，明显地假定了$E(u_i|X_i)=0$。见第 3.2 节。

[10] 还有一个困难，即性别、教育、宗教等变量难以数量化。

[11] Milton Friedman, *A Theory of the Consumption Function*, Princeton University Press, Princeton, N.J., 1957.

[12] “叙述应在被证明为不妥之前尽可能保持简单”，*The World of Mathematics*, vol.2, J.R. Newman(ed), Simon & Schuster, New York, 1956, p.1247。或“非必要勿添加名目”，Donald F. Morrison, *Applied Linear Statistical Methods*, Prentice Hall, Englewood Cliffs, N.J., 1983, p58.

[13] 如在引言中所指出的，在一个变量的上方加一个帽形，表示对有关总体值的一个估计量。

[14] Ernst R. Berndt, *The Practice of Econometrics: Classical and Contemporary*, Addison Wesley, Reading, Mass., 1991. 顺便提一句，这是一本读者能从中发现计量经济学如何用于做研究的优秀教材。

第 3 章　双变量回归模型：估计问题

58　如在第 2 章中所指明的，我们的第一件事是根据样本回归函数尽可能准确地估计总体回归函数。在附录 A 中，我们讨论了两个通常所用到的两种估计方法：(1) **普通最小二乘法**（Ordinary Least Squares，OLS）和 (2) **最大似然法**（Maximum Likelihood，ML）。一般来说，普通最小二乘法在回归分析中应用得较广泛，因为它更具吸引力，并且也比最大似然法简单得多。除此之外，如我们在后面所示，联系两种方法的线性回归通常都会得到相同的结果。

§3.1　普通最小二乘法

普通最小二乘法归功于德国数学家高斯（Carl Friedrich Gauss）。在一定的假定下（见 3.2 节），最小二乘法有一些非常令人向往的统计性质，从而使它成为回归分析中最有功效的和最为流行的方法之一。为了说明这个方法，我们先解释最小二乘原理。

回顾双变量 PRF：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.4.2}$$

58 然而，如在第 2 章所提到的，这个 PRF 不是直接可观测的。我们通过 SRF 去估计它：

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i & (2.6.2) \\ &= \hat{Y}_i + \hat{u}_i & (2.6.3) \end{aligned}$$

其中 $\hat{Y}_i$ 是 Y_i 的估计值（条件均值）。

但 SRF 又是怎样决定的呢？为了看清楚这个问题，让我们一步步解说如下。首先把（2.6.3）写成：

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y}_i \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i & (3.1.1) \end{aligned}$$

借以表明 $\hat{u}_i$（即残差）不过是实际 Y 值与估计值 $\hat{Y}$ 之差。

对于给定的 Y 和 X 的 n 对观测值，我们希望这样决定 SRF，使得它尽可能靠近实际的 Y。为达到此目的，我们可以采用如下的准则：选择这样的 SRF，使得残差和 $\sum \hat{u}_i = \sum (Y_i - \hat{Y}_i)$ 尽可能小。这看来尽管有直观上的说服力，却不是一个很好的准则。这可以从图 3.1 中的一个人为的散点图中看出。

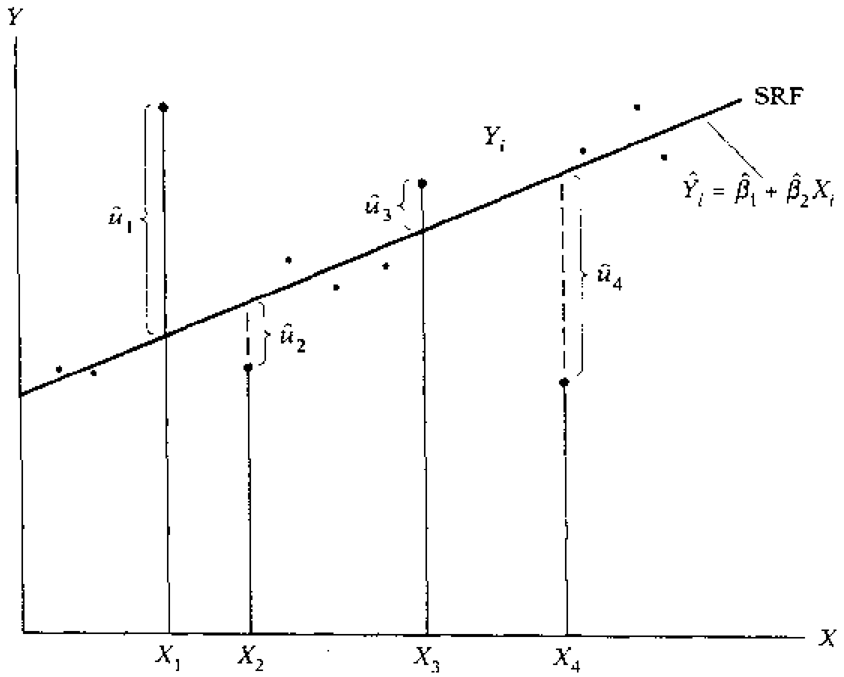


图 3.1　最小二乘准则

如果采纳 $\sum \hat{u}_i$ 最小化的准则，那么在总和 $(\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4)$ 中残差

60 $\hat{u}_2$ 及 $\hat{u}_3$ 得到的权重和 $\hat{u}_1$ 及 $\hat{u}_4$ 得到的权重一样多。尽管前两个残差比后两个残差要靠近 SRF 多得多。换言之，不管个别的观测点的散布离开 SRF 多远，所有残差都受到同样的重视。结果，很可能 $\hat{u}_i$ 离开 SRF 而散布得很

远，但 $\hat{u}_i$ 的代数和却很小（甚至是零）。为了看清楚这点，假定图 3.1 中的 $\hat{u}_1$，$\hat{u}_2$，$\hat{u}_3$ 和 $\hat{u}_4$ 分别取值 10，-2，$+2$ 和 -10，虽然 $\hat{u}_1$ 和 $\hat{u}_4$ 比 $\hat{u}_2$ 和 $\hat{u}_3$ 离开 SRF 远得多，这些残差的代数和却是零。如果我们采取最小二乘准则，就可避免这种问题。最小二乘准则是要确定 SRF 使得：

$$\begin{aligned}\sum \hat{u}_i^2 &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \end{aligned} \tag{3.1.2}$$

尽可能地小，其中 $\hat{u}_i^2$ 是残差的平方。该方法通过对 $\hat{u}_i$ 平方而赋予诸如图 3.1 中的 $\hat{u}_1$ 和 $\hat{u}_4$ 比 $\hat{u}_2$ 和 $\hat{u}_3$ 更多的权重。如前所说，在 $\sum \hat{u}_i$ 最小化的准则下，虽然 $\hat{u}_i$ 在 SRF 周围散布得很宽，但其总和可能很小。但在最小二乘方法中，这是不可能的，因为 $\hat{u}_i$（在绝对值上）越大，$\sum \hat{u}_i^2$ 也越大。采用最小二乘法的理由，如我们即将看到的，还在于由它得出的估计量有一些很好的统计性质。

由（3.1.2）明显地看到：

$$\sum \hat{u}_i^2 = f(\hat{\beta}_1, \hat{\beta}_2) \tag{3.1.3}$$

就是说，残差平方和是估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的某个函数。对任意给定的一组数据，选择不同的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 值将给出不同的 $\hat{u}$，从而有不同的 $\sum \hat{u}_i^2$ 值。为了清楚地看到这点，考虑由表 3.1 头两列给出的、关于 Y 和 X 的一些人为数据。
61 现在做两个实验。在实验 1 中，取 $\hat{\beta}_1 = 1.572$ 和 $\hat{\beta}_2 = 1.357$（暂不必问这两个数值是怎样得来的，就算是一种猜测好了）。[11]利用这些 $\hat{\beta}$ 值和表3.1第 2 列的 X 值，便容易算出该表第 3 列对 Y_i 的估计值，记作 $\hat{Y}_{1i}$（下标 1 表示第 1 个实验）。然后再做另一个实验，但这回利用 $\hat{\beta}_1 = 3$ 和 $\hat{\beta}_2 = 1$ 两个值。

表 3.1　　SRF 的实验决定法

Y_i (1)	X_i (2)	$\hat{Y}_{1i}$ (3)	$\hat{u}_{1i}$ (4)	$\hat{u}_{1i}^2$ (5)	$\hat{Y}_{2i}$ (6)	$\hat{u}_{2i}$ (7)	$\hat{u}_{2i}^2$ (8)
4	1	2.929	1.071	1.147	4	0	0
5	4	7.000	−2.000	4.000	7	−2	4
7	5	8.357	−1.357	1.841	8	−1	1
12	6	9.714	2.286	5.226	9	3	9
总和：28	16		0.0	12.214		0	14

注：$\hat{Y}_{1i} = 1.572 + 1.357X_i$（即 $\hat{\beta}_1 = 1.572$ 和 $\hat{\beta}_2 = 1.357$）

$\hat{Y}_{2i} = 3.0 + 1.0X_i$（即 $\hat{\beta}_1 = 3$ 和 $\hat{\beta}_2 = 1.0$）

$\hat{u}_{1i} = (Y_i - \hat{Y}_{1i})$

$\hat{u}_{2i} = (Y_i - \hat{Y}_{2i})$

把对 Y_i 的估计记为 $\hat{Y}_{2i}$，列在表 3.1 的第 6 列上。由于两个实验的 $\hat{\beta}$ 值有所不同，所估计的残差也有所不同，如表所示；$\hat{u}_{1i}$是得自第 1 个实验的残差而 $\hat{u}_{2i}$是得自第 2 个实验的残差，其平方列于（5）和（8）两列。显然，可从（3.1.3）预测到，由于所依据的 $\hat{\beta}$ 值的不同，这些残差的平方和是不同的。

那么，我们应选取哪一组 $\hat{\beta}$ 值呢？因为第 1 个实验的 $\hat{\beta}$ 值比第 2 个实验的 $\hat{\beta}$ 值给出一个更低的 $\sum \hat{u}_i^2$（12.214 小于 14）。所以说第 1 个实验的 $\hat{\beta}$ 值是“最优”值。但怎样知道是最优呢？如果我们的时间和耐心都是无限的，我们就能做许多类似的实验，每次选择不同的 β 值，然后比较所得的 $\sum u_i^2$，并从中选择给出最可能小的 $\sum \hat{u}_i^2$ 值的那组$\hat{\beta}$ 值。当然，这里假定了我们已经考虑过所有可想像的 β_1 和 β_2 值。但由于时间和耐心无疑是有限的，我们必须考虑这种尝试与纠错法的某种捷径。幸而，最小二乘法为我们提供了这一捷径。由最小二乘法原理或方法选出的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$，将使得对于给定的样本或一组数据，$\sum \hat{u}_i^2$ 是尽可能小的。换言之，对一给定的样本，最小二乘法为我们提供使得 $\sum \hat{u}_i^2$ 达到最可能小的 β_1 和 β_2 估计。怎样做到这点？这是微积分学的一个直接的运算，如附录 3A 的 3A.1 节所表明的。通过微分法将得到用于估计 β_1 和 β_2 的下列方程：

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \tag{3.1.4}$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \tag{3.1.5}$$

其中 n 是样本大小。这组联立方程被称为**正规方程**（normal equations）。

62 解此联立方程得：

$$\begin{aligned}\hat{\beta}_2 &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \\ &= \frac{\sum x_i y_i}{\sum x_i^2}\end{aligned} \tag{3.1.6}$$

其中 $\bar{X}$ 和 $\bar{Y}$ 是 X 和 Y 的样本均值，并且定义 $x_i = (X_i - \bar{X})$ 和 $y_i = (Y_i - \bar{Y})$。从此以后，我们将遵循一个惯例：用小写字母表示对均值的离差。

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ &= \bar{Y} - \hat{\beta}_2 \bar{X}\end{aligned} \tag{3.1.7}$$

（3.1.7）中最后一步是通过简单的代数运算直接从（3.1.4）得到的。

顺便指出，利用简单的代数恒等式，用于估计 β_2 的公式（3.1.6）可另表述为：

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum X_i^2 - n\bar{X}^2} \\ &= \frac{\sum X_i y_i}{\sum X_i^2 - n\bar{X}^2}\end{aligned} \qquad (3.1.8)^{[2]}$$

这样，如果我们用一部手头计算器去解一个仅涉及一小组数据的回归问题的话，就可以减轻计算上的负担。

上面得到的估计量是从最小二乘原理演算出来的，所以叫做**最小二乘估计量**。注意以下由普通最小二乘法得到的估计量的**数值性质**："数值性质是
63 指由于运用普通最小二乘法而得以成立的那些性质，而不管数据是怎样产生的。"[3]稍后，我们还将考虑 *OLS* 估计量的**统计性质**，也就是"仅在数据产生的方式满足一定的假设下才得以成立的"性质。[4]（参看 3.2 节中的经典线性回归模型。）

Ⅰ.OLS 估计量是纯粹由可观测的（即样本）量（指 X 和 Y）表达的，因此这些量是容易计算的。

Ⅱ. 这些量是**点估计量**（point estimators），即对于给定的样本，每一估计量仅提供有关总体参数的一个（点）值。［在第 5 章中，我们将考虑所谓**区间估计量**（interval estimators），后者对未知的总体参数的可能值提供了一个范围。］

Ⅲ. 一旦从样本数据得到 OLS 估计值，便容易画出样本回归线（图 3.1）。这样得到的回归线有如下性质：

1. 它通过 Y 和 X 的样本均值。这是从（3.1.7）显见的事实，因该式可写为 $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$，如图 3.2 所示。

64 2. 估计的 $Y(=\hat{Y}_i)$ 均值等于实测的 Y 均值。因为：

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}) + \hat{\beta}_2 X_i \\ &= \bar{Y} + \hat{\beta}_2 (X_i - \bar{X})\end{aligned} \qquad (3.1.9)$$

将最后一个等式两边对样本值求和并同除以样本大小 n，即得：

$$\bar{\hat{Y}} = \bar{Y} \qquad (3.1.10)^{[5]}$$

这里利用了等式 $\sum (X_i - \bar{X}) = 0$。(为什么?)

3. 残差 $\hat{u}_i$ 的均值为零。由附录 3A，第 3A.1 节，第一个方程是：

$$-2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0$$

但 $\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$，故上述方程化为 $-2\sum \hat{u}_i = 0$，从而 $\bar{\hat{u}} = 0$。[6]

作为上述性质的一个结果，样本回归

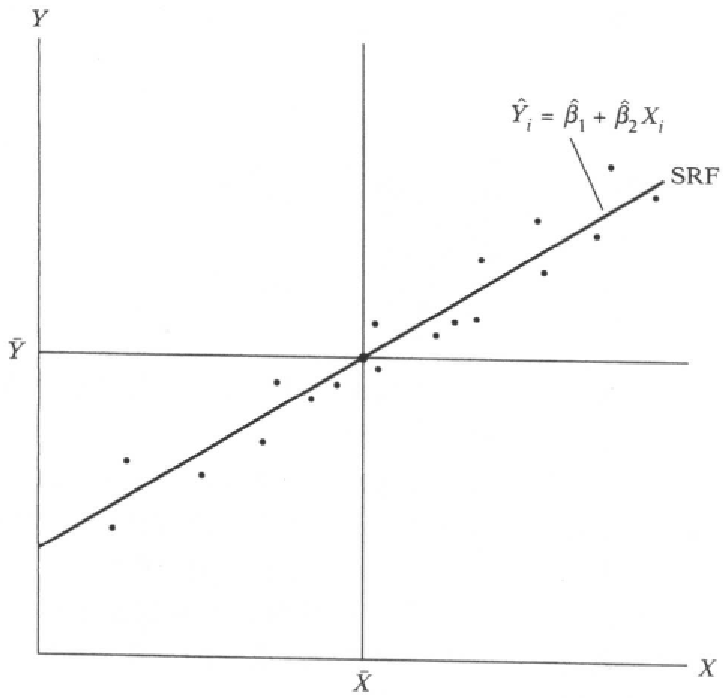


图 3.2　本图解表明样本回归线通过 Y 和 X 的样本均值

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{2.6.2}$$

可表达为另一种形式，其中 Y 和 X 都表示为对其均值的离差。为了看清楚这一点，对（2.6.2）两边求和得：

$$\begin{aligned}\sum Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \quad 因 \sum \hat{u}_i = 0\end{aligned} \tag{3.1.11}$$

方程（3.1.11）两边同时除以 n 得：

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X} \tag{3.1.12}$$

这无异于（3.1.7）。从方程（2.6.2）减去（3.1.12）就得到：

$$Y_i - \bar{Y} = \hat{\beta}_2 (X_i - \bar{X}) + \hat{u}_i$$

65 或

$$y_i = \hat{\beta}_2 x_i + \hat{u}_i \tag{3.1.13}$$

其中，按照惯例，y_i 和 x_i 分别是对其（样本）均值的离差。

方程（3.1.13）称为**离差形式**（deviation form）。注意，在此形式中截距项 β_1 不再出现。然而截距项总可以从（3.1.7）估计出来，也就是从样本回归线通过 Y 和 X 的样本均值这一事实估计出来。离差形式的好处是，当我们用台式计算机做计算时，它往往能简化算术计算。但在今天的电脑时代里，这种好处也许微不足道。

顺便指出，按照离差形式，SRF 可写为：

$$\hat{y}_i = \hat{\beta}_2 x_i \tag{3.1.14}$$

而按照原来的测算单位则是 $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$，如（2.6.1）所示。

4. 残差 $\hat{u}_i$ 和预测的 Y_i 值不相关。这一陈述可验证如下，利用离差形式，可推出：

$$\begin{aligned}\sum \hat{y}_i \hat{u}_i &= \hat{\beta}_2 \sum x_i \hat{u}_i \\ &= \hat{\beta}_2 \sum x_i (y_i - \hat{\beta}_2 x_i) \\ &= \hat{\beta}_2 \sum x_i y_i - \hat{\beta}_2^2 \sum x_i^2 \\ &= \hat{\beta}_2^2 \sum x_i^2 - \hat{\beta}_2^2 \sum x_i^2 \\ &= 0\end{aligned} \tag{3.1.15}$$

其中我们利用了 $\hat{\beta}_2 = \sum x_i y_i / \sum x_i^2$ 这一事实。

5. 残差 $\hat{u}_i$ 和 X_i 不相关；就是说 $\sum \hat{u}_i X_i = 0$。这一事实可从附录 3A，第 3A.1 节的方程（2）推知。

§3.2 经典线性回归模型：最小二乘法的基本假定

如果我们的目的仅是估计 β_1 和 β_2，那么前节所讨论的 OLS 法就足够用了。但回顾一下第 2 章，在回归分析中我们的目的不仅仅是获得 $\hat{\beta}_1$ 和 $\hat{\beta}_2$，而且要对真实的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 做出推断。例如，我们想知道 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 离它们相应的总体值有多接近或者 $\hat{Y}_i$ 与其期望值$E(Y|X_i)$ 多接近。为达到这一目的，我们不仅要如同（2.4.2）那样有确定的模型的函数形式，还要对 Y_i 的产生
66 方式做出某些假定。为了看清楚为什么需要做出这种要求，让我们看一下 PRF：$Y_i = \beta_1 + \beta_2 X_i + u_i$，它表明 Y_i 依赖于 X_i 和 u_i。因此，除非我们明确 X_i 和 u_i 是怎样产生的，否则我们将无法对 Y_i 做出任何统计推断，而且我们也无法对 β_1 和 β_2 做出任何统计推断。就是说，为了回归估计的有效解释，对 X_i 变量（一个或多个）和误差项做出假定是极其重要的。

经典（又称高斯或标准）线性回归模型（记为 CLRM）。这一模型已成为大部分计量经济学理论的奠基石，它有 10 个假定。我们先在双变量回归模型的框架下来讨论这些假定；等到第 7 章我们再把这些假定推广到多变量回归模型，即含有多于一个回归元的复回归模型。

假定 1：线性回归模型。回归模型**对参数而言是线性的，**有如（2.4.2）

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.4.2}$$

我们已在第 2 章中讨论过模型（2.4.2）。因对参数为线性的回归模型是

CLRM的出发点，我们在本书中将始终坚持这一假定。如第2章所讨论的，回归子 Y 和回归元 X 本身可以是非线性的。[8]

> **假定2：在重复抽样中X值是固定的。**在重复的样本中，回归元 X 所取的值被认为是固定的。说得更专门些，假定 X 是非随机的。

在第2章关于PRF的讨论中即隐含着这一假定。"重复抽样中的固定值"这一概念可通过表2.1中的例子加以解释。对它的理解非常重要。考虑与表中诸收入水平相对应的各个 Y 总体，比如说，把收入值 X 固定在80美元的水平上，我们随机地抽取一个家庭，并观测到它的每周家庭消费支出 Y，比方说为60美元，仍然把 X 固定在80美元，而随机地另抽取一个家庭并观测到它的 Y 值是75美元。在每次抽取（即重复抽样）中，X 值都固定在80美元。我们可以对表中的全部 X 值重复这一过程。事实上，表2.4和表2.5中的样本数据就是这样抽取得来的。

所有这些都意味着我们的回归分析是**条件回归分析**（conditional regression analysis），就是以回归元 X 的给定值作为条件的。

67

> **假定3：干扰项 u_i 的均值为零。**对给定的 X 值，随机干扰项 u_i 的均值或期望值为零，专业地讲，u_i 的条件均值为零，符号上记为：
>
> $$E(u_i \mid X_i)=0 \tag{3.2.1}$$

假定3是说，以给定的 X_i 为条件的 u_i 的均值为零，其几何意义可由图3.3描绘出来。图中显示了变量 X 的几个值以及与每一 X 值相对应的一个 Y 总体。如图所示，对应于给定的 X，每一个 Y 总体都是围绕其均值（由打上圆圈的点来表示）分布的；一些 Y 值位于均值之上，另一些则位于均值之下。离开均值的上方和下方距离就是 u_i。(3.2.1) 所要求的，无非是这些对应于任一给定 X 的离差的均值应等于零。[9]

鉴于2.4节（参看方程2.4.5）的讨论，这一假定应是不难理解的。这一假定无非是说，凡是模型不含的因而归属于 u_i 的因素，对 Y 的均值都没有系统的影响；这样说来，正的 u_i 值就抵消了负的 u_i 值，以致它们对 Y 的

68 平均影响为零。[10]

顺便指出，假定 $E(u_i \mid X_i)=0$ 意味着假定 $E(Y_i \mid X_i)=\beta_1+\beta_2X_i$。（为什么？）因此这两个假定是等价的。

> **假定4：同方差性或 u_i 的方差相等。**给定 X 值，对所有的观测，u_i 的方差都是相同的。就是说 u_i 的条件方差是恒定的。用符号表示为：
>
> $$\begin{aligned}\mathbf{var}(u_i \mid X_i) &= E[u_i - E(u_i \mid X_i)]^2 \\ &= E(u_i^2 \mid X_i) \qquad \text{（由于假定 3）} \\ &= \sigma^2 \end{aligned} \tag{3.2.2}$$

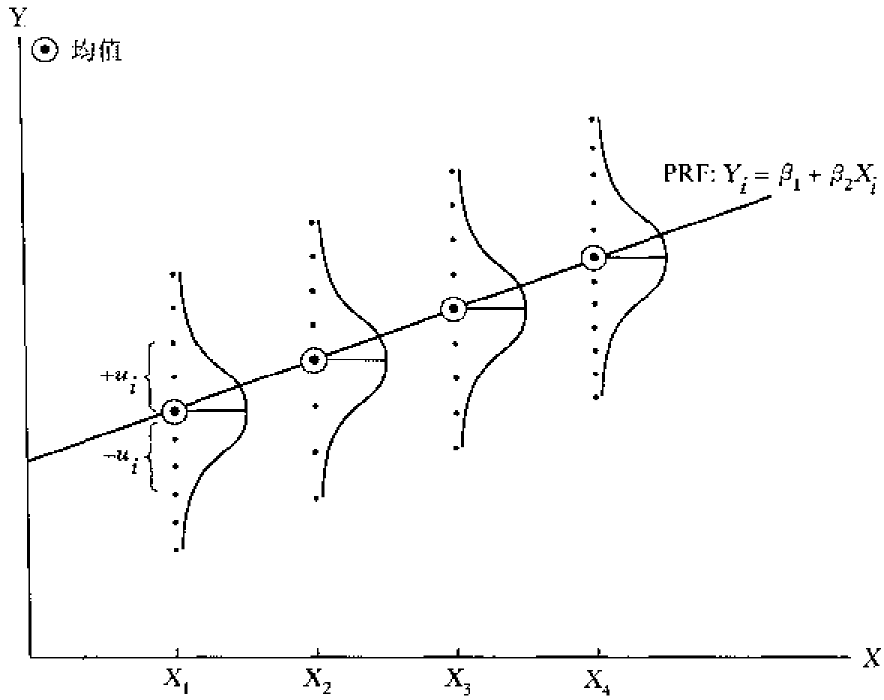


图 3.3　干扰项 u_i 的条件分布

其中 **var** 表示方差。

方程（3.2.2）是说，对每个 u_i 的方差（即 u_i 的条件方差）都是某个等于 σ^2 的正的常数。用专业术语说，（3.2.2）代表**同方差性**（homoscedasticity）或者说相同的散布或相等的方差。换言之，（3.2.2）意味着，对应于不同 X 值的 Y 总体均有同样的方差。图 3.4 展示了这种情形。

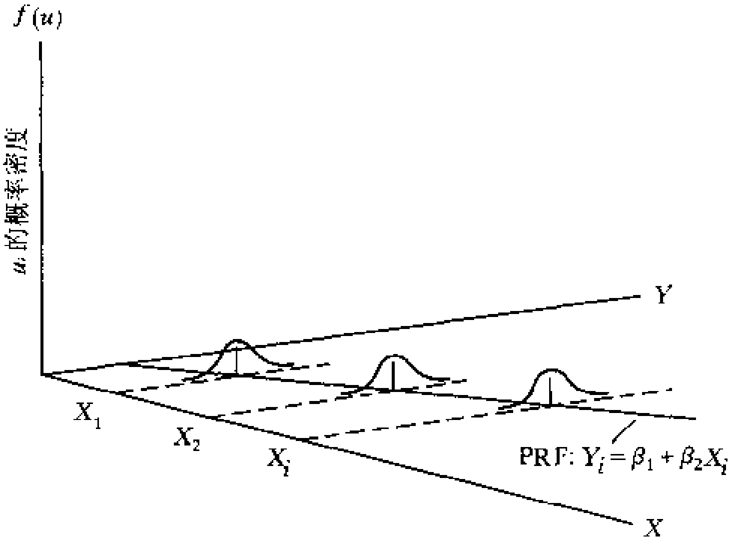


图 3.4　同方差性

69　与此对照，图 3.5 表示 Y 总体的条件方差随 X 而变。这种情形的相应

名称是**异方差性**（heteroscedasticity）或者说非相同的散布（unequal spread）或非相等的方差（variance）。用符号表示，这时（3.2.2）可写为：

$$\operatorname{var}(u_i \mid X_i)=\sigma_i^2 \tag{3.2.3}$$

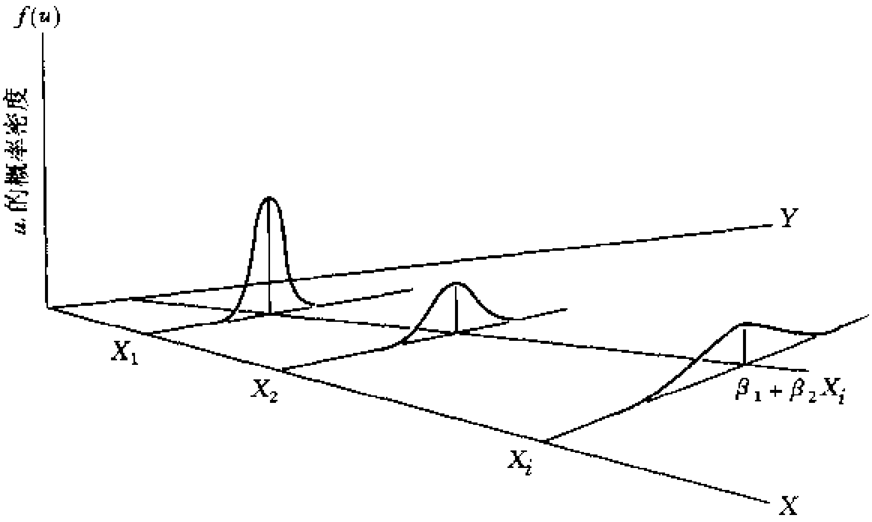


图 3.5 异方差性

注意方程（3.2.3）中的 σ^2 的下标，它表示 Y 总体的方差不再是恒定不变的了。

为了清楚地区分这两种情形，令 Y 代表每周消费支出而 X 代表每周收入。图 3.4 和 3.5 都表示随着收入增加，平均消费支出也增加。但在图 3.4 中，消费支出的方差在所有的收入水平上都保持不变，而在图 3.5 中，这个方差随收入的增加而增加，换句话说，富有的家庭比贫穷的家庭平均消费更多，但前者的消费支出也有更大的变异。

为了了解其中的道理，请看图 3.5。如该图所表明，$\operatorname{var}(u \mid X_1)<\operatorname{var}(u \mid X_2)$，…，$<\operatorname{var}(u \mid X_i)$。因此，很有可能，来自 $X=X_1$ 时的 Y 总体的观测值比来自 $X=X_2$、$X=X_3$ 等等时的 Y 总体的观测值会更靠近 PRF。简言之，并不是对应于不同 X 的所有 Y 值都是同样可靠的。要根据 Y 值多么靠近或多么远离其均值（也就是 PRF 上的点）分布以判断其可靠程度。如果这种想法符合实际，难道我们不认为那些离均值较近的 Y 总体的样本比远为分散的 Y 总体的样本更为可取吗？但这样做会限制 X 值的变动。

70 现在，借助于假定 4，我们说，对应于不同 X 值的全部 Y 值都有同样的重要性。在第 11 章中，我们将看到如果情况不是这样，亦即存在有异方差性，那么又会发生什么情况。

顺便提请注意，假定 4 意味着 Y_i 的条件方差也是同方差的，就是说：

$$\operatorname{var}(Y_i \mid X_i)=\sigma^2 \tag{3.2.4}$$

当然，Y 的无条件方差为 σ_Y^2。以后我们将会看到区分 Y 的条件和无条件方差是重要的。（关于条件和无条件方差的细节，可参看附录 A。）

假定 5：各个干扰之间无自相关性。 给定任意两个 X 值：X_i 和 X_j（$i \neq j$），u_i 和 u_j 之间的相关性为零。用符号表示：

$$
\begin{aligned}
\text{cov}(u_i, u_j \mid X_i, X_j) &= E\{[u_i - E(u_i)] \mid X_i\}\{[u_j - E(u_j)] \mid X_j\} \\
&= E(u_i \mid X_i)(u_j \mid X_j) \quad \text{（为什么？）} \\
&= 0 \qquad (3.2.5)
\end{aligned}
$$

其中 i 和 j 为两次不同的观测，而 cov 表示**协方差**。

通俗地说，(3.2.5) 设定干扰 u_i 和 u_j 不相关。用专门术语来说，这是**无序列相关**（no serial correlation）或**无自相关**（no auto correlation）假定。也就是说，给定 X_i，任意两个 Y 值对它们的均值的离差都不会表现出如图 3.6（a）和 3.6（b）那样的模式。在图 3.6（a）中的 u 值是**正相关的**，即正的 u 伴随着正的 u，或负的 u 伴随着负的 u。图 3.6（b）的 u 值则是**负相关的**，即正（负）的 u 伴随着负（正）的 u。

如果这些干扰（离差）展现出某种系统性模式，如图 3.6（a）和 3.6（b）那样，就表明它们有自相关或序列相关。假定 5 所要求的，就是不要有这种相关。图 3.6（c）表示 u 之间没有这种系统性的模式，即指零相关。

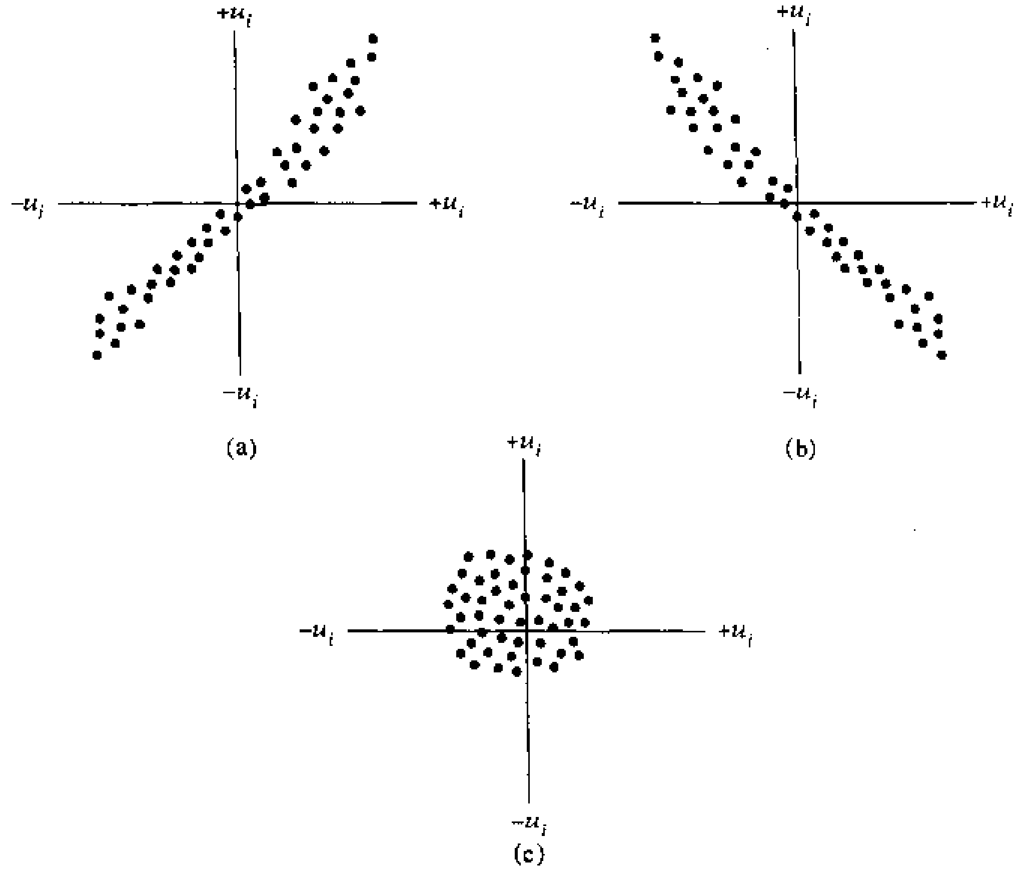


图 3.6　干扰之间的相关模式：
(a) 正序列相关；(b) 负序列相关；(c) 零相关

在第 12 章里，我们将透彻地解释这一假定的全部涵义。但直观上，我们可以对此假定作如下解释：设想在我们的 PRF（$Y_t = \beta_1 + \beta_2 X_t + u_t$）中，$u_t$ 和 u_{t-1} 正相关，那么，Y_t 不仅依赖于 X_t，而且依赖于 u_{t-1}，因为 u_{t-1} 在一定程度上决定了 u_t。在讨论这一主题的现阶段，我们要利用假定 5，就是说，我们将只考虑 X_t 对 Y_t 的系统性影响和是否有影响，而不去担心由于 u 之间的可能的交互相关而造成的其他可能作用于 Y 的影响。但是，如同在第 12 章中所表明的，我们能对这些干扰之间的相关性做出分析并查看其后果。

71 **假定 6：u_i 和 X_i 的协方差为零**，或 $E(u_iX_i)=0$。形式上，

$$
\begin{aligned}
\operatorname{cov}(u_i, X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\
&= E[u_i(X_i - E(X_i))] \qquad \text{因 } E(u_i)=0 \\
&= E(u_iX_i) - E(X_i)E(u_i) \qquad \text{因 } E(X_i)\text{非随机} \\
&= E(u_iX_i) \qquad \text{因 } E(u_i)=0 \\
&= 0 \qquad \text{由假定} \qquad (3.2.6)
\end{aligned}
$$

假定 6 是说，干扰 u 和解释变量 X 是不相关的。作此假定的理由如下：当我们把 PRF 表述为（2.4.2）时，我们假定了 X 和 u（后者代表所有被省略的变量的影响）对 Y 有各自的（并且可加的）影响。但若 X 和 u 是相关的，就不可能评估它们各自对 Y 的影响。例如，若 X 和 u 正相关，则当 u
72 增加时 X 也增加，而当 u 减小时 X 也减小。类似地，若 X 和 u 为负相关，则当 u 减小时 X 增加，而当 u 增加时 X 减小。无论是哪一种情形，要分开 X 和 u 对 Y 的影响都是困难的。

如果 X 变量是非随机的，并且假定 3 也成立，假定 6 就自动地得到满足。因为，这时 $\operatorname{cov}(u_i, X_i) = [X_i - E(X_i)]E[u_i - E(u_i)] = 0$。（为什么？）但因我们已假定 X 变量不仅是非随机的，而且是在重复样本中取固定值的[11]，故假定 6 对我们来说并不是什么关键性的假定。这里把它叙述出来，只是为了表明，即使这些 X 是随机的，只要它们独立于干扰项 u_i 或至少与 u_i 无关，下面讲的回归理论就是真实的。[12]（在本书第 2 部分中，我们将研究放弃假定 6 的后果。）

假定 7：观测次数 n 必须大于待估计的参数个数。另一种说法是，观测次数 n 必须大于解释变量的个数。

设立这一假定，并非是可有可无的。在表 3.1 的人为例子中，不妨设想我们只有对 Y 和 X 的第一对观测值（4 和 1），这样就无法从单一的观测去估计两个未知数 β_1 和 β_2。我们至少需要两双观测值来估计两个未知数。在后面的一章里，我们将会认识到这一假定的关键作用。

假定 8：X 值要有变异性。在一个给定的样本中，X 值不可以全是相同的。用专门术语来说，$\mathrm{var}(X)$ 必须是一个有限的正数。[13]

这一假定也不是但说无妨。且看方程（3.1.6）。如果全部 X 值都相同，则 $X_i = \bar{X}$。（为什么？）该方程的分母就变为零，从而无法估计 β_2，也就无法估计 β_1。凭直觉我们就能看出此假定为什么重要。看看我们第 2 章中的
73 家庭消费支出例子，如果家庭收入很少变动，我们就不怎么能解释消费支出的变化。读者应该记住，要把回归分析作为一种研究工具来使用，Y 和 X 两者均有变化是最为重要的。简言之，变量必须在变！

假定 9：正确地设定了回归模型。另一种说法是，在经验分析中所用的模型没有**设定偏误**（specification bias or error）。

如引言中所讨论过的，经典计量经济方法论不是明显地就是隐含地假定，用于检验经济理论的模型是"正确地设定的"。对这一假定可作如下的非正式解释。一项计量经济研究开始时，要先对感兴趣的现象设定其计量经济模型。在模型的设定中出现的一些重要问题包括：（1）模型应包括哪些变量？（2）模型的函数形式为何？它是不是对参数，对变量或对两者为线性？（3）对进入模型的 Y_i，X_i 和 u_i 要做些什么概率上的假定？

这些都是极为重要的问题。在第 13 章中，我们将说明，如果模型漏掉一些重要变量，或选择了错误的函数形式，或者对所含变量做出错误的随机假定，那么，要对所估计的回归作有效的解释就大有疑问了。为了对此有所感知，现在回到图 1.3 中的菲利普斯曲线。假使我们选择下列两个模型去描述货币工资变化率和失业率的理论关系：

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \tag{3.2.7}$$

$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \tag{3.2.8}$$

其中 Y_i = 货币工资变化率，而 X_i = 失业率。

回归模型（3.2.7）对参数和变量都是线性的，而（3.2.8）则对参数为线性（因此按照我们的定义是线性回归模型），但对变量 X 为非线性。现在考虑图 3.7。

如果模型（3.2.8）是"正确"或"真实"模型，而用了模型（3.2.7）去拟合散布点，如图 3.7 所示，这样就导致出错误的预测：对于 A 和 B 两点之间任给的 X_i，模型（3.2.7）都将过高地估计真实的 Y 均值；而对于 A 以左和 B 以右的 X_i，则过低地估计了真实的 Y 均值。

74 上述例子是所谓**设定偏误**的一例；其偏误在于选择了错误的函数形式。在第 13 章中我们还将看到其他类型的设定偏误。

不幸的是，实际上人们很少知道模型应包括的正确变量、模型的正确函数形式或关于进入模型变量的正确的概率假定是什么。因为，关于某一具体研究（例如菲利普斯类型的货币工资变化—失业率的关系）的基础理论未必

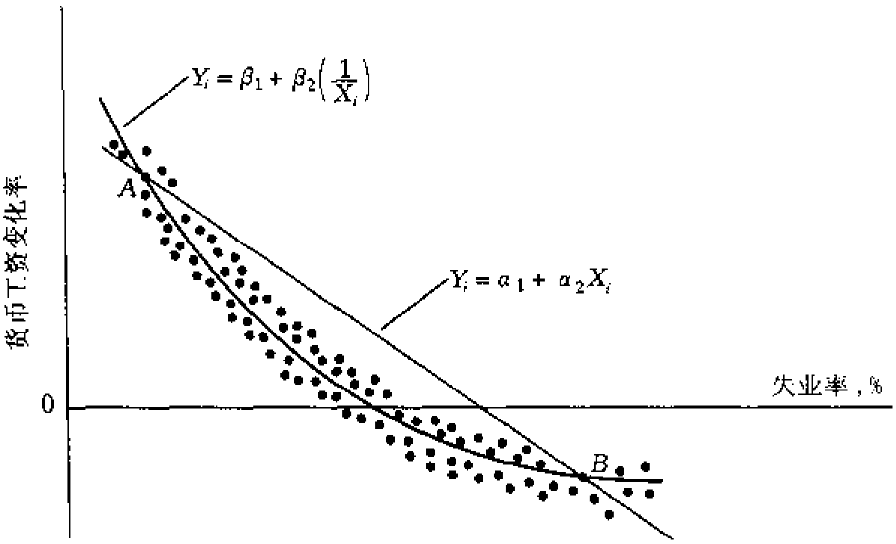


图 3.7　线性和非线性菲利普斯曲线

牢靠或扎实到足以回答所有这些问题。因此，在实践中，计量经济学家在选择进入模型的变量个数、模型的函数形式以及关于模型所含变量的概率性质的假定时，必须做出一些自己的判断。为做经验分析而选择"正确"模型在一定程度上涉及一些尝试与纠错过程。[14]

如果在选择模型时需要做出判断，那么假定 9 还有什么用呢？暂且不详细讨论这个问题（参看第 13 章），假定 9 摆在这里是为了提醒我们，回归分析以及由分析而得到的结果，是以所选的模型为条件的，从而警惕我们，在
75 建立计量经济模型时必须十分审慎，特别是对某些经济现象如通货膨胀率、货币需求或者一种股票或债券的合理或均衡定价如何解释，常存在多种富有争议的理论。由此可见，计量经济的模型建造，与其说是一种科学，毋宁说是一种艺术。

至此，我们完成了关于经典线性回归模型的基本假定的讨论。重要的是要注意到所有这些假定都仅仅是关于 PRF 的，而不是关于 SRF 的。但有趣的是，我们看到前面讨论的最小二乘法有一些性质，和我们对 PRF 所作的假定相类似。例如，我们发现$\sum\hat{u}_i=0$因而$\bar{\hat{u}}=0$，这和假定$E(u_i|X_i)=0$相近。又如，我们发现$\sum\hat{u}_iX_i=0$类似于假定$\text{cov}(u_i,X_i)=0$。注意到最小二乘法似乎试图"复制"我们赋予 PRF 的一些假定，颇使人有心安理得之感。

当然，SRF 并不复制对 CLRM 的全部假定。以后我们将表明，虽然根据假定$\text{cov}(u_i,u_j)=0(i\neq j)$，但样本$\text{cov}(\hat{u}_i,\hat{u}_j)=0(i\neq j)$并非真实。事实上，我们以后将看到，这些残差不仅是自相关的而且是异方差的（见第 12 章）。

当我们走出双变量模型的范围，而进入多变量复回归模型，也就是模型含有多个回归元时，我们增补下面的一个假定。

假定 10：没有完全的多重共线性。就是说，**解释变量之间没有完全的线性关系。**

我们在第 7 章讲复回归模型时再来讨论这一假定。

对这些假定的总结

所有这些假定有多真实？这个千钧分量的问题是科学哲学中的一个古老问题。有些人称辩说假定是否真实是无关重要的。重要的是基于这些假设的预测。以"假定无关重要论"著称的有弗里德曼，对他来说，假定的非真实性有着积极的意义。"为了有意义，……一个假设在其假定中从描述上看必定是错误的。"[15]

我们可以不完全赞同这一观点，但回想一下在任何科学研究中我们做出
76
某些假定，是因为它们便于逐步开展我们的主题研究，并不因为它们在准确地复制了现实的意义下必然是真实的。正如一位作者所说，"……如果简单性是好的理论所盼望的一个准则，那么所有好的理论都将毫无禁忌地理想化和简单化。"[16]

这里作一个类比也许是有益的。经济学学生通常先熟悉完全竞争模型，再去接触诸如独占（垄断）和寡头等非完全竞争模型。因为从完全竞争模型引申出来的含义能使我们更好地领会非完全竞争模型，并不是因为完全竞争模型一定是真实的。计量经济学中的 CLRM 就相当于价格理论中的完全竞争模型！

我们的计划是先透彻地研究 CLRM 的性质，然后在以后的篇章里深入分析如果 CLRM 的一个或多个假定不成立时会出现什么情况。在本章末我们在表 3.4 中给出一个指南，告诉读者到哪里去寻找当某一特定的假定不被满足时将会发生的情况。

正如一位同僚向我指出的，当我们评阅他人的研究工作时，我们需要考虑研究者所作的假定是否切合他的数据和问题。经常出现的情况是，已发表的研究论文有赖于对问题和数据的隐含的假定。这些假定可能不正确，而所做出的估计却是以它们为依据的。显然，有见识的读者看到这种问题，应对研究工作持有怀疑态度，所以表 3.4 中所列的假定，为了指导我们自己的研究和评价别人的研究，都是一份参照目录。

有了这些背景，我们现在来研究 CLRM 就已准备就绪。具体地说，我们想来比较 OLS 的**统计性质**和先前讲的纯**数值性质**。OLS 的统计性质，是以 CLRM 的诸假定为依据的，并且在著名的**高斯-马尔可夫定理**（Gauss-Markov theorem）中被奉若神明。但在我们转到这个定理——为 OLS 广为应用提供理论说明的定理之前，我们需要先考虑最小二乘估计的**精度**（precision）或**标准误差**（standard errors）。

§3.3 最小二乘估计的精度或标准误差

由方程（3.1.6）和（3.1.7）显见，最小二乘估计是样本数据的函数。但因数据会从一个样本变到另一个样本，估计值也必定随之改变。因此需要有关于估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的“可靠性”或**精密度**的某种度量。在统计学中，一
77 个估计量的精密度由它的标准误（se）来衡量。[17]附录 3A 第 3A.3 节中证明，在高斯的假定下，OLS 估计量的标准误可求得如下：

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\operatorname{se}(\hat{\beta}_2)=\frac{\sigma}{\sqrt{\sum x_i^2}} \tag{3.3.2}$$

$$\operatorname{var}(\hat{\beta}_1)=\frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \tag{3.3.3}$$

$$\operatorname{se}(\hat{\beta}_1)=\sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\sigma \tag{3.3.4}$$

其中，var 代表方差，而 se 代表标准误，并且 σ^2 为常数或假定 4 中的 u_i 的共同方差。

除 σ^2 以外，上述方程中的一切变量均可从数据中估计出来。如附录 3A 第 3A.5 节所推导，σ^2 由下列公式来估算：

$$\hat{\sigma}^2=\frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

其中 $\hat{\sigma}^2$ 是真正的但未知的 σ^2 的 OLS 估计量，而表达式 $n-2$ 是被称为**自由度**（number of degrees of freedom，df）**的个数**，$\sum \hat{u}_i^2$ 则表示**残差平方的总和或剩（残）余平方和**（residual sum of squares，RSS）。[18]

一旦获知 $\sum \hat{u}_i^2$，$\hat{\sigma}^2$ 就可容易算出。$\sum \hat{u}_i^2$ 既可从（3.1.2）算出，也可从下面的表达式（证明见 3.5 节）计算：

$$\sum \hat{u}_i^2=\sum y_i^2-\hat{\beta}_2^2\sum x_i^2 \tag{3.3.6}$$

和方程（3.1.2）相比，方程（3.3.6）易于计算，而且并不要求计算每次观测的 $\hat{u}_i$，尽管这种计算本身有它的用处（在第 11 和 12 章中我们将看到这点）。

因

$$\hat{\beta}_2=\frac{\sum x_i y_i}{\sum x_i^2}$$

78 故，计算 $\sum \hat{u}_i^2$ 的另一表达式是：

$$\sum \hat{u}_i^2 = \sum y_i^2 - \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2} \tag{3.3.7}$$

顺便提一下，$\hat{\sigma}^2$ 的正的平方根为：

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \tag{3.3.8}$$

被称为**估计的标准误**（standard error of estimate）。它不外是 Y 对估计的回归线的离差的标准差。通常用于衡量所估计的回归线的“拟合优度（goodness of fit）”，这是 3.5 节要讨论的一个题目。

前些时候我们曾注明，对于给定的 X_i，σ^2 兼代表 u_i 和 Y_i 的（条件）方差。因此，估计的标准误也可叫做 u_i 和 Y_i 的（条件）标准差。当然，和平常一样，σ_Y^2 和 σ_Y 分别代表 Y 的无条件方差和无条件标准差。

注意 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差（以及它们的标准误）有如下特点。

1. $\hat{\beta}_2$ 的方差与 σ^2 成正比，而与 $\sum x_i^2$ 成反比。就是说，对给定的 σ^2，X 值的变化越大，$\hat{\beta}_2$ 的方差越小，从而 β_2 得以更大的精密度加以估计。简言之，给定 σ^2，X 值有大的变化时比没有大的变化时（回忆假定 8），β_2 的测算更为精准。而且，对给定的 $\sum x_i^2$，方差 σ^2 越大，$\hat{\beta}_2$ 的方差也越大。注意，随着样本含量 n 的增大，总和 $\sum x_i^2$ 中的项数将增加。β_2 的估计的精密度能够随 n 的增加而增加。(为什么?)

2. $\hat{\beta}_1$ 的方差与 σ^2 和 $\sum X_i^2$ 成正比，而与 $\sum x_i^2$ 和样本大小 n 成反比。

3. 由于 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 是估计量，它们不仅从一个样本变到另一个样本，而且对给定的一个样本，它们还可能是互相依赖的。这种依赖性将由它们之间的协方差来衡量。在附录 3A 第 3A.4 节中，我们证明：

$$\begin{aligned} \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) &= -\bar{X} \operatorname{var}(\hat{\beta}_2) \\ &= -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) \end{aligned} \tag{3.3.9}$$

79 因为，和任何变量的方差一样，$\operatorname{var}(\hat{\beta}_2)$ 总是正的，所以 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 之间的协方差是正是负与 $\bar{X}$ 的符号有关。如果 $\bar{X}$ 是正的，那么从公式可以看出，协方差将是负的。这时，如果斜率系数 β_2 **被过高估计**（即斜率被估计得太陡），则截距系数 β_1 将**被过低估计**（即截距被估计得太小）。以后（特别是讨论多重共线的这一章即第 10 章里）我们将看到，研究所估计的回归系数之间的协方差是有用的。

回归系数的估计量（$\hat{\beta}_1$ 和 $\hat{\beta}_2$）的方差又怎样能够用来判断这些估计的可靠性呢？这是一个统计推断问题，将在第 4 章和第 5 章里继续讨论。

§3.4 最小二乘估计量的性质：高斯-马尔可夫定理

如前面所指出的，在给定经典线性回归模型的假定条件下，最小二乘估计具有一些理想的或最优的性质。这些性质包含在著名的**高斯-马尔可夫定理**之中。为阐明此定理，需要考虑一个估计量的**最优线性无偏性质**（best linear unbiasedness property，BLUE）。[20] 如在附录 A 中所解释的，一个估计量，比方说，OLS 估计量 $\hat{\beta}_2$，是 β_2 的最优线性无偏估计量，假定下列条件成立：

1. 它是**线性的**，即它是一个随机变量，如回归模型中的因变量 Y 的线性函数。

2. 它是**无偏的**，即它的均值或期望值 $E(\hat{\beta}_2)$ 等于真实值 β_2。

3. 它在所有这样的线性无偏估计量一类中有最小方差；有最小方差的无偏估计量叫做**有效**（或优效）**估计量**（efficient estimator）。

在回归的论述中，可以证明，OLS 的估计量是 BLUE。这就是著名的高斯-马尔可夫定理的精髓，可叙述为：

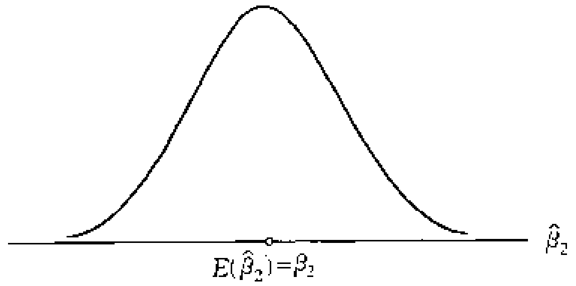
> **高斯-马尔可夫定理**：在给定经典线性回归模型的假定下，最小二乘估计量，在无偏线性估计量一类中，有最小方差，就是说，它们是 BLUE。

在附录 3A 第 3A.6 中勾勒出这一定理的证明。随着本书的进展，高斯-
80 马尔可夫定理的全部含义将渐见分明。这里只须指明，该定理在理论上和实际上都是重要的。[21]

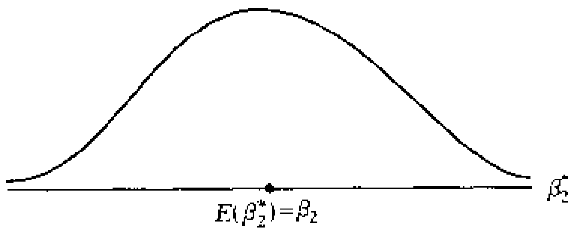
所有这些可以通过图 3.8 加以解释。

在图 3.8（a）中我们展示 OLS 估计量 $\hat{\beta}_2$ 的**抽样分布**，即在重复试验中 $\hat{\beta}_2$ 所取的值的分布（回忆表 3.1）。为方便起见，我们假定了 $\hat{\beta}_2$ 是对称分布的（进一步的讨论见第 4 章）。如图所示，$\hat{\beta}_2$ 的均值等于 β_2。这时我们说 $\hat{\beta}_2$
81 是 β_2 的一个无偏估计量。在图 3.8（b）中我们展示了来自另一种方法（不是 OLS 法）的 β_2 的一个估计量 β_2^* 的抽样分布。为方便起见，仍假定 β_2^* 和 $\hat{\beta}_2$ 同样是无偏的，即其均值或期望值等于 β_2。再假定 $\hat{\beta}_2$ 和 β_2^* 都是线性估计量，即它们都是 Y 的线性函数。试问你会选取哪一个估计量，$\hat{\beta}_2$ 还是 β_2^*？

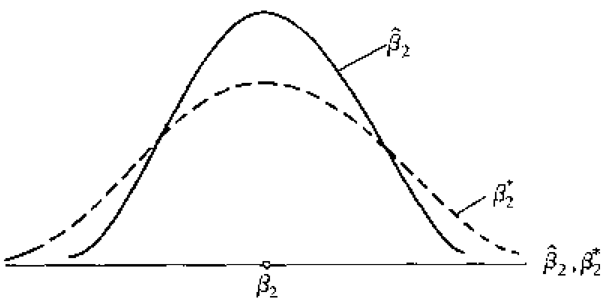
要回答这一问题，把两个图形重叠起来，如图 3.8（c）那样。显然，尽管 $\hat{\beta}_2$ 和 β_2^* 两者都是无偏的，但 β_2^* 的分布比 $\hat{\beta}_2$ 的分布围绕均值扩散得更广。换句话说，β_2^* 的方差比 $\hat{\beta}_2$ 的方差要大。既然两个估计量都是线性和无偏的，人们就会选择有较小方差的估计量，因为它比另一个估计量更可能



(a) $\hat{\beta}_2$ 的抽样分布



(b) β_2^* 的抽样分布



(c) $\hat{\beta}_2$ 和 β_2^* 的抽样分布

图 3.8　OLS 估计量 $\hat{\beta}_2$ 和其他估计量 β_2^* 的抽样分布

接近β_2。简单地说，人们会选择 BLUE 估计量。

方才讨论的统计性质称为**有限样本性质**（finite sample properties）：这些性质不管估计量所依据的样本大小如何都能成立。以后我们还有机会考虑**渐近性质**（asymptotic properties），即仅当样本非常大（技术性地说，无穷大）时才会成立的性质。附录 A 对估计量的有限样本和大样本性质有一个一般性的讨论。

§3.5　判定系数 r^2："拟合优度"的一个度量

直到现在，我们考虑的是估计回归系数的问题，它们的标准误以及其他一些性质。现在我们来考虑对一组数据所拟合的回归线的**拟合优度**（goodness of fit）。也就是，我们要说出这条样本回归线对数据拟合得有多么好。由图 3.1 看出，如果全部观测点都落在样本回归线上，我们就得到一个"完

美”的拟合。但是这种情形很少发生。一般的情形是，总有一些正的 $\hat{u}_i$ 和一些负的 $\hat{u}_i$。我们所能希望的仅是这些围绕着回归线的残差尽可能小。**判定系数** r^2（双变量情形）或 R^2（多变量情形）就是告诉人们这条样本回归线对数据的拟合有多么好的一个总度量。

82 在我们说明 r^2 怎样计算之前，先通过图形对 r^2 作一个直观的解释。这一图形称为**维恩图**（Venn diagram）或**巴伦坦**（Ballentine），如图 3.9 所示。[22]

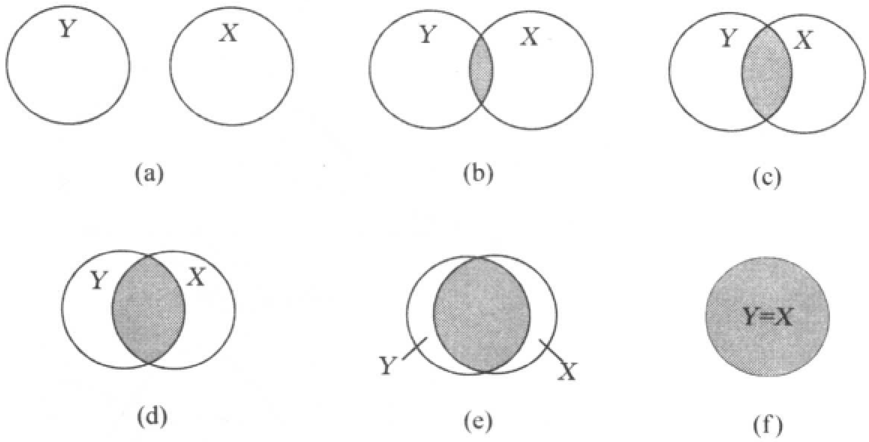


图 3.9　通过韦恩图看 r^2：(a) $r^2=0$；(f) $r^2=1$

在此图中，圆圈 Y 代表因变量 Y 的变异，圆圈 X 代表解释变量 X 的变异。[23] 两圆的重叠部分代表 Y 的变异可由 X 的变异（比如说，通过 OLS 回归）来解释的程度。重叠程度越大，Y 的变异被 X 解释得越多。r^2 不外是这一重叠的一个数值度量。在此图中，从左到右，重叠部分渐增，即 Y 的变异被 X 解释的部分依次变大。在无重叠时，r^2 显然为零，但若全部重叠则 r^2 为 1。此时 Y 的变异百分之百地被 X 解释了。以下我们将要简短地展示，r^2 落在 0 和 1 之间。

计算 r^2 的步骤如下：回顾

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{2.6.3}$$

或利用（3.1.13）和（3.1.14）写成离差形式：

$$y_i = \hat{y}_i + \hat{u}_i \tag{3.5.1}$$

两边平方并对样本求和，便得：

83

$$\begin{aligned}\sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i\hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \\ &= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2\end{aligned} \tag{3.5.2}$$

因为 $\sum \hat{y}_i\hat{u}_i = 0$（为什么？）且 $\hat{y}_i = \hat{\beta}_2 x_i$。

出现在（3.5.2）中的各项平方和可描述为：$\sum y_i^2 = \sum(Y_i - \bar{Y})^2 =$ 实测的 Y 值围绕其均值的总变异，称为**总平方和**（Total Sum of Squares, TSS）。$\sum \hat{y}_i^2 = \sum(\hat{Y}_i - \bar{\hat{Y}})^2 = \sum(\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 \sum x_i^2 =$ 估计的 Y 值围绕其均值（$\bar{\hat{Y}} = \bar{Y}$）的变异，可适当地称为回归（即来自解释变量的）平方和，或者说，由回归解释的平方和，或简称**解释平方和**（Explained Sum of Squares, ESS）。$\sum \hat{u}_i^2 =$ 残差（或剩余）或**未被解释**的围绕回归线的 Y 值的变异，或简称**残差平方和**（Residual Sum of Squares, RSS）。这样，（3.5.2）就是：

$$\text{TSS} = \text{ESS} + \text{RSS} \tag{3.5.3}$$

这说明 Y 的观测值围绕其均值的总变异可分解为两部分，一部分来自回归线，而另一部分则来自随机势力，因为并不是所有的实际 Y 观测值都落在所拟合的直线上。从几何意义上看，可画出图 3.10。

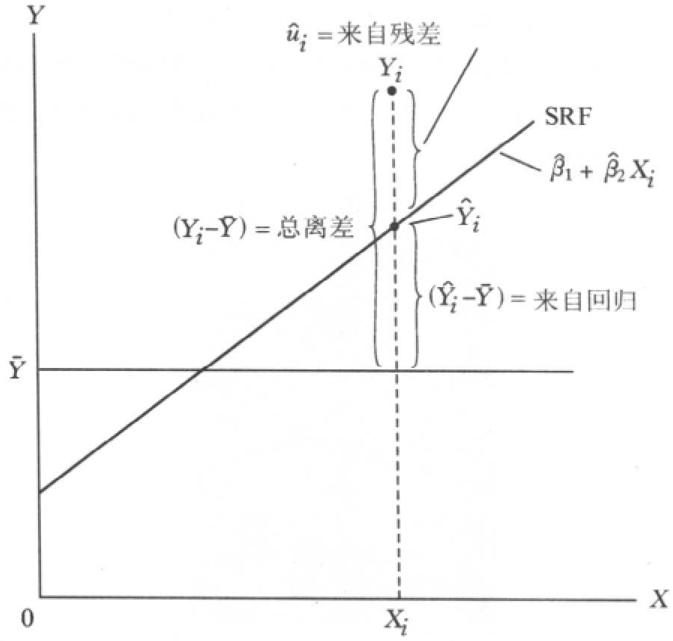


图 3.10　Y_i 的变异分解为两个成分

84　现用 TSS 除（3.5.3）的两边得：

$$1 = \frac{\text{ESS}}{\text{TSS}} + \frac{\text{RSS}}{\text{TSS}}$$
$$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} \tag{3.5.4}$$

我们定义 r^2 为：

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\text{ESS}}{\text{TSS}} \tag{3.5.5}$$

或者写成另一形式：

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{3.5.5a}$$

如上定义的数量 r^2 称之为（样本）**判定系数**，它是对回归线拟合优度的最为常用的度量，一般来说，r^2 测度了在Y 的总变异中由回归模型解释的那个部分所占的比例或百分比。

注意 r^2 的两个性质：

1. 它是一个非负量。(为什么?)

2. 它的界限为 $0 \leqslant r^2 \leqslant 1$。等于 1 的 r^2 意味着一个完美的拟合，即对每个 i 都有 $\hat{Y}_i = Y_i$。另一方面，等于 0 的 r^2 意味着回归值与回归元之间无任何关系（即 $\hat{\beta}_2 = 0$）。这时，如（3.1.9）所表明，$\hat{Y}_i = \hat{\beta}_1 = \bar{Y}$，就是说，对任一 Y 值的最优预测值都是它的均值。从而回归线平行于 X 轴。

虽然 r^2 可按（3.5.5）所给的定义直接计算，但利用下面的公式能更快捷地求得：

$$r^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} = \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right) \tag{3.5.6}$$

85 若用样本大小 n（或者对小样本用 $n-1$）同除（3.5.6）中的分子与分母，就有：

$$r^2 = \hat{\beta}_2^2 \left(\frac{S_x^2}{S_y^2} \right) \tag{3.5.7}$$

其中 S_y^2 和 S_x^2 分别是 Y 和 X 的样本方差。

由于 $\hat{\beta}_2 = \sum x_i y_i / \sum x_i^2$，方程（3.5.6）还可表达成：

$$r^2 = \frac{\left(\sum x_i y_i \right)^2}{\sum x_i^2 \sum y_i^2} \tag{3.5.8}$$

这是一个容易计算的表达式。

给定 r^2 的定义，可将上面讨论过的 ESS 和 RSS 表示如下：

$$\begin{aligned} \text{ESS} &= r^2 \cdot \text{TSS} \\ &= r^2 \sum y_i^2 \end{aligned} \tag{3.5.9}$$

$$\begin{aligned} \text{RSS} &= \text{TSS} - \text{ESS} \\ &= \text{TSS}(1 - \text{ESS}/\text{TSS}) \\ &= \sum y_i^2 \cdot (1 - r^2) \end{aligned} \tag{3.5.10}$$

因此可写：

$$\begin{gathered} \text{TSS} = \text{ESS} + \text{RSS} \\ \sum y_i^2 = r^2 \sum y_i^2 + (1 - r^2) \sum y_i^2 \end{gathered} \tag{3.5.11}$$

这是我们以后非常有用的表达式。

与 r^2 关系紧密但概念与 r^2 很不相同的一个数量是**相关系数**（coefficient of correlation），它测出两个变量之间的关联度，如第 1 章所述，它既可由：

$$r = \pm \sqrt{r^2} \tag{3.5.12}$$

得出，也可从它的定义：

$$\begin{aligned} r &= \frac{\sum x_i y_i}{\sqrt{\left(\sum x_i^2\right)\left(\sum y_i^2\right)}} \\ &= \frac{n \sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{\left[n \sum X_i^2 - \left(\sum X_i\right)^2\right]\left[n \sum Y_i^2 - \left(\sum Y_i\right)^2\right]}} \end{aligned} \tag{3.5.13}$$

86 算出。该定义被称为**样本相关系数**（Sample Correlation Coefficient）。[24]

下面是 r 的一些性质（见图 3.11）：

1. 它可正可负，其符号与（3.5.13）的分子即两变量的协变异（covariation）的符号相同。

2. 它落在极限 -1 和 $+1$ 之间；即 $-1 \leqslant r \leqslant 1$。

3. 它有对称性；即 X 与 Y 之间的相关系数（r_{XY}）和 Y 与 X 之间的相关系数（r_{YX}）相同。

4. 它与原点和尺度都无关；即如果定义 $X_i^* = aX_i + c$ 和 $Y_i^* = bY_i + d$，
87 其中 $a>0$，$b>0$，且 c 和 d 都是常数，则 X^* 和 Y^* 之间的 r 无异于原始变量 X 与 Y 之间的 r。

5. 如果 X 与 Y 统计上独立（参看附录 A 中的定义），则它们之间的相关系数为零；但反过来 $r=0$ 不等于说两个变量是独立的。换句话说，**零相关**

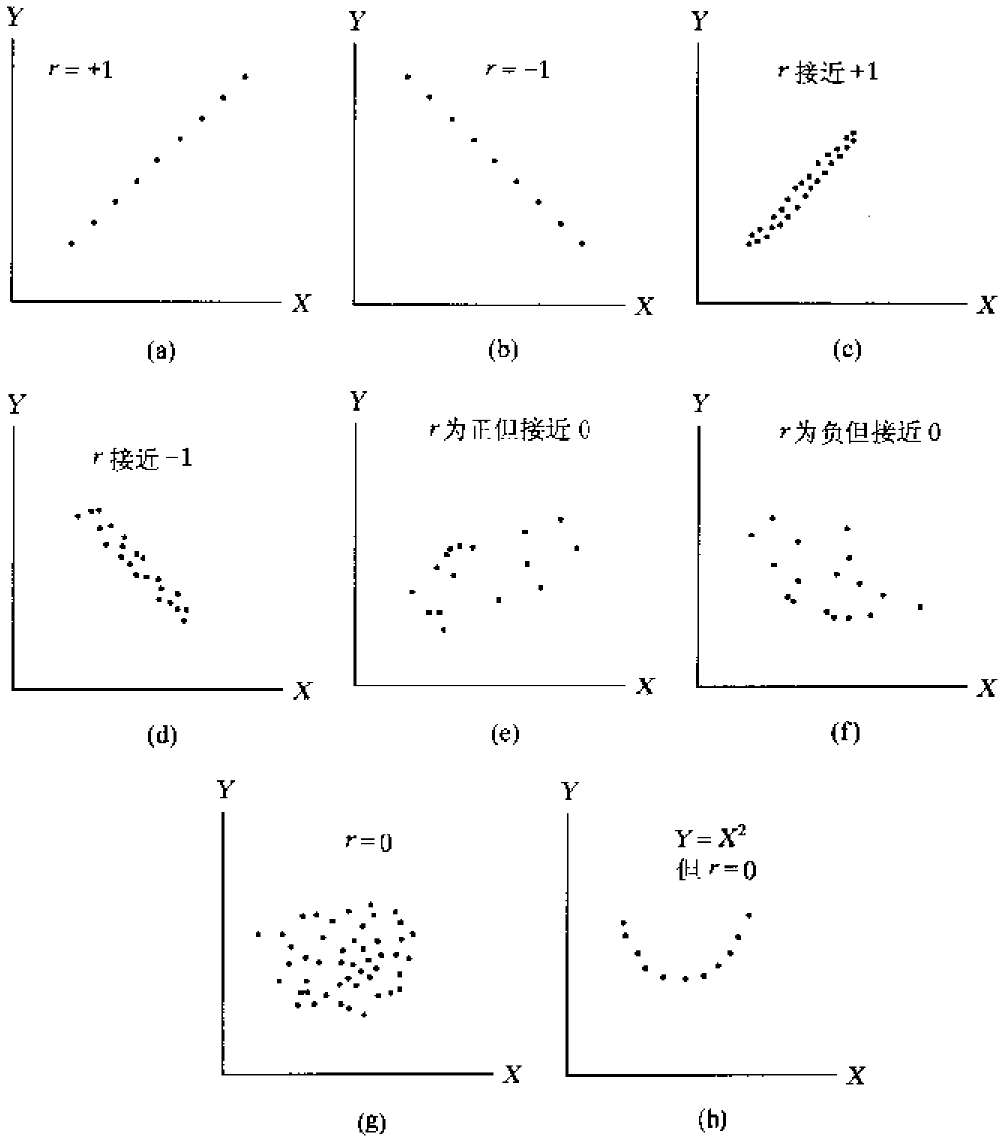


图 3.11　相关式样

资料来源：Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1978, p.86.

并不一定意味着独立性。[见图 3.11 (h)。]

6. 它仅是线性关联或线性相依的一个度量；它不能用于描述非线性关系。例如，在图 3.11 (h) 中 $Y = X^2$ 是一个准确的关系式，然而 r 为零。(为什么?)

7. 如第1章中所指出的，虽然它是两个变量之间的线性关联的一个度量，却不一定有因果关系的含义。

在回归分析中，r^2 是一个比 r 更有意义的度量，因为前者告诉我们在因变量的变异中由解释变量解释的部分占怎样一个比例，因而对一个变量的变异在多大程度上决定另一个变量的变异，提供了一个总的度量。而后者则没有这种价值。[25]此外，如下面我们将会看到的，在一个复回归模型中，对 r $(=R)$ 做的解释有多大价值，是个疑问。不管怎样，等到第 7 章，我们

还要进一步讨论 r^2。

顺便指出，前面定义的 r^2 还可作为实测的 Y_i 与估计的 Y_i（即 $\hat{Y}_i$）之间的相关系数的平方来计算。也就是可利用（3.5.13）把它写为：

$$r^2 = \frac{\left[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})\right]^2}{\sum(Y_i - \bar{Y})^2 \sum(\hat{Y}_i - \bar{Y})^2}$$

即：

$$r^2 = \frac{\left(\sum y_i \hat{y}_i\right)^2}{\left(\sum y_i^2\right)\left(\sum \hat{y}_i^2\right)} \tag{3.5.14}$$

其中 Y_i = 实际的 Y，$\hat{Y}_i$ = 估计的 Y，而 $\bar{Y} = \bar{\hat{Y}} = Y$ 的均值。有关证明见习题 3.15。表达式（3.5.14）解释了为什么把 r^2 描述成为拟合优度的一个度量，这是因为它告诉我们 Y 的估计值和它的真实值相距多近。

§ 3.6 一个数值例子

我们一直在用引言中讨论过的凯恩斯消费函数来阐明所讲的计量经济理论。记得凯恩斯曾说："基本的心理定律是……作为一种规律和平均情形，
88 人们随着他们收入的增加而倾向于增加其消费，但不如收入增加那么多。"这就是说，边际消费倾向（MPC）大于零而小于 1。虽然凯恩斯没有明确规定，消费与收入之间的关系的准确函数形式，为简单起见，我们假定这个关系式如同（2.4.2）那样是线性的。作为对凯恩斯消费函数的一种检验，我们利用表 2.4 中的数据，并为方便起见，将其复制于表 3.2。为估计回归系数，其标准误等等所需的原始数据均列在表 3.3 中。根据这些原始数据得到如下计算，建议读者加以核实。

表 3.2　　每周家庭消费支出 Y 和每周家庭收入 X 的假想数据

Y（美元）	X（美元）	Y（美元）	X（美元）
70	80	115	180
65	100	120	200
90	120	140	220
95	140	155	240
110	160	150	260

89 **表 3.3** **根据表 3.2 的原始数据**

	Y_i (1)	X_i (2)	Y_iX_i (3)	X_i^2 (4)	$x_i=X_i-\bar{X}$ (5)	$y_i=Y_i-\bar{Y}$ (6)	x_i^2 (7)	x_iy_i (8)	$\hat{Y}_i$ (9)	$\hat{u}_i=Y_i-\hat{Y}_i$ (10)	$\hat{Y}_i\hat{u}_i$ (11)
	70	80	5 600	6 400	−90	−41	8 100	3 690	65.181 8	4.818 1	314.052 4
	65	100	6 500	10 000	−70	46	4 900	3 220	75.363 6	−10.363 6	−781.038 2
	90	120	10 800	14 400	−50	−21	2 500	1 050	85.545 4	4.454 5	381.062 0
	95	140	13 300	19 600	−30	16	900	480	95.727 2	−0.727 2	−69.612 8
	110	160	17 600	25 600	−10	−1	100	10	105.909 0	4.090 9	433.263 1
	115	180	20 700	32 400	10	4	100	40	116.090 9	−1.090 9	−126.643 4
	120	200	24 000	40 000	30	9	900	270	125.272 7	−6.272 7	−792.070 8
	140	220	30 800	48 400	50	29	2 500	1 450	136.454 5	3.545 4	483.785 8
	155	240	37 200	57 600	70	44	4 900	3 080	145.636 3	8.363 6	1 226.407 3
	150	260	39 000	67 600	90	39	8 100	3 510	156.818 1	−6.818 1	−1 069.201 4
总和	1 110	1 700	205 500	322 000	0	0	33 000	16 800	1 109.999 5 ≈1 110.0	0	0.004 0 ≈0.0
均值	111	170	nc	nc	0	0	nc	nc	110	0	0

$$\hat{\beta}_2=\frac{\sum x_iy_i}{\sum x_i^2}=16\ 800/33\ 000=0.509\ 1$$

$$\hat{\beta}_1=\bar{Y}-\hat{\beta}_2\bar{X}=111-0.509\ 1\ (170)=24.454\ 5$$

注：≈表示"近似地等于"；nc 表示"未作计算"（not computed）。

$$\begin{aligned}&\hat{\beta}_1=24.454\ 5 \quad \mathrm{var}(\hat{\beta}_1)=41.137\ 0 \quad 和 \quad \mathrm{se}(\hat{\beta}_1)=6.413\ 8\\&\hat{\beta}_2=0.509\ 1 \quad \mathrm{var}(\hat{\beta}_2)=0.001\ 3 \quad 和 \quad \mathrm{se}(\hat{\beta}_2)=0.035\ 7\end{aligned} \tag{3.6.1}$$

$$\mathrm{cov}(\hat{\beta}_1,\hat{\beta}_2)=-0.217\ 2 \quad \hat{\sigma}^2=42.159\ 1$$

$$r^2=0.962\ 1 \quad r=0.980\ 9 \quad \mathrm{df}=8$$

因此所估回归线是：

$$\hat{Y}_i=24.454\ 5+0.509\ 1X_i \tag{3.6.2}$$

其几何图形见图 3.12。

由第 2 章，对 SRF［方程（3.6.2）］及相应的回归线可解释如下：回归线上的每一点都给出与选定的 X 值相对应的 Y 期望值或均值的一个估计值；即 $\hat{Y}_i$ 是 $E(Y|X_i)$ 的估计值。代表回归线的斜率的 $\hat{\beta}_2=0.509\ 1$，表示在 80 美元到 260 美元这个 X 的样本范围内，X 每增加 1 美元，平均每周消费支出估计增加 51 美分。代表回归线的截距的 $\hat{\beta}_1=24.454\ 5$，则表示每周收入

90 为零时的每周消费支出的平均水平，不过这是对截距项的一种机械式的解释。在回归分析中，对截距项的这种字面的解释也许没有什么意义，尽管人们可以称辩，没有任何收入的家庭（由于失业或被解雇等原因），仍要通过借债或动用储蓄维持某一最低消费支出水平。但因 X 值的样本变化范围（极差）常常不包括零这样一个观测值，通常都要借助于常识来解释截距项。

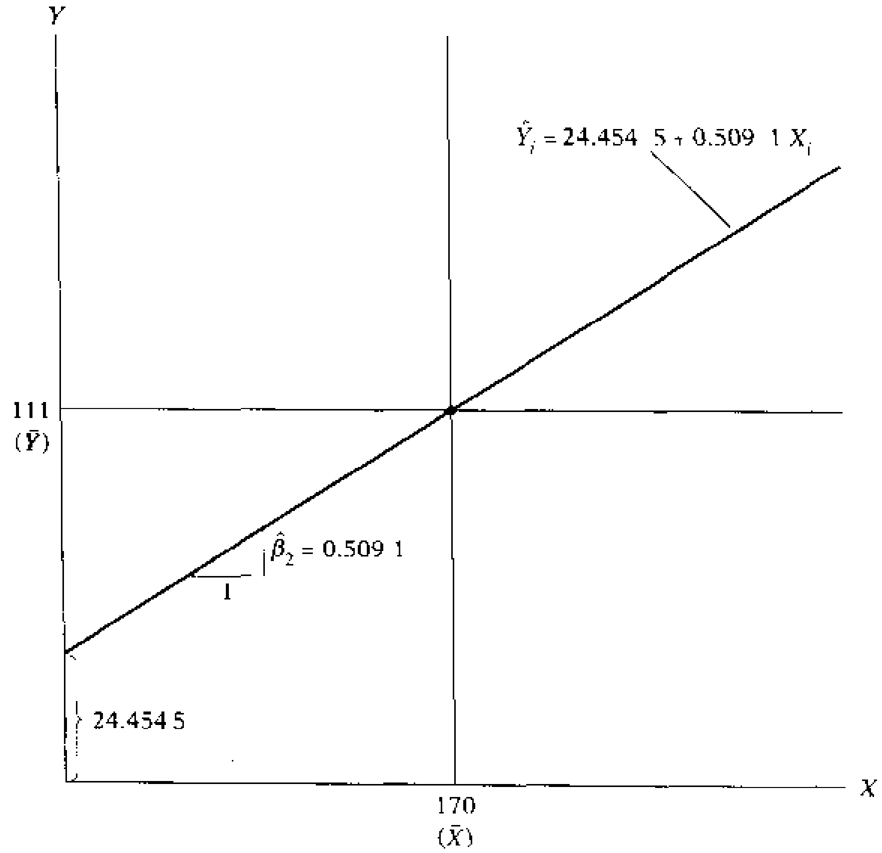


图 3.12　根据表 3.2 的数据得到的样本回归线

也许最好把截距项解释为所有从回归模型中略去的变量对 Y 的平均影响。r^2 值等于0.962 1是说，约有 96% 的每周消费支出的变异，能由收入来说明。由于 r^2 充其量等于 1，所测算的 r^2 表明样本回归线对数据拟合得非常好。[26] 相关系数为0.980 9表示消费支出和收入两个变量是高度正相关的。所估计的回归系数的标准误将在第 5 章再作解释。

§ 3.7　说明性例子

例 3.1　美国消费—收入关系，1982—1996

让我们回到引言的表 I.1 中给出的消费和收入数据。我们已经在图 I.3

及估计的回归线（I.3.3）中展示了这些数据。现在我们给出基本的 OLS 回归结果。（此结果得自统计软件 Eview 3。）注：Y 表示个人消费支出（PCE），X 表示国内生产总值（GDP），均以 1992 年 10 亿美元计。此例中的数据是时间序列数据。

$$\hat{Y}_i = -184.078\,0 + 0.706\,4X_i \qquad (3.7.1)$$

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= 2\,140.170\,7 & \text{se}(\hat{\beta}_1) &= 46.261\,9 \\ \text{var}(\hat{\beta}_2) &= 0.000\,061 & \text{se}(\hat{\beta}_2) &= 0.007\,827 \\ r^2 &= 0.998\,406 & \hat{\sigma}^2 &= 411.491\,3 \end{aligned}$$

方程（3.7.1）是总量（即对整个国家而言）凯恩斯消费函数。如此方程所示，边际消费倾向约为 0.71，它表明如果收入增加 1 美元，平均个人消费支出（PCE）约上升 71 美分。按凯恩斯理论，MPC 小于 1。截距值 −184告诉我们，如果收入为零，PCE 将约为 −1 840亿美元。当然，由于收入值为零远超出了我们考虑的范围之外，又不代表一种可能的结果（见表 I.1），所以对截距项做这种机械的解释在目前没有任何经济意义。如我们多数情况下所见，截距项通常都没有多少经济意义。因此，尽管我们偶尔在某些说明性例子中看到截距项很有意义，但通常都不会很有意义。更有意义的值是斜率系数，即本例中的 MPC。

r^2 的值为0.998 4意味着，PCE 变化的 99% 都可由 GDP 的变化来解释。由于 r^2 最多等于 1，所以我们可以说，图 I.3 所示的回归线（3.7.1）对数据拟合得非常好；如图中所见，实际数据点十分密集地散布在估计的回归线周围。如我们将在本书中所见到的那样，对时间序列数据的回归通常都能得到很高的 r^2 值。在自相关一章中，我们将看到这一现象背后的原因。

91

例 3.2 印度的食物支出

查阅习题 2.15 中表 2.8 所给出的数据。数据涉及印度 55 个农户的一个样本。本例中的回归子是食物支出，而回归元则是收入的代理变量总支出，都以卢比为单位。因此本例中的数据为横截面数据。

基于给定数据，我们得到如下回归：

$$\text{食物支出}_i = 94.208\,7 + 0.436\,8\,\text{总支出}_i \qquad (3.7.2)$$

$$\begin{aligned} \text{var}(\hat{\beta}_1) &= 2\,560.940\,1 & \text{se}(\hat{\beta}_1) &= 50.856\,3 \\ \text{var}(\hat{\beta}_2) &= 0.006\,1 & \text{se}(\hat{\beta}_2) &= 0.078\,3 \\ r^2 &= 0.369\,8 & \hat{\sigma}^2 &= 4\,469.691\,3 \end{aligned}$$

我们从（3.7.2）可见，如果总支出增加 1 卢比，那么平均食物支出将增加 44 派沙（1 卢比 = 100 派沙）。如果总支出为零，则平均的食物支出为 94 卢比。同样，对截距项的这种机械解释可能没有意义。但在本例中，人们可以认为，即使总支出为零（如因为失业），人们仍可能通过借贷或动用储蓄来在某个最低水平维持食物支出。

r^2 值约为 0.37 表明，食物支出变动中的 37% 由总支出来解释。看上去

这是一个相当低的值，但如我们通书所见，在横截面数据中，通常获得低 r^2 值都可能是因为样本单位的分散性所致。我们将在异方差一章（第11章）中讨论这一专题。

例 3.3　平均小时工资与受教育水平之间的关系

我们在表2.6中看到有关平均小时工资和受教育水平（以读书年数度量）的数据。利用那些数据，如果我们将平均小时工资（Y）对受教育水平（X）做回归，那我们得到如下结果：

$$\hat{Y}_i = 0.743\ 7 + 0.641\ 6X_i \tag{3.7.3}$$
$$\mathrm{var}(\hat{\beta}_1) = 0.698\ 0 \qquad \mathrm{se}(\hat{\beta}_1) = 0.835\ 5$$
$$\mathrm{var}(\hat{\beta}_2) = 0.004\ 4 \qquad \mathrm{se}(\hat{\beta}_2) = 0.066\ 4$$
$$r^2 = 0.894\ 4 \qquad \hat{\sigma}^2 = 0.804\ 0$$

如回归结果所示，受教育水平与工资之间存在正相关联系，这是一个无足为奇的结论。每多读1年的书，平均小时工资约增加64美分。截距项为正的那些组可能没有经济意义。r^2 值表明，平均小时工资变化中的89%可由受教育水平来解释。对横截面数据而言，这么高的 r^2 相当不同寻常。

§3.8　关于蒙特卡罗实验的一个注记

本章说过，在CLRM的假定下，最小二乘估计量有某些良好的、可归结为BLUE性质的统计特性。在本章的附录中，我们给出了这一性质较正式的证明。但实际上我们怎样才能知道这一BLUE性质是否成立？比如，怎样能知道OLS估计量是否无偏？所谓的**蒙特卡罗**实验，基本上是一种计算机模拟或抽样实验法。也许该实验能提供这一答案。

92

为了介绍基本概念，且考虑我们的双变量PRF：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3.8.1}$$

蒙特卡罗实验的程序如下：

1. 假定参数有如下的真值：$\beta_1 = 20$ 和 $\beta_2 = 0.6$。

2. 选定样本大小，比方说 $n = 25$。

3. 每次观测固定一个 X 值，这样共有25个 X 值。

4. 从一张随机数表选出25个数值，且称它们为 u_i（当今的统计包软件大多含有内在随机数发生器）。[28]

5. β_1，β_2，X_i 和 u_i 已知，便可利用（3.9.1）得到25个 Y_i 值。

6. 现在利用如此产生的25个 Y_i 值，对在第3步中所选的25个 X 值做回归，求得最小二乘估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。

7. 假使重复这一实验99次，每次都用相同的 β_1、β_2 和 X 值。当然，u_i 在每次实验中有所变化，因而在一共100次的实验中，就产生 β_1 和 β_2 的各100个值。（实际上，人们做过许多这样的实验，有时重复到1 000～2 000次。）

8. 取这100个估计值的均值，并称它们为 $\bar{\hat{\beta}}_1$ 和 $\bar{\hat{\beta}}_2$。

9. 如果这些均值和在第1步中所假定的 β_1 和 β_2 的真值（20，0.6）相差不多，那么蒙特卡罗法就"证实"了最小二乘估计量确实是无偏的。回想在CLRM假定下，$E(\hat{\beta}_1)=\beta_1$ 和 $E(\hat{\beta}_2)=\beta_2$。

以上步骤刻画了蒙特卡罗试验的一般性质。这种实验常被用来研究各种估计总体参数方法的统计性质。它们在研究小样本或有限样本的估计量的性态时尤其有用。这些实验对于彻底掌握**重复抽样**（repeated sampling）的概念也是绝好的手段。重复抽样的概念，如在第5章中我们将会看到的，是大部分经典统计推断的基础。我们将通过课堂作业的一些练习题提供蒙特卡罗实验的若干例子。（见习题3.27。）

§3.9 要点与结论

本章讨论的重要问题和概念可概括如下：

1. 回归分析的基本构架是 **CLRM**。

2. CLRM是以一组假定为基础的。

3. 基于这些假定，最小二乘估计量便具有一些可归结为高斯-马尔可夫定理的性质。该定理说，在线性无偏估计量一类中，最小二乘估计量有最小的方差。简单地说，这些估计量是BLUE。

4. OLS估计量的精度由其**标准误（差）**来衡量，在第4和5章中，我们将看到这些标准误怎样用来推断总体参数——β系数。

5. 回归模型的总拟合优度由**判定系数** r^2 来衡量。它表明在因变量或回归子的变异中，由解释变量或回归元解释的部分所占的比例。r^2 落在0与1之间；它愈靠近1，回归拟合得愈好。

6. 与判定系数相连的一个概念是**相关系数** r。它是两个变量之间的线性关联的一个度量，并落在 −1与+1之间。

7. CLRM仅是一个理论上的构想或抽象。因为它是以一组严谨的或者说"不真实"的假定作为依据的。但是这种抽象不管在哪一知识领域的研究中，在其初始阶段都常常是必需的。一旦掌握了CLRM，就能发现如果某一或某些假定不成立，将会出现什么情况。本书的第1篇专门讨论CLRM。其他篇则考虑CLRM的精细化。表3.4是一幅学习索引图。

表 3.4 **违反 CLRM 假定的种种情况**

假定编号	违反类型	在何处查询
1	对参数非线性	第 14 章
2	随机回归元（一个或多个）	第 2 篇篇首语
3	u_i 有非零均值	第 2 篇篇首语
4	异方差性	第 11 章
5	干扰自相关	第 12 章
6	干扰与回归元有非零协方差	第 4 篇篇首语
7	样本观测次数小于回归元个数	第 10 章
8	回归元缺少变异	第 10 章
9	设定偏误	第 13、14 章
10	多重共线性	第 10 章
11*	干扰的非正态性	第 2 篇篇首语

* 干扰项 u_i 为正态分布的假定，不属于 CLRM。进一步的讨论见第 4 章。

习 题

94 **问答题**

3.1 给定下表第 1 列中的假定，说明第 2 列中的假定是与之等效的。

关于经典模型的假定

(1)	(2)
$E(u_i \mid X_i)=0$	$E(Y_i \mid X_i)=\beta_1+\beta_2 X$
$\operatorname{cov}(u_i, u_j)=0 \quad i \neq j$	$\operatorname{cov}(Y_i, Y_j)=0 \quad i \neq j$
$\operatorname{var}(u_i \mid X_i)=\sigma^2$	$\operatorname{var}(Y_i \mid X_i)=\sigma^2$

3.2 说明表 3.1 的第 1 个实验所用的估计值 $\hat{\beta}_1=1.572$ 和 $\hat{\beta}_2=1.357$ 事实上是 OLS 估计量。

3.3 按照马林伏特（Malinvaud）（参看章末注 11）的意见，假设 $E(u_i \mid X_i)=0$ 是相当重要的。为了看到这点，考虑 PRF：$Y=\beta_1+\beta_2 X_i+u_i$。现区分两种情形：（ⅰ）$\beta_1=0$，$\beta_2=1$ 及 $E(u_i)=0$；和（ⅱ）$\beta_1=1$，$\beta_2=0$ 及 $E(u_i)=(X_i-1)$。然后对这两种情形求以 X 为条件的 PRF 的数学期望，看你是否同意马林伏特关于假定 $E(u_i \mid X_i)=0$ 的重要性的观点。

3.4 考虑样本回归：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

求在约束条件（ⅰ）$\sum \hat{u}_i = 0$ 和（ⅱ）$\sum \hat{u}_i X_i = 0$ 下的估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$，并阐明它们无异于（3.1.6）和（3.1.7）中所给出的最小二乘估计量。这种求估计量的方法叫做**类比原理**（analogy principle）。试述施加约束条件（ⅰ）和（ⅱ）的直觉理由。（提示：回顾关于 u_i 的 CLRM 假定。）顺便指出，估计未知参数的类比原理又叫做**矩法**（method of moments）即用样本矩（例如样本均值）去估计总体矩（例如总体均值）。如在附录 A 中所指出的，矩是概率分布的一个摘要统计量，诸如期望值和方差。

3.5 阐明由（3.5.5）定义的 r^2 落在 0 与 1 之间。你可以利用柯西-施瓦茨（Cauchy-Schwarz）不等式，就是，对任意随机变量 X 和 Y，下列关系式总是成立的：

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

3.6 令 $\hat{\beta}_{YX}$ 和 $\hat{\beta}_{XY}$ 分别为 Y 对 X 回归和 X 对 Y 回归中的斜率。说明

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

其中 r 为 X 与 Y 之间的相关系数。

3.7 假使在题 3.6 中 $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$。那么求 Y 对 X 的回归和求 X 对 Y 的回归有什么差别？请细心解释。

95 **3.8** 斯皮尔曼（Spearman）等级相关系数 r_s 的定义如下：

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

其中 d = 编排给同一单元或现象的等级差

n = 参与等级编排的单元或现象个数

试从（3.5.13）中 r 的定义推出 r_s。提示：从 1 到 n 将 X 和 Y 编排等级。注意 X 和 Y 的等级和各为 $n(n+1)/2$，因而它们的均值都是 $(n+1)/2$。

3.9 建立如下的双变量 PRF：

$$模型\ \mathrm{I}: Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$模型\ \mathrm{II}: Y_i = \alpha_1 + \alpha_2(X_i - \bar{X}) + u_i$$

a. 求 β_1 和 α_1 的估计量。它们是否相同？它们的方差是否相同？

b. 求 β_2 和 α_2 的估计量，它们是否相同？它们的方差是否相同？

c. 如果模型 Ⅱ 比模型 Ⅰ 好，好在哪里？

3.10 假使你做了如下的回归：

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$$

其中 y_i 和 x_i 照例是偏离它们各自均值的离差，问 $\hat{\beta}_1$ 将取何值？

为什么？$\hat{\beta}_2$ 会不会和方程（3.1.6）的 $\hat{\beta}_2$ 一样？为什么？

3.11 令 $r_1 = n$ 为对（Y_i，X_i）值的相关系数，而 $r_2 = n$ 为对（$aX_i + b$，$cY_i + d$）值的相关系数，其中 a，b，c 和 d 为常数。说明 $r_1 = r_2$，从而证实相关系数对尺度和原点的改变有不变性。

提示：应用（3.5.13）中所给的 r 定义。

注：运算 aX_i，$X_i + b$ 和 $aX_i + b$ 分别叫做尺度变换，原点变换和尺度与原点同时变换。

3.12 如果 n 对（X_i，Y_i）值的相关系数是正的，试判别以下各个陈述的对错：

a.（$-X_i$，$-Y_i$）之间的 r 也是正的。

b.（$-X_i$，Y_i）之间以及（X_i，$-Y_i$）之间的 r 可正可负。

c. 斜率系数 β_{yx} 和 β_{xy} 都是正的，其中 β_{yx} = Y 对 X 的回归斜率系数，而 β_{xy} = X 对 Y 回归的斜率系数。

3.13 如果 X_1，X_2 和 X_3 是有同样方差但互不相关的变量，试说明 $X_1 + X_2$ 和 $X_2 + X_3$ 之间的相关系数等于 1/2。为什么这个相关系数不是零？

3.14 假使在回归 $Y_i = \beta_1 + \beta_2 X_i + u_i$ 中我们用常数 2 去乘每一 X 值。问这会不会改变 Y 的残差及拟合值？为什么？如果我们给每个 X 值都加上一个常数 2，又会怎样？

96 **3.15** 说明（3.5.14）事实上就是判定系数。

提示：应用(3.5.13)中所给的 r 定义并回忆 $\sum y_i\hat{y}_i = \sum(\hat{y}_i + \hat{u}_i)\hat{y}_i = \sum\hat{y}_i^2$，同时记住(3.5.6)。

3.16 解释以下命题的正误或不确定，并给出原因。

a. 由于两个变量 Y 和 X 之间的相关系数取值范围为 [-1，1]，所以这也意味着，cov（Y，X）也在此范围内。

b. 如果两个变量之间的相关系数为零，那就意味着这两个变量之间无论如何也没有什么关系。

c. 如果你将 Y_i 对 $\hat{Y}_i$ 回归（即实际的 Y 对估计的 Y 回归），那么截距和斜率的值分别为 0 和 1。

3.17 **不含回归元的回归**。假如给你一个模型 $Y_i = \beta_1 + u_i$。利用 OLS 求出 β_1 的估计量。其方差和 RSS 是多少？估计的 β_1 有直觉上的意义吗？现在考虑双变量模型 $Y_i = \beta_1 + \beta_2 X_i + u_i$。在此模型中增加 X_i 值得吗？若不然，为什么要进行回归分析呢？

解答题

3.18 表3.5 给出了 10 名学生在统计学期中和期末考试中的等级（名次）。计算斯皮尔曼等级相关系数并加以解释。

表 3.5

等级	学生									
	A	B	C	D	E	F	G	H	I	J
期中	1	3	7	10	9	5	4	8	2	6
期末	3	2	8	7	9	6	5	10	1	4

3.19 **名义汇率与相对价格之间的关系。**1980—1994 年的年度观测值，可得到如下回归结果，其中 Y 表示德国马克与美元的汇率（GM/\$），X 表示美国消费者价格指数与德国消费者价格指数之比；即 X 代表了这两个国家的相对价格：

$$\hat{Y}_t = 6.682 - 4.318X_t \qquad r^2 = 0.528$$
$$\text{se} = (1.22)(1.333)$$

a. 解释这个回归。你如何解释 r^2？

b. X_t 的负值有经济意义吗？其背后的经济理论是什么？

c. 假若我们将 X 定义为德国 CPI 与美国 CPI 之比。X 的符号会改变吗？为什么？

3.20 表3.6 就美国 1980—1997 年商业和非农商业部门给出了每小时产出指数（X）和真实时薪（Y）的数据。基年（1992）指数为 100，且指数经过了季节性调整。

97 **表 3.6** **商业部门的生产率及相关数据，1959—1998 年［指数 1992 = 100；季度数据经过季节性调整］**

年份	所有人的每小时产出[1]		时薪[2]	
	商业部门	非农商业部门	商业部门	非农商业部门
1959……	50.5	54.2	13.1	13.7
1960……	51.4	54.8	13.7	14.3
1961……	53.2	56.6	14.2	14.8
1962……	55.7	59.2	14.8	15.4
1963……	57.9	61.2	15.4	15.9
1964……	60.6	63.8	16.2	16.7
1965……	62.7	65.8	16.8	17.2
1966……	65.2	68.0	17.9	18.2
1967……	66.6	69.2	18.9	19.3
1968……	68.9	71.6	20.5	20.8

1969……	69.2	71.7	21.9	22.2
1970……	70.6	72.7	23.6	23.8
1971……	73.6	75.7	25.1	25.4
1972……	76.0	78.3	26.7	27.0
1973……	78.4	80.7	29.0	29.2
1974……	77.1	79.4	31.8	32.1
1975……	79.8	81.6	35.1	35.3
1976……	82.5	84.5	38.2	38.4
1977……	84.0	85.8	41.2	41.5
1978……	84.9	87.0	44.9	45.2
1979……	84.5	86.3	49.2	49.5
1980……	84.2	86.0	54.5	54.8
1981……	85.8	87.0	59.6	60.2
1982……	85.3	88.3	64.1	64.6
1983……	88.0	89.9	66.8	67.3
1984……	90.2	91.4	69.7	70.2
1985……	91.7	92.3	73.1	73.4
1986……	94.1	94.7	76.8	77.2
1987……	94.0	94.5	79.8	80.1
1988……	94.7	95.3	83.6	83.7
1989……	95.5	95.8	85.9	86.0
1990……	96.1	96.3	90.8	90.7
1991……	96.7	97.0	95.1	95.1
1992……	100.0	100.0	100.0	100.0
1993……	100.1	100.1	102.5	102.2
1994……	100.7	100.6	104.4	104.2
1995……	101.0	101.2	106.8	106.7
1996……	103.7	103.7	110.7	110.4
1997……	105.4	105.1	114.9	114.5

[1] 产出指该部门真实 GDP。

[2] 雇员的工资和薪水加上雇主对社会保障和私人救济方案的支付。还包括对自雇族工资、薪水和补充收入的估计。

资料来源：*Economic Report of the President*, 1999, Table B-49, p.384.

98

a. 分别按两个部门将 Y 对 X 描点。

b. 这两个变量间关系的背后有什么经济理论？散点图支持该理论吗？

c. 估计 Y 对 X 的 OLS 回归，在学完第 5 章后，再回头看一下你的结果。

3.21 根据 10 次观测值的一个样本，得到如下的结果：

$$\sum Y_i = 1\ 110 \quad \sum X_i = 1\ 700 \quad \sum X_i Y_i = 205\ 500$$
$$\sum X_i^2 = 322\ 000 \qquad \sum Y_i^2 = 132\ 100$$

并且相关系数 r=0.975 8。但在重新核对这些计算时，发现有两双观测值的记录是：

Y	X		Y	X
90	120	而不是	80	110
140	220		150	210

但后者才是正确的。问这一错误对 r 有何影响？求正确的 r。

3.22 表 3.7 给出 1977—1991 年期间美国的黄金价格、消费者价格指数和纽约股票交易所指数数据。NYSE 指数包括在 NYSE 上市的大多数股票，约有 1 500 多种。

表 3.7

年份	在纽约每盎司黄金的美元价格	消费者价格指数 1982—1984 = 100	纽约股票交易所指数 1965.12.31 = 100
1977	147.98	60.6	53.69
1978	193.44	65.2	53.70
1979	307.62	72.6	58.32
1980	612.51	82.4	68.10
1981	459.61	90.9	74.02
1982	376.01	96.5	68.93
1983	423.83	99.6	92.63
1984	360.29	103.9	92.46
1985	317.30	107.6	108.90
1986	367.87	109.6	136.00
1987	446.50	113.6	161.70
1988	436.93	118.3	149.91
1989	381.28	124.0	180.02
1990	384.08	130.7	183.46
1991	362.04	136.2	206.33

资料来源：Data on CPI and NYSE index are from the *Economic Report of the President*, January 1993, Tables B-59 and B-91, respectively. Data on gold prices are *from U. S. Department of Commerce*, *Bureau of Economic Analysis*, *Business Statistics*, 1963－1991, p.68.

99 a. 在同一散布图中描绘黄金价格，CPI 和 NYSE 指数。

b. 一种投资，如果它的价格和（或）回报率至少赶得上通货膨胀，就被认为是（对通货膨胀）保值（能抵御通货膨胀）的。为检验这一假设：投资是保值的，假定 a 中的散点图表明拟合以下模型是适宜的：

$$黄金价格_t = \beta_1 + \beta_2 \text{CPI}_t + u_t$$
$$\text{NYSE 指数}_t = \beta_1 + \beta_2 \text{CPI}_t + u_t$$

3.23 表 3.8 给出 1959—1997 年美国国内总产值数据。

表 3.8 **美国 1959—1997 年间名义和真实 GDP**

年份	NGDP	RGDP	年份	NGDP	RGDP
1959	507.200 0	2 210.200	1979	2 557.500	4 630.600
1960	526.600 0	2 262.900	1980	2 784.200	4 615.000
1961	544.800 0	2 314.300	1981	3 115.900	4 720.700
1962	585.200 0	2 454.800	1982	3 242.100	4 620.300
1963	617.400 0	2 559.400	1983	3 514.500	4 803.700
1964	663.000 0	2 708.400	1984	3 902.400	5 140.100
1965	719.100 0	2 881.100	1985	4 180.700	5 323.500
1966	787.700 0	3 069.200	1986	4 422.200	5 487.700
1967	833.600 0	3 147.200	1987	4 692.300	5 649.500
1968	910.600 0	3 293.900	1988	5 049.600	5 865.200
1969	982.200 0	3 393.600	1989	5 438.700	6 062.000
1970	1 035.600	3 397.600	1990	5 743.800	6 136.300
1971	1 125.400	3 510.000	1991	5 916.700	6 079.400
1972	1 237.300	3 702.300	1992	6 244.400	6 244.400
1973	1 382.600	3 916.300	1993	6 558.100	6 389.600
1974	1 496.900	3 891.200	1994	6 947.000	6 610.700
1975	1 630.600	3 873.900	1995	7 269.600	6 761.700
1976	1 819.000	4 082.900	1996	7 661.600	6 994.800
1977	2 026.900	4 273.600	1997	8 110.900	7 269.800
1978	2 291.400	4 503.000			

注：NGDP = 名义 GDP（以当年 10 亿美元计）；RGDP = 真实 GDP（以 1992 年的 10 亿美元计）。

资料来源：*Economic Report of the President*，1999，Tables B－1 and B－2，pp.326－328.

a. 将当年美元和不变（即 1992 年）美元数据对时间描图。

b. 用 Y 表示 GDP，X 表示时间（按年历从 1 代表 1959，2 代表 1960 开始，直至 39 代表 1997）。看以下模型是否适合 GDP 数据：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

试用当年美元和不变美元两种数据分别估计此模型。

c. 你会怎样解释 β_2？

d. 如果用当年美元估计的 β_2 和用不变美元估计的 GDP 有所不同，怎样解释这个差异？

100 e. 从你计算的结果，你能对样本时期美国通货膨胀的性质得出什么评论？

3.24 利用引言中表 I.1 所给数据，证实方程（3.7.1）。

3.25 对习题 2.16 中 SAT 一例做以下的练习：

a. 将女生语文分数对应于男生语文分数描点。

b. 如果散点图表明两者似有线性关系，试求女生语文分数对男生语文分数的回归。

c. 如果这两种语文分数之间有某种关系，它是不是因果关系？

3.26 用数学分数代替语文分数，重复解释习题 3.25 中的练习。

3.27 蒙特卡罗研究课堂作业：

回到表 3.2 中所列的 10 个 X 值。令 $\beta_1 = 25$ 和 $\beta_2 = 0.5$。假定 $u_i \sim N(0, 9)$，即 u_i 遵从均值为 0、方差为 9 的正态分布。用这两个参数值去产生 100 个样本，求出 β_1 和 β_2 的 100 个估计值，然后对这些估计值描图。从这一蒙特卡罗研究中，你能得出什么结论？注：当今大多数统计包都能从一些最熟知的概率分布中产生随机数。如果你在产生这些随机数时遇到困难，请向你的老师求助。

附录 3A

3A.1 最小二乘估计的推导

将（3.1.2）对 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 求偏导数得：

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_1} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum \hat{u}_i \tag{1}$$

$$\frac{\partial\left(\sum \hat{u}_i^2\right)}{\partial \hat{\beta}_2} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i = -2\sum \hat{u}_i X_i \tag{2}$$

令这些导数为零。经过代数上的简化和运算，便得到方程（3.1.6）和（3.1.7）所给的估计量。

3A.2　最小二乘估计量的线性和无偏性质

由（3.1.8）得：

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i \tag{3}$$

其中：

$$k_i = \frac{x_i}{\left(\sum x_i^2\right)}$$

101 这说明 $\hat{\beta}_2$ 是 Y_i 的一个线性函数；它是以 k_i 为权数的 Y_i 的一个加权平均，从而它是一个**线性估计量**。同理，$\hat{\beta}_1$ 也是一个线性估计量。

顺便指出权数 k_i 的一些性质：

1. 因 X_i 被假定为非随机的，故 k_i 也是非随机的。
2. $\sum k_i = 0$。
3. $\sum k_i^2 = 1/\sum x_i^2$。
4. $\sum k_i x_i = \sum k_i X_i = 1$。这些性质均可直接从 k_i 的定义中验证。

例如，

$$\begin{aligned}\sum k_i &= \sum \left(\frac{x_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2}\sum x_i, \\ &= 0\end{aligned}$$

$$\left(\begin{array}{l}\text{因为对一个给定的样本，} \sum x_i^2 \text{ 为已知。} \\ \text{因为偏离均值的离差总和} \sum x_i \text{ 必等于零。}\end{array}\right)$$

现将 PRF　$Y_i = \beta_1 + \beta_2 X_i + u_i$ 代入（3）得：

$$\begin{aligned}\hat{\beta}_2 &= \sum k_i(\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i \\ &= \beta_2 + \sum k_i u_i\end{aligned} \tag{4}$$

其中利用了前面提到的 k_i 的性质。

对（4）两边求数学期望并注意到 k_i 是非随机的，即可视同常数，于是得：

$$\begin{aligned}E(\hat{\beta}_2) &= \beta_2 + \sum k_i E(u_i) \\ &= \beta_2\end{aligned} \tag{5}$$

这是因为假定$E(u_i)=0$。因此，$\hat{\beta}_2$ 是 β_2 的一个无偏估计量。同理，可证 $\hat{\beta}_1$ 也是 β_1 的一个无偏估计量。

3A.3 最小二乘估计量的方差和标准误

按方差定义，可写：

$$\begin{aligned}\mathrm{var}(\hat{\beta}_2) &= E[\hat{\beta}_2 - E(\hat{\beta}_2)]^2\\ &= E(\hat{\beta}_2 - \beta_2)^2, \text{因 } E(\hat{\beta}_2) = \beta_2\\ &= E(\sum k_i u_i)^2, \text{利用上面的方程(4)}\\ &= E(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2 + 2k_1k_2u_1u_2 + \cdots\\ &\quad + 2k_{n-1}k_nu_{n-1}u_n)\end{aligned} \tag{6}$$

102 因由假定$E(u_i^2)=\sigma^2$，对每一 i 和$E(u_iu_j)=0$，且 $i\neq j$，故：

$$\begin{aligned}\mathrm{var}(\hat{\beta}_2) &= \sigma^2\sum k_i^2\\ &= \frac{\sigma^2}{\sum x_i^2}, \text{（利用 } k_i^2 \text{ 的定义）}\\ &= \text{方程 (3.3.1)}\end{aligned} \tag{7}$$

按照同样的推理可求得 $\hat{\beta}_1$ 的方差。一旦得到 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差，取其正的平方根即是相应的标准误。

3A.4 $\hat{\boldsymbol{\beta}}_1$ 与 $\hat{\boldsymbol{\beta}}_2$ 的协方差

由定义得，

$$\begin{aligned}\mathrm{cov}(\hat{\beta}_1,\hat{\beta}_2) &= E\{[\hat{\beta}_1 - E(\hat{\beta}_1)][\hat{\beta}_2 - E(\hat{\beta}_2)]\}\\ &= E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \qquad \text{（为什么?）}\\ &= -\bar{X}E(\hat{\beta}_2 - \beta_2)^2\\ &= -\bar{X}\,\mathrm{var}(\hat{\beta}_2)\\ &= \text{方程(3.3.9)}\end{aligned} \tag{8}$$

这里用到关系式 $\hat{\beta}_1=\bar{Y}-\hat{\beta}_2\bar{X}$ 及$E(\hat{\beta}_1)=\bar{Y}-\beta_2\bar{X}$，并由此给出 $\hat{\beta}_1-E(\hat{\beta}_1)=-\bar{X}(\hat{\beta}_2-\beta_2)$。注：$\mathrm{var}(\hat{\beta}_2)$由(3.3.1)给出。

3A.5 $\boldsymbol{\sigma}^2$ 的最小二乘估计量

回顾：

$$Y_i = \beta_1 + \beta_2X_i + u_i \tag{9}$$

因此

$$Y = \beta_1 + \beta_2 X + u \tag{10}$$

从（9）减去（10）得：

$$y_i = \beta_2 x_i + (u_i - \bar{u}) \tag{11}$$

再回顾：

$$\hat{u} = y_i - \hat{\beta}_2 x_i \tag{12}$$

从而把（11）代入（12）得：

$$\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i \tag{13}$$

103 合并同类项，平方，然后两边求和得：

$$\sum \hat{u}_i^2 = (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + \sum (u_i - u)^2 - 2(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u}) \tag{14}$$

两边取数学期望得：

$$\begin{aligned} E\left(\sum \hat{u}_i^2\right) &= \sum x_i^2 E(\hat{\beta}_2 - \beta_2)^2 + E\left[\sum (u_i - \bar{u})^2\right] \\ &\quad - 2E\left[(\hat{\beta}_2 - \beta_2)\sum x_i (u_i - \bar{u})\right] \\ &= \sum x_i^2 \operatorname{var}(\hat{\beta}_2) + (n-1)\operatorname{var}(u_i) - 2E\left[\sum k_i u_i (x_i u_i)\right] \\ &= A + B + C \\ &= (n-2)\sigma^2 \end{aligned} \tag{15}$$

由经典线性回归模型的假定以及方才建立的一些结果可以验证：

$$\begin{aligned} A &= \sigma^2 \\ B &= (n-1)\sigma^2 \\ C &= -2E\left[\sum k_i x_i u_i^2\right] = -2\sigma^2 \end{aligned}$$

因此，将这些值代入(15)便得：

$$E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2 \tag{16}$$

若定义：

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{17}$$

利用(16)，它的期望值就是：

$$E(\hat{\sigma}^2) = \frac{1}{n-2} E\left(\sum \hat{u}_i^2\right) = \sigma^2 \tag{18}$$

这表明 $\hat{\sigma}^2$ 是真实 σ^2 的一个无偏估计量。

3A.6 最小二乘估计量的最小方差性质

104 在附录 3A，3A.2 节中曾表明，最小二乘估计量 $\hat{\beta}_2$ 是线性和无偏的（而且 $\hat{\beta}_1$ 也如此）。为了表明这些估计量在所有线性无偏估计量一类中有最小方差，且考虑最小二乘估计量 $\hat{\beta}_2$：

$$\hat{\beta}_2 = \sum k_i Y_i$$

其中：
$$k_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} = \frac{x_i}{\sum x_i^2} \quad \text{（见附录 3A.2）} \tag{19}$$

这表明 $\hat{\beta}_2$ 是以 k_i 为权重的 Y_i 的加权平均。

定义 β_2 的另一线性估计量 β_2^* 如下：

$$\beta_2^* = \sum w_i Y_i \tag{20}$$

其中权 w_i 不一定等于 k_i，于是：

$$\begin{aligned} E(\beta_2^*) &= \sum w_i E(Y_i) \\ &= \sum w_i (\beta_1 + \beta_2 X_i) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i X_i \end{aligned} \tag{21}$$

为要 β_2^* 无偏，必须：

$$\sum w_i = 0 \tag{22}$$

及：
$$\sum w_i X_i = 1 \tag{23}$$

而且，可以写：

$$\begin{aligned} \operatorname{var}(\beta_2^*) &= \operatorname{var} \sum w_i Y_i \\ &= \sum w_i^2 \operatorname{var} Y_i \quad [\text{注：} \operatorname{var} Y_i = \operatorname{var} u_i = \sigma^2] \\ &= \sigma^2 \sum w_i^2 \quad [\text{注：} \operatorname{cov}(Y_i, Y_j) = 0 (i \neq j)] \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2 \quad \text{（注意其中的数学技巧）} \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} \\ &\quad + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right) \left(\frac{x_i}{\sum x_i^2} \right) \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right) \end{aligned} \tag{24}$$

因为倒数第二步的最后一项消失了。（为什么？）

105 由于（24）中的最后一项是常数，（β_2^*）的方差只能通过第一项的处理

使之最小化。若令：

$$w_i = \frac{x_i}{\sum x_i^2}$$

则方程（24）简化为：

$$\begin{aligned}\operatorname{var}(\beta_2^*) &= \frac{\sigma^2}{\sum x_i^2} \\ &= \operatorname{var}(\hat{\beta}_2)\end{aligned} \tag{25}$$

一般来说，当权 w_i = 最小二乘的权 k_i 时，线性估计量 β_2^* 的方差等于最小二乘估计量 $\hat{\beta}_2$ 的方差；不然的话，$\operatorname{var}(\beta_2^*) > \operatorname{var}(\hat{\beta}_2)$。也就是说，如果存在 β_2 的一个最小方差线性无偏估计量，那么，它必定是最小二乘估计量。同理，可以表明，$\hat{\beta}_1$ 是 β_1 的最小方差线性无偏估计量。

3A.7 最小二乘估计量的一致性

我们在经典线性回归模型的框架中已经证明，最小二乘估计量在无论大样本或小样本容量的情况下都是无偏的（和有效的）。但如附录 A 中所讨论的那样，一个估计量有时候可能不满足一个或多个优良的小样本统计性质。但随着样本容量无限扩大，这些估计量就具有几个优良的统计性质。这些性质被称为**大样本或渐近性质**（large sample or asymptotic properties）。在此附录中，我们将讨论一个大样本性质——**一致性**（consistency），在附录 A 中将更充分地对此展开讨论。对双变量模型，我们已经证明了 OLS 估计量 $\hat{\beta}_2$ 是真实 β_2 的一个无偏估计量。现在我们来证明，$\hat{\beta}_2$ 也是 β_2 的一个一致估计量。如附录 A 所证明，一致性的一个充分条件是，$\hat{\beta}_2$ 是无偏的，且随着样本容量趋于无穷，其方差趋于零。

既然我们已经证明了无偏性，那现在就只需证明 $\hat{\beta}_2$ 的方差在 n 无限增加时趋于零。我们知道

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2/n}{\sum x_i^2/n} \tag{26}$$

将分子和分母同时除以 n，不会改变这个等式。

现在

$$\underbrace{\lim \operatorname{var}(\hat{\beta}_2)}_{n\to\infty} = \underbrace{\lim\left(\frac{\sigma^2/n}{\sum x_i^2/n}\right)}_{n\to\infty} = 0 \tag{27}$$

106 其中用到如下事实：(1) 比率的极限等于分子的极限与分母的极限之比（参考任何一本微积分方面的书）；(2) 随着 n 趋于无穷，σ^2/n 因 σ^2 是一个有限的数而趋于零；而根据 CLRM 的假定 8，X 的方差是一个有限的正数，所

以 $[(\sum x_i^2)/n] \neq 0$。

上一段讨论的结果是，OLS 估计量 $\hat{\beta}_2$ 是真实 β_2 的一个一致估计量。与此相仿，我们可以证明，$\hat{\beta}_1$ 也是一个一致估计量。因此，在重复（或小）样本下 OLS 估计量是无偏的，而随着样本容量无限增加，OLS 估计量是一致的。如以后所见，即使 CLRM 的某些假定不满足，我们在某几种情况下也能得到回归系数的一致估计量。

【注释】

[1] 对好奇的读者补充一句，这些值是由下面即将讨论的最小二乘法得到的。见方程（3.1.6）和（3.1.7）。

[2] 注 1：由于 $\bar{X}$ 为一常数，$\sum x_i^2 = \sum(X_i - \bar{X})^2 = \sum X_i^2 - 2\sum X_i\bar{X} + \sum \bar{X}^2 = \sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2$。而且 $\sum X_i = n\bar{X}$ 和 $\sum \bar{X}^2 = n\bar{X}^2$，最后便得到 $\sum x_i^2 = \sum X_i^2 - n\bar{X}^2$。

注 2：由于 $\bar{Y}$ 为一常数，并且一个变量对其均值的离差的总和恒为零［例如，$\sum(X_i - \bar{X}) = 0$］，故有

$$\begin{aligned}\sum x_i y_i &= \sum x_i(Y_i - \bar{Y}) \\ &= \sum x_i Y_i - \bar{Y}\sum x_i \\ &= \sum x_i Y_i - \bar{Y}\sum(X_i - \bar{X}) \\ &= \sum x_i Y_i\end{aligned}$$

类似地，$\sum y_i = \sum(Y_i - \bar{Y}) = 0$。

[3] Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, p.3.

[4] 同上所引。

[5] 注意，这个结果仅当回归模型含有截距 β_1 时才是真实的。如附录 6A，第 6A.1 所示，当模型不含有 β_1 时，这个结果不一定成立。

[6] 这一结果还要求截距项 β_1 必须在模型中出现（参看附录 6A，第 6A.1 节）。

[7] 经典的意义在于，它于 1821 年由高斯首创，并且从那以后都以它作为标准来比较那些不满足高斯假定的回归模型。

[8] 第 14 章中将对参数为非线性的回归模型做简要讨论。

[9] 为了说明，这里仅假定 u 是对称地分布的，如图 3.3 所示。但在第 4 章中，我们将假定 u 是正态分布的。

[10] 关于假定 3 为什么是必要的，还有一个更为技术性的理由，可参看 E. Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1996, p.75。还可参看本书习题 3.3。

[11] 请回忆，在获取表 2.4 和 2.5 中的样本时，我们保持着相同的 X 值。

[12] 如同我们在第 2 部分将要讨论的，如果诸 X 是随机的，但却独立于 u_i 而分布，那么，下面很快将要讨论的最小二乘估计量的性质仍然成立。但若随机变量 X 仅仅与 u_i 无关，则仅当样本含量很大时，OLS 估计量的性质才会成立。然而，在现阶段的讨论中没有必要拘泥于这一理论问题。

[13] X 的样本方差是:

$$\operatorname{var}(X)=\frac{\sum(X_i-\bar{X})^2}{n-1}$$

其中 n 是样本含量（或样本大小）。

[14] 但应避免通称的**“数据开采”**（data mining），即对所有可能的模型一一尝试，希望至少从中找出一个与数据拟合良好的模型。这说明为什么模型的选择必须有一些经济学的理论基础，以及模型的修改必须有一些经济学方面的理由。纯粹地为拟合而拟合的模型很难在先验的理论上站得住脚。简言之，理论是估计的基础。

[15] Milton Friedman, *Essays in Positive Economics*, University of Chicago Press, Chicago, 1953, p.14.

[16] Mark Blaug, *The Methodology of Economics: Or How Economists Explain*, 2d ed., Cambridge University Press, New York, 1992, p.92.

[17] **标准误**无非是估计量的抽样分布的标准差，而一个估计量的抽样分布，就是该估计量的概率或频率分布，也就是得自一定总体的同样大小（容量）的所有可能样本的估计值集合的一个分布。抽样分布的使用，是为了能够根据从一个或多个样本计算出来的估计值去推断总体的参数值（详见附录 A）。

[18] **自由度**（number of degrees of freedom）一词指样本中观测值的总数（$=n$）减去对它们的独立的（线性）约束或限制的个数。换句话说，它是观测值的总个数中独立的观测值的个数。例如，在能算出 RSS（3.1.2）之前必须先算出 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。这两个估计值就是附加给 RSS 的两个约束。因此，在计算 RSS 时，就只有 $n-2$ 而不是 n 个独立观测值。按照这一逻辑，在 3 变量回归中 RSS 将有 $n-3$ 个自由度。至于 k 变量模型，它就有 $n-k$ 个 df。**一般的规律是** df =（n − 待估参数的个数）。

[19] 虽然以高斯-马尔可夫定理命名，但高斯的最小二乘研究方法（1821）要早于马尔可夫的最小方差方法（1900）。

[20] 关于线性估计量的重要意义以及关于统计估计量的优良性质的一般讨论，可参考附录 A。

[21] 例如，可以证明诸 β 的任意线性组合，如（$\beta_1-2\beta_2$），可由（$\hat{\beta}_1-2\hat{\beta}_2$）来估计，并且这一估计是 BLUE。详见 Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp.401–402.

[22] 参阅 Peter Kennedy, “Ballentine: A Graphical Aid for Econometrics,”, *Australian Economics Papers*, vol.20, 1981, pp.414–416。名字 Ballentine 来自著名的 Ballantine 啤酒的圆圈标徽。

[23] 变异一词和方差有所不同。变异指一个变量对其均值的离差平方和，而方差指此平方和除以适当的自由度。简言之，方差＝变异/自由度。

[24] 总体相关系数，记为 ρ，其定义见附录 A。

[25] 回归建模时所依据的基础理论将表明，Y 与 X 间的因果方向在简单回归中一般地说是从 X 到 Y 的。

[26] 我们将在第 8 章中正式讲解 r^2 的显著性检验。

[27] 每个研究领域都有其行话。“将 Y 对 X 回归”一语仅意味着将 Y 作为回归子而将 X 作为回归元。

[28] 实际上，假定 u_i 遵从给定参数（如均值和方差）的某一概率分布，如正态分布。一旦给定了参数，就容易用统计包来产生 u_i。

第 4 章　经典正态线性回归模型

107 所谓**统计推断的经典理论**由两个分支构成，即**估计**和**假设检验**。我们到目前为止已讨论了（双变量）线性回归模型的参数估计问题。用 OLS 的方法，我们能估计参数 β_1、β_2 和 σ^2。在**经典线性回归模型**的假定下，我们可以证明，$\hat{\beta}_1$、$\hat{\beta}_2$ 和 $\hat{\sigma}^2$ 这些参数的估计量都满足几个理想的统计性质，如无偏性和最小方差等。（回顾 BLUE 性质。）注意，既然它们都是估计量，所以它们的值将随样本的变化而变化。因此，这些估计量都是**随机变量**。

但估计是成功的一半。假设检验是另一半。回想我们在回归分析中的目标，不仅仅是估计样本回归函数，而是像第 2 章所强调的那样，我们要用估计来对总体回归函数进行推断。于是，我们想知道，$\hat{\beta}_1$ 和 $\hat{\sigma}^2$ 与真实的 β_1 和 σ^2 到底有多接近。比如，在例 3.2 中，我们在方程（3.7.2）中估计了 SRF。但这个回归只是基于一个由 55 个家庭构成的样本，我们怎么知道估计的 MPC 为0.436 8代表了整个总体（真实）的 MPC 呢？

因此，由于 $\hat{\beta}_1$、$\hat{\beta}_2$ 和 $\hat{\sigma}^2$ 是随机变量，所以我们需要清楚它们的概率分布，若不知其概率分布，那我们就无法将它们与其真实值相联系。

§4.1 干扰项 u_i 的概率分布

108 为得到 OLS 估计量的概率分布，我们将如下进行：专门考虑 $\hat{\beta}_2$。如我们在附录 3A.2 中所证明，

$$\hat{\beta}_2 = \sum k_i Y_i \tag{4.1.1}$$

其中 $k_i = x_i / \sum x_i^2$。但由于假定 X 为固定或非随机的，所以我们的条件回归分析就以 X_i 的固定值为条件。方程（4.1.1）表明，$\hat{\beta}_2$ 是 Y_i 的一个线性函数，Y_i 根据假定也是随机的。但由于 $Y_i = \beta_1 + \beta_2 X_i + u_i$，所以我们可以把（4.1.1）写成

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) \tag{4.1.2}$$

由于 k_i，β 系数和 X_i 都是固定的，所以 $\hat{\beta}_2$ 最终是 u_i（u_i 假定为随机变量）的一个线性函数。因此，$\hat{\beta}_2$（及 $\hat{\beta}_1$）的概率分布将取决于对 u_i 的概率分布所做的假定。由于对 OLS 估计量的概率分布知识足以对其总体做出推断，所以 u_i 概率分布的性质在假设检验中就起到极为重要的作用。

记得当我们把普通最小二乘法应用于经典线性回归模型时，我们并没有对干扰 u_i 的概率分布作任何假定。对这些 u_i 所作的假定仅是：它们的期望值为零，它们是不相关的，并且有一个不变的方差。有了这些假定，便看到（第 3 章）OLS 中估计量 $\hat{\beta}_1$，$\hat{\beta}_2$ 和 $\hat{\sigma}^2$ 满足若干优良的统计性质，诸如无偏性和最小方差。如果我们的目的仅是做点估计，则 OLS 法已经足够好了。但点估计只是统计推断的一个方面，另一个方面则是假设检验。

就是说，我们的兴趣不仅是要得到（比如说）$\hat{\beta}_2$，而且要利用它对真值 β_2 做出推断。说得更一般些，我们的目的不仅是要得到样本回归函数，而且如同第 2 章所强调的，还要用它来推测总体回归函数。

因为我们的目标既是估计又是假设检验，我们就有必要规定干扰 u_i 的概率分布。为什么呢？答案并不难给出。在附录 3A，第 3A.2 节中，我们阐明，OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 都是 u_i 的线性函数，而按假定 u_i 是随机的，因此，OLS 估计量的抽样或概率分布将依赖于 u_i 假定的概率分布。又因为必须知道这些估计量的概率分布，方能对它们的总体值进行推断，所以 u_i 的概率分布是什么性质，在假设检验中就有着非常重要的作用。

尽管有了高斯-马尔可夫定理，但由于 OLS 法不对 u_i 的概率性质做任何假定，仍难于从 SRF 去推断 PRF。对这一不足，如果我们愿意假定 u_i 遵循某种概率分布的话，就可加以补充。在回归分析中，人们常常假定 u_i 遵循正态分布，这是有理由的，稍后即明。在第 3 章中讨论的经典线性回归模型的假定中增加 u_i 的正态性假定，我们就得到了所谓的**经典正态线性回归模型**（classical normal linear regression model，CNLRM）。

§4.2 关于 u_i 的正态性假定

经典正态线性回归假定每个 u_i 都是正态分布的，且其

$$均值：E(u_i)=0 \tag{4.2.1}$$

$$方差：E[u_i-E(u_i)]^2=E(u_i^2)=\sigma^2 \tag{4.2.2}$$

协方差 $\mathrm{cov}(u_i,u_j)$：

$$E\{[u_i-E(u_i)][u_j-E(u_j)]\}=E(u_i u_j)=0 \quad i\neq j \tag{4.2.3}$$

这些假定可更简洁地叙述为：

$$u_i \sim N(0,\sigma^2) \tag{4.2.4}$$

109 其中～表示"其分布为"，而 N 代表"正态分布"，括号中的项代表正态分布的两个参数：均值与方差。

就像附录 A 所提到的，**对两个正态分布变量来说，零协方差或零相关就意味着两个变量互相独立。**因此，在正态性假定下，(4.2.4) 不仅意味着 u_i 与 u_j 不相关，而且它们是独立地分布的。

于是，可将 (4.2.4) 写为：

$$u_i \sim \mathrm{NID}(0,\sigma^2) \tag{4.2.5}$$

这里 **NID** 表示正态且独立分布。

为什么是正态假定?

我们为什么采用正态性假定呢？有如下几个理由：

1. 如第2.5 节所指出的，u_i 代表回归模型中未明显引进的许多自变量（对因变量）的总影响。如同已注明的那样，我们希望这些被忽略的变量所起的影响是微小的，而且最多是随机的。于是，利用统计学中著名的**中心极限定理**（central limit theorem，CLT）就能证明（细节见附录 A)，如果存在大量独立且相同分布的随机变量，那么，除了少数例外情形，随着这些变量的个数无限地增大，它们的总和将趋向正态分布。[1]正是这个中心极限定理，为 u_i 的正态性假定提供了理论基础。

2. 中心极限定理的另一解说是，即使变量个数并不很大或这些变量还不是严格独立的，它们的总和仍旧是正态分布的。[2]

3. 如附录 A 中所言，**正态分布的一个性质是，正态分布变量的任何线性函数都是正态分布的。**因此，在正态性假定下，OLS 估计量的概率分布是容易导出的。前面曾讨论过，OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 是 u_i 的线性函数。因此，若 u_i 是正态分布的，则 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 也是正态分布的，这就将使我们的假设检验工作十分简单。

4. 正态分布是一个比较简单的、仅涉及两个参数（均值和方差）的分
110 布；它为人们所熟知，它的理论性质在数理统计学中受到广泛的研究。除此

以外，看上去许多现象也都为正态分布。

5. 最后，如果我们在处理小样本或有限容量样本时，比如说数据少于100次的观测，那么正态假定就起到关键作用。它不仅有助于我们推导出OLS估计量精确的概率分布，而且使我们能用 t、F 和 χ^2（卡方）来对回归模型进行统计检验。在附录A中讨论了 t、F 和 χ^2 概率分布的统计性质。如我们稍后所见，如果样本容量大到合理的程度，我们或许能够放松正态性假定。

一句忠告：既然我们"施加了"正态性假定，那我们就有必要在一些涉及小样本容量数据的实际应用中注意正态性假定是否适当。稍后，我们将对此做一些检验。此外，我们以后还会偶尔遇到一些正态性假定不适当的情况。但出于前面讨论的原因，我们仍继续做正态性假定，除非我们看到这个假定并不适当。

§4.3 在正态性假定下OLS估计量的性质

在如（4.2.5）中的 u_i 为正态分布的假定下，OLS估计量有如下统计性质；附录A中对估计量的统计性质做了一个大致的讨论。

1. 它们是无偏的。

2. 它们有最小方差。连同性质1，就意味着它们是**最小方差无偏的**（minimum-variance unbiased）或者说它们是**有效（优效）估计量**（efficient estimators）。

3. 一致性。就是说，随着样本含量无限地增大，估计量将收敛到它们的真值。

4. $\hat{\beta}_1$（u_i 的线性函数）是正态分布的，

$$\text{其均值：}\quad E(\hat{\beta}_1)=\beta_1 \tag{4.3.1}$$

$$\text{方差}\quad \operatorname{var}(\hat{\beta}_1):\quad \sigma^2_{\hat{\beta}_1}=\frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \qquad =(3.3.3)\ (4.3.2)$$

或更简洁地写作：

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

然后，利用正态分布的性质，定义：

$$Z=\frac{\hat{\beta}_1-\beta_1}{\sigma_{\hat{\beta}_1}} \tag{4.3.3}$$

11 便知变量 Z 遵循**标准正态分布**（standard normal distribution），即零均值和单位方差（=1）的正态分布，或写作：

$$Z \sim N(0,\ 1)$$

图 4.1 从几何图形上描绘了 $\hat{\beta}_1$ 的概率分布。

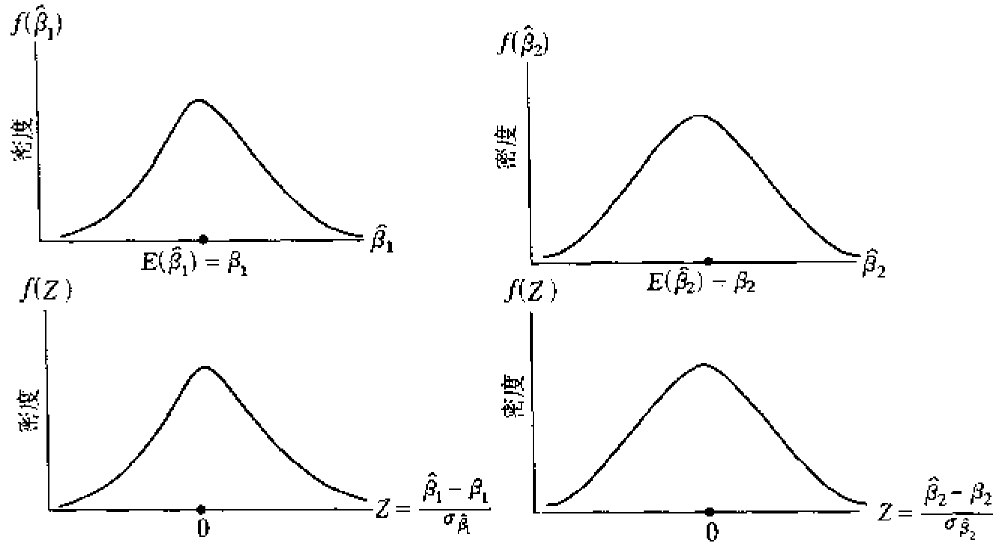


图 4.1　$\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的概率分布

5. $\hat{\beta}_2$（u_i 的线性函数）是正态分布的，其

均值：　$E(\hat{\beta}_2)=\beta_2$　　　　(4.3.4)

方差　$\text{var}(\hat{\beta}_2)$：　$\sigma^2_{\hat{\beta}_2}=\frac{\sigma^2}{\sum x_i^2}$　　　　= (3.3.1) (4.3.5)

或，更简洁地写作：

$$\hat{\beta}_2 \sim N(\beta_2,\sigma^2_{\hat{\beta}_2})$$

然后，如同 (4.3.3) 那样，

$$Z=\frac{\hat{\beta}_2-\beta_2}{\sigma_{\hat{\beta}_2}} \tag{4.3.6}$$

也遵循标准正态分布。

图 4.1 也展示了 $\hat{\beta}_2$ 的概率分布的几何图形。

112 **6.** $(n-2)\hat{\sigma}^2/\sigma^2$ 遵循 $n-2$ 个自由度的 χ^2 分布。第 5 章有它的一个应用。[3]如我们在第 5 章将看到的那样，这一点有助于我们从估计的 σ^2 中对真实的 σ^2 做出推断。(关于 χ^2 分布及其性质的讨论见附录 A。)

7. $(\hat{\beta}_1,\hat{\beta}_2)$ 的分布独立于 $\hat{\sigma}^2$。对这点的重要性将在下一章做出解释。

8. $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 在整个无偏估计序列中，无论是线性或非线性估计，都有最小方差。这一结果归功于拉奥 (Rao)，是非常之强有力的。它和高斯-马尔可夫定理不一样，并不仅限于线性估计量一类。[4]因此，我们可以说最小二乘估计量是**最优无偏估计量** (BUE)。那就是说，在整个无偏估计序列中，这些估计量具有最小方差。

总结：重要的是要看到，正态性假定使我们能推导出 $\hat{\beta}_1$(正态)，$\hat{\beta}_2$(正态)以及 $\hat{\sigma}^2$(与 χ^2 相关的)的概率或抽样分布。在下一章中我们将看到，这

将简化建立置信区间和假设检验（统计的）的工作。

顺便指出，如果假定 u_i 遵从以 0 为均值 σ^2 为方差的正态分布，则 Y_i 本身也遵循正态分布，其均值和方差依次为：

$$E(Y_i)=\beta_1+\beta_2X_i \tag{4.3.7}$$

$$\operatorname{var}(Y_i)=\sigma^2 \tag{4.3.8}$$

或更简洁地写为：

$$Y_i \sim N(\beta_1+\beta_2X_i,\sigma^2) \tag{4.3.9}$$

§4.4 最大似然法

和 OLS 相比，一种具有某些更强的理论性的点估计方法是**最大似然**（maximum likelihood，ML）法。此法因较复杂，故放在本章的附录中讨论。对一般读者来说，只需知道：如果假定 u_i 是正态分布的（假定的理由已讨论在前），则回归系数 β 的 ML 估计量和 OLS 估计量是相同的，无论所考虑的是简单（二元）回归还是复（多元）回归皆然。但 σ^2 的 ML 估计量是
113 $\sum\hat{u}_i^2/n$。这是一个有偏差的估计量，而 σ^2 的 OLS 估计量为 $\sum\hat{u}_i^2/(n-2)$ 则是无偏的。比较 σ^2 的这两种估计量，可知随着样本容量 n 的变大，两者趋于相等。因此说，σ^2 的 ML 估计量是渐近（即随 n 无限增大）无偏的。

既然补充了 u_i 的正态性假定，最小二乘法便为我们提供了对线性回归模型进行估计和假设检验的全部必备工具。即使读者由于最大似然法在数学上略为复杂而不愿意探讨它，也不致蒙受什么损失。

§4.5 要点与结论

1. 本章讨论了经典**正态**线性回归模型。

2. 此模型与经典线性回归模型的差异在于它特意假定了进入回归模型的干扰项 u_i 是正态分布的。CLRM 则不要求对 u_i 的概率分布作任何假定；而仅要求 u_i 的均值为零以及方差为一个有限的常数。

3. 正态性假定的理论依据是**中心极限定理**。

4. 正态性假定除外，在第 3 章中所讨论的其他假定下，高斯-马尔可夫定理表明，OLS 估计量是最优线性无偏估计量。

5. 加上正态性假定，OLS 估计量就不仅是**最优无偏估计量**，而且遵循熟知的概率分布。截距和斜率的 OLS 估计量本身是正态分布的，并且 u_i 的方差的 OLS 估计量（$=\hat{\sigma}^2$）与 χ^2 分布相关。

6. 在第 5 章和第 8 章中，我们将说明怎样把这些知识用于推断样本总体参数的真值。

7. 取代最小二乘法的另一方法是**最大似然**（**ML**）法。然而，为了使用此法，必须对干扰项 u_i 的概率分布作一假定。在回归分析中，最常做的假定就是 u_i 服从正态分布。

8. 在正态性假定下，截距和斜率参数的 ML 估计量和 OLS 估计量是完全相同的。但是，u_i 的方差的 OLS 和 ML 两个估计量却有差别。然而，在大样本中，这两个估计量趋于一致。

9. 因此，通常称 ML 法为大样本方法。ML 法有更为广泛的应用。意思是，它可应用于参数为非线性的回归模型。在非线性情况下，一般都不用 OLS。有关于此的更多内容见第 14 章。

114 10. 在本教材中，我们基本上用的是 OLS 法，其具体理由如下：(a) 相对于 ML 来说，OLS 易于应用；(b) β_1 和 β_2 的 ML 估计量和 OLS 估计量是相同的（对多元回归也是如此）；(c) 即使样本容量不很大，σ^2 的 OLS 和 ML 估计量也相差不多。

然而，为了方便有数学爱好的读者，本章附录以及附录 A 中给出了 ML 的一个简要的导论。

附录 4A

4A.1 双变量回归模型的最大似然估计

假定在双变量模型 $Y_i = \beta_1 + \beta_2 X_i + u_i$ 中，Y_i 是正态且独立分布的，其均值 $= \beta_1 + \beta_2 X_i$，其方差 $= \sigma^2$。[参看方程 (4.3.9)。] 从而，给定上述均值和方差，Y_1，Y_2，…，Y_n 的联合概率密度函数就可写为：

$$f(Y_1, Y_2, \cdots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2)$$

由于诸 Y 的独立性，此联合概率密度函数可写为 n 个单独的密度函数之积：

$$\begin{aligned} &f(Y_1, Y_2, \cdots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \\ &= f(Y_1 \mid \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 \mid \beta_1 + \beta_2 X_i, \sigma^2) \cdots f(Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \quad (1) \end{aligned}$$

其中：

$$f(Y_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \right\} \quad (2)$$

这是给定均值和方差的一个正态分布变量的密度函数。

（注：exp 指以 { } 中的表达式为幂指数的 e。）

将 (2) 代入 (1) 中的每个 Y_i 得：

$$f(Y_1, Y_2, \cdots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2)$$
$$= \frac{1}{\sigma^n(\sqrt{2\pi})^n} \exp\left\{ -\frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \right\} \tag{3}$$

若 Y_1, Y_2, …, Y_n 为已知或给定，而 β_1, β_2 和 σ^2 为未知，则称（3）为**似然函数**（likelihood function），记为 LF（β_1, β_2, σ^2）并写为[1]：

15 $$\mathrm{LF}(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n(\sqrt{2\pi})^n} \exp\left\{ -\frac{1}{2} \sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \right\} \tag{4}$$

最大似然法，顾名思义，就是要在估计未知参数时使得观测到给定的这些 Y_i 的概率尽可能大。因此，有必要求函数（4）的最大值点。这不过是微分运算中的一个简单的练习题。为了求微分，将（4）表示成如下的对数形式更为容易运算。[2]（注：$\ln = \log_e$ 自然对数。）

$$\ln \mathrm{LF} = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$$
$$= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \tag{5}$$

将（5）对 β_1, γ_2 和 σ^2 求偏导数得：

$$\frac{\partial \ln \mathrm{LF}}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-1) \tag{6}$$

$$\frac{\partial \ln \mathrm{LF}}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_i)(-X_i) \tag{7}$$

$$\frac{\partial \ln \mathrm{LF}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_i)^2 \tag{8}$$

令这些方程为零（取最值的一阶条件），并记 ML 估计量为 $\tilde{\beta}_1$, $\tilde{\beta}_2$ 和 $\tilde{\sigma}^2$，便得[3]：

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) = 0 \tag{9}$$

$$\frac{1}{\tilde{\sigma}^2}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)X_i = 0 \tag{10}$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum (Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = 0 \tag{11}$$

116 经过简化，方程（9）和（10）给出：

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \tag{12}$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \tag{13}$$

这正是在（3.1.4）和（3.1.5）中得到的最小二乘理论的正规方程。由此可见，ML 估计量 $\tilde{\beta}_i$ 和由（3.1.6）和（3.1.7）给出的 OLS 估计量 $\hat{\beta}_i$（为了表示复数，故将 $\hat{\beta}$ 写为 $\hat{\beta}_i$。——译者注）是相同的，这个等同的结果并非偶然。分析一下似然函数（5），我们看到最后一项是带有负号的。因此，（5）

的最大化就是这一项的最小化，而后者如同我们能从（3.1.2）看到的，正是最小二乘法所采取的方法。

将 ML（=OLS）估计量代入（11）并加以简化，就得到 σ^2 的 ML 估计量为：

$$
\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{n}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 \\
&= \frac{1}{n}\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \\
&= \frac{1}{n}\sum \hat{u}_i^2 \qquad (14)
\end{aligned}
$$

从（14）明显看出，ML 估计量 $\tilde{\sigma}^2$ 不同于 OLS 估计量 $\hat{\sigma}^2 = [1/(n-2)]\sum\hat{u}_i^2$，后者在附录 3A 第 3A.5 节中已被证明是 σ^2 的一个无偏估计量，因此，σ^2 的 ML 估计量是有偏差的。偏差的程度用如下方法是不难决定的。

对（14）两边取数学期望得：

$$
\begin{aligned}
E(\tilde{\sigma}^2) &= \frac{1}{n}E(\sum\hat{u}_i^2) \\
&= \left(\frac{n-2}{n}\right)\sigma^2 \qquad \text{利用附录 3A 第 3A.5 节的方程 (16)} \\
&= \sigma^2 - \frac{2}{n}\sigma^2 \qquad (15)
\end{aligned}
$$

这表明，在小样本中，$\tilde{\sigma}^2$ 偏之于过小（即低估了真实的 σ^2）。但应看到，随着样本大小 n 无限地增大，（15）中的第二项即偏误因子将趋于零。因此 $\tilde{\sigma}^2$ 是**渐近**（即在很大的样本中）**无偏的**，也就是 $\lim E(\tilde{\sigma}^2) = \sigma^2$，当
117 $n \to \infty$。还可进一步证明，$\tilde{\sigma}^2$ 也是**一致性估计量**。[4] 即当 n 无限增大时，$\tilde{\sigma}^2$ 收敛于其真值 σ^2。

4A.2 印度食物支出的最大似然估计

回到例 3.2 和回归（3.7.2），那里就印度 55 个农户的食物支出数据对总支出做了回归。由于在正态性假定下，回归系数的 OLS 和 ML 估计量相同，所以我们得到 ML 估计量为 $\tilde{\beta}_1 = \hat{\beta}_1 = 94.208\,7$ 和 $\tilde{\beta}_2 = \hat{\beta}_2 = 0.438\,6$。$\sigma^2$ 的 OLS 估计量为 $\hat{\sigma}^2 = 4\,469.691\,3$，但 ML 估计量 $\tilde{\sigma}^2 = 4\,407.156\,3$ 却小于 OLS 估计量。如注意到的那样，ML 估计量在小样本情形下有向下的偏误；即它低估了总体的真实方差 σ^2。当然，如你所料，随着样本容量的扩大，这两个估计量之间的差别将越来越小。将估计量的值放到似然函数中，我们得到的值为 −308.162 5。若需要 LF 的最大值，只须取 −308.162 5的反对数即可。没有其他任何参数值能以更高的概率得到你用以分析的样本。

附录 4A　习题

4.1 “若两个随机变量在统计上独立，则两者的相关系数为零。但反之未必成立。就是说，零相关不意味着统计独立性。然而，如果两个变量都是正态分布的，则零相关必然意味着统计独立性。”试利用下面的两个正态分布变量 Y_1 和 Y_2 的联合概率密度函数（又称**双变量正态概率密度函数**）来证实这一论述：

$$f(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\times\left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{Y_2-\mu_2}{\sigma^2}\right)^2\right]\right\}$$

118 其中 $\mu_1 = Y_1$ 的均值

$\mu_2 = Y_2$ 的均值

$\sigma_1 = Y_1$ 的标准差

$\sigma_2 = Y_2$ 的标准差

$\rho = Y_1$ 与 Y_2 之间的相关系数

4.2 试用取最值的二阶条件（即二阶导数检验），证明通过解方程（9）、（10）和（11）而得到的 β_1，β_2 和 σ^2 的 ML 估计量，确实可使似然函数取最大值。

4.3 随机变量 X 服从**指数分布**（exponential distribution），如果它有如下的概率密度函数：

$$f(X) = (1/\theta)e^{-X/\theta} \quad \text{当 } X > 0$$
$$= 0 \quad \text{当 } X \text{ 取其他值}$$

其中 $\theta > 0$ 是此分布的参数。试用 ML 法证明 θ 的 ML 估计量是 $\tilde{\theta} = \sum X_i/n$。（原书误为 $\hat{\theta} = \sum X_i/n$。——译者注）其中 n 为样本容量。也就是证明 θ 的 ML 估计量是样本均值 $\bar{X}$。

【附录注释】

[1] 当然，若 β_1，β_2 和 σ^2 已知而诸 Y_i 未知，则（4）代表联合概率密度函数——指联合地测到 Y_i 的概率。

[2] 因对数函数是单调函数，故 ln LF 和 LF 在同一点上达到最大。

[3] 我们用～（波纹号）表示 ML 估计量；用＾（帽号）表示 OLS 估计量。

[4] 关于最大似然估计量性质的一般讨论，以及渐近无偏性与一致性之间的差别所在，参看附录 A。粗略地说，对于渐近无偏性，我们要设法求出当 n 趋于无穷大时的 $\lim E(\tilde{\sigma}_n^2)$，其中 n 代表估计量所依据的样本大小，而对于一致性，我们要设法求出当 n 无限增大时 $\tilde{\sigma}_n^2$ 的变化情况。注意，无偏性是指基于给定样本容量的一

个估计量的重复抽样性质，而一致性是关于一个估计量在样本含量无限增大过程中所表现的性态。

【注释】

[1] Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, 2d ed, Harcourt Academic Press, New York, 2000, pp.193 - 194。对此定理有一个简单直接的讨论。本定理的一个例外情形是柯西分布；见 M.G.Kendall and A.Stuart, *The Advanced Theory of Statistics*, Charles Griffin & Co., London, 1960, vol.1, pp.248 - 249。

[2] Harald Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946, Chap.17, 叙述了中心极限定理的各种形式。

[3] 例如，可参阅 Robert V.Hogg and Allen T.Craig, *Introduction to Mathematical Statistics*, 2d ed, Macmillan, New York, 1965, p.144.

[4] C.R.Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, 1965, p.258.

第 5 章 双变量回归：区间估计与假设检验

119 警惕过多地检验假设；你对数据愈苛求，数据会愈多地向你供认，但在威逼下得到的供词，在科学询查的法庭上是不容许的。[1]

如第 4 章所指出，估计与假设检验构成经典统计学的两个主要分支。估计理论由两部分组成：点估计与区间估计。在前面两章中我们介绍 OLS 和 ML 点估计方法时已透彻地讨论过点估计。在本章中，我们先考虑区间估计，然后再讨论假设检验的问题。后者是一个与区间估计有紧密关系的问题。

§5.1 统计学的预备知识

在讲解如何建立置信区间与检验统计假设的具体步骤之前，我们假定读者已熟悉概率与统计学的基本概念。附录 A 虽然不能代替一门统计学的基础课程，却给出了读者所必须掌握的统计学要义。一些基本概念如**概率、概率分布、第Ⅰ类和第Ⅱ类错误、显著（性）水平、统计检验的功效**以及**置信区间**等等，对于理解本章和以后各章的内容都起着关键的作用。

§5.2 区间估计：一些基本概念

120 为了建立概念，不妨考虑第3章的人为消费—收入一例。方程（3.6.2）给出边际消费倾向 β_2 的估计值为0.509 1，这是对未知的总体MPC β_2 的一个单一的点估计。这个估计值有多可靠？如第3章所指出的，虽然在重复抽样中估计值的均值可能会等于真值［注：$E(\hat{\beta}_2)=\beta_2$］，但由于抽样波动，单一估计值很可能不同于真值。在统计学中，一个点估计量的可靠性由它的标准误来衡量。因此，我们不能完全信赖一个点估计值，而是要围绕点估计量构造一个区间。比方说，在点估计量的两旁各划出宽为2或3个标准误的一个区间，使得它有95%的概率包含着真实的参数值。这就是**区间估计**的粗略概念。

说得更确切些，假定我们想知道究竟，比方说，$\hat{\beta}_2$ 离 β_2 有多"近"。为了这个目的，我们试求两个正数 δ 和 α，α 位于0与1之间，使得**随机区间**（random interval）$(\hat{\beta}_2-\delta,\ \hat{\beta}_2+\delta)$ 包含 β_2 的概率为 $1-\alpha$。用符号表示：

$$\Pr(\hat{\beta}_2-\delta\leqslant\beta_2\leqslant\hat{\beta}_2+\delta)=1-\alpha \tag{5.2.1}$$

这样的一个区间，如果存在的话，就称之为**置信区间**（confidence interval）；$1-\alpha$ 称**置信系数**（confidence coefficient）；而 α $(0<\alpha<1)$ 称**显著（性）水平**（level of significance）。[2]置信区间的端点称**置信限**（confidence limits），也称**临界值**（critical values）。$\hat{\beta}_2-\delta$ 为**置信下限**，而 $\hat{\beta}_2+\delta$ 为**置信上限**。顺便指出，在实践中，α 和 $1-\alpha$ 常用百分数表示，例如，若 $\alpha=0.05$，则表示为5%和95%的形式。

方程（5.2.1）表明，和点估计量相对照，区间估计量是一个构造出来的区间，要使得它把参数的真值包括在区间的界限内有一个特定的概率 $1-\alpha$。比方说，$\alpha=0.05$ 或5%，那么（5.2.1）就可读为：式中的（随机）区间包含真实 β_2 的概率为0.95或95%。从而区间估计量给出了一个真实 β_2 会落入其中的数值范围。

理解区间估计所涉及的下列疑难是非常重要的：

1. 方程（5.2.1）并没有说 β_2 落入给定界限内的概率是 $1-\alpha$。因为 β_2 虽然未知，但被假定为某个定数，或者落在区间内，或者落在区间外。（5.2.1）所表述的是，使用本章所描述的方法构造出来的一个区间包含 β_2 的概率为 $1-\alpha$。

2. （5.2.1）中的区间是一个**随机区间**；它从一个样本变到另一个样本，因为它是根据 $\hat{\beta}_2$ 来构造的，而 $\hat{\beta}_2$ 是随机的。（为什么？）

3. 既然置信区间是随机的，对它做的概率表述就应从长远的意义上，也就是从重复抽样的意义上加以理解。说得更具体些，（5.2.1）是说：如果在重复抽样中，像（5.2.1）那样，在 $1-\alpha$ 的概率基础上构造置信区间多

次，那么，从长期看，平均地说，这些区间将有 100%（1 - α）次包含着参数的真值。

4. 如在上面 2 中所看到的，只要 $\hat{\beta}_2$ 尚不知道，区间（5.2.1）就是随机的。但是，一旦我们有了一个特定的样本并获得 $\hat{\beta}_2$ 的一个特定的数值，区间（5.2.1）就不再是随机的，而是固定的了。这时我们**不可**做如同（5.2.1）那样的表述；也就是，我们不能说一个给定了的**固定**区间包含真实 β_2 的概率是 $1-\alpha$。在这种情况下，β_2 或者落入这个固定区间内或者落在区间之外，从而概率只能是 1 或 0。这样，拿我们的人为消费—收入例子来说，如果我们求得的 95% 置信区间是（$0.426\,8 \leqslant \beta_2 \leqslant 0.591\,4$），如同我们很快就要看到的（5.3.9）那样，我们就**不可以**说这个区间包含真实 β_2 的概率是 95%。这个概率不是 1 就是 0。

怎样构造置信区间？从上面的讨论，读者也许预想到，如果估计量的**抽样或概率分布**已知，就可以做出如同（5.2.1）那样的置信区间的表达式。在第 4 章中，我们曾看到，在干扰项 u_i 的正态性假定下，OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 本身是正态分布的，而 OLS 估计量 $\hat{\sigma}^2$ 则与 χ^2 分布有关。这样看来，构造置信区间是一桩简单的事情。确实如此！

§5.3 回归系数 β_1 和 β_2 的置信区间

β_2 的置信区间

第 4 章 4.3 节已表明，在 u_i 的正态性假定下，OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 本身就是正态分布的，其均值和方差已随之列出。因此，以 $\hat{\beta}_2$ 为例，变量

$$
\begin{aligned}
Z &= \frac{\hat{\beta}_2-\beta_2}{\mathrm{se}(\hat{\beta}_2)} \\
&= \frac{(\hat{\beta}_2-\beta_2)\sqrt{\sum x_i^2}}{\sigma}
\end{aligned} \tag{5.3.1}
$$

122 如在（4.3.6）中所表明的，这是一个标准化正态变量。因此，如果真实的总体方差 σ^2 已知，就可利用正态分布对 β_2 作概率性表达。当 σ^2 已知时，以 μ 为均值，σ^2 为方差的正态分布变量有一个重要性质，就是正态曲线下 $\mu \pm \sigma$ 之间的面积，约占 68%；在 $\mu \pm 2\sigma$ 之间的面积约占 95%；在 $\mu \pm 3\sigma$ 之间的面积约占 99.7%。

但是 σ^2 很少能知道，在实践中是用无偏估计量 $\hat{\sigma}^2$ 来测定。如果我们用 $\hat{\sigma}$ 代替 σ，（5.3.1）就可写为：

$$
t=\frac{\hat{\beta}_2-\beta_2}{\mathrm{se}(\hat{\beta}_2)}=\frac{\text{估计量}-\text{参数}}{\text{估计量的标准误的估计值}}
$$

$$= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \tag{5.3.2}$$

其中$\mathrm{se}(\hat{\beta}_2)$在这里用来表示**估计量的标准误**（estimated standard error）。可以证明（见附录5A，第5A.1节），这样定义的t变量遵循自由度为$n-2$的t分布。[注意（5.3.1）与（5.3.2）之间的区别。] 因此，我们不用正态分布，而是要用t分布来建立β_2的置信区间，如下所示：

$$\Pr(-t_{\alpha/2} \leqslant t \leqslant t_{\alpha/2}) = 1 - \alpha \tag{5.3.3}$$

其中位于这个双重不等式中间的t值就是由（5.3.2）给出的t值，而$t_{\alpha/2}$是由显著水平为$\alpha/2$和自由度为$n-2$的t分布给出的t变量值，常常被称为在$\alpha/2$显著水平上的**临界**（critical）值。将（5.3.2）代入（5.3.3）得：

$$\Pr\left[-t_{\alpha/2} \leqslant \frac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)} \leqslant t_{\alpha/2}\right] = 1 - \alpha \tag{5.3.4}$$

重新整理（5.3.4）得[3]：

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\mathrm{se}(\hat{\beta}_2) \leqslant \beta_2 \leqslant \hat{\beta}_2 + t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)] = 1 - \alpha \tag{5.3.5}$$

123 方程（5.3.5）给出β_2的一个$100(1-\alpha)\%$**置信区间**，可更简洁地把它写成：

β_2的$100(1-\alpha)\%$置信区间：

$$\hat{\beta}_2 \pm t_{\alpha/2}\mathrm{se}(\hat{\beta}_2) \tag{5.3.6}$$

利用（4.3.1）和（4.3.2），类似地推理，就能写出：

$$\Pr[\hat{\beta}_1 - t_{\alpha/2}\mathrm{se}(\hat{\beta}_1) \leqslant \beta_1 \leqslant \hat{\beta}_1 + t_{\alpha/2}\mathrm{se}(\hat{\beta}_1)] = 1 - \alpha \tag{5.3.7}$$

或更简练地写为：

β_1的$100(1-\alpha)\%$置信区间：

$$\hat{\beta}_1 \pm t_{\alpha/2}\mathrm{se}(\hat{\beta}_1) \tag{5.3.8}$$

注意，由（5.3.6）和（5.3.8）给出的置信区间有一重要特点：在这两个方程中**置信区间的宽度都与估计量的标准误成比例**。就是说，标准误越大，置信区间越宽。换句话说，估计量的标准误越大，对未知参数的真值进行估计的不确定性越大。因此，估计量的标准误常被喻为估计量的**精度**。就是说，用估计量去测定真实的总体值有多精确。

回到我们用以说明方法的消费—收入例子。在第3章中（3.6节），我们求得$\hat{\beta}_2 = 0.509\,1$，$\mathrm{se}(\hat{\beta}_2) = 0.035\,7$并且自由度$=8$。若取$\alpha = 5\%$，也就是取95%置信系数，则$t$表告诉我们，自由度为8的临界值$t_{\alpha/2} = t_{0.025} = 2.306$。读者将这些值代入（5.3.5），即可证实$\beta_2$的95%置信区间为：

$$0.426\,8 \leqslant \beta_2 \leqslant 0.591\,4 \tag{5.3.9}$$

或者，按照（5.3.6）的形式把它写为：

$$0.509\,1 \pm 2.306(0.035\,7)$$

即：

$$0.509\,1 \pm 0.082\,3 \tag{5.3.10}$$

对这个置信区间的解释是：给定置信系数为95%，从长远看，在类似
124 于（0.426 8，0.591 4）的每100个区间中，将有95个包含着真实的 β_2 值。但要注意，如前面所告诫的，我们不可以说，这个特定的区间有95%的概率包含着真实的 β_2，因为这个区间已经固定而不再是随机的了，那么，β_2 要么落入其中，要么落在其外。因此，这个给定的固定区间包含着真实的 β_2 的概率不是1就是0。

$\boldsymbol{\beta}_1$ 的置信区间

仿照（5.3.7），读者容易证实，在消费—收入一例中，β_1 的95%置信区间是：

$$9.664\,3 \leqslant \beta_1 \leqslant 39.244\,8 \tag{5.3.11}$$

或利用（5.3.8），求出它是：

$$24.454\,5 \pm 2.306(6.413\,8)$$

即：

$$24.454\,5 \pm 14.790\,2 \tag{5.3.12}$$

在你解释这个置信区间时，仍然需要留意，从长远看，在类似于（5.3.11）的每100个区间中有95个将包含真实的 β_1；但这个特殊的固定区间含有真实 β_1 的概率则是1或0。

$\boldsymbol{\beta}_1$ 和 $\boldsymbol{\beta}_2$ 的联合置信区域

有时我们需要构造 β_1 和 β_2 的这样一个联合置信域，使得 β_1 和 β_2 同时落在其中的置信系数（$1-\alpha$）等于，比方说，95%。由于这个问题比较复杂，有兴趣的读者可参考有关文献。[4]我们将在第8章和第10章简要地提到这个主题。

§5.4 σ^2 的置信区间

如在第 4 章第 4.3 节中所指出的，在正态性假定下，变量：

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

125 遵循自由度为 $n-2$ 的 χ^2 分布。[5]故可利用 χ^2 分布来建立 σ^2 的置信区间：

$$\Pr(\chi^2_{1-\alpha/2} \leqslant \chi^2 \leqslant \chi^2_{\alpha/2}) = 1-\alpha \tag{5.4.2}$$

其中居于双重不等式中间的 χ^2 值由 (5.4.1) 给出，而 $\chi^2_{1-\alpha/2}$和 $\chi^2_{\alpha/2}$是得自 χ^2 数值表中自由度为 $n-2$ 的两个 χ^2 值（**临界** χ^2 值），使得它们各切去 χ^2 分布的 100 $(\alpha/2)$% 的尾部面积，如图 5.1 所示。

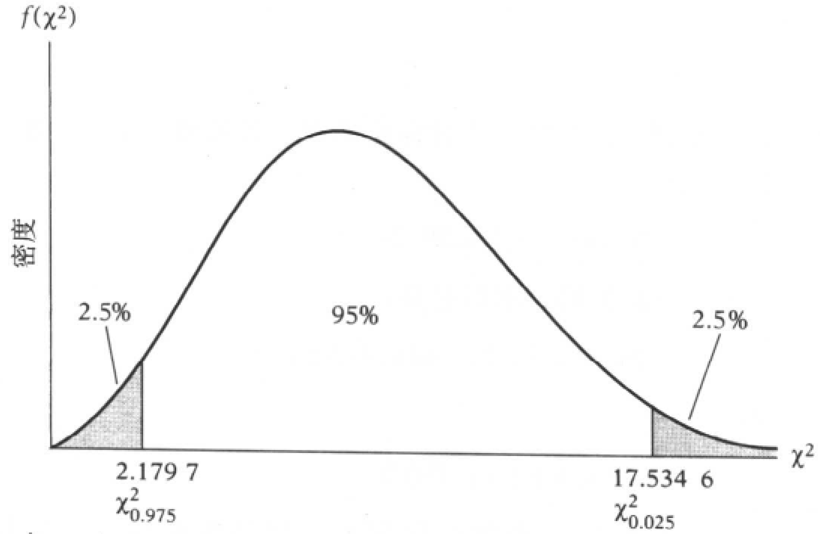


图 5.1　χ^2 的 95% 置信区间（8 个自由度）

将 (5.4.1) 的 χ^2 代入 (5.4.2)，并加整理得

$$\Pr\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leqslant \sigma^2 \leqslant (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha \tag{5.4.3}$$

这就给出 σ^2 的 100 $(1-\alpha)$% 置信区间。

作为说明，考虑下面的例子。由第 3 章 3.6 节我们求得 $\hat{\sigma}^2 = 42.159\ 1$和自由度 = 8。如果取 α 为 5%，则对于自由度为 8 的 χ^2 表给出下列临界值：$\chi^2_{0.025} = 17.534\ 6$ 和 $\chi^2_{0.975} = 2.179\ 7$。这些值表示 χ^2 值超过17.534 6的概率是 2.5%；超过 2.179 7 的概率是 97.5%。因此，这两值之间的区间构成 χ^2 的一个 95% 置信区间，如图 5.1 所示。（注意 χ^2 分布的偏态。）

126 读者将比例的数据代入 (5.4.3)，便能证实 σ^2 的 95% 置信区间为：

$$19.2347 \leqslant \sigma^2 \leqslant 154.7336 \tag{5.4.4}$$

对这个区间的解释是：如果我们建立了 σ^2 的95%置信（界）限，并且事先声称这些界限将包含真实的 σ^2，那么从长远看，我们将有95%的机会是正确的。

§5.5 假设检验：概述

我们已经讨论了点和区间估计的问题。现在我们考虑假设检验的问题。在本节中我们只对这个问题作一个简要的概述；更多的一些细节见附录A。

统计假设检验的问题可简单地叙述如下：某一给定的观测或发现是否与某声称的假设相符？这里用"相符"（compatible）一词表示与假设的值"足够相近"，因而我们不拒绝（摒弃）所声称的假设。例如，如果某种理论或先前经验使我们相信消费—收入一例的真实斜率系数 β_2 等于1，那么从表3.2的样本得到的观测值 $\hat{\beta}_2=0.5091$ 是否与声称的假设值1相一致呢？如果是，我们不拒绝该假设；否则就可拒绝它。

用统计学的语言说，这个声称的假设叫做**虚拟假设**（null hypothesis）[通常代表一种信以为真或意在维护的所谓**维持假设**（maintained hypothesis）]，并用符号 H_0 来表示。通常在检验虚拟假设时要有一个**对立假设**（alternative hypothesis），常记为 H_1。比如，H_1 表示真实的 β_2 不等于1。对立假设可以是**简单的**或**复合的**。[6] 例如，H_1：$\beta_2=1.5$ 是一个简单假设，但 H_1：$\beta_2 \neq 1.5$ 则是一个复合假设。

假设检验理论是要研制出一个观测或程序，以便决定拒绝抑或不拒绝一个虚拟假设。为了设计这样的规则有两种**互为补充的**方法，就是**置信区间**和**显著性检验**（test of significance）。两种方法都宣称所考虑的变量（统计量或估计量）遵循某种概率分布，并且做假设检验就在于对这个分布的参数值发表意见或做出判断。例如，我们知道在正态性假定下，$\hat{\beta}_2$ 是正态分布的，其均值等于 β_2；其方差由（4.3.4）给出。当我们假设 $\beta_2=1$ 时，我们就是在对这个正态分布的两参数之一即均值做出判断。本书中所遇到的统计假设，大多数都属于这种类型，都是对某些假定的概率分布诸如正态，F，t
127 或 χ^2 分布中的一个或多个参数的值做出判断。这些判断是怎样做出的将在下面两节讨论。

§5.6 假设检验：置信区间的方法

双侧或双尾检验

为了说明置信区间的方法，再次回到消费—收入一例。我们知道，所估计的边际消费倾向 $\hat{\beta}_2$ 是 0.509 1。假使我们公设：

$$H_0:\ \beta_2 = 0.3$$
$$H_1:\ \beta_2 \neq 0.3$$

就是说在虚拟假设下 MPC 是 0.3，而在对立假设下 MPC 大于或小于 0.3。虚拟假设是一个简单假设，而对立假设则是一个复合假设；实际上这就是人们说的**双侧假设**（two-sided hypothesis）。这样的双侧对立假设，常常反映着我们对于对立假设偏离虚拟假设的方向没有一个强有力的先验性或理论性期望。

所观测的 $\hat{\beta}_2$ 是否与 H_0 相符？为了回答此问题，不妨引用置信区间（5.3.9）。我们知道，从长远来看，像（0.426 8，0.591 4）这样的许许多多区间将有 95%的概率包含真实的 β_2。因此，在长远（即重复抽样）意义下，这样的区间以（比方说）95%的置信系数给出的真值 β_2 落入其中的一个范围或界限，从而置信区间给出了可信的虚拟假设的一个集合。因此，如果虚拟假设的 β_2 落入这个 100（1－α）%置信区间，我们就不拒绝虚拟假设；如果它落在区间之外，我们就可拒绝虚拟假设。[7] 在图 5.2 中，我们勾画出了这一区间范围。

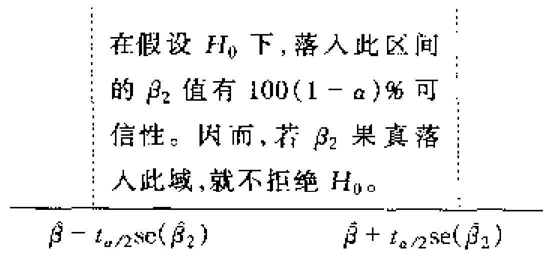


图 5.2　β_2 的一个 100（1－α）%置信区间

> **决策规则**：构造一个 β_2 的 100（1－α）%置信区间。如果 β_2 在假设 H_0 下落入此区间，就不要拒绝 H_0。但如果它落在此区间之外，就要拒绝 H_0。

遵照此规则，拿我们的人为例子来说，$H_0: \beta_2 = 0.3$。显然落在由

128 (5.3.9) 给出的95%置信区间之外，因此我们能以95%的置信度拒绝MPC的真值是0.3的假设。即使虚拟假设是真的，我们得到一个大到0.5091的MPC值，最多也只有5%的机会，这是一个小概率。

在统计学中，当我们拒绝虚拟假设时，我们说我们的发现是**统计上显著的**。反之，当我们不拒绝虚拟假设时，我们说我们的发现**不是统计上显著的**。

一些作者使用“统计上高度显著”一词。该词通常是指，当他们拒绝虚拟假设时，犯第Ⅰ类错误的概率（即 α）是一个小数，通常指1%。但在第5.8节中，我们对 ***p* 值**的讨论将表明，较好的做法是，让研究者自己去决定一个统计上的发现，究竟是“显著的”、“中度显著的”，还是“高度显著的”。

单侧或单尾检验

有时，我们因有一种强的先验性或理论性预期（也许根据某些先前的经验性工作）而把对立假设取为单侧或单向的，而不是才讨论过的双侧假设。比如，拿我们的消费—收入例子说，我们可以公设：

$$H_0:\ \beta_2 \leqslant 0.3 \text{ 和 } H_1:\ \beta_2 > 0.3$$

也许是经济理论或先前的经验提示了我们边际消费倾向大于0.3。尽管检验上述假设的程序容易从(5.3.5)导出，而实际的步骤最好通过下面即将讨论的显著性检验方法予以说明。

§5.7 假设检验：显著性检验法

检验回归系数的显著性：t 检验

129 检验统计假设的另一方法，同时也是对置信区间法的一种补充，是分别独立地由费希尔（R.A.Fisher）以及由内曼（Neyman）和皮尔逊合作研究出来的**显著性检验法**（test-of-significance approach）。[9] **概括地说，显著性检验是利用样本结果，来证实一个虚拟假设的真伪的一种检验程序。**显著性检验的基本思想在于一个**检验统计量**（作为估计量）以及在虚拟假设下，这个统计量的抽样分布。根据手中数据算出的统计量的值决定是否接受 H_0。

作为方法说明，回忆在正态性假设下的变量：

$$t=\frac{\hat{\beta}_2-\beta_2}{\operatorname{se}(\hat{\beta}_2)}$$

$$= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \tag{5.3.2}$$

遵循自由度为 $n-2$ 的 t 分布。如果在虚拟假设下 β_2 的真值被设定，则容易从现有的样本中算出（5.3.2）的 t 值。因此这个 t 变量就可作为一个统计量。并且由于这个统计量遵循 t 分布，故可做出如下的置信区间表述：

$$\Pr\left[-t_{\alpha/2} \leqslant \frac{\hat{\beta}_2 - \beta_2^*}{\mathrm{se}(\hat{\beta}_2)} \leqslant t_{\alpha/2}\right] = 1 - \alpha \tag{5.7.1}$$

其中 β_2^* 是在 H_0 下的 β_2 值，而 $-t_{\alpha/2}$ 和 $t_{\alpha/2}$ 是来自 t 表中相对于 $(\alpha/2)$ 显著水平和 $n-2$ 个自由度的 t 值（t 临界值）[参考(5.3.4)]。t 表见于附录 D。

将（5.7.1）整理得：

$$\Pr[\beta_2^* - t_{\alpha/2}\mathrm{se}(\hat{\beta}_2) \leqslant \hat{\beta}_2 \leqslant \beta_2^* + t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)] = 1 - \alpha \tag{5.7.2}$$

此式给出在给定 $\beta_2 = \beta_2^*$ 时，$\hat{\beta}_2$ 以概率 $1-\alpha$ 落入其中的区间。用假设检验的语言说，（5.7.2）中建立的 $100(1-\alpha)\%$ 置信区间叫做（虚拟假设的）
130 **接受域**，而置信区间以外的（一个或多个）区域叫做（H_0 的）**拒绝域或临界域**。如前所说，置信限，即置信区间的端点，又叫做**临界值**。

现在比较（5.3.5）和（5.7.2）就能看清假设检验的置信区间法和显著性检验法之间的密切联系。在置信区间程序中，我们试图建立一个以某种概率包含有真实但未知的 β_2 的一个范围或区间，而在显著性检验步骤中，我们假设 β_2 为某值，然后来看所计算的 $\hat{\beta}_2$ 是否位于该假设值周围的某个合理（置信）范围之内。

让我们再次回到消费—收入一例。我们知道 $\hat{\beta}_2 = 0.509\,1$，$\mathrm{se}(\hat{\beta}_2) = 0.035\,7$ 和自由度 $= 8$。若取 $\alpha = 5\%$，则 $t_{\alpha/2} = 2.306$。若令 H_0：$\beta_2 = \beta_2^* = 0.3$ 和 H_1：$\beta_2 \neq 0.3$，（5.7.2）即为：

$$\Pr(0.217\,7 \leqslant \hat{\beta}_2 \leqslant 0.382\,3) = 0.95 \tag{5.7.3}$$ [10]

如图 5.3 所示。因所测的 $\hat{\beta}_2$ 落在临界域中，故拒绝真实 $\beta_2 = 0.3$ 的虚拟假设。

在实践中，并不需要明显地估计（5.7.2），而可按（5.7.1）给出的双重不等式计算居中的 t 值，然后看它是落在两个 t 临界值之间还是之外。对于我们的例子

$$t = \frac{0.509\,1 - 0.3}{0.035\,7} = 5.86 \tag{5.7.4}$$

131 这清楚地落在图 5.4 的临界域内。结论仍然是一样的：拒绝 H_0。

注意，如果估计的 β_2（$=\hat{\beta}_2$）等于假设的 β_2，（5.7.4）中的 t 将为零。然而，随着估计的 β_2 值远离假设的 β_2 值，$|t|$（即 t 的绝对值；注意 t 可

正可负）将越来越大。因此，一个"大"的 $|t|$ 值便是与虚拟假设相抵触的迹象。当然，我们总可以利用 t 表来决定一个特殊的 t 值是大还是小；我们知道，这个答案依赖于自由度数和我们愿意接受的第Ⅰ类错误概率。如果你翻阅一下附录 D 中的 t 表，你将发现，对给定的自由度，得到的 $|t|$ 越大，其概率越小。比如说，对 20 个自由度来说，得到一个 1.725 或更大的 $|t|$ 值的概率是 0.10 或 10%。但对于同样的自由度，得到一个 3.552 或更大的 $|t|$ 值的概率仅是 0.002 或 0.2%。

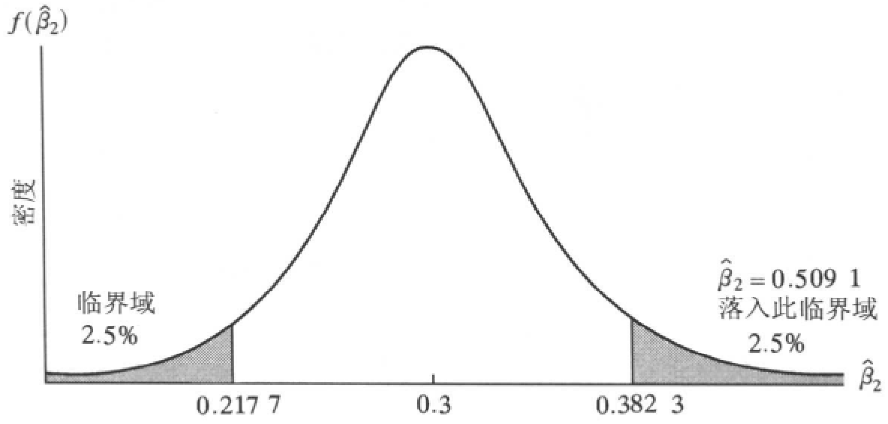


图 5.3　在假设 $\beta_2=0.3$ 下 $\hat{\beta}_2$ 的 95%置信区间

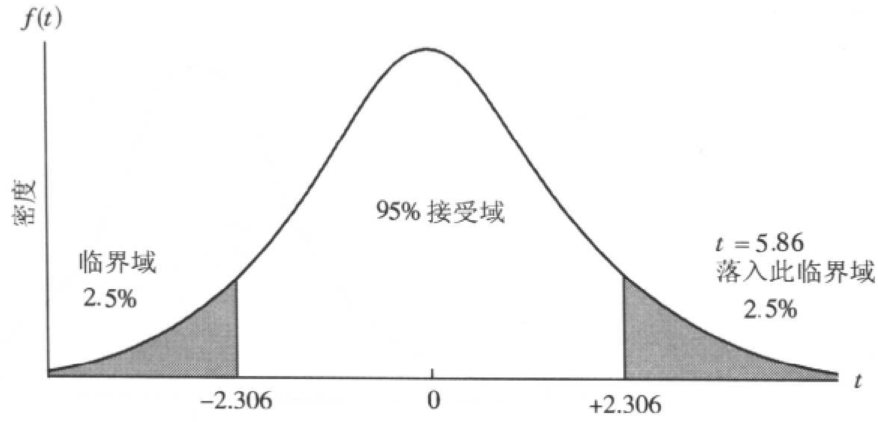


图 5.4　t（8 个自由度）的 95%置信区间

因为我们应用了 t 分布，所以前述检验程序适合称为 **t 检验**。**用显著性检验的语言说，如果一个统计量的值落在临界域内，这个统计量是统计上显著的，这时我们拒绝虚拟假设。如果一个统计量的值落在接受域中，这个检验是统计上不显著的，这时我们不拒绝虚拟假设。**在我们的例子中 t 是显著的，因而我们拒绝虚拟假设。

在结束我们对假设检验的讨论前，请注意我们刚才描述的检验程序，是一种**双侧**或**双尾**显著性检验程序，因为我们把有关概率分布的两个尾端当作拒绝域；如果虚拟假设值落入任一尾端，就拒绝该假设，这样做的原因是我

132 们的 H_1 是一个双侧复合假设；$\beta_2 \neq 0.3$ 表示 β_2 或者大于 0.3，或者小于 0.3。但是，假使先前的经验提示我们，MPC 预期要比 0.3 大，这样，我们就有：H_0：$\beta_2 \leqslant 0.3$ 和 H_1：$\beta_2 > 0.3$。虽然 H_1 仍是一个复合假设，但它却是单侧的。为了检验此假设，我们利用**单尾**（右尾部）**检验**，如图 5.5 所示。(并参考第 5.6 节的讨论。)

除了上端置信限或临界值现在是 $t_\alpha = t_{0.5}$ 即 5% 的水平外，检验的程序同前。如图 5.5 所示，在此例中我们并不需要考虑 t 分布的下尾端。究竟使用双尾还是单尾显著性检验，要看对立假设是怎样构成的。而后者又有赖于某种先验性思考或先前的实际经验。(进一步的讨论见第 5.8 节。)

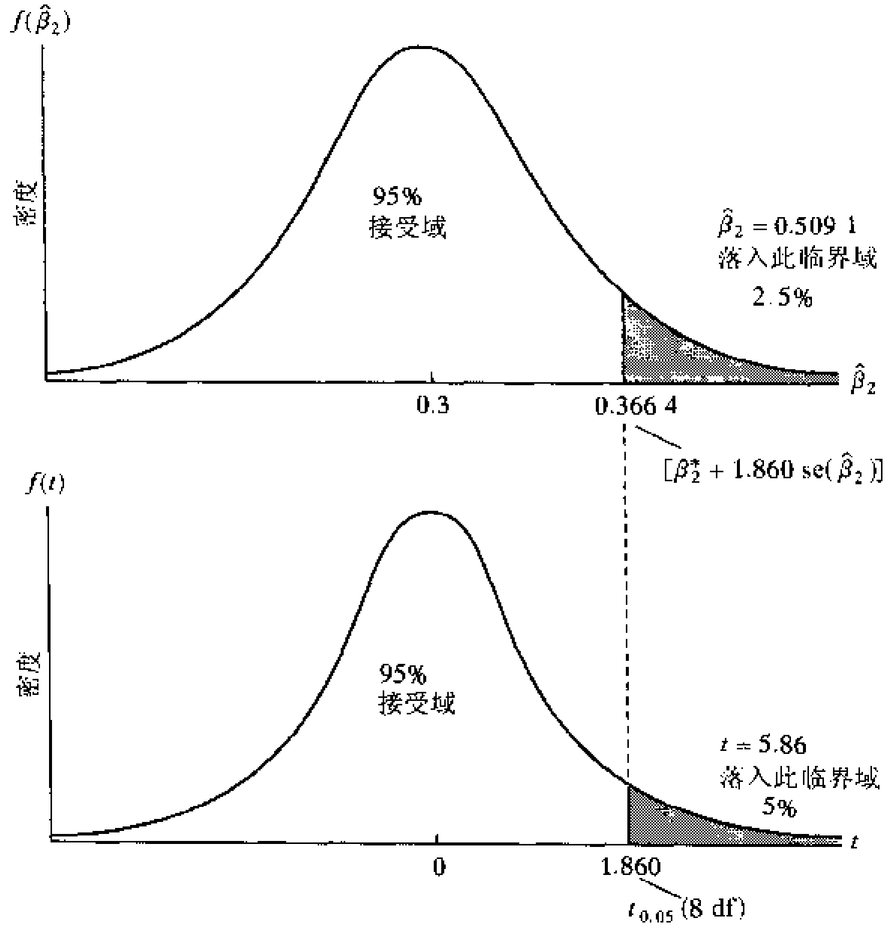


图 5.5　单尾显著性检验

假设检验的显著性 t 检验法可概括为表 5.1。

133 **表 5.1　　显著性 t 检验：决策规则**

假设类型	H_0：虚拟假设	H_1：对立假设	决策规则：拒绝 H_0 如果
双尾	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{\alpha/2}$，df

右尾	$\beta_2 \leqslant \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_\alpha$, df
左尾	$\beta_2 \geqslant \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_\alpha$, df

注：β_2^* 是 β_2 的假设数值。

$|t|$指 t 的绝对值。

t_α 或 $t_{\alpha/2}$指在 α 或 $\alpha/2$ 显著水平上的临界 t 值。

df：自由度，对双变量模型是（$n-2$），对三变量模型是（$n-3$）。依此类推，同样的程序适用于 β_1 的假设检验。

检验 σ^2 的显著性：χ^2 检验

作为检验方法论的另一说明，考虑以下变量：

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

前面已指出，这个变量遵循自由度为 $n-2$ 的 χ^2 分布。对于我们的人为例子，$\hat{\sigma}^2 = 42.159\,1$并且自由度=8。如果假设：$H_0$：$\sigma^2 = 85$ 则 H_1：$\sigma^2 \neq 85$，方程（5.4.1）便给出关于 H_0 的检验统计量。把相应的数值代入（5.4.1）就能求出在 H_0 下 $\chi^2 = 3.97$。如果我们取 $\alpha = 5\%$，χ^2 的两个临界值便是2.179 7和17.534 6。由于计算出来的 χ^2 落在这两个界限之间，表明数据支持虚拟假设，因此我们不去拒绝它（见图 5.1）。这一检验程序叫做 χ^2 **显著性检验**。假设检验的 χ^2 显著性检验法可概括为表 5.2。

表 5.2　χ^2 检验概要

H_0：虚拟假设	H_1：对立假设	临界域：拒绝 H_0 如果
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\mathrm{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\mathrm{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\mathrm{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\mathrm{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\mathrm{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\mathrm{df}}$ 或 $< \chi^2_{(1-\alpha/2),\mathrm{df}}$

注：σ_0^2 是在虚拟假设下的 σ^2 值。最后一列中 χ^2 的第一个下标指显著水平，而第二个下标指自由度。这些 χ^2 均是临界值。注意，对双变量回归模型，自由度为（$n-2$），对三变量回归模型自由度为（$n-3$）。依此类推。

§5.8 假设检验：一些实际操作问题

“接受”或“拒绝”假设的含义

134 在显著性检验，比如说 t 检验的基础上，如果我们决定“接受”虚拟假设，我们其实是说，根据样本证据，我们还没有理由去拒绝它；而不是说，虚拟假设毫无疑问是真的。为什么？为了回答此问题，让我们回到消费—收入一例，并假定 H_0：β_2（MPC）$=0.50$。既然 MPC 的估计值是 $\hat{\beta}_2=0.5091$，并且 $\mathrm{se}(\hat{\beta}_2)=0.0357$，那么，根据 t 检验，我们求得 $t=(0.5091-0.50)/0.0357=0.25$，这在 $\alpha=5\%$ 的水平上是不显著的。因此，我们说“接受”H_0。但是，让我们再假定 H_0：$\beta_2=0.48$，应用 t 检验又得 $t=(0.5091-0.48)/0.0357=0.82$，这仍然是统计上不显著的。于是，再说“接受”此 H_0。但这两个虚拟假设哪一个“真实”呢？我们不知道。所以，在“接受”一个虚拟假设时，应时刻警觉到另一个虚拟假设也会同样地与数据相符。有鉴于此，我们宁可说可以接受一个虚拟假设，而不说我们（确实）接受它。更好的说法是：

> ……正如一个法庭宣告某一判决为“无罪”而不是“清白”，统计检验的结论也应为“不拒绝”而不是“接受”。[11]

“零”虚拟假设与“2－t”经验法则

在经验工作中经常检验的一个虚拟假设是 H_0：$\beta_2=0$，即斜率系数是零。这个“零”虚拟假设像是一个草人，目的是要明确 Y 是否与解释变量 X 有任何关系。如果要从 Y 和 X 之间无任何关系开始，那么检验诸如 $\beta_2=0.3$ 或任何其他值的虚拟假设就没有意义。

可以非常容易地用前几节所讨论的置信区间法或 t 检验法去检验这个虚拟假设。但常常可采用“2 倍 t”显著性法则将这类按部就班的检验方法加以简化如下：

“2 倍 t”经验法则（“2-t” Rule of Thumb）：如果自由度≥20 且显著水平定在 0.05，那么，从（5.3.2）算得 t 值 $[=\hat{\beta}_2/\mathrm{se}(\hat{\beta}_2)]$ 在绝对值上超过 2 时，就可拒绝虚拟假设 $\beta_2=0$。

此准则的合理性是不难领会的。由（5.7.1）知，对于适当的自由度，

我们将拒绝 H_0：$\beta_2=0$，如果

$$t=\hat{\beta}_2/\mathrm{se}(\hat{\beta}_2)>t_{\alpha/2} \qquad \text{当} \quad \hat{\beta}_2>0$$

或

135 $$t=\hat{\beta}_2/\mathrm{se}(\hat{\beta}_2)< \quad t_{\alpha/2} \qquad \text{当} \quad \hat{\beta}_2<0$$

也就是，如果

$$|t|=\left|\frac{\hat{\beta}_2}{\mathrm{se}(\hat{\beta}_2)}\right|>t_{\alpha/2} \tag{5.8.1}$$

检查一下附录 D 中的 t 表便可看到，当自由度约为 20 或更大时，计算的 t 值在绝对值上超过 2，比如说 2.1，在 5% 的水平是统计上显著的，即意味着对虚拟假设的拒绝。因此，对于 20 或更多的自由度，如果计算的 t 值比如说是 2.5 或 3，我们就不需要查阅 t 表以评定所估的斜率系数的显著性。当然，为了得知准确的显著水平，我们可随时查阅 t 表，而当自由度小于 20 时，我们一定要查阅 t 表。

顺便指出，如果我们检验单侧假设 $\beta_2=0$，对 $\beta_2>0$ 或 $\beta_2<0$，则当：

$$|t|=\left|\frac{\hat{\beta}_2}{\mathrm{se}(\hat{\beta}_2)}\right|>t_{\alpha} \tag{5.8.2}$$

时，应拒绝虚拟假设。如果把 α 定在 0.05，则从 t 表我们看到，对于 20 或更多的自由度，一个超过 1.73 的 t 值在 5% 显著水平上（单尾）是统计上显著的。因而，每当 t 值超过比如说 1.8（在绝对值上）且自由度为 20 或更大，就不需要查阅 t 表以评定所测系数的统计显著性。当然，如果 α 选定在 0.01 或任何其他水平上，则还必须决定适当的 t 值作为临界参考点。至此读者们应该知道怎样做了。

建立虚拟与对立假设[12]

给定了虚拟和对立假设，如何检验它们的统计显著性就不再是什么神秘的事了。但是怎样建立这些假设呢？并没有什么一成不变的规则。常常是我们所研究的现象会提示我们虚拟和对立假设的性质。例如，习题 5.16 要求你去估计证券组合理论中的资本市场线（CML），该理论假设 $E_i=\beta_1+\beta_2\sigma_i$，其中 E＝组合证券的期望回报，而 σ＝回报的标准差。后者是风险的一个度量。因为人们预期着回报与风险有正向的关系——风险愈大，回报愈高——
136 相对于虚拟假设 $\beta_2=0$ 的自然对立假设便是 $\beta_2>0$。就是说，人们不会做出考虑 β_2 为负值的选择。

然而，试看对货币的需求问题。以后我们将表明，货币需求的重要决定因素之一是收入。先前的关于货币需求函数的研究曾表明，货币需求的收入弹性（指收入变化 1% 时货币需求变化的百分数）典型地位于 0.7 和 1.3 之间。因此，如果在一项新的货币需求研究中公设收入弹性系数为 1，则对立

假设可取为 $\beta_2 \neq 1$，即为双侧对立假设。

可见，理论预期或经验工作或同时两者都可作为建立假设的依据，但不管怎样建立这些假设，一件极为重要的事，是研究者要在进行经验调查研究之前先建立这些假设。不然的话，他或她就犯了迂回推理或自欺欺人的错误。就是说，如果先分析经验结果再作假设的话，就不免受到一种诱惑，要建立一种假设来维护自己所得的结果。要不惜一切代价去避免这种做法，至少为了科学事业要如此，请牢记本章一开头就引用的施蒂格勒（Stigler）的话！

选择显著性水平 α

讨论至此，应该清楚，拒绝或不拒绝虚拟假设，关键在于 α 这个显著性水平或犯**第Ⅰ类错误**的概率——拒绝了真值的假设的概率。在附录 A 中我们充分地讨论了第Ⅰ类错误的性质，它和**第Ⅱ类**错误（接受了错误的假设）的关系，以及为什么经典统计学通常都集中于讨论第Ⅰ类错误。但是，即使我们讨论了这些问题，人们仍会问，为什么 α 通常都固定在 1%、5%，也许还有 10%的水平上？其实，这些值并不是神圣不可侵犯的；任何其他值也都是可以的。

在像本书这样的以介绍性为主的书中，不可能深入地讨论为什么人们选择 1%，5%或 10%的显著性水平。这样做就会把我们引入到其本身就是一个学科分支的统计决策领域。然而，这里可以作一个简短的概括。如同在附录 A 中所讨论的，对于给定的样本大小，如果我们要减少第Ⅰ类错误，第Ⅱ类错误就要增加；反之亦然。就是说，给定了样本大小，如果我们企图减少拒绝真实假设的概率，我们就同时增加了接受错误假设的概率，因此，对于给定的样本大小，这两种错误类型之间有一种替换关系。解决这一替换关系的惟一途径是找出两类错误的相对代价。于是，

137 如果错误地拒绝一个其实是真实的虚拟假设（第Ⅰ类错误）的代价比起错误地接受一个其实是错误的虚拟假设（第Ⅱ类错误）的代价相对昂贵，那么把第Ⅰ类错误的概率定得低些将是合理的。反之，如果犯第Ⅰ类错误的代价比犯第Ⅱ类错误的代价相对低廉，就值得把Ⅰ类错误的概率定得高些（从而使第Ⅱ类错误的概率低些）。[13]

当然，困难在于我们很少知道犯这两类错误的代价。因此，应用计量经济学家一般都是跟随大多数，把 α 定在 1%、5%甚至 10%水平上。然后选择一个能使犯第Ⅱ类错误的概率尽可能小的检验统计量。用 1 减去犯第Ⅱ类错误的概率被称为**检验的功效**，这一程序相当于求检验功效的最大化。（关于检验功效的讨论，参看附录 A。）

但是，如果我们采用下节将讨论的检验统计量的 ***p* 值**，则所有有关选

择适当的 α 值的问题均可避免。

精确的显著性水平：p 值

如方才指出的，经典假设检验方法的痛处在于选择 α 时的武断性。当我们对给定的样本算出一个检验统计量（如 t 统计量）的值时，为什么不干脆查阅适当的统计表，看看得到一个大到和从样本得到的检验统计量那样大或者更大的数值的确切概率？这个概率就叫做 **p 值，即概率值**（probability value），也叫做**观测或精确显著性水平**，或犯第Ⅰ类错误的精确概率。用更专业化的语言说，p 值被定义为**一个虚拟假设可被拒绝的最低显著水平**。

为说明起见，仍回到消费—收入一例。给定虚拟假设：真实 MPC 是 0.3，我们得到（5.7.4）中的 t 值为 5.86。得到一个大到 5.86 或更大的 t 值的 p 值是什么？查阅附录 D 中的 t 表，我们看到，对于自由度为 8，得到这样的 t 值的概率一定比 0.001（单尾）或 0.002（双尾）小得多。通过计算机可以得出获得 5.86 或更大的 t 值（对于 8 个自由度）的概率约为 0.000 189。[14]这就是所测 t 统计量的 p 值。这一观测或精确的 t 统计量显著水平，比起习惯地并且是武断地固定的显著水平，如 1%，5%或 10%，要
138 小得多。事实上，如果我们真的使用刚才算的 p 值来拒绝“真实 MPC 是 0.3”的虚拟假设，那么我们犯第Ⅰ类错误的概率只有 0.02%，也就是万分之二！

我们在前面曾指出，如果数据不支持虚拟假设，则在虚拟假设下得到的 $|t|$ 值将会很“大”，得到这样一个 $|t|$ 值的 p 值因而就很“小”。换言之，对于给定的样本容量，随着 $|t|$ 的增加，p 值会不断下降，我们也就越来越有信心拒绝虚拟假设。

p 值和显著性水平 α 是怎样一种关系？如果我们养成一种习惯，把 α 固定在一个检验统计量（如 t 统计量）的 p 值上，这两个值就没有任何矛盾。换句话说，**与其人为地把 α 固定在某一水平，不如干脆选取检验统计量的 p 值**。让读者自己去决定是否在给定的 p 值水平上拒绝虚拟假设好了。如果在一项应用中，检验统计量的 p 值正好是 0.145 或 14.5%，并且如果读者想要在这一精确显著性水平上拒绝虚拟假设，就让他这样做好了。采用一个 14.5%，犯（第Ⅰ类）错误（拒绝了真实的虚拟假设的错误）的机会，并没有任何过错。同样，在我们的消费—收入例子中，研究者如要采用一个约为 0.02%的 p 值，而不愿接受比万分之二更大的犯错误机会，也没有什么过错。毕竟有一些研究者会是风险爱好者，而另一些则是风险厌恶者！

在本书的其余部分，我们一般都标出给定检验统计量的 p 值。读者可以把 α 固定在某一水平上，并在 p 值小于 α 时拒绝虚拟假设。这是他们的选择自由。

统计显著性与实际显著性

回到我们的消费—收入例子。现假设真实的 MPC 是 0.61($H_0: \beta_2 = 0.61$)。根据样本结果 $\hat{\beta}_2 = 0.509\,1$，我们曾求得 95%置信区间为(0.426 8，0.591 4)。由于此区间不包含 0.61，我们可以说，有 95%的把握我们的估计值(0.509 1)是统计上显著的，就是说，它显著地异于 0.61。

但是我们的这个发现有什么实际或实质显著性呢？也就是问，我们把 MPC 当作 0.61 而不是0.509 1，会有什么差别呢？两个 MPC 之间的0.100 9 的差别有什么实际重要意义？

问题的回答有赖于我们要用这些估计量来干什么。例如，从宏观经济学中我们得知，收入乘数是1/(1 − MPC)。因此，如果 MPC 是0.509 1，这个乘数就是 2.04；但如果 MPC 是 0.61，它就是 2.56。这就是说，如果政策打算增加 1 美元的开支，以拯救经济萧条，如果 MPC 是0.509 1，收入将最终增加 2.04 美元；而如果 MPC 是 0.61，则收入最终增加 2.56 美元。而这一差异对于经济的复苏也许很重要。

139 **全部讨论的要点在于：不要把统计上的显著性和实际上或经济上的显著性混同起来**，正如戈德伯格（Goldberger）所说：

> 当人们设定一个虚拟假设，比如说，H_0: $\beta_j = 1$ 时，其用意很可能是说 β_j 接近于 1，且接近到这样一个程度，以致为了一切实际的目的，都可以把它看做 1。然而，1.1 是否“实际上无异于”1.0？这是一个经济学的问题，而不是统计学的问题。我们不能靠假设检验来解决这个问题。因为，检验统计量[$t =$]$(b_j - 1) / \hat{\sigma}_{bj}$是用标准误做单位来衡量所估的系数的，而标准误并不是衡量经济参数 $\beta_j - 1$ 的一个有意义的单位。一个好的办法，也许是把“显著性”一词留给统计学概念用，而在经济学概念中使用“实在性”一词。[15]

戈德伯格的论点是重要的。当样本含量变得非常之大时，统计显著性的问题会变得黯然失色，而经济显著性的问题会变得至关重要。的确，在样本非常大的情况下，几乎任何虚拟假设都会被拒绝，点估计的大小就会成为惟一可研究的问题。

假设检验的置信区间法和显著性检验法的选择

在大多数应用性的经济分析中，虚拟假设的建立犹如一个草人的竖立。经验研究工作的目的，是要把它打倒，也就是，拒绝这个虚拟假设。以我们的消费—收入关系为例，虚拟假设 MPC 即 $\beta_2 = 0$ 显然是荒谬的，可是我们

却常用它来把经验上的成果戏剧化。看来著名期刊的编辑都发觉发表一篇不拒绝虚拟假设的经验性文章是不足以激动人心的。发现 MPC 在统计上异于零，多少要比发现它等于 0.7 更值得作为新闻报道！

因此，J·布拉德福康·德·朗（J.Bradford De Long）和凯文·兰（Kevin Lang）称辩说，对经济学家说，较好的做法是，

> ……集中讨论系数的大小并报告其置信水平，而不去提显著性检验。如果全部或几乎全部虚拟假设都是错误的，讨论一个估计值是否无异于它在虚拟假设下的预测值，就是无意义的，反而，我们也许想探明什么模型可充当良好的逼近式，这就需要知道为经验估计所排除的参数值域。[16]

简言之，这些作者认为，置信区间法优于显著性检验法。读者不妨把这一忠告铭记在心中。[17]

§5.9 回归分析与方差分析

140 本节我们从方差分析的观点研究回归分析，从而为读者介绍一种对待统计推断问题的、有启迪的和有补充作用的方法。

在第 3 章 3.5 节中，我们曾导出如下的恒等式：

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \tag{3.5.2}$$

这就是，TSS = ESS + RSS，它把总平方和分解为两个构成部分：解释平方和与残差或剩余平方和。对 TSS 的这些构成部分进行研究就叫做从回归的观点做**方差分析**（analysis of variance，ANOVA）。

同任一个平方和联系在一起的是它所依据的自由度，即独立观测值的个数。因为在计算样本均值 $\bar{Y}$ 时，我们失去一个自由度，故 TSS 有 $n-1$ 个自由度。RSS 有 $n-2$ 个自由度。（为什么？）（注：仅对有截距 β_1 的双变量回归模型才是对的。）ESS 有 1 个自由度（也仅对双变量情形才对），这是因为 ESS = $\hat{\beta}_2^2 \sum x_i^2$，当 $\sum x_i^2$ 为已知时仅是 $\hat{\beta}_2$ 的函数。

把各项平方和及其相应的 df，引入表 5.3，成为 ANOVA 表的标准型，有时又称 ANOVA 表。给定表 5.3 中的条目，现考虑以下变量：

$$\begin{aligned} F &= \frac{\text{MSS of ESS}}{\text{MSS of RSS}} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum \hat{u}_i^2/(n-2)} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2} \end{aligned} \tag{5.9.1}$$

表 5.3　　双变量回归模型的 ANOVA

变异来源	SS*	df	MSS+
由于回归（ESS）	$\sum \hat{y}_i^2 = \hat{\beta}_2^2 \sum x_i^2$	1	$\hat{\beta}_2^2 \sum x_i^2$
由于剩余（RSS）	$\sum \hat{u}_i^2$	$n-2$	$\frac{\sum \hat{u}_i^2}{n-2} = \hat{\sigma}^2$
TSS	$\sum y_i^2$	$n-1$	

* SS 指平方和。

\+ 均方和，得自 SS 除以其自由度。

假定干扰项 u_i 是正态分布的且 H_0：$\beta_2=0$，就可证明（5.9.1）的 F 满足定
141 理 4.6 的条件，从而遵从自由度为 1 和 $n-2$ 的 F 分布。（见附录 5A，5A.3 节。）

上述 F 比有什么用处？可以证明[18]：

$$E(\hat{\beta}_2^2 \sum x_i^2) = \sigma^2 + \beta_2^2 \sum x_i^2 \tag{5.9.2}$$

和

$$E \frac{\sum \hat{u}_i^2}{n-2} = E(\hat{\sigma}^2) = \sigma^2 \tag{5.9.3}$$

（注意出现在这些方程右端的 β_2 和 σ^2 是真实的参数。）因此，若 β_2 确实是零，方程（5.9.2）和（5.9.3）两者都给出相同的真实 σ^2 的估计，这时解释变量 X 与 Y 没有任何线性影响，Y 的全部变异均由随机干扰项 u_i 来解释。而另一方面，若 β_2 不是零，（5.9.2）和（5.9.3）将有所不同，从而 Y 的部分变异将归因于 X。于是，（5.9.1）的 F 比提供了对虚拟假设 H_0：$\beta_2=0$ 的一个检验。由于此方程中的每一个量都可从已有的样本算得，这个 F 比就为检验虚拟假设真实 β_2 是零提供了一个检验统计量。我们所需做的无非是算出 F 比，再拿它同从 F 表在选定显著水平上读出的 F 临界值相比较，或者是查找所算 F 统计量的 p 值。

为了说明，继续用我们的消费—收入例子。此例的 ANOVA 表见于表 5.4。我们看到计算的 F 值为 202.87。这个 F 值对应于 1 和 8 个自由度的 p 值不能从附录 D 给出的 F 表读出。但利用电子统计表可以显示 p 值是 0.000 000 1，确实是一个极其小的概率。如果你决定选择假设检验的显著性水平法，并把 α 固定在 0.01 或 1% 的水平上，你便能看到所算的 202.87 这一 F 值在此水平上明显是显著的。所以，如果我们拒绝虚拟假设 $\beta_2=0$，
142 犯第Ⅰ类错误的概率就是非常之小的。从一切实际意义来考虑，我们的样本都不可能来自一个 β_2 为零的总体，从而我们能够以高度的可信性做出收入 X 对消费支出有影响的结论。

表 5.4　　消费—收入一例的 ANOVA 表

变异来源	SS	df	MSS	
由于回归（ESS）	8 552.73	1	8 552.73	$F=\frac{8\,552.73}{42.159}$
由于剩余（RSS）	337.27	8	42.159	$=202.87$
TSS	8 890.00	9		

回顾附录 5A.1 的定理 5.7，该定理说，自由度为 k 的 t 值的平方是一个分子自由度为 1，分母自由度为 k 的 F 值。对消费—收入一例来说，如果作假设 H_0：$\beta_2=0$，则由（5.3.2）容易验证，估计 t 值是 14.26，这个 t 值有 8 个自由度。在同样的虚拟假设下，F 值曾算出是 202.87 且有自由度 1 和 8。可见，$(14.24)^2=F$ 值，不计四舍五入进位误差。

由此可知，t 检验和 F 检验是检验虚拟假设 $\beta_2=0$ 的两个互为补充的备选方法。果然如此，又为什么不是仅仅用 t 检验就够了，还要麻烦到用 F 检验以及伴随的方差分析？对于双变量模型，确实不需要用 F 检验。但当我们考虑复（多元）回归问题时，我们将看到 F 的一些有趣的应用，使得它成为检验统计假设的非常有用和有效的方法。

§5.10　回归分析的应用：预测问题

根据表 3.2 的样本数据，我们曾得到如下的样本回归：

$$\hat{Y}_i=24.454\,5+0.509\,1X_i \tag{3.6.2}$$

其中 $\hat{Y}_i$ 是对应于给定 X 的真实 $E(Y_i)$ 的估计量。这一描述**历史的回归**（historical regression）能有什么用处？一个用途是"预测"或"预报"对应于某给定收入水平 X 的未来消费支出 Y。现有两种预测：（1）对应于选定的 X 比如说 X_0，预测 Y 的条件均值，也就是预测总体回归线本身上的点（见图 2.2），以及(2)预测对应于 X_0 的 Y 的一个个别值。我们将把这两种预测分别称为**均值预测**(mean prediction)和**个值预测**(individual prediction)。

均值预测[19]

为了明确概念，假定 $X_0=100$，我们要预测 $E(Y\mid X_0=100)$。可以表明，历史上的回归（3.6.2）给出这个均值预测的点估计如下：

$$\hat{Y}_0=\hat{\beta}_1+\hat{\beta}_2X_0$$

$$= 24.4545 + 0.5091(100)$$
$$= 75.3645 \tag{5.10.1}$$

143 其中 $\hat{Y}_0 = E(Y|X_0)$的估计量。可以证明，这个点预测量是一个最优线性无偏估计量。

$\hat{Y}_0$ 既是一个估计量，就可能不同于它的真值。两值之差将给出预测或预报误差的某种概念。为了评估这个误差，我们需要求出 $\hat{Y}_0$ 的抽样分布。附录 5A，5A.4 节表明了（5.10.1）中的 $\hat{Y}_0$ 是正态分布的，其均值为（$\beta_1 + \beta_2 X_0$）而方差由下式给出：

$$\operatorname{var}(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \tag{5.10.2}$$

将未知的 σ^2 代以它的无偏估计量 $\hat{\sigma}^2$，就可推知变量：

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\operatorname{se}(\hat{Y}_0)} \tag{5.10.3}$$

遵循 $n-2$ 个自由度的 t 分布。因而 t 分布可用来推导真实$E(Y_0|X_0)$的置信区间，并且用惯常的方式去检验关于它的假设，即：

$$\Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2}\operatorname{se}(\hat{Y}_0) \leqslant \beta_1 + \beta_2 X_0 \leqslant \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2}\operatorname{se}(\hat{Y}_0)] = 1 - \alpha \tag{5.10.4}$$

其中$\operatorname{se}(\hat{Y}_0)$由(5.10.2)求得。

对于我们的数据（见表 3.3），

$$\operatorname{var}(\hat{Y}_0) = 42.159\left[\frac{1}{10} + \frac{(100-170)^2}{33\,000}\right]$$
$$= 10.4759$$

和

$$\operatorname{se}(\hat{Y}_0) = 3.2366$$

因此，真实 $E(Y|X_0) = \beta_1 + \beta_2 X_0$ 的 95%置信区间由下式给出：

$$75.3645 - 2.306(3.2366) \leqslant E(Y_0|X = 100) \leqslant 75.3645 + 2.306(3.2366)$$

就是

$$67.9010 \leqslant E(Y|X = 100) \leqslant 82.8381 \tag{5.10.5}$$

144 这是说，给定 $X_0 = 100$，在重复抽样中，每 100 个类似于（5.10.5）的区间将有 95 个包含着真实的均值；真实均值的单一最优估计，当然是点估计值 75.3645。

如果我们对表 3.2 中的每一个 X 值求类似于（5.10.5）的 95%置信区间，把这些区间的端点连结起来，我们就得到如图 5.6 所展示的一个关于总体回归函数的所谓**置信带（域）**(confidence band)。

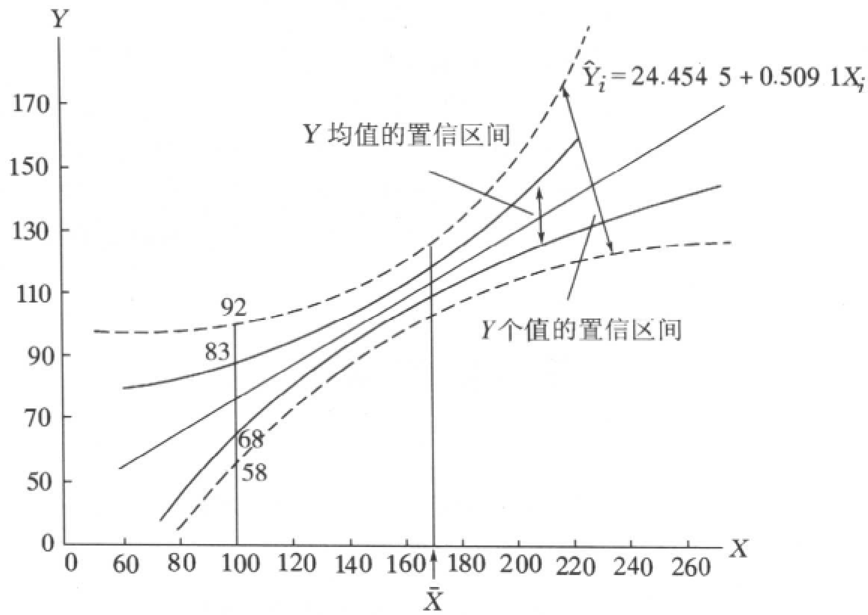


图 5.6　Y 均值与 Y 个值的置信区间（带）

个值预测

如果我们的兴趣在于预测对应于给定 X 值（比方说 X_0）的单个 Y 值（Y_0），那么，如附录 5，5A.3 节所表明，Y_0 的一个最优线性无偏估计量仍由（5.10.1）给出，但是它的方差：

$$\operatorname{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \tag{5.10.6}$$

可以进一步证明 Y_0 也遵循正态分布，其均值和方差分别由（5.10.1）和（5.10.6）给出。用 $\hat{\sigma}^2$ 代 σ^2，即推出：

145

$$t = \frac{Y_0 - \hat{Y}_0}{\operatorname{se}(Y_0 - \hat{Y}_0)}$$

也遵循 t 分布。因此 t 分布可用来对真实 Y_0 作推断。继续用我们的消费—收入例子，我们看到，Y_0 的点预测和 $\hat{Y}_0$ 的点预测一样，同是75.364 5。但它的方差是52.634 9（读者可验证这一计算）。由此可见，对应于 $X_0 = 100$ 的 Y_0 的 95%置信区间是：

$$(58.634\ 5 \leqslant Y_0 \mid X_0 = 100 \leqslant 92.094\ 5) \tag{5.10.7}$$

拿此区间同（5.10.5）相比，即看出 Y_0 的置信区间比 Y_0 的均值的置信区间要宽。（为什么?）以表 3.2 所给的诸 X 值为条件计算类似于（5.10.7）的诸多置信区间，连结起来，就得到对应于这些 X 值的单个 Y 值的 95%置信带。图 5.6 同时展示了对应于同样 X 的这一（估计 Y 个值——

译者注）置信带和 $\hat{Y}_0$（即估计 Y 均值——译者注）的置信带。

注意图 5.6 展示的置信带的一个重要特点。这些带的宽度当 $X_0=\bar{X}$ 时达到最小。（为什么?）随着 X_0 远离 $\bar{X}$，这个宽度急剧地变大。（为什么?）这种变化说明历史的样本回归线的预测能力随着 X_0 越来越远离 X 而显著下降。**因此，当 X_0 远离 $\bar{X}$ 时，人们凭借"外推"历史回归线来预测对应于给定 X_0 的$E(Y|X_0)$或 Y_0 时，必须保持高度警觉。**

§5.11 报告回归分析的结果

报告回归分析的结果有许多方式，本书将采用以下方式，仍以第 3 章的消费—收入例子为例：

$$
\begin{aligned}
\hat{Y}_i &= 24.454\,5 + 0.509\,1X_i \\
\text{se} &= (6.413\,8) \quad (0.035\,7) \qquad r^2=0.962\,1 \\
t &= (3.812\,8) \quad (14.260\,5) \qquad \text{df}=8 \\
p &= (0.002\,571) \quad (0.000\,000\,289) \qquad F_{1,8}=202.87
\end{aligned}
\tag{5.11.1}
$$

146 (5.11.1) 中第一组括号内的数字代表回归系数（指估计的系数——译者注）的估计的标准误（指这个标准误的估计值——译者注），第二组的数字代表在每个回归系数的真实总体值都是零的虚拟假设下，由（5.3.2）计算出来的 t 估计值（例如 3.812 8=24.454 5÷6.413 8），而第三组数字代表估计的 p 值。比如，当自由度为 8 时，得到一个等于3.812 8或更大的 t 值的概率是0.002 6；得到一个等于或大于14.260 5的 t 值的概率约合0.000 000 3。

把这些估计的 t 系数的p 值显示出来，我们就能马上看到每一个 t 估计值的精确显著性水平。例如，在真实总体截距值为零的虚拟假设下，得到一个大到3.812 8或更大的 t 值的精确概率（即 p 值）仅约为0.002 6。因此，如果我们拒绝这个虚拟假设，我们犯第 1 类错误的概率仅约合 1 万次中有 26 次，确实是一个很小的概率。从一切实际目的考虑，我们都能说真实总体截距不是零。同理，所估斜率系数的 p 值从一切实际方面看，都可等同于零。如果真实 MPC 确实是零的话，我们得到一个0.509 1的 MPC 的机会，实际上将是零。从而我们能拒绝真实 MPC 是零的虚拟假设。

我们在前面曾指明 F 和 t 两统计量有密切关系，即 $F_{1,k}=t_k^2$。在真实 $\beta_2=0$ 的虚拟假设下（5.11.1）表明 F 值为 202.87（对于 1 个分子自由度和 8 个分母自由度），而 t 值约为 14.24（8 个自由度）。如所预料，前一数值正是后一数值的平方，且不计其进位误差。这个问题的 ANOVA 表前面已讨论过。

§5.12 评价回归分析的结果

在引言的表 I.4 中，我们对计量经济模型的建立作过一个简要的剖析。现在既已在（5.11.1）中给出了我们的消费—收入一例的回归分析结果，我们不免要提问，这个拟合的模型的适宜性如何。这个拟合的模型有多“好”？为了回答这个问题需要有一些准则。

第一，所估系数的符号是否与理论或事前预期相一致？先验地说，消费函数中的边际消费倾向 β_2 应是正的。在本例中，确是如此。第二，如果理论上认为这个关系式不仅是正的，而且是统计上显著的，那么在本例中是这样的吗？如第 5.11 节中所讨论的，MPC 不仅是正的，而且统计上显著地异于零，t 估计值的 p 值极小，同样的评语还适用于截距系数。第三，回归模型在多大的程度上解释了消费支出的变异？可以用 r^2 来回答此问题。本例中 r^2 约为 0.96，考虑到 r^2 最多只能大到 1，这就是一个很高的值了。

如此看来，为了解释消费支出行为，我们选用的模型算是够好的了。但在我们结束讨论之前，还想来看看我们的模型是否满足 CNLRM 的假定。因
147 为模型简单得如此明显，我们现在不去审查这许多的假定。但有一个假定是我们想要检查的，就是关于干扰项 u_i 的正态性。回想一下，前面所用的 t 和 F 检验都要求误差项遵循正态分布。否则，在小样本或有限样本中，检验的程序将是无效的。

正态性检验

虽然文献中有多种正态性检验，我们只想讨论三种：（1）残差直方图；（2）正态概率图（NPP）；（3）**雅克-贝拉检验**（Jarque-Bera test）。

残差直方图（Histogram of Residuals）。残差直方图是用于了解随机变量的 PDF 形状的一个简单图示。在水平轴上，我们将所关注变量的值（比如 OLS 残差）分成适当的区间，在每个区间里，我们做垂直的矩形，并让矩形的高等于落在该区间内的观测次数（即频数）。如果你从心里在直方图上估画一条钟形的正态分布曲线，那你就对正态（PDF）近似是否适当有些认识。在第 5.13 节给出了一个简明的例子（见图 5.8）。用画残差直方图作为检验正态性假定的一种粗略和灵便的方法总是一种好的做法。

正态概率图（Normal Probability Plot，NPP）。要研究随机变量的概率密度函数的形状，一个相对简单的图示法就是，利用正态概率纸（一种专门设计的坐标纸）做出**正态概率图**。在水平轴或 x 轴上，我们描上所关注变量的值（如 OLS 残差 $\hat{u}_i$），而在纵轴或 y 轴上，我们标出这个变量遵循正态分布时的期望值。因此，如果这个变量实际上来自正态总体，那么 NPP 将近

似为一条直线。从消费—收入回归得到的残差的 NPP 示于图 5.7，它是从 MINITAB 软件包（13 版）得到的。如前面所讲，如果 NPP 的拟合线近似为一条直线，那人们就可以断定这个变量是正态分布的。在图 5.7 中，由于一条直线相当好地拟合了这些数据，所以我们说明性例子中的残差是渐近正态分布的。

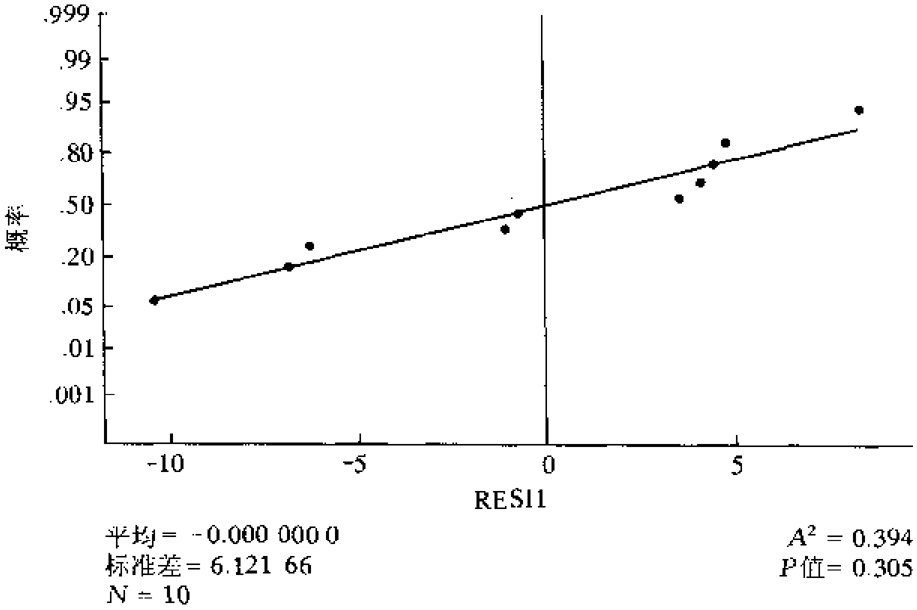


图 5.7　来自消费—收入一例的残差分布，零以上和零以下的标准差(σ)的倍数

MINITAB 还给出**安德森-达林正态性检验**（Anderson-Darling normality test），称 **A^2 统计量**（A^2 Statistic）。基本的虚拟假设是，所考虑的变量是正态分布的。如图 5.7 所示，就我们的例子而言，计算出来的 A^2 统计量为 0.394。得到这样一个 A^2 值的 p 值为 0.305，这是相当高的。因此，我们不
148 能拒绝此例中残差为正态分布的假设。顺便提一句，图 5.7 显示了此（正态）分布的两个参数：均值近似为 0，标准差约为 6.12。

正态性的雅克-贝拉（JB）检验。[20]正态性的 JB 检验是一项渐近或大样本检验。它仍以 OLS 残差为依据。此检验先计算 OLS 残差的**偏态**（skewness）和**峰态**（kurtosis）（附录 A 中有所描述），再使用下列检验统计量：

$$\mathrm{JB}=n\left[\frac{S^2}{6}+\frac{(K-3)^2}{24}\right] \tag{5.12.1}$$

其中 S 代表**偏态**，K 代表**峰态**。

因为对一个正态分布来说偏态取值零而峰态取值 3，故（5.12.1）中的（$K-3$）代表超额峰态（excess kurtosis）。在残差为正态分布的虚拟假设下，雅克和贝拉证明了（5.12.1）所给的JB统计量渐近地（即在大样本中）遵循自由度为2的 χ^2 分布。如果在一项应用中算出来的 χ^2 统计量的 p 值充分地低，就可拒绝残差为正态分布的假设。但如果 p 值合理地高，就不要拒绝正态性假定。

回到我们的消费—收入一例，我们发现（使用 **SHAZAM**，**TSP** 或 **ET** 软
149 件包）JB 值为0.776 9。如果样本合理地大，得到 2 个自由度的这一 χ^2 值的 p 值约为 0.678 1，这是一个相当大的概率。所以，在渐近的意义上，我们不拒绝正态性假定。

在消费—收入例子中，样本容量相当小。因此，严格地讲，不应该用 JB 统计量。如果我们在此例中机械地应用 JB 公式，我们就会看到最终的 JB 统计量为0.776 9。从自由度为 2 的 χ^2 分布看，得到这样一个值相应的 p 值约为 0.68，这个 p 值相当高。换句话说，我们恐怕不能拒绝此例子中的正态性假定。当然，要警惕样本容量的问题。

模型适宜性的其他检验

CNLRM 除误差项的正态性外还做了许多假定。随着我们进一步分析我们的计量经济理论，我们将考虑模型适宜性的若干其他检验。在这之前，请记住我们的回归模型是建立在一些不一定总是正确的简化的假定上的。

一个结论性的例子

让我们回到关于印度食物支出的例 3.2。利用（3.7.2）中给出的数据并采用（5.11.1）的格式，我们得到如下方程

$$
\begin{aligned}
&\text{食物支出}\ i = 94.208\ 7 + 0.436\ 8\ \text{总支出}\ i \\
&\quad \text{se} = (50.856\ 3) \qquad (0.078\ 3) \\
&\quad t = (1.852\ 4) \qquad (5.577\ 0) \\
&\quad p = (0.069\ 5) \qquad (0.000\ 0)^{*} \qquad\qquad (5.12.2)\\
&\quad r^2 = 0.369\ 8 \qquad \text{df} = 53 \\
&\quad F_{1,53} = 31.103\ 4 \qquad (p\ \text{值} = 0.000\ 0)^{*}
\end{aligned}
$$

其中 * 表示极小。

首先，让我们解释这个回归，食物支出与总支出之间存在着预期的正相关。如果总支出增加 1 卢比，食物支出平均增加约 44 派沙。如果总支出为零，食物支出约为 94 卢比。当然，对截距的这种机械解释可能没有多大经济意义。约为 0.37 的 r^2 值意味着，食物支出的变异中有 37% 可由收入的代理变量总支出来解释。

假设我们想检验食物支出与总支出之间没有关系的虚拟假设，即真正的斜率系数 $\beta_2 = 0$。β_2 的估计值为0.436 8。如果虚拟假设正确，那么得到 0.436 8这样一个值的概率是多大？在这个虚拟假设下，我们从（5.12.2）中观测到，t 值为5.577 0，得到这样一个 t 值的 p 值实际上为零。换句话说，我们完全可以拒绝这个虚拟假设。但假若虚拟假设是 $\beta_2 = 0.5$ 又怎样呢？利用 t 检验，我们得到

$$t = \frac{0.436\ 8 - 0.5}{0.078\ 3} = -0.807\ 1$$

得到 $|t|=0.8071$ 的概率大于 20%。因此我们不能拒绝真正的 β_2 为 0.5 的假设。

注意，在真正的斜率系数为零的虚拟假设下，F 值如（5.12.2）所示为 31.103 4。在同样的假设下，我们得到一个为5.577 0的 t 值。将它平方则得到31.102 9，约等于 F 值，这又再次表明了 t 统计量与 F 统计量之间的密切关系。（注：F 统计量的分子 df 必须为 1，这里正是如此。）

利用回归的估计残差，我们对误差项的概率分布有何见解呢？图 5.8 给出了这方面的信息。如图所示，食物支出回归的残差看来是对称分布的。应用雅克-贝拉检验表明，JB 统计量约为0.257 6，在正态性假定下，得到这样一个统计量的概率约为 88%。因此我们不能拒绝误差项正态分布的假设。但要记住，55 个观测的样本容量可能不算太大。

至于构造回归系数的置信区间并得出正态概率图，以及做均值和个值预测，则留给读者自己完成。

150

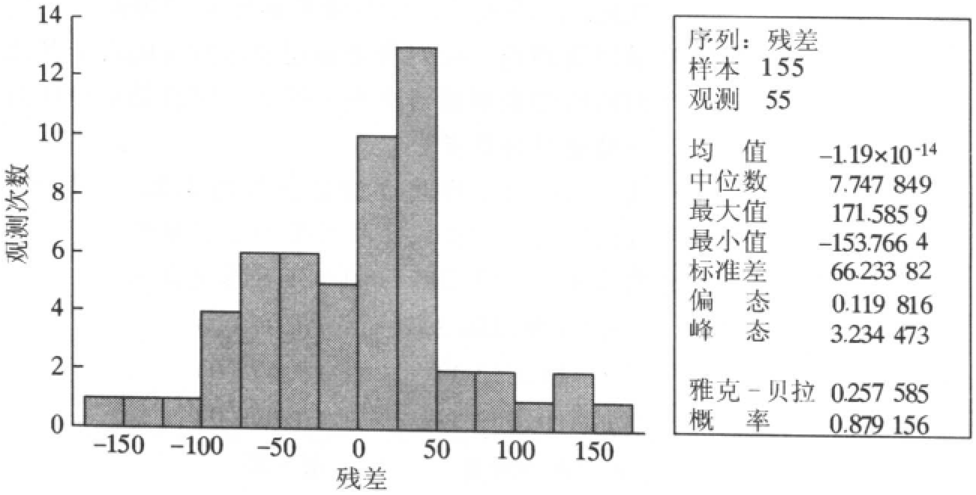


图 5.8　食物支出回归的残差

§5.13　要点与结论

1. 估计与假设检验是经典统计学的两个主要分支。在第 3 章和第 4 章讨论了估计问题之后，本章讨论假设检验的问题。

2. 假设检验要回答这样的问题：一个给定的发现是否与声称的假设相符？

3. 为回答上述问题，有两个互为补充的方法：**置信区间与显著性检验**。

4. 置信区间法建立在**区间估计**的概念上。一个区间估计量是指一个区间或变化域的构造，要使得它把未知参数的真值包含在其界限内有预定的概

率。如此构造的区间称为**置信区间**，这个区间常用百分数比如说 90% 或 95% 的形式来表述。置信区间对未知参数的取值提供了一个可信假设集。如果虚拟假设值落入置信区间，就不拒绝假设；如果它落在此区间之外，就可拒绝虚拟假设。

5. 在**显著性检验**程序中，我们找出一个**检验统计量**（test statistic），并研究它在虚拟假设下的抽样分布。通常这个检验统计量都遵从一个有明确定义的概率分布，诸如正态、t、F 或 χ^2 方分布。一旦从现有的数据算出某个
151 检验统计量（如 t 统计量），就容易求出它的 p 值。这个 p 值给出在虚拟假设下得到所估算的检验统计量的精确概率。如果这个 p 值小，就可拒绝虚拟假设；但如果它大，就不可拒绝。什么才算小的或大的 p 值，应由研究者来决定。在选择 p 值时，研究者要切实考虑犯**第Ⅰ类和第Ⅱ类错误**的概率。

6. 在实践中，**第Ⅰ类错误**概率 α 被选定在 1%、5% 或 10% 这些人为数值上的问题，要仔细考虑。较好的做法是引用检验统计量的 ***p* 值**。而且，不要把统计显著性和实际显著性混同起来。

7. 当然，假设检验事先就认定，选用来做经验分析的模型，不违反经典正态线性回归模型中的任一个或多个假定，所以是适宜的。因此，应把模型适宜性的检验放在假设检验之前来做。本章介绍了适宜性的一个检验，就是**正态性检验**，借以发现误差项是否遵从正态分布。因为在小的或有限的样本中，t、F 和 χ^2 检验都需要有正态性假定，所以对此假定正式地加以核实是重要的。

8. 如果模型被认为实用上适宜，就可用于预报。但在预报回归子的未来值时，切勿超出回归元的样本取值范围太远。否则预测误差会戏剧性地增大。

习 题

问答题

5.1 下面的陈述是正确的，错误的或者不能肯定？试着精确地阐述理由。

a. 本章所讨论的显著性 t 检验要求估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的抽样分布是正态分布。

b. 即使 CLRM 中的干扰项不是正态分布的，OLS 估计量仍然是无偏的。

c. 如果回归模型中没有截距项，u_i 的估计值（$=\hat{u}_i$）的总和将不为零。

d. p 值和检验统计量的尺度指的是一回事。

e. 在一个含有截距的回归模型中，残差的总和必定为零。

f. 如果一个虚拟假设不被拒绝，它就是真实的。

g. σ^2 的值越大，(3.3.1) 所给的 $\hat{\beta}_2$ 的方差也越大。

h. 一个随机变量的条件均值和无条件均值是一样的。

i. 在双变量 PRF 中，如果斜率系数 β_2 是零，则截距 β_1 由样本均值 $\bar{Y}$ 来估计。

j. 如果 X 对 Y 无影响，条件方差 $\operatorname{var}(Y_i \mid X_i)=\sigma^2$ 和 Y 的无条件方差 $\operatorname{var}(Y)=\sigma_Y^2$ 将是一样的。

152 **5.2** 对 (3.7.2) 所给的回归模型建立像表 5.4 那样的 ANOVA 表，并检验假设：印度的食物支出和总支出之间没有关系。

5.3 从表 2.6 所给出的关于收入与教育水平的数据，我们得到以下回归 [见方程 (3.7.3)]：

$$\text{平均收入}_i = 0.7437 + 0.6416\ \text{教育水平}_i$$
$$\text{se} = (0.8355) \qquad (\qquad)$$
$$t = (\qquad) \qquad (9.6536) \qquad r^2 = 0.8944 \quad n = 13$$

a. 将缺数填入 (　　) 中。

b. 怎样解释系数 0.641 6?

c. 你会不会拒绝真实斜率系数为零的假设？你使用哪一种检验？为什么？你的检验统计量的 p 值是多少？

d. 建立本例的 ANOVA 表，并检验真实斜率系数为零的假设。你使用哪一种检验？为什么？

e. 假如在刚才给定的回归中没有给出 r^2 值。你能否从回归的其他（数值）结果中得到它？

5.4 令 ρ^2 代表真实的总体判定系数，假定你想检验假设：$\rho^2=0$，用文字说明你会怎样检验此假设。提示：利用等式 (3.5.11)，还可参看习题 5.7。

5.5 现代投资分析中所谓的**特征线** (characteristic line) 就是得自以下模型的回归线：

$$r_{it} = \alpha_i + \beta_i r_{mt} + u_t$$

其中　r_{it} = 第 i 种证券在时间 t 的回报率

r_{mt} = 市场组合证券在时间 t 的回报率

u_t = 随机干扰项

在此模型中，β_i 被称为第 i 种证券的 **β 系数** (beta coefficient)，是对证券的市场（或系统）风险的一种度量。[1]

福格勒 (Fogler) 和加纳帕赛 (Ganapathy)，根据 1956—1976 年期间的 240 个月回报率，算得 IBM 股票相对于芝加哥大学编制的市场组合证券指数的特征线如下[2]：

$$r_{it} = 0.7264 + 1.0598 r_{mt} \qquad r^2 = 0.4710$$
$$\text{se} = (0.3001)\ (0.0728) \qquad \text{df} = 238$$
$$F_{1,238} = 211.896$$

a. β系数大于 1 的证券称**易变**（volatile）或**进取**（aggressive）证券。问在此研究期间 IBM 是易变证券吗？

153 b. 问截距系数是否显著地异于零？如果是，它的实际意义何在？

5.6 方程（5.3.5）还可写为：

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\text{se}(\hat{\beta}_2) < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}\text{se}(\hat{\beta}_2)] = 1 - \alpha$$

就是说，弱不等式（≤）可代之以强不等式（<）。为什么？

5.7 费希尔曾导出由（3.5.13）定义的相关系数的抽样分布。如果假定变量 X 和 Y 是联合正态分布的，即如果它们来自双变量正态分布（见附录 4A，习题 4.1），则在总体相关系数 ρ 为零的假定下，可以证明 $t = r\sqrt{n-2}/\sqrt{1-r^2}$ 遵循自由度为 $n-2$ 的学生氏 t 分布。[3]试说明这个 t 值等同于在虚拟假设 $\beta_2 = 0$ 下由（5.3.2）给出的 t 值，从而确立在相同的虚拟假设下 $F = t^2$。（见第 5.9 节。）

解答题

5.8 考虑如下回归结果[4]：

$$\hat{Y}_i = 0.2033 + 0.6560 X_t$$
$$\text{se} = (0.0976)\ (0.1961)$$
$$r^2 = 0.397 \quad \text{RSS} = 0.0544 \quad \text{ESS} = 0.0358$$

其中 Y = 1972 年的妇女劳动力参与率（labor force participation rate，LFPR），X = 1968 年的妇女 LFPR。回归结果从美国 19 个城市构成的一个样本得到。

a. 你如何解释这个回归？

b. 相对 H_1：$\beta_2 > 1$ 检验假设 H_0：$\beta_2 = 1$。你用哪一个检验？为什么？你用的检验所依据的假定是什么？

c. 假设 1968 年的 LFPR 为 0.58（或 58%）。基于上述回归结果，1972 年 LFPR 的均值是多少，构建均值预测的一个 95% 置信区间。

d. 你如何检验总体回归中的误差项为正态分布这一假设？给出必要的计算。

5.9 表 5.5 给出了 1985 年 50 个州和哥伦比亚特区公共教师的平均工资（以美元计年薪）和对公立学校每个学生的支出（美元）方面
154 的数据。为了探明公立学校中教师工资与对每个学生的支出之间是否存在某种关系，有人提出如下模型：$\text{Pay}_i = \beta_1 + \beta_2 \text{Spend}_i + u_i$，其中 Pay 表示教师工资，Spend 表示对每个学生的支出。

表 5.5　　1985 年的教师工资与学生支出（美元）

观测	年薪	支出	观测	年薪	支出
1	19 583	3 346	27	22 795	3 366
2	20 263	3 114	28	21 570	2 920
3	20 325	3 554	29	22 080	2 980
4	26 800	4 642	30	22 250	3 731
5	29 470	4 669	31	20 940	2 853
6	26 610	4 888	32	21 800	2 533
7	30 678	5 710	33	22 934	2 729
8	27 170	5 536	34	18 443	2 305
9	25 853	4 168	35	19 538	2 642
10	24 500	3 547	36	20 460	3 124
11	24 274	3 159	37	21 419	2 752
12	27 170	3 621	38	25 160	3 429
13	30 168	3 782	39	22 482	3 947
14	26 525	4 247	40	20 969	2 509
15	27 360	3 982	41	27 224	5 440
16	21 690	3 568	42	25 892	4 042
17	21 974	3 155	43	22 644	3 402
18	20 816	3 059	44	24 640	2 829
19	18 095	2 967	45	22 341	2 297
20	20 939	3 285	46	25 610	2 932
21	22 644	3 914	47	26 015	3 705
22	24 624	4 517	48	25 788	4 123
23	27 186	4 349	49	29 132	3 608
24	33 990	5 020	50	41 480	8 349
25	23 382	3 594	51	25 845	3 766
26	20 627	2 821			

资料来源：National Education Association, as reported by *Albuquerque Tribune*, Nov.7, 1986.

a. 描出这些数据点并目测一条回归线。

b. 假设你想根据 a 估计上述回归模型。求参数估计值及其标准误、r^2、RSS 和 ESS。

c. 解释这个回归。它有经济意义吗?

d. 构造 β_2 的一个 95% 置信区间。你会拒绝真实的斜率系数为 3.0 的假设吗?

e. 若对每个学生的支出为 5 000 美元，求 Pay 的均值和个值预测值。同样也分别构造它们的 95% 置信区间。

f. 你如何检验误差项的正态性假定? 说明你所用的检验。

5.10 参照习题 3.20，建立 ANOVA 表，以检验假设：生产率与真实
155 工资报酬之间没有关系。对商业部门和非农商业部门分别做这个检验。

5.11 参照习题1.7。

a. 以印象为纵轴和广告支出为横轴描点。你观察到哪种关系?

b. 对数据拟合一个双变量线性回归模型合适吗? 为什么? 若不合适，你将用哪种类型的回归模型来拟合数据? 我们有拟合这样一个模型的必要工具吗?

c. 假设你不描点而简单地对数据拟合一个双变量回归模型。给出通常的回归结果。留存结果等以后再看这个问题。

5.12 参照习题1.1。

a. 将美国的消费者价格指数 CPI 相对加拿大的 CPI 描图，这个图说明了什么?

b. 假设你想基于加拿大的 CPI 来预测美国的 CPI，给出一个适当的模型。

c. 检验这两个 CPI 之间没有关系的假设 ($\alpha = 5\%$)。如果拒绝了虚拟假设，是否意味着加拿大的 CPI "导致了" 美国的 CPI? 为什么?

5.13 参照习题 3.22。

a. 估计那里的两个回归，以获得通常的输出 (结果)，如标准误等等。

b. 检验两个回归模型的干扰都是正态分布的假设。

c. 在黄金价格回归中，检验假设 $\beta_2 = 1$，就是黄金价格和 CPI 之间有 1 比 1 的关系 (也就是，黄金是一种完美的保值工具)。所估计的检验统计量的 p 值为何?

d. 再对纽约股票交易所指数回归按 (c) 题要求做一遍。投资于股票市场是防范通货膨胀的完美保值手段吗? 你检验的是什么虚拟假设? 它的 p 值为何?

e. 在黄金与股票之间，你会选择哪一种投资? 你的决策依据是什么?

5.14 表 5.6 给出 1970—1983 年美国的 GNP 和 4 种定义的货币存量。

表 5.6 **GNP 和货币存量的四种度量**

年份	GNP，10 亿美元	货币存量度量，10 亿美元计			
		M1	M2	M3	L
1970	992.70	216.6	628.2	677.5	816.3
1971	1 077.6	230.8	712.8	776.2	903.1
1972	1 185.9	252.0	805.2	886.0	1 023.0
1973	1 326.4	265.9	861.0	985.0	1 141.7
1974	1 434.2	277.6	908.5	1 070.5	1 249.3
1975	1 549.2	291.2	1 023.3	1 174.2	1 367.9
1976	1 718.0	310.4	1 163.6	1 311.9	1 516.6
1977	1 918.3	335.4	1 286.7	1 472.9	1 704.7
1978	2 163.9	363.1	1 389.1	1 647.1	1 910.6
1979	2 417.8	389.1	1 498.5	1 804.8	2 117.1
1980	2 631.7	414.9	1 632.6	1 990.0	2 326.2
1981	2 957.8	441.9	1 796.6	2 238.2	2 599.8
1982	3 069.3	480.5	1 965.4	2 462.5	2 870.8
1983	3 304.8	525.4	2 196.3	2 710.4	3 183.1

定义：M1 = 现金 + 活期存款 + 旅行支票 + 其他支票存款（OCD）

M2 = M1 + 隔日 RP 及欧元 + 货币市场共同基金（MMMF）结余 + 货币市场存款账户 + 储蓄及小额存款

M3 = M2 + 大额定期存款 + 定期 RP + 机构 MMMF

L = M3 + 其他流动资产

资料来源：*Economic Report of the President*，1985，GNP 数据取自表 B-1，第 232 页；货币存量数据取自表 B-61，第 303 页。

将 GNP 对各种定义的货币作回归，并将所得结果列入表 5.7：

表 5.7 **GNP—货币存量回归，1970—1983**

1）$\text{GNP}_t = -787.472\,3 + 8.086\,3M_{1t}$ $\qquad r^2 = 0.991\,2$

$\qquad(77.966\,4)\ (0.219\,7)$

2）$\text{GNP}_t = -44.062\,6 + 1.587\,5M_{2t}$ $\qquad r^2 = 0.990\,5$

$\qquad(61.013\,4)\ (0.044\,8)$

3) $GNP_t = 159.1366 + 1.2034 M_{3t}$ $\quad r^2 = 0.9943$

$(42.9882)\ (0.0262)$

4) $GNP_t = 164.2071 + 1.0290 L_t$ $\quad r^2 = 0.9938$

$(44.7658)\ (0.0234)$

注：括号中的数字是估计的标准误（差）。

货币主义者或货币数量理论家声称，名义收入（即名义GNP）主要由货币存量的数量变化决定，虽然什么是货币的“合适”定义，尚无一致意见。给定上表中的结果，试考虑如下问题：

a. 哪一种货币定义似与名义 GNP 有密切关系？

b. 既然 r^2 项都一致地较高，这是否意味着无论怎样选择货币定义都关系不大？

c. 如果联邦储备银行想控制货币供给，那么为此目的，这些货币度量中的哪一种可作为较好的目标？你能从这些回归结果看出来吗？

5.15 假使两种物品的**无差异曲线**（indifference curve）方程是：

$$X_i Y_i = \beta_1 + \beta_2 X_i$$

你会怎样估计此模型的参数？将此模型应用于表 5.8 中的数据并评述你所得的结果：

表 5.8

消费品 X：	1	2	3	4	5
消费品 Y：	4	3.5	2.8	1.9	0.8

5.16 自从 1986 年以来《经济学人》（*Economist*）出版 Big Mac 指数，作为一种粗糙而喧嚷一时的指标，用以评价国际货币是否定在符合**购买力平价**（Purchasing Power Parity，PPP）学说的“正确”汇率上。PPP 意为，一单位货币能在所有国家买到同样一包商品（the same bundle of goods）。PPP 的支持者声称，从长远看，各国货币都朝向它们的 PPP 移动。《经济学人》把麦当劳餐厅的巨无霸（Big Mac）当作代表性商品包并给出以下的信息。

157

表 5.9 **汉堡包标准**

	巨无霸价格		美元的隐含PPP[*]	一美元的实际汇率4/17/2001	当地货币过低(－)/过高(＋)定价,%
	以当地货币计	以美元计			
美国	2.54 美元	2.54	—	—	—
阿根廷	2.50 比索	2.50	0.98	1.00	－2
澳大利亚	3.00 澳大利亚元	1.52	1.18	1.98	－40
巴西	3.60 瑞亚尔	1.64	1.42	2.19	35
英国	1.99 英镑	2.85	1.28[‡]	1.43[‡]	12
加拿大	3.33 加拿大元	2.14	1.31	1.56	16
智利	1 260 智利比索	2.10	496	601	－17
中国大陆	9.90 元	1.20	3.90	8.28	－53
捷克	56.00 克朗	1.43	22.0	39.0	44
丹麦	24.75 丹麦克朗	2.93	9.74	8.46	15
欧元区	2.57 欧元	2.27	0.99[§]	0.88[§]	－11
法国	18.5 法郎	2.49	7.28	7.44	－2
德国	5.10 德罗马克	2.30	2.01	2.22	－9
意大利	4 300 里拉	1.96	1693	2195	－23
西班牙	395 比塞塔	2.09	156	189	18
中国香港	10.70 港币	1.37	4.21	7.80	－46
匈牙利	399 福林	1.32	157	303	－48
印度尼西亚	14 700 盾	1.35	5787	10855	－47
日本	294 日元	2.38	116	124	－6
马亚西亚	4.52 林吉特	1.19	1.78	3.80	－53
墨西哥	21.9 墨西哥比索	2.36	8.62	9.29	－7
新西兰	3.60 新西兰元	1.46	1.42	2.47	43
菲律宾	59.00 菲律宾比索	1.17	23.2	50.3	－54
波兰	5.90 兹罗提	1.46	2.32	4.03	42
俄罗斯	35.00 卢布	1.21	13.8	28.9	－52
新加坡	3.30 新加坡元	1.82	1.30	1.81	－28
南非	9.70 南非兰特	1.19	3.82	8.13	－53
韩国	3 000 圆	2.27	1 181	1 325	－11
瑞典	24.0 瑞典克朗	2.33	9.45	10.28	－8
瑞士	6.30 瑞士法郎	3.65	2.48	1.73	44
中国台湾	70.0 新台币	2.13	27.6	32.9	－16
泰国	55.0 泰铢	1.21	21.7	45.5	－52

* PPP：当地价除以美国价。

+ 取纽约、芝加哥、三藩市和亚特兰大的平均。

‡ 每英镑合美元数。

§ 每欧元合美元数。

资料来源：McDonald's；*The Economist*，April 21，2001.

考虑如下回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

158 其中 Y = 实际汇率而 X = 美元的隐含 PPP。

a. 若 PPP 成立，你会先验地预期 β_1 和 β_2 取什么值？

b. 回归的结果是否支持你的预期？你用什么形式的检验去检验你的假设？

c.《经济学人》是否应继续出版 Big Mac 指数？为什么？或为什么不？

5.17 参照习题 2.16 中所给的 S.A.T. 数据。假使你想根据女生的数学得分，通过做以下回归的方法去预测男生的数学得分：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

a. 估计上述模型。

b. 从所估计的残差看，正态性假定是否可以维系？

c. 检验假设：$\beta_2 = 1$，即假设：男生和女生的数学得分有一个 1 比 1 的对应关系。

d. 建立此问题的 ANOVA 表。

5.18 把前面的解答题重做一遍，但这一回令 Y 和 X 分别代表男生和女生的语文得分。

5.19 表5.10给出了美国1960—1999 年的消费者价格指数（CPI）和批发价格指数（WPI），或称为生产者价格指数（PPI）。

a. 以 WPI 为横轴、CPI 为纵轴描点。据经验，你预期这两个指数之间有何种关系？为什么？

159 b. 假设你想基于一个指数预测另一个指数。你会用哪个指数为回归元，哪个指数为回归子？为什么？

c. 做 b 部分你确定的回归。给出标准的结果，检验这两个指数有 1:1 变化关系的假设。

d. 从 c 部分回归所得到的残差，你能接受真正的误差项为正态分布的假设吗？说明你所用的检验。

表 5.10　　美国 1960—1999 年的 CPI 和 WPI

年份	CPI	WPI	年份	CPI	WPI
1960	29.8	31.7	1980	86.3	93.8
1961	30.0	31.6	1981	94.0	98.8
1962	30.4	31.6	1982	97.6	100.5
1963	30.9	31.6	1983	101.3	102.3
1964	31.2	31.7	1984	105.3	103.5
1965	31.8	32.8	1985	109.3	103.6
1966	32.9	33.3	1986	110.5	99.70
1967	33.9	33.7	1987	115.4	104.2
1968	35.5	34.6	1988	120.5	109.0
1969	37.7	36.3	1989	126.1	113.0
1970	39.8	37.1	1990	133.8	118.7
1971	41.1	38.6	1991	137.9	115.9
1972	42.5	41.1	1992	141.9	117.6
1973	46.2	47.4	1993	145.8	118.6
1974	51.9	57.3	1994	149.7	121.9
1975	55.5	59.7	1995	153.5	125.7
1976	58.2	62.5	1996	158.6	128.8
1977	62.1	66.2	1997	161.3	126.7
1978	67.7	72.7	1998	163.9	122.7
1979	76.7	83.4	1999	168.3	128.0

资料来源：*Economic Report of the President*，2000，pp.373 and 379.

【习题注释】

[1] 参看 Haim Levy and Marshall Sarnat，*Portfolio and Investment Seleltion：Theory and Practice*，Prentice-Hall International，Englewood Cliffs，N.J.，1984，Chap.12。

[2] 见 H.Russell Fogler and Sundaram Ganapathy，*Financial Econometrics*，Prentice Hall，Englewood Cliffs，N.J.，1982，p.13。

[3] 如 ρ 确实是零，费希尔曾证明，只要 X 或 Y 遵循正态分布，r 便遵循同样的 t 分布。但若 ρ 不等于零，则两个变量必须都是正态分布的。参看 R.L.Anderson and T.A.Bancroft，*Statistical Theory in Research*，McGraw-Hill，New York，1952，pp.87－88。

[4] 节选自 Samprit Chatterjee, Ali S.Hadi, and Bertram Price, *Regression Analysis by Example*, 3d ed., Wiley Interscience, New York, 2000, pp.46–47。

附录 5A

5A.1 与正态分布有关的概率分布

在附录 A 中讨论了 t、χ^2 和 F 概率分布，它们与正态分布都有内在的联系。由于我们在以后章节要大量使用这些概率分布，所以我在下面的一些定理中归纳出它们与正态分布之间的关系；至于定理的证明，则超出本书的范围，可以在参考书目中查找。[1]

定理 5.1 如果 Z_1, Z_2, …, Z_n 都是服从 $Z_i \sim N(\mu_i, \sigma_i^2)$ 的独立正态分布变量，那么它们的线性和 $Z = \sum k_i Z_i$ 也服从均值为 $\sum k_i\mu_i$ 和方差为 $\sum k_i^2\sigma_i^2$ 的正态分布，即 $Z \sim N(\sum k_i\mu_i, \sum k_i^2\sigma_i^2)$。其中 k_i 是不全为零的常数。注：μ 表示均值。

简言之，正态变量的线性组合本身还是正态分布的。比如 Z_1 和 Z_2 是正态独立分布变量：$Z_1 \sim N(10,2)$, $Z_2 \sim N(8,1.5)$，那么线性组合 $Z = 0.8Z_1 + 0.2Z_2$ 也是正态分布的，均值为 $0.8(10) + 0.2(8) = 9.6$，方差为 $0.64(2) + 0.04(1.5) = 1.34$，即 $Z \sim (9.6, 1.34)$。

定理 5.2 若 Z_1, Z_2, …, Z_n 都为正态分布但不独立，则和 $Z = \sum k_i Z_i$ 也服从均值为 $\sum k_i\mu_i$ 和方差为 $[\sum k_i^2\sigma_i^2 + 2\sum k_i k_j \text{cov}(Z_i, Z_j), i \neq j]$ 的正态分布，其中 k_i 是不全为零的常数。

因此，若 $Z_1 \sim N(6,2)$, $Z_2 \sim N(7,3)$ 和 $\text{cov}(Z_1, Z_2) = 0.8$，则线性组合 $0.6Z_1 + 0.4Z_2$ 也是正态分布的，均值为 $0.6(6) + 0.4(7) = 6.4$、方差为 $[0.36(2) + 0.16(3) + 2(0.6)(0.4)(0.8)] = 1.584$。

160 **定理 5.3** 若 Z_1, Z_2, …, Z_n 都服从 $Z_i \sim N(0, 1)$ 的独立正态分布，即标准正态分布，则 $\sum Z_i^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ 服从自由度为 n 的 χ^2 分布。用符号表示即 $\sum Z_i^2 \sim \chi_n^2$，其中 n 表示自由度，df。

简言之，"独立标准正态变量的平方和服从自由度等于正态变量个数的 χ^2 分布。"[2]

定理 5.4 若 Z_1, Z_2, …, Z_n 为服从自由度 $\text{df} = k_i$ 的 χ^2 分布的独立分布随机变量，则 $\sum Z_i = Z_1 + Z_2 + \cdots + Z_n$ 也服从自由度为 $\sum k_i$ 的 χ^2 分布。

于是，若 Z_1 和 Z_2 为自由度分别是 k_1 和 k_2 的独立 χ^2 变量，则 $Z =$

Z_1+Z_2 也是一个自由度为 (k_1+k_2) 的 χ^2 变量，称为 χ^2 分布的**再生性质**(reproductive property)。

定理 5.5 若 Z_1 是一个标准化的正态变量 $[Z_1 \sim N(0,\ 1)]$，另一变量 Z_2 服从自由度为 k 的 χ^2 分布并独立于 Z_1，则如下定义的变量

$$t=\frac{Z_1}{\sqrt{Z_2}/\sqrt{k}}=\frac{Z_1\sqrt{k}}{\sqrt{Z_2}}=\frac{\text{标准正态变量}}{\sqrt{\text{独立}\ \chi^2\ \text{分布变量}/\mathrm{df}}}\sim t_k$$

即服从 df $=k$ 的学生氏 t 分布。注：此分布在附录 A 中讨论，并在第 5 章中加以说明。

顺便一提，注意随着自由度 k 无限增加（即 $k\to\infty$），学生氏 t 分布趋近于标准正态分布。[3]作为惯例，记号 t_k 表示自由度为 k 的学生氏 t 分布或变量。

定理 5.6 若 Z_1 和 Z_2 分别是自由度为 k_1 和 k_2 的独立分布 χ^2 变量，则变量

$$F=\frac{Z_1/k_1}{Z_2/k_2}\sim F_{k_1,k_2}$$

即服从自由度为 k_1 和 k_2 的 F 分布，其中 k_1 表示**分子自由度**（numerator degrees of freedom），k_2 表示**分母自由度**（denominator degrees of freedom）。

161 同样出于习惯，记号 F_{k_1,k_2} 表示自由度为 k_1 和 k_2 的 F 变量，分子自由度放在前面。

换言之，定理 5.6 说明，F 变量无非就是两个独立分布的 χ^2 变量分别除以其自由度后的比率。

定理 5.7 自由度为 k 的 t 变量的平方具有分子自由度 $k_1=1$ 和分母自由度 $k_2=k$ 的 F 分布。[4]即

$$F_{1,k}=t_k^2$$

注意，欲使此等式成立，F 变量的分子自由度必须为 1。于是 $F_{1,4}=t_4^2$ 或 $F_{1,23}=t_{23}^2$ 等。

我们以后将逐渐看到前面这些定理的实际用处。

定理 5.8 对很大的分母自由度，分子自由度乘以 F 值就近似等于具有分子自由度的 χ^2 值。即

$$mF_{m,n}=\chi_m^2 \qquad \text{当}\ n\to\infty$$

定理 5.9 在自由度充分大时，χ^2 分布可由标准正态分布近似如下：

$$Z=\sqrt{2\chi^2}-\sqrt{2k-1}\sim N(0,\ 1)$$

其中 k 表示自由度。

5A.2 方程（5.3.2）的推导

令：

$$Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\mathrm{se}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{x_i^2}}{\sigma} \tag{1}$$

和

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{2}$$

如果 σ 已知，则 Z_1 遵循标准正态分布；也就是，$Z_1 \sim N(0, 1)$。（为
162 什么?）Z_2 遵循（$n-2$）个自由度的 χ^2 分布。[5]而且，可以证明 Z_2 的分布独立于 Z_1。[6]因此，借助于定理 5.5，变量

$$t = \frac{Z_1\sqrt{n-2}}{\sqrt{Z_2}} \tag{3}$$

遵循 $n-2$ 个自由度的 t 分布。将（1）和（2）代入（3）即给出方程（5.3.2）。

5A.3 方程（5.9.1）的推导

方程（1）表明 $Z_1 \sim N(0, 1)$。因此，由定理 5.3，这个量的平方：

$$Z_1^2 = \frac{(\hat{\beta}_2 - \beta_2)^2\sum x_i^2}{\sigma^2}$$

遵循 1 个自由度的 χ^2 分布。如第 5A.1 节所指出：

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum \hat{u}_i^2}{\sigma^2}$$

也遵循 $n-2$ 个自由度的 χ^2 分布。此外，如第 4.3 节所提到的，Z_2 的分布独立于 Z_1。然后应用定理 5.6 就推出：

$$F = \frac{Z_1^2/1}{Z_2/(n-2)} = \frac{(\hat{\beta}_2 - \beta_2)^2(\sum x_i^2)}{\sum \hat{u}_i^2/(n-2)}$$

遵从自由度分别是 1 和 $n-2$ 的 F 分布。在虚拟假设 H_0：$\beta_2 = 0$ 下，上述 F 比率简化为方程（5.9.1）。

5A.4 方程（5.10.2）和（5.10.6）的推导

均值预测的方差

给定 $X_i = X_0$，对真实均值的预测$E(Y_0 | X_0)$由下式给出：

$$E(Y_0 | X_0) = \beta_1 + \beta_2 X_0 \tag{1}$$

我们用

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{2}$$

来估计（1）。给定 X_0，取（2）的数学期望得：

$$\begin{aligned} E(\hat{Y}_0) &= E(\hat{\beta}_1) + E(\hat{\beta}_2) X_0 \\ &= \beta_1 + \beta_2 X_0 \end{aligned}$$

163 这是因为 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 都是无偏估计量。因此，

$$E(\hat{Y}_0) = E(Y_0 | X_0) = \beta_1 + \beta_2 X_0 \tag{3}$$

就是说，$\hat{Y}_0$ 是 $E(Y_0 | X_0)$的一个无偏误的预测元(unbiased predictor)。

现利用性质：$\mathrm{var}(a + b) = \mathrm{var}(a) + \mathrm{var}(b) + 2\mathrm{cov}(a, b)$，我们得到：

$$\mathrm{var}(\hat{Y}_0) = \mathrm{var}(\hat{\beta}_1) + \mathrm{var}(\hat{\beta}_2) X_0^2 + 2\mathrm{cov}(\hat{\beta}_1 \hat{\beta}_2) X_0 \tag{4}$$

利用（3.3.1），（3.3.3）和（3.3.9）中所给 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差与协方差公式，将各项归并整理即得：

$$\mathrm{var}(\hat{Y}_0) = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right] \quad = (5.10.2)$$

个值预测的方差

我们要预测对应于 $X = X_0$ 的个值 Y，也就是，我们要得到：

$$Y_0 = \beta_1 + \beta_2 X_0 + u_0 \tag{5}$$

我们把 Y_0 预测为：

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{6}$$

预测误差 $Y_0 - \hat{Y}_0$ 是：

$$\begin{aligned} Y_0 - \hat{Y}_0 &= \beta_1 + \beta_2 X_0 + u_0 - (\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2) X_0 + u_0 \end{aligned} \tag{7}$$

因此，

$$E(Y_0-\hat{Y}_0)=E(\beta_1-\hat{\beta}_1)+E(\beta_2-\hat{\beta}_2)X_0-E(u_0)$$
$$=0$$

这是因为 $\hat{\beta}_1$，$\hat{\beta}_2$ 是无偏的，X_0 是固定数，而 $E(u_0)$按假定为零。

对(7)两边平方再取期望值，就得到 $\text{var}(Y_0-\hat{Y}_0)=\text{var}(\hat{\beta}_1)+X_0^2\text{var}(\hat{\beta}_2)+2X_0\text{cov}(\beta_1,\beta_2)+\text{var}(u_0)$。利用先前给出的 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差公式并注意到 $\text{var}(u_0)=\sigma^2$，最后得到：

$$\text{var}(Y_0-\hat{Y}_0)=\sigma^2\left[1+\frac{1}{n}+\frac{(X_0-\bar{X}^2)}{\sum x_i^2}\right] \qquad =(5.10.6)$$

【附录注释】

[1] 至于各个定理的证明，见 Alexander M.Mood，Franklin A.Graybill，and Duane C.Bose，*Introduction to the Theory of Statistics*，3d ed.，McGraw-Hill，New York 1974，pp.239－249。

[2] 前引文献 p.243。

[3] 证明参见 Henri Theil，*Introduction to Econometrics*，Prentice-Hall，Englewood Cliffs，N.J.，1978，pp.237 245。

[4] 证明参见方程 (5.3.2) 和 (5.9.1)。

[5] 证明参见 Robert V.Hogg and Allen T.Craig，*Introduction to Mathematical Statistics*，2d ed.，Macmillan，New York，1965，p.144。

[6] 证明参见 J.Johnston，*Econometric Methods*，McGraw-Hill，3d.ed.，New York，1984，pp.181－182。(但要读懂它，需要有矩阵代数的知识。)

【注释】

[1] Stephen M.Stigler，"Testing Hypothesis of Fitting Models? Another Look at Mass Extinctions"，in Matthew H.Nitecki and Antoni Hoffman，eds.，*Neutral Models in Biology*，Oxford University Press，Oxford，1987，p.148.

[2] 又称**犯第Ⅰ类错误的概率**。第Ⅰ类错误，是指拒绝一个正确的假设。而第Ⅱ类错误，则指接受一个错误的假设。(对这个问题，附录 A 中有充分的讨论。) 符号 α 又称 **(统计) 检验的尺度** [size of the (statistical) test]。

[3] 一些作者喜欢在 (5.3.5) 中把自由度显示出来，把 (5.3.5) 写为：

$$\Pr[\hat{\beta}_2-t_{(n-2),\alpha/2}\,\text{se}(\hat{\beta}_2)\leqslant\beta_2\leqslant\hat{\beta}_2+t_{(n-2),\alpha/2}\,\text{se}(\hat{\beta}_2)]=1-\alpha$$

但为简单起见，我们将保持我们的记号；有关的自由度在行文中加以澄清。

[4] 一个容易读懂的论述，见于 John Neter，William Wasserman，and Michael H.Kutner，*Applied Linear Regression Models*，Richard D.Irwin，Homewood，Ill.，1983，Chap.5。

[5] 证明见 Robert V.Hogg and Allen T.Craig，*Introduction to Mathematical Statistics*，2d ed.，Macmillan，New York，1965，p.144。

[6] 一个统计假设如果规定了一个概率分布的参数的准确值，就叫做**简单假设** (simple hypothesis)；否则就叫做**复合假设** (composite hypothesis)。例如在正态概率

密度函数$(1/\sigma\sqrt{2\pi})\exp\{-\frac{1}{2}[(X-\mu)/\sigma]^2\}$中，如果我们断言 H_1：$\mu=15$ 和 $\sigma=2$，它就是一个简单假设；但如果 H_1：$\mu=15$ 和 $\sigma>15$，由于标准差没有一个准确值，它就是一个复合假设。

[7] 请牢记，即使假设 H_0 正确，在 H_0 下仍有 $100\alpha\%$ 的机会，这个区间不包含 β_2。简单地说，有 $100\alpha\%$ 的机会犯第Ⅰ类错误。比方说，如果 $\alpha=0.05$，就会有 5% 的机会我们拒绝了一个其实是真的虚拟假设。

[8] 如果你想采用置信区间方法，就要构造一个 β_2 的 $(100-\alpha)\%$ 单侧或单尾置信区间。为什么？

[9] 详见 E.L.Lehman, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, 1959。

[10] 5.2 节第 4 点曾表明，我们不可以说固定的区间（0.426 8，0.591 4）包含有真实 β_2 的概率是 95%。但是作为一个估计量的 $\hat{\beta}_2$ 是一个随机变量。我们就可做出（5.7.3）中的概率性表述。

[11] Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, p.114.

[12] 关于如何建立假设的一个饶有趣味的讨论，见 J.Bradford De Long and Kevin Lang, "Are All Economic Hypotheses False?", *Journal of Political Economy*, vol.100, no.6, 1992, pp.1257 – 1272。

[13] Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, pp.126 – 127.

[14] 利用电子统计表可以得到好几位小数的 p 值。可惜，常用的统计表，因限于篇幅，做不到那般精细。大多数统计软件包现在都一律打印出这些 p 值。

[15] Arthur S.Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1991, p.240. 注：b_j 是 β_j 的 OLS 估计量，而 $\hat{\sigma}_{bj}$ 是它的标准误。赞同的观点还见于 D.N.McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review*, vol.75, 1985, pp.201 – 205. 还可见 D.N.McCloskey and S.T.Ziliak, "The Standard Error of Regression," *Journal of Economic Literature*, Vol.37, 1996, pp.97 – 114。

[16] 见注释［12］所引的他们的文章，第 1271 页。

[17] 至于有些不同的观点，参见 Carter Hill, William Griffiths, and George Judge, *Undergraduate Econometrics*, Wiley & Sons, New York, 2001, p.108。

[18] 证明见 K.A.Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, New York, 1960, pp.278 – 280。

[19] 关于所作的各种陈述的证明，见附录 5A，第 5A.4 节。

[20] 见 C.M.Jarque and A.K.Bera, "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, *vol*.55, 1987, pp.163 – 172。

第 6 章 双变量线性回归模型的延伸

164 线性回归分析的某些方面的问题，能够很容易地纳入我们至今已讨论过的双变量线性回归模型的框架之中。首先让我们考虑**过原点的回归**情形。也就是，模型中没有截距项 β_1。然后，我们考虑**度量单位**的问题，即 Y 和 X 变量用什么单位来度量，度量单位的变化会不会影响回归的结果。最后，我们考虑线性回归模型的**函数形式**问题。至今我们所考虑的模型既对参数为线性，又对变量为线性。但请回顾，前面各章所讲的回归理论，仅要求参数是线性的；进入模型的变量则可以是或可以不是线性的。本章中我们将表明，考虑参数为线性、变量不一定为线性的模型能处理一些令人感兴趣的实际问题。

我们一旦掌握好本章所引入的思想，当我们在第 7 章和第 8 章中看到这些思想被推广到复回归模型中时，我们就了如指掌了。

§6.1 过原点回归

有时双变量 PRF 采取如下形式：

$$Y_i = \beta_2 X_i + u_i \tag{6.1.1}$$

在此模型中截距项不出现或者为零。因此取名为**过原点回归**（regression through the origin）。

165 作为一个说明性例子，考虑现代证券组合理论中的资本资产定价模型（Capital Asset Pricing Model，CAPM）。可用风险溢价或升水（risk-premium）的形式把它表述为[1]：

$$(\mathrm{ER}_i - r_f) = \beta_i(\mathrm{ER}_m - r_f) \tag{6.1.2}$$

其中 ER_i = 第 i 种证券的期望回报率

ER_m = 比方说，由标准普尔 500（S&P）复合股指所代表的市场证券组合的期望回报率

r_f = 无风险回报率，比方说，90 天国库券的回报率

β_i = Beta 系数，指不能通过分散化而消除的系统风险（systematic risk）的一种度量，又指第 i 种证券回报率与市场互动程度的一种度量。一个大于 1 的 β_i 意味着一种（高）波动性或进攻型（aggressive）证券，而一个小于 1 的 β_i 则意味着一种防御型（defensive）证券。（注：不要把这个 β_i 和双变量回归的斜率系数 β_2 混同起来。）

如果资本市场有效地运行，则 CAPM 要求：证券 i 的期望风险溢价（$= \mathrm{ER}_i - r_f$）等于期望市场风险溢价（$= \mathrm{ER}_m - r_f$）乘以该证券的 β 系数。如果 CAPM 成立，我们就得到图 6.1 所描述的情形。图中所展示的直线叫做**证券市场线**（security market line，SML）。

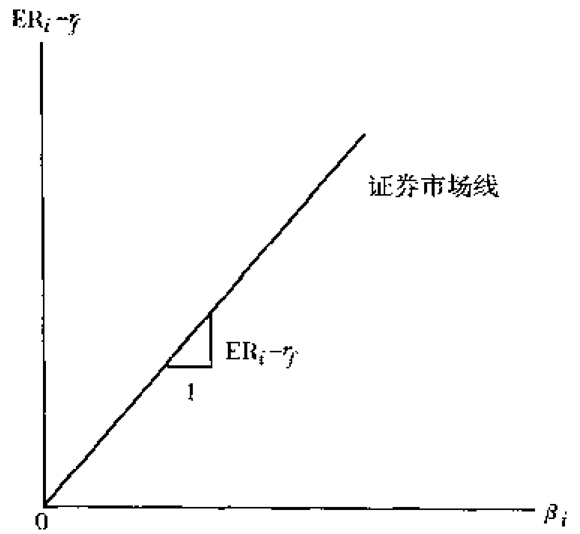


图 6.1 系统风险

为了经验研究的目的，(6.1.2) 常表达为：

$$R_i - r_f = \beta_i(R_m - r_f) + u_i \tag{6.1.3}$$

或：

$$R_i - r_f = \alpha_i + \beta_i(R_m - r_f) + u_i \tag{6.1.4}$$

166 后一个式子叫做**市场模型**（Market Model）。[2] 如果 CAPM 成立，则预期 α_i 为零。（见图 6.2。）

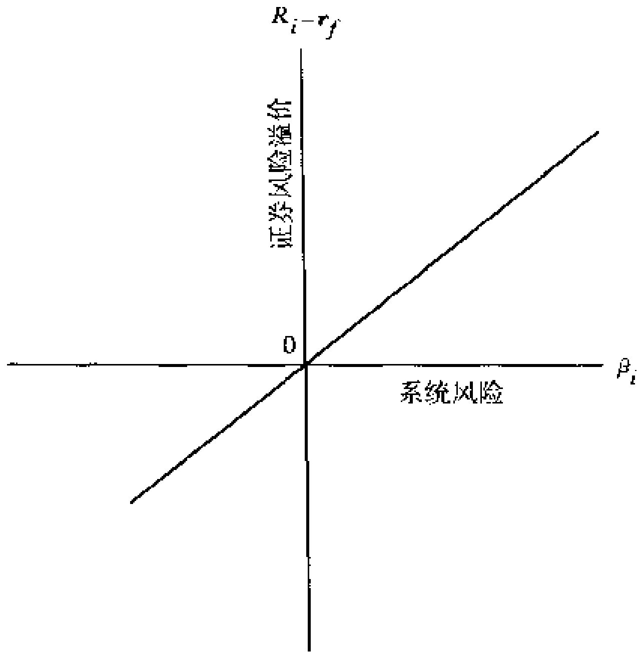


图 6.2　证券组合理论的市场模型（假定 $\alpha_i = 0$）

要注意：(6.1.4) 中的因变量 Y 是 $(R_i - r_f)$，但解释变量 X 是波动性系数 β_i，而不是 $(R_m - r_f)$。因此，为了做回归 (6.1.4)，必须先估计 β_i，后者如同习题 5.5 所描述的，通常要从**特征线**导出。（进一步的细节，见于习题 8.28。）

如本例所表明的，有时基础理论能断定某个模型没有截距项。其他适合于零截距模型的例子还有弗里德曼的永久收入假说（permanent income hypothesis）：永久消费正比于永久收入；成本分析理论：生产的可变成本正比于产出；以及货币主义者理论的某些解说如：价格变化率（即通货膨胀率）正比于货币供给变化率。

像 (6.1.1) 这样的模型怎样估计？这类模型提出了什么特殊问题？为了回答这些问题，可先把 (6.1.1) 的 SRF 写成：

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i \tag{6.1.5}$$

现对 (6.1.5) 应用 OLS 法，得到 $\hat{\beta}_2$ 的如下公式及其方差（证明见附录 6A，6A.1 节）：

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{6.1.6}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2} \tag{6.1.7}$$

167 其中 σ^2 估计为：

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1} \tag{6.1.8}$$

将这些公式同含有截距项的模型的公式相比是有趣的，后者是：

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{3.1.6}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{3.3.1}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{3.3.5}$$

两组公式之间的差异应是明显的：在没有截距项的模型中，我们使用粗或原生平方和及交叉乘积和，而在有截距项的模型中，我们使用偏离均值的平方和及交叉乘积和。其次，在计算 $\hat{\sigma}^2$ 时，前者的自由度是（$n-1$），而后者是($n-2$)。(为什么?)

虽然无截距或零截距模型在某些情况下是适宜的，但要注意这种模型的一些特点。第一，对有截距项的模型（习惯的模型）来说，$\sum \hat{u}_i = 0$ 总是成立的。但当截距项不出现时，$\sum \hat{u}_i = 0$ 就不一定成立。简言之，在过原点回归中 $\sum \hat{u}_i$ 不一定是零。第二，第 3 章所介绍的判定系数 r^2 对传统的模型来说总是非负的。但对无截距模型而言，有时可能出现负值！这些反常结果的出现，是因为第 3 章中所介绍的 r^2，明确地假定了模型包含有截距。因此，按习惯计算的 r^2 未必适合于过原点的回归模型。[3]

过原点回归模型的 r^2

如刚才所指出并在附录 6A 第 6A.1 节中作进一步讨论的，第 3 章给出的习惯上的 r^2 并不适合于不含有截距的回归。但是我们可以对这样的模型计算以粗 r^2 为名的、定义如下的 r^2：

$$\text{raw}\, r^2 = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2} \tag{6.1.9}$$

168 注：这些是原生（而不是经过均值校正的）平方和及交叉乘积和。

虽然粗 r^2 满足关系 $0 < r^2 < 1$，却不能直接同习惯的 r^2 值相比。因此，一些作者并不对零截距回归模型报告 r^2 值。

由于此模型的这些异常特性，在使用零截距回归模型时须特别小心。除非有非常强的先验性预期，否则以采取含有截距的模型为好。这样做有两方面的好处。第一，尽管模型含有截距项，但若该项的出现是统计上不显著的(即统计上等于零)，则从任何实际方面考虑，都可认为这个结果是一个过原点的回归模型。[4]第二，也是更为重要的，如果在模型中确实有截距，而我

们却执意拟合一个过原点回归，我们就犯了**设定的错误**（specification error），从而违背了经典线性回归模型的假定 9。

一个说明性例子：证券组合理论的特征线

表 6.1 给出了 1971—1980 年期间 Afuture 基金——一种以资本收益（capital gain）最大化为主要投资目标的共同基金的年回报率（%）和用费希尔指数来衡量的市场证券组合的年回报率（%）。

表 6.1　1971—1980 年 Afuture 基金和费希尔指数（市场证券组合）的年回报率

年份	Afuture 基金回报（%）Y	费希尔指数回报（%）X
1971	67.5	19.5
1972	19.2	8.5
1973	−35.2	−29.3
1974	−42.0	−26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

资料来源：Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N. J., 1984, pp. 730 and 738. These data were obtained by the authors from Weisenberg Investment Service, *Investment Companies*, 1981 edition.

$$Y_i = \alpha_i + \beta_i X_i + u_i \tag{6.1.10}$$

其中　Y_i = Afuture 基金的年回报率（%）

X_i = 市场证券组合的年回报率（%）

β_i = 斜率系数，又称证券组合理论中的 Beta 系数

α_i = 截距

[（6.1.10）应读如 $Y_t = \alpha_i + \beta_i X_t + u_t$，以避免误解。——译者注]

关于 α_i 的取值，文献中并无一致意见。一些经验结果曾表示 α_i 是正数且统计上显著，而另一些则表示它并不统计上显著地异于零；对于后者就可

把模型写为：

$$Y_i = \beta_i X_i + u_i \tag{6.1.11}$$

即为一过原点回归。

169 如果我们决定使用模型（6.1.11），就得到如下回归结果：

$$\begin{aligned} \hat{Y}_i &= 1.089\ 9X_i \\ &\quad (0.191\ 6) \qquad \text{raw}r^2 = 0.782\ 5 \\ t &= (5.688\ 4) \end{aligned} \tag{6.1.12}$$

这表明 β_i 显著地大于零。其意义是，市场回报率每增加 1%，平均而言，导致 Afuture 基金回报率增加约 1.09%。

怎样能肯定适宜的模型是（6.1.11）而不是（6.1.10），尤其是当我们看到人们对 α_i 事实上为零的假设并没有任何强烈的先验性信念时？这可通过对（6.1.10）的回归加以检查。用表 6.1 的数据可得到如下回归结果：

$$\begin{aligned} \hat{Y}_i &= \ 1.279\ 7 \quad + \quad 1.069\ 1X_i \\ &\quad (7.688\ 6) \qquad (0.238\ 3) \\ t &= \ (0.166\ 4) \qquad (4.486\ 0) \qquad r^2 = 0.715\ 5 \end{aligned} \tag{6.1.13}$$

注：（6.1.12）的 r^2 值和（6.1.13）的 r^2 值不是直接可比的。从这些结果看来，我们还不能拒绝真实截距等于零的假设，因此就说明使用过原点回归（6.1.1）是有道理的。

顺便提出：（6.1.12）和（6.1.13）的结果并没有很大的差异，尽管从过原点回归模型中估计出来的 $\hat{\beta}$ 的标准误略低，从而支持了章末注 4 中泰尔（Theil）的声称，即如果 α_i 确实是零，斜率系数就可以测算准确得多：读者利用表 6.1 的数据及回归的结果，容易证实过原点回归的斜率系数的 95% 置信区间是（0.656 6，1.523 2），而对模型（6.1.13）置信区间则是（0.519 5，1.618 6）。就是说，前者比后者狭窄些。

§6.2 尺度与测量单位

为领会本节所讲的意思，考虑表 6.2 中的数据。表中数据是关于 1988—1997 年期间，以 1992 年美元计算的私人国内总投资（GPDI）和国内生产总值。第一列给出以 10 亿美元计的 GPDI 数据；第二列给出以百万美元计的同样数据；第三和第四列分别给出以 10 亿美元和百万美元计的 GDP 数据。（都以 1992 年美元为标准计。）

170 假使在 GPDI 对 GDP 的回归中某一研究者使用以 10 亿美元计的数据，而另一研究者使用以百万美元计的同样变量的数据。这两种情形的回归结果会不会是一样的？如果不一样，哪一种结果应被采用？简言之，Y 和 X 的

测量单位会造成回归结果的任何差异吗？果然如此的话，在选择回归分析的测量单位时要采取哪些合理的途径？为了回答这些问题，让我们系统地进行分析。令：

表 6.2　　1988—1997 年美国私人国内总投资与国内生产总值

年份	GPDIBL	GPDIM	GDPB	GDPM
1988	828.200 0	828 200.0	5 865.200	5 865 200
1989	863.500 0	863 500.0	6 062.000	6 062 000
1990	815.000 0	815 000.0	6 136.300	6 136 300
1991	738.100 0	738 100.0	6 079.400	6 079 400
1992	790.400 0	790 400.0	6 244.400	6 244 400
1993	863.600 0	863 600.0	6 389.600	6 389 600
1994	975.700 0	975 700.0	6 610.700	6 610 700
1995	996.100 0	996 100.0	6 761.600	6 761 600
1996	1 084.100 0	1 084 100.0	6 994.800	6 994 800
1997	1 206.400 0	1 206 400.0	7 269.800	7 269 800

注：GPDIBL = 以 1992 年 10 亿美元计私人国内总投资。
GPDIM = 以 1992 年百万美元计私人国内总投资。
GDPB = 以 1992 年 10 亿美元计国内生产总值。
GDPM = 以 1992 年百万美元计国内生产总值。
资料来源：*Economic Report of the President* 1999，Table B-2，p. 328.

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \tag{6.2.1}$$

其中 Y = GPDI，X = GDP。定义：

$$Y_i^* = w_1 Y_i \tag{6.2.2}$$

$$X_i^* = w_2 X_i \tag{6.2.3}$$

其中 w_1 和 w_2 为常数，称做**尺度因子**（scale factors）；w_1 和 w_2 可以相等或相异。

由（6.2.2）和（6.2.3）显见 Y_i^* 和 X_i^* 是重新度量（rescaled）的 Y_i 和 X_i。例如，Y_i 和 X_i 是以 10 亿美元计量的而某人想改用百万美元去表示它们，于是有 $Y_i^* = 1\ 000\ Y_i$，$X_i^* = 1\ 000 X_i$；$w_1 = w_2 = 1\ 000$。

现考虑使用变量 Y_i^* 和 X_i^* 的回归：

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{6.2.4}$$

其中：$Y_i^* = w_1 Y_i$，$X_i^* = w_2 X_i$，并且 $\hat{u}_i^* = w_1 \hat{u}_i$。（为什么？）

我们要找出以下每组变量之间的关系式：

1. $\hat{\beta}_1$ 和 $\hat{\beta}_1^*$；
2. $\hat{\beta}_2$ 和 $\hat{\beta}_2^*$；
3. $\operatorname{var}(\hat{\beta}_1)$和 $\operatorname{var}(\hat{\beta}_1^*)$；
4. $\operatorname{var}(\hat{\beta}_2)$和 $\operatorname{var}(\hat{\beta}_2^*)$；
5. $\hat{\sigma}^2$ 和 $\hat{\sigma}^{*2}$；
6. r_{xy}^2和 $r_{x^*y^*}^2$。

由最小二乘理论，我们知道（见第 3 章）：

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \tag{6.2.5}$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{6.2.6}$$

$$\operatorname{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 \tag{6.2.7}$$

171

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{6.2.8}$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \tag{6.2.9}$$

类似地，把 OLS 法应用于(6.2.4)，我们得到：

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6.2.10}$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \tag{6.2.11}$$

$$\operatorname{var}(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \tag{6.2.12}$$

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \tag{6.2.13}$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \tag{6.2.14}$$

根据这些结果，建立这两组参数估计之间的关系式是不难的。所要做的仅是回忆如下定义的关系式：$Y_i^* = w_1 Y_i$（或 $y_i^* = w_1 y_i$）；$X_i^* = w_2 X_i$（或 $x_i^* = w_2 x_i$）；$\hat{u}_i^* = w_1 \hat{u}_i$；$\bar{Y}^* = w_1 \bar{Y}$ 和$\bar{X}^* = w_2 \bar{X}$。读者利用这些定义便能容易地证明：

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\hat{\beta}_2 \tag{6.2.15}$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \tag{6.2.16}$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \tag{6.2.17}$$

$$\operatorname{var}(\hat{\beta}_1^*) = w_1^2 \operatorname{var}(\hat{\beta}_1) \tag{6.2.18}$$

$$\operatorname{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \operatorname{var}(\hat{\beta}_2) \tag{6.2.19}$$

$$r_{xy}^2 = r_{x^*y^*}^2 \tag{6.2.20}$$

由上述结果显见，一旦尺度因子 w 为已知，给定了一种测量尺度的回归结果，便可导出另一种测量尺度的回归结果。在实践中，虽然我们应该合理地选择测量单位，但要用许多零去把数字表达成百万或 10 亿是没有多少意义的。

172 从（6.2.15）到（6.2.20）这些结果容易导出一些特例。例如，当 $w_1 = w_2$ 即尺度因子相同时，斜率系数及其标准误不受尺度从（Y_i，X_i）变到（Y_i^*，X_i^*）的影响。这点是直观明了的。然而，截距及其标准误却放大或缩小了 w_1 倍。但若 X 尺度不变（即 $w_2 = 1$），而 Y 尺度按因子 w_1 改变，则斜率和截距系数以及它们各自的标准误都要乘以相同的 w_1 因子。最后，如果 Y 尺度不变（即 $w_1 = 1$），而 X 尺度按因子 w_2 改变，则斜率系数及其标准误要乘以因子（$1/w_2$），但截距系数及其标准误不变。

然而，应该知道从（Y，X）到（Y^*，X^*）的尺度变换并不影响前面各章所讨论的 OLS 估计量的性质。

一个数值例子：1988—1997 年美国 GPDI 与 GDP 的关系

为了证实前述理论结果，让我们回到表 6.2 的例子，并检查下述回归结果（括号中的数字为估计标准误。）

GPDI 和 GDP 都以10亿美元计算：

$$\begin{aligned} \widehat{\text{GPDI}}_t &= -1\,026.498 + 0.301\,6\ \text{GDP}_t \\ \text{se} &= (257.587\,4)\quad (0.039\,9) \\ r^2 &= 0.877\,2 \end{aligned} \tag{6.2.21}$$

GPDI 和 GDP 都以百万美元计算：

$$\begin{aligned} \widehat{\text{GPDI}}_t &= -1\,026\,498 + 0.301\,6\ \text{GDP}_t \\ \text{se} &= (257\,587.4)\quad (0.039\,9) \\ r^2 &= 0.877\,2 \end{aligned} \tag{6.2.22}$$

注意，如理论所示，截距及其标准误都是回归（6.2.21）中相应值的 1 000 倍（即从 10 亿美元变到百万美元的 $w_1 = 1\,000$），但斜率系数及其标准误均不变。

GPDI以10亿美元计算而 GDP 以百万美元计算：

$$\begin{aligned} \widehat{\text{GPDI}}_t &= -1\,026.498 + 0.000\,301\text{GDP}_t \\ \text{se} &= (257.587\,4)\quad (0.000\,039\,9) \\ r^2 &= 0.877\,2 \end{aligned} \tag{6.2.23}$$

如所预料，因为仅仅 X 或 GDP 改变了尺度，所以斜率系数及其标准误是它们在（6.2.21）中的值的（1/1 000）倍。

GPDI以百万美元计算而 GDP 以10亿美元计算：

$$\widehat{\text{GPDI}}_t = -1\,026\,498 + 301.582\,6\ \text{GDP}_t$$
$$\text{se} = (257\,587.4)\ (39.899\,89) \qquad (6.2.24)$$
$$r^2 = 0.877\,2$$

再次看到如理论结果所示，截距和斜率系数以及它们各自的标准误都是它们在（6.2.21）中的值的 1 000 倍。

注意，上面给出所有回归的 r^2 值都将保持不变，由于 r^2 值是一个纯数或没有维度，所以它不随度量单位而变化，上面的 r^2 都相同也就无足为奇了。

为结果的解释进一言[14]

173 因为斜率系数 β_2 无非就是变化率，它的单位就是如下比率的单位[5]：

$$\frac{\text{因变量 } Y \text{ 的单位}}{\text{解释变量 } X \text{ 的单位}}$$

例如，在回归（6.2.21）中，斜率系数 0.301 6 的意义是，GDP 每改变一个单位，即 10 亿美元，GPDI 平均改变 0.301 6 个 10 亿美元。在回归（6.2.23）中，GDP 的一个单位即 1 百万美元的变化，平均导致 GPDI 的 0.000 302 个 10 亿美元的变化。当然，这两个结果从它们的 GDP 对 GPDI 的影响看是完全相同的；只不过用不同的测量单位来表示而已。

§6.3 标准化变量的回归

我们在上一节看到，回归子和回归元的单位会影响到回归系数的截距。如果我们愿意把回归子和回归元表示成**标准化变量**，这种影响就可得以避免。如果将一个变量在减去其均值后再除以其标准差，我们就说把这个变量标准化了。

于是，在 Y 对 X 的回归中，如果我们把这些变量重新定义为：

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \qquad (6.3.1)$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \qquad (6.3.2)$$

其中 $\bar{Y} = Y$ 的样本均值，$S_Y = Y$ 的样本标准差，$\bar{X} = X$ 的样本均值，$S_X =$

X 的样本标准差；变量 Y_i^* 和 X_i^* 则被称为**标准化变量**（standardized variables）。

标准化变量的一个有趣特征是，其均值总是 0，标准差总是 1。（证明见附录 6A，第 6A.2 节。）

结果，回归子和回归元如何度量就无所谓了。于是，不再做标准（双变量）回归：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{6.3.3}$$

我们对标准化变量做回归

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* \tag{6.3.4}$$

$$= \beta_2^* X_i^* + u_i^* \tag{6.3.5}$$

174 由于对标准化的回归子和回归元回归，所以截距项总是零。[15] 标准化变量的回归系数（由 β_1^* 和 β_2^* 表示）在文献中被称为 **β 系数**。[16] 顺便指出，(6.3.5) 是一个过原点的回归。

如何解释这些 β 系数呢？其解释是，如果（标准化）回归元增加一个单位的标准差，则（标准化）回归子平均增加 β_2^* 单位个标准差。于是，与传统模型 (6.3.3) 不同，我们度量的变量影响不再使用原来表示 Y 和 X 的单位，而是用其标准差为单位。

为说明 (6.3.3) 与 (6.3.5) 之间的差别，让我们回到上一节中讨论的 GPDI 和 GDP 一例。为便于讨论，将上一节得到的结果再次给出如下：

$$\begin{aligned} \text{GPDI}_t &= -1\,026.498 + 0.301\,6\text{GDP}_t \\ \text{se} &= \quad (257.587\,4) \quad (0.039\,9) \qquad r^2 = 0.887\,2 \end{aligned} \tag{6.3.6}$$

其中 GPDI 和 GDP 均以 10 亿美元计。

对应于 (6.3.5) 的结果如下，其中带星号的变量为标准化变量：

$$\begin{aligned} \text{GPDI}_t^* &= 0.938\,7\text{GDP}_t^* \\ \text{se} &= (0.114\,9) \end{aligned} \tag{6.3.7}$$

我们知道如何解释 (6.3.6)：若 GDP 提高 1 美元，则 GPDI 平均提高 30 美分。(6.3.7) 该如何解释呢？这里的解释是，若（标准化）GDP 增加一个标准差，则（标准化）GPDI 平均约增加 0.94 个标准差。

标准化回归模型与传统模型相比有些什么优势呢？若不止一个回归，则优势会更明显，我们在第 7 章将讨论这个论题。通过将回归元标准化，我们就能将它们放到同等地位并直接进行比较。如果一个标准化回归元的系数比模型中另一个标准化回归元的系数大，那么前者就能比后者更多地解释回归子。换言之，我们可以用 β 系数作为各个回归元相对解释力的一种度量。在接下来的两章有更多的说明。

175 在结束本论题之前，须注意两点。第一点，由于标准化回归 (6.3.7) 是一个过原点回归，而我们在第 6.1 节已经指出通常的 r^2 就能使用，所以我们就没有给出其 r^2 值。第二点，传统模型的 β 系数与这里的 β 系数之间

存在一种有趣的关系。在双变量情形中，这种关系如下：

$$\hat{\beta}_2^* = \hat{\beta}_2\left(\frac{S_x}{S_y}\right) \tag{6.3.8}$$

其中 $S_x = X$ 回归元的样本标准差，$S_y =$ 回归子的样本标准差。因此，若我们知道回归元和回归子的（样本）标准差，则可以将两个 β 系数相互转换。我们在下一章中会看到，在多元回归中，这一关系仍然成立。针对我们的说明性例子验证（6.3.8）的成立，则留给读者作为一个练习。

§6.4 回归模型的函数形式

如第 2 章中所指出的，本书主要考虑对参数为线性的模型；对变量则可以是或不是线性的。在下面的几节中，我们考虑一些常用的回归模型，它们也许对变量是非线性的，但对参数则是线性的。或者可通过适当的变量代换而转变为对参数线性。具体地说，我们讨论如下的回归模型：

1. 对数线性模型；
2. 半对数模型；
3. 倒数模型；
4. 对数倒数模型。

我们讨论每一种模型的特点，这些模型在什么场合适用，以及怎样来估计它们。每一种模型都用适当的例子加以说明。

§6.5 怎样测度弹性：对数线性模型

考虑以**指数回归模型**（exponential regression model）命名的如下模型：

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \tag{6.5.1}$$

这可另行表达为[7]：

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + u_i \tag{6.5.2}$$

其中 ln= 自然对数（即以 e 为底的对数，$e=2.718$）。[8]

如果将（6.5.2）写成：

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \tag{6.5.3}$$

176 其中 $\alpha = \ln\beta_1$，这个模型就是对参数 α 和 β_2 为线性，并且对变量 Y 和 X 的对数为线性，从而可用 OLS 回归来估计。由于这样的线性性质，这种模型被称为**对数—对数**（log-log），**双对数**（double-log）或**对数线性**（log-linear）模型。

如果经典线性回归模型的假定均得到满足，则可用 OLS 法估计（6.5.3）中的参数。令：

$$Y_i^* = \alpha + \beta_2 X_i^* + u_i \tag{6.5.4}$$

其中 $Y_i^* = \ln Y_i$，而 $X_i^* = \ln X_i$。所得的 OLS 估计量 $\hat{\alpha}$ 和 $\hat{\beta}_2$ 将分别是 α 和 β_2 的最优线性无偏估计量。

对数—对数模型的一个诱人并且是使它获致普遍应用的特点，是斜率系数 β_2 测度了 Y 对 X 的弹性，也就是由给定的（小的）X 的百分比变化引起的 Y 的百分比变化。[9] 比如说，Y 代表对某一商品的需求量，X 代表其单位
177 价格，则 β_2 测度了需求的价格弹性，这是一个颇具经济含义的参数。如果需求量与价格的关系如图 6.3（a）所示，则由图 6.3（b）显示的双对数变换将给出价格弹性（$-\beta_2$）的估计。

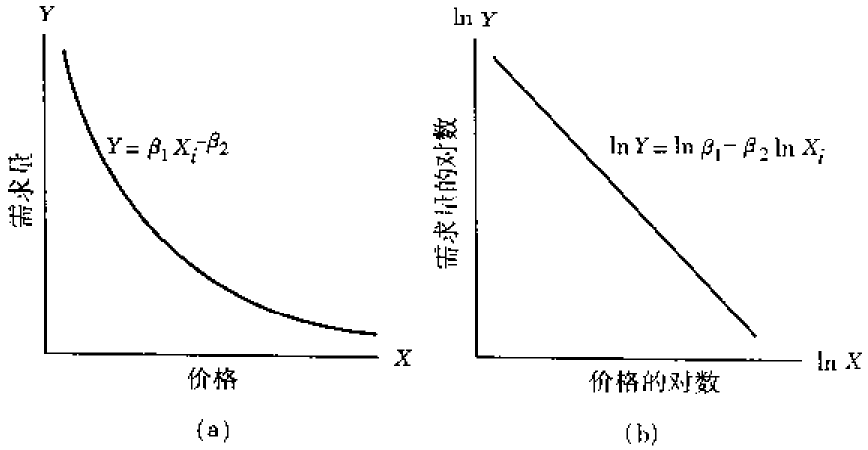


图 6.3　不变弹性模型

可以指出对数线性模型的两个特点：该模型假定 Y 与 X 之间的弹性系数 β_2 在整个研究范围内保持不变（为什么?），因此又名**不变弹性模型**（constant elasticity model）。[10] 换言之，如同图 6.3（b）所示，不管在 $\ln X$ 哪一处测度弹性，$\ln X$ 每变化一个单位所引起的 $\ln Y$ 的变化都是一样的。模型的另一特点是，虽然 $\hat{\alpha}$ 和 $\hat{\beta}_2$ 是 α 和 β_2 的无偏估计量，β_1 进入原始模型参数的估计 $\hat{\beta}_1 = \hat{\alpha}$ 的反对数本身却是一个有偏误的估计量。然而在大多数的实际问题中，截距项都居于次要地位，我们没有必要为得到一个无偏估计量而发愁。[11]

在双变量模型中，决定对数线性模型能否拟合好数据的最简单方法，是描绘出 $\ln Y_i$ 对 $\ln X_i$ 的散点图，看看这些散点是否差不多落在一条如同图 6.3（b）那样的直线上。

一个说明性例子：耐用品支出与个人消费总支出之间的关系

表 6.3 给出了个人消费总支出（personal consumption expenditure，PCEXP）、耐用品支出（expenditure on durable goods，EXPDUR）、非耐用品支出（expenditure on nondurable goods，EXPNONDUR）和劳务支出（expenditure on services，EXPSERVICES）方面的数据，均以 1992 年的 10 亿美元计。[12]

178

表 6.3　个人消费总支出及其类别

观测	EXPSERVICES	EXPDUR	EXPNONDUR	PCEXP
1993-Ⅰ	2 445.3	504.0	1 337.0	4 286.8
1993-Ⅱ	2 455.9	519.3	1 347.8	4 322.8
1993-Ⅲ	2 480.0	529.9	1 356.8	4 366.6
1993-Ⅳ	2 494.4	542.1	1 361.8	4 398.0
1994-Ⅰ	2 510.9	550.7	1 378.4	4 439.4
1994-Ⅱ	2 531.4	558.8	1 385.5	4 472.2
1994-Ⅲ	2 543.8	561.7	1 393.2	4 498.2
1994-Ⅳ	2 555.9	576.6	1 402.5	4 534.1
1995-Ⅰ	2 570.4	575.2	1 410.4	4 555.3
1995-Ⅱ	2 594.8	583.5	1 415.9	4 593.6
1995-Ⅲ	2 610.3	595.3	1 418.5	4 623.4
1995-Ⅳ	2 622.9	602.4	1 425.6	4 650.0
1996-Ⅰ	2 648.5	611.0	1 433.5	4 692.1
1996-Ⅱ	2 668.4	629.5	1 450.4	4 746.6
1996-Ⅲ	2 688.1	626.5	1 454.7	4 768.3
1996-Ⅳ	2 701.7	637.5	1 465.1	4 802.6
1997-Ⅰ	2 722.1	656.3	1 477.9	4 853.4
1997-Ⅱ	2 743.6	653.8	1 477.1	4 872.7
1997-Ⅲ	2 775.4	679.6	1 495.7	4 947.0
1997-Ⅳ	2 804.8	648.8	1 494.3	4 981.0
1998-Ⅰ	2 829.3	710.3	1 521.2	5 055.1
1998-Ⅱ	2 866.8	729.4	1 540.9	5 130.2
1998-Ⅲ	2 904.8	733.7	1 549.1	5 181.8

注：EXPSERVICES = 劳务支出，以 1992 年 10 亿美元计。
EXPDUR = 耐用品支出，以 1992 年 10 亿美元计。
EXPNONDUR = 非耐用品支出，以 1992 年 10 亿美元计。
PCEXP = 个人消费总支出，以 1992 年 10 亿美元计。
资料来源：*Economic Report of the President*，1999，Table B－17，p.347.

假设我们想求出耐用品支出对个人消费总支出的弹性。将耐用品支出的对数相对个人消费总支出的对数描点，你将看到二者之间存在线性关系。因此，双对数模型可以适用。回归结果如下：

$$
\begin{aligned}
\ln \widehat{\text{EXDUR}}_t &= -9.6971 + 1.9056 \ln \text{PCEX}_t \\
\text{se} &= \quad (0.4341) \quad (0.0514) \qquad\qquad (6.5.5)\\
t &= (-22.3370)^{*} \; (37.0962)^{*} \qquad r^2 = 0.9849
\end{aligned}
$$

其中 * 表示 p 值极小。

如这些结果所示，EXPDUR 对 PCEX 的弹性约为 1.90，这表明，若个人消费总支出提高 1%，耐用品支出则提高约 1.9%。因此，耐用品支出很容易受到个人消费总支出变动的影响。这就是耐用品生产者总是关注个人收入和个人消费支出变动的原因之一。在习题 6.17 和 6.18 中，要求读者针对非耐用品支出和劳务支出做类似的练习。

§6.6 半对数模型：线性到对数与对数到线性模型

怎样测量增长率：线性到对数模型

经济学家，企业人员与政府常常对找出某些经济变量，如人口、GNP、货币供给、就业、生产力、贸易赤字等的增长率感兴趣。

假设我们想对表 6.3 中的数据求出个人劳务消费支出的增长率。令 Y_t 表示 t 时期对劳务的真实支出，Y_0 表示劳务支出的初始值（即 1992 年第 4 季度末的值）。回忆你在经济学入门课程中学到的如下的著名复利公式：

$$Y_t = Y_0 (1 + r)^t \qquad (6.6.1)$$

179 其中 r 是 Y 的复合（即在时间上）增长率。取（6.6.1）的自然对数，则可写：

$$\ln Y_t = \ln Y_0 + t\ln(1 + r) \qquad (6.6.2)$$

现假设：

$$\beta_1 = \ln Y_0 \qquad (6.6.3)$$

$$\beta_2 = \ln(1 + r) \qquad (6.6.4)$$

就可把（6.6.2）写为：

$$\ln Y_t = \beta_1 + \beta_2 t \tag{6.6.5}$$

在（6.6.5）中加一个干扰项得[13]：

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \tag{6.6.6}$$

此模型和任何其他线性模型一样，也是对参数 β_1 和 β_2 为线性的。惟一的区别，在于回归子是 Y 的对数而回归元是取值 1、2、3 等等的"时间"。

像（6.6.6）那样的模型叫做**半对数模型**（semilog models），因为只有一个变量（在本例中为回归子）以对数形式出现。为了便于叙述，只是回归子取对数的模型叫做**线性到对数**（log-lin）**模型**。〔以区别于上面讲的对数线性（log-linear）（意谓取对数后即为线性）模型。——译者注〕稍后我们将考虑回归子是线性而（诸）回归元是对数的另一种模型，并称之为**对数到线性**（lin-log）**模型**。

在我们列出回归的结果之前，先来检查模型（6.6.5）的一些性质。在此模型中，斜率系数度量了给定回归元（在本例中为时间变量 t）取值的绝对改变量时 Y 的恒定比例或相对改变量，也就是[14]：

$$\beta_2 = \frac{\text{回归子的相对改变量}}{\text{回归元的绝对改变量}} \tag{6.6.7}$$

如果将 Y 的相对改变量乘以 100，则（6.6.7）将给出相对于回归元 X 的绝对改变量的、Y 的百分比变化或增长率。即 100 乘以 β_2 给出 Y 的增长率；
180 100 乘以 β_2 在文献中被称为 Y 对 X 的**半弹性**（semielasticity）。（问题：要得到弹性，我们该怎么做？）

一个说明性例子：劳务支出的增长率

为了说明增长模型（6.6.6），考虑表 6.3 中给出的劳务支出数据。回归结果如下：

$$\begin{aligned} \widehat{\ln \text{EXS}_t} &= 7.789\,0 + 0.007\,43t \\ \text{se} &= \quad (0.002\,3)\quad(0.000\,17) \\ t &= \quad (3\,387.619)^*\quad(44.282\,6)^* \qquad r^2 = 0.989\,4 \end{aligned} \tag{6.6.8}$$

注：EXS 表示劳务支出，而 * 表示 p 值极小。

对方程（6.6.8）的解释是，1993 年第 1 季度到 1998 年第 3 季度期间，劳务支出以（每季度）0.743% 的速度增加。粗略地讲，这等于 2.97% 的年增长率。由于 7.789 0 = 研究期初 EXS 的对数，所以取其反对数则得到 EXS 期初值（即 1992 年第 4 季度末）2 413.90（单位为 10 亿美元）。图 6.4 勾勒了方程（6.6.8）中得到的回归线。

瞬时与复合增长率。增长模型（6.6.6）中趋势变量的系数 β_2 给出**瞬时**（指一个时点的）增长率而不是**复合**（指一个时期的）增长率。然而后者是容易从（6.6.4）求出的：只须取 β_2 估计值的反对数，再从中减去 1，然后用 100 去乘这个差数。于是，对我们的说明性例子，估计的斜率系数为

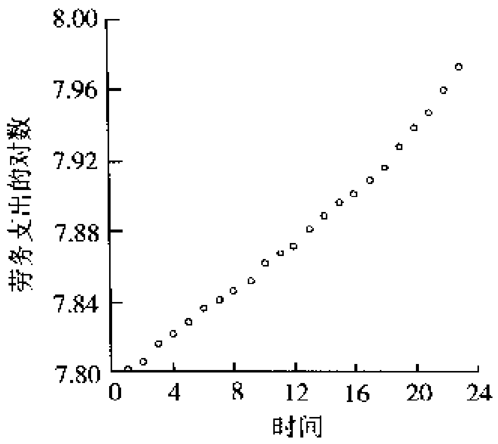


图 6.4　1972—1991 年，美国实际 GDP 的增长：半对数模型

0.007 43。因此[antilog(0.007 43) − 1] = 0.007 46 或 0.746%。因此，在说明性例子中，劳务支出的**复合增长率**约为每季度 0.746%，略高于 0.743% 的瞬时增长率。这当然是由于复合效应所致。

线性趋势模型。有时研究者不去估计模型 (6.6.6)，而代之以如下的模型：

$$Y_t = \beta_1 + \beta_2 t + u_t \tag{6.6.9}$$

就是，不做 ln Y 对时间的回归，而是做 Y 对时间的回归。这样的模型叫做**线性趋势模型** (linear trend model)，并且把时间变量 t 取名为**趋势变量** (trend variable)。趋势的意思，是指一个变量的行为中的一种持续上升或下降的运动。如果 (6.6.9) 中的斜率系数是正的，则 Y 有一上升趋势；反之，如果它是负的，则 Y 有一下降趋势。

181 对于我们前面考虑的劳务支出数据，拟合线性趋势模型 (6.6.9) 的结果如下：

$$\begin{aligned}\widehat{\text{EXS}}_t &= 2\,405.848 \;+\; 19.692\,0t\\ t &= (322.985\,5)\;(36.247\,9)\\ r^2 &= \quad 0.984\,3\end{aligned} \tag{6.6.10}$$

对照 (6.6.8)，对方程 (6.6.10) 的释义如下：在 1993 年第 1 季度至 1998 年第 3 季度期间，劳务支出以每季度约 200 亿美元的速度增加。即劳务支出有上涨的趋势。

增长模型 (6.6.8) 与线性趋势模型 (6.6.10) 之间的取舍，有赖于人们对实际 GDP 的相对或绝对变化的偏好。尽管对许多研究目的来说，相对变化是更为重要的。顺便提请注意，因为 (6.6.8) 和 (6.6.10) 两个模型的回归子不相同，所以不能比较它们的 r^2 值。我们在第 7 章将说明，该如何比较像模型 (6.6.8) 和 (6.6.10) 的 R^2。

对数到线性模型

在刚才讨论的增长模型中，我们感兴趣的是，对 X 的一个绝对的单位变化，找出 Y 的百分比增长。但现在我们感兴趣的是对 X 的一个百分比变化，找出 Y 的绝对变化量。能达到这一目的的模型可写为：

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \tag{6.6.11}$$

为便于描述，我们把这个模型叫做**对数到线性模型**。

让我们来解释斜率系数 β_2。[15] 如同平常，

$$\begin{aligned}\beta_2 &= \frac{Y\text{ 的变化}}{\ln X\text{ 的变化}}\\ &= \frac{Y\text{ 的变化}}{X\text{ 的相对变化}}\end{aligned}$$

上面第二步是因为一个数的对数变化就是它的相对变化。

182 用符号表示，我们有：

$$\beta_2 = \frac{\Delta Y}{\Delta X / X} \tag{6.6.12}$$

其中，Δ 照例表示一个小的变化。方程（6.6.12）又可等效地写为：

$$\Delta Y = \beta_2(\Delta X / X) \tag{6.6.13}$$

这个方程是说，Y 的绝对变化（ΔY）等于 β_2 乘以 X 的相对变化。如果后者乘以 100，则（6.6.13）给出了 X 变化 1% 时 Y 的绝对变化量。例如，$\Delta X/X$ 改变 0.01 个单位（或 1%）时，Y 的绝对变化量是 0.01（β_2）。比如，人们在某一应用中求得 $\beta_2 = 500$，那么，Y 的绝对变化量是(0.01)(500) = 5.0。因此，当人们用 OLS 来估计类似于（6.6.11）的回归时［比较（6.6.14）。——译者注］，要将斜率系数 β_2 的估计值乘以 0.01，或者除以 100 也是一样的。如果不牢记这一点，你在应用中的解释将具有高度的误导性。

实际的问题是，像（6.6.11）这样的一个对数到线性模型什么时候有用？一个有趣的应用在于所谓的**恩格尔支出**（Engel expenditure）模型——以德国统计学家厄恩斯特·恩格尔（Ernst Engel，1821—1896）命名。（见习题 6.10。）恩格尔写道："用于食物的总支出以算术级数增加，而总支出以几何级数增加。"[16]

一个说明性例子

作为对对数到线性模型的一个说明，让我们回顾有关印度食物支出的例 3.2。我们在那里拟合了一个线性于变量的模型作为初步的近似。但若描点

则得到图 6.5 中的描点图。如此图所示，食物支出比总支出增加的更缓慢，这可能是恩格尔法则的凭证。将数据拟合到对数到线性模型的结果如下：

$$食物支出_i = -1\,283.912 + 257.270\,0 \ln 总支出_i \qquad (6.6.14)$$

$$t = (-4.384\,8)^* \qquad (5.662\,5)^* \qquad r^2 = 0.376\,9$$

注：* 号表示 p 值极小。

183 按前面描述的方法来解释，约等于 257 的斜率系数意味着总支出每提高 1%，导致样本中包括的 55 个家庭的食物支出平均增加约 2.57 卢比。(注：我们已将估计系数除以 100。)

在进一步说明之前，注意，如果你想计算对数到线性或线性到对数模型的弹性系数，你可以从前面给出的弹性定义来计算，即

$$弹性 = \frac{\mathrm{d}Y}{\mathrm{d}X} \cdot \frac{X}{Y}$$

事实上，一旦一个模型的函数形式已知，就能利用上述定义来计算弹性。(后面将给出的表 6.6 对各种模型的弹性系数进行了总结概括。)

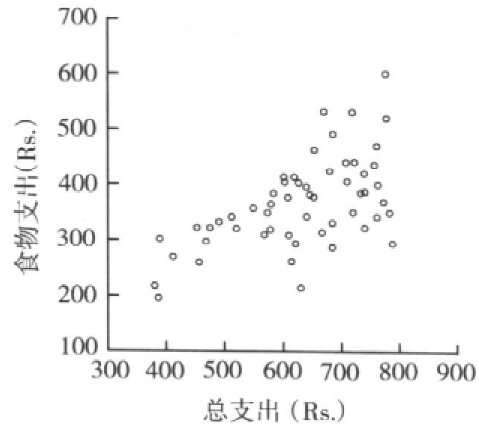


图 6.5

§6.7 倒数模型

属于以下类型的模型均称**倒数** (reciprocal) 模型。

$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \qquad (6.7.1)$$

虽然此模型对变量 X 而言是非线性的 (因为它以倒数形式进入模型)，但模型对 β_1 和 β_2 而言却是线性的，因此它是一个线性回归模型。[17]

此模型有这样一些特点：随着 X 无限地增大，$\beta_2(1/X)$ 项趋于零（注：β_2 是一常数），而 Y 趋于极限或**渐近值** β_1。因此，像（6.7.1）这样的模型在结构上有一内在的**渐近线**或极限值。当变量 X 值无限增大时因变量将取此极限值。[18]

图 6.6 展示出与（6.7.1）相符的一些可能的曲线形状。作为对图 6.6（a）的一个说明，考虑表 6.4 中给出的数据，即 64 个国家的儿童死亡率及其他变量数据。目前主要考虑儿童死亡率（CM）和人均 GNP 这两个变量，并在图 6.7 中描出相应的点。

如你所见，此图与图 6.6（a）相似：假定所有其他变量保持不变，随着人均 GNP 的提高，预计儿童死亡率会因人们能承担更多的健康医疗费用而下降。但这种关系不是一条直线：随着人均 GNP 的增加，CM 首先明显下降，但随着人均 GNP 继续增加，CM 的下降逐渐减弱。

184 如果我们试图拟合倒数模型（6.7.1），我们得到如下回归结果：

$$
\begin{aligned}
\widehat{\mathrm{CM}}_i &= 81.794\,36 + 27\,237.17(1/\mathrm{PGNP}_i) \\
\mathrm{se} &= (10.832\,1) \quad (3\,759.999) \\
t &= (7.551\,1) \qquad (7.253\,5) \quad r^2 = 0.459\,0
\end{aligned}
\tag{6.7.2}
$$

随着人均 GNP 无限增加，儿童死亡率趋近其渐近值，每千人中死亡 82 人。如章末注 18 所解释的那样，$(1/\mathrm{PGNP}_i)$ 的正系数意味着 CM 随着 PGNP 反向变化。

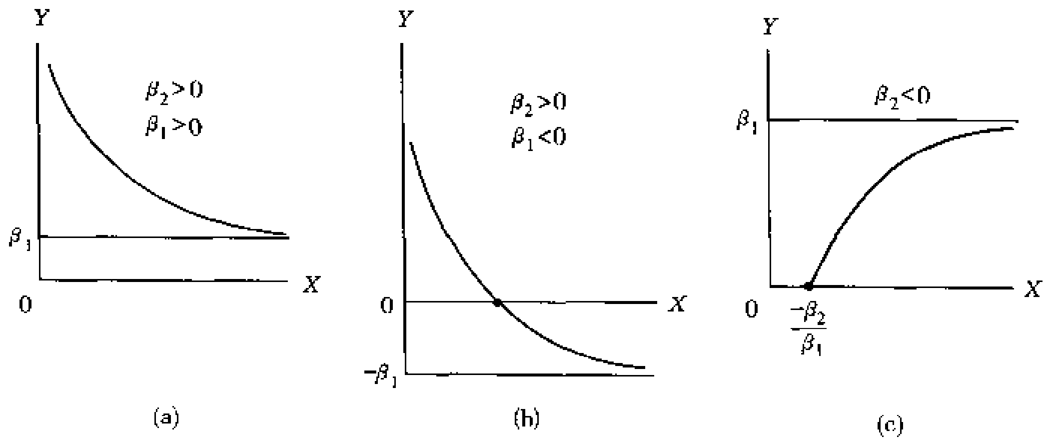


图 6.6　倒数模型：$Y=\beta_1+\beta_2\left(\frac{1}{X}\right)$

表 6.4　64 个国家的生育率及其他数据

观测	CM	FLFP	PGNP	TFR	观测	CM	FLFP	PGNP	TFR
1	128	37	1 870	6.66	33	142	50	8 640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00

4	197	65	570	6.25	36	41	66	1 620	3.91
5	96	76	2 050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2 090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1 180	3.93	43	191	31	1 010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1 730	3.50	45	37	88	1 730	3.46
14	165	31	1 150	7.41	46	103	35	780	5.66
15	94	77	1 160	4.21	47	67	85	1 300	4.82
16	96	80	1 270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1 080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4 430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1 340	7.17
24	12	81	4 240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4 370	2.86
27	107	87	3 020	6.66	59	121	41	1 310	4.88
28	72	63	1 420	7.28	60	115	62	1 470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19 830	5.23	62	47	85	3 630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

注:CM－儿童死亡率,每千名儿童中每年不足 5 岁便死亡的儿童人数。

FLFP＝妇女识字率,%。

PGNP＝1980 年的人均 GNP。

TFR＝1980—1985 年的总生育率,即一位妇女生育的平均子女数,用给定年份按年龄划分的生育率表示。

资料来源:Chandan Mukherjee, Howard White, and Marc Whyte, *Econometrics and Data Analysis for Developing Countries*, Routledge, London, 1998, p.456.

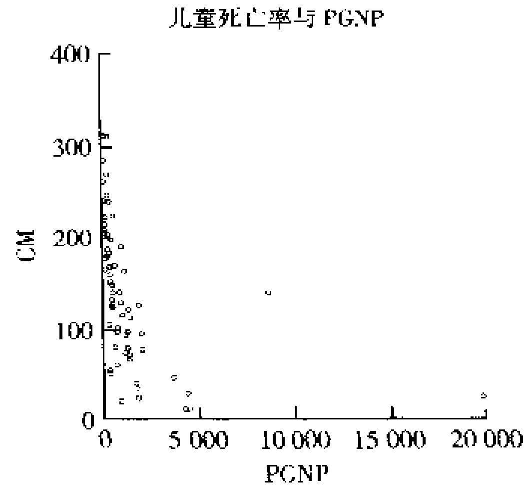


图 6.7　64个国家中儿童死亡率与人均GNP的关系

185 图6.6（b）的重要应用之一，是宏观经济学中著名的菲利普斯曲线(Phillips curve)。根据1861—1957年期间英国货币工资的百分率变化（Y）和失业的百分比（X）数据，菲利普斯得到一条在一般形状上类似于图6.6（b）的曲线（图6.8）。[19]

如图6.8表明，工资变化对失业水平的反应，存在有不对称性：当失业率低于经济学家所称的自然失业率［被定义为保持（工资）通货膨胀不变所需要的失业率］U^N时，由失业的单位变化引起的工资上升，要快于当失业率高于自然水平时，由失业的同样变化引起的工资下降。而β_1表示工资变化的渐近底限（asymptotic floor）。菲利普斯曲线的这一具体特征可能缘于工会的讨价还价能力、最低工资规定、失业补贴等制度因素。

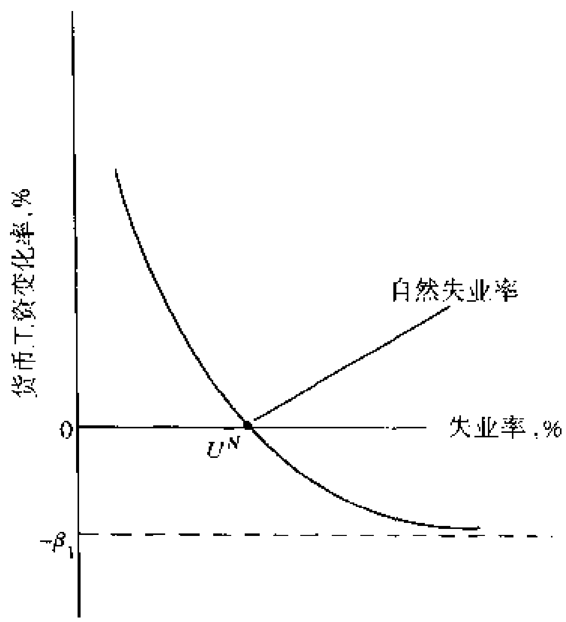


图 6.8　菲利普斯曲线

自从菲利普斯的论文发表以来，从理论和经验水平上对菲利普斯曲线的研究十分广泛。篇幅不容许我们详细介绍围绕着菲利普斯曲线而展开的争辩。菲利普斯曲线本身也几经演变。一个相对近期的表述由布兰查德（Olivier Blanchard）提供。[20]如果我们令 π_t 表示 t 时期的通货膨胀率，其定义是价格水平（由一个有代表性的价格数来度量，如消费者价格指数 CPI）的百分比变化，令 UN_t 表示 t 时期的失业率，那么现代版的菲利普斯曲线表示为如下格式：

$$\pi_t - \pi_t^e = \beta_2(\mathrm{UN}_t - U^n) + u_t \tag{6.7.3}$$

187 其中 π_t = 第 t 年的实际通货膨胀率

π_t^e = 在第（$t-1$）年对第 t 年通货膨胀率的预期

UN_t = 第 t 年的实际失业率

U^n = 第 t 年的自然失业率

u_t = 随机误差项[21]

由于 π_t^e 不能直接观测，所以可以从简化假定 $\pi_t^e = \pi_{t-1}$ 开始；即今年的预期通货膨胀率为去年的通货膨胀率；当然，在形成预期时也可以做更复杂的假定，我们在讨论分布滞后模型的第 17 章中讨论这个问题。

将这个假定代入（6.7.3），并将回归模型写成标准形式，我们就得到如下估计方程：

$$\pi_t - \pi_{t-1} = \beta_1 + \beta_2\mathrm{UN}_t + u_t \tag{6.7.4}$$

其中 $\beta_1 = -\beta_2 U^n$。方程（6.7.4）说明，两个时期之间通货膨胀率的变化与当前的失业率线性相关。据经验，预计 β_2 为负（为什么?），而 β_1 为正（这是因为 β_2 为负且 U^n 为正）。

顺便提一下，（6.7.3）中的菲利普斯曲线在文献中被称为**修正的菲利普斯曲线**（modified Phillips curve），或**附加预期的菲利普斯曲线**（expectation-augmented Phillips curve）（意味着 π_{t-1} 表示预期的通货膨胀）或加速主义者菲利普斯曲线（acceleratonist Phillips curve）（表明低失业率导致通货膨胀上升，并因而成为价格水平的加速器）。

作为对修正的菲利普斯曲线的一个说明，我们表 6.5 中给出 1960—1998 年间的通货膨胀和失业率数据，其中通货膨胀由消费者价格指数的逐年百分比变化（CPI 膨胀）来度量，失业率指城市失业率。我们从这些数据可以得到通货膨胀率的变化（$\pi_t - \pi_{t-1}$），并相对城市失业率描点；我们用 CPI 作为对通货膨胀的一种度量。由此得到图 6.9。

表 6.5　　　美国通货膨胀率与失业率：1960—1998

观测	通货膨胀率	失业率	观测	通货膨胀率	失业率
1960	1.7	5.5	1980	13.5	7.1
1961	1.0	6.7	1981	10.3	7.6

1962	1.0	5.5	1982	6.2	9.7
1963	1.3	5.7	1983	3.2	9.6
1964	1.3	5.2	1984	4.3	7.5
1965	1.6	4.5	1985	3.6	7.2
1966	2.9	3.8	1986	1.9	7.0
1967	3.1	3.8	1987	3.6	6.2
1968	4.2	3.6	1988	4.1	5.5
1969	5.5	3.5	1989	4.8	5.3
1970	5.7	4.9	1990	5.4	5.6
1971	4.4	5.9	1991	4.2	6.8
1972	3.2	5.6	1992	3.0	7.5
1973	6.2	4.9	1993	3.0	6.9
1974	11.0	5.6	1994	2.6	6.1
1975	9.1	8.5	1995	2.8	5.6
1976	5.8	7.7	1996	3.0	5.4
1977	6.5	7.1	1997	2.3	4.9
1978	7.6	6.1	1998	1.6	4.5
1979	11.3	5.8			

注：通货膨胀率为 CPI 的逐年百分比变化。失业率为城市失业率。

资料来源：*Economic Report of the President*，1999，CPI 变化见表 B-63，p.399；失业率见表 B-42，p.376。

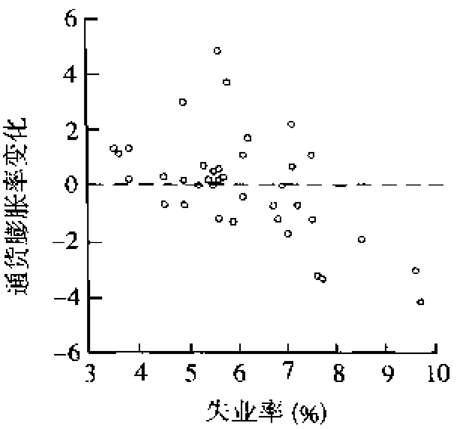


图 6.9　修正的菲利普斯曲线

恰如所料，通货膨胀率变化和失业率之间存在负向关系——低失业率导致通货膨胀率的提高，并因此使价格水平加速上升，加速主义者菲利普斯曲线由此得名。从图 6.9 来看，该数据是一个线性（直线）回归模型还是一个倒数模型拟合数据并不明显；这两个变量之间也可能存在曲线关系。我们基于这两个模型给出如下回归。但记住，如章末注 18 所指出的那样，倒数模

型的截距项预计为负，斜率为正。

线性模型：$(\pi_t - \pi_{t-1}) = 4.178\,1 - 0.689\,5\mathrm{UN}_t$

$$t = (3.952\,1)\ (-4.069\,2) \quad r^2 = 0.315\,0 \qquad (6.7.5)$$

倒数模型：$(\pi_t - \pi_{t-1}) = -3.251\,4 + 18.550\,8\left(\frac{1}{\mathrm{UN}_t}\right)$

$$t = (-2.971\,5)\ (3.062\,5) \quad r^2 = 0.206\,7 \qquad (6.7.6)$$

189 这两个模型的所有估计系数都是个别统计显著的，所有的 p 值都低于 0.005 的水平。

模型（6.7.5）表明，若失业率下降 1 个百分点，则通货膨胀率平均上升约 0.7 个百分点，反之亦然。模型（6.7.6）表明，即便失业率无限增加，通货膨胀率的最大变化也就是下降约 3.25 个百分点。顺便提一句，我们从方程（6.7.5）可以计算出其背后的自然失业率为

$$U^n = \frac{\hat{\beta}_1}{-\hat{\beta}_2} = \frac{4.178\,1}{0.689\,5} = 6.059\,6 \qquad (6.7.7)$$

即自然失业率约为 6.06%。经济学家认为自然失业率介于 5% 与 6% 之间，尽管美国最近的实际失业率远远低于这个数字。

对数双曲线或对数倒数模型

通过考虑形如

$$\ln Y_i = \beta_1 - \beta_2\left(\frac{1}{X_i}\right) + u_i \qquad (6.7.8)$$

的对数倒数模型，我们结束对倒数模型的讨论。其形状如图 6.10 所示。如此图所示，Y 首先以递增的速度增加（即这个曲线先是凸的），然后以递减

190 的速度增加（即变成凹的）。[22]这样的模型可能因此适合于短期生产函数模型。回想在微观经济学中，如果劳动和资本是一个生产函数的投入，而且我们保持资本投入不变但增加劳动投入，那么产出与劳动之间的短期关系就类似图 6.10。（见第 7 章例 7.4。）

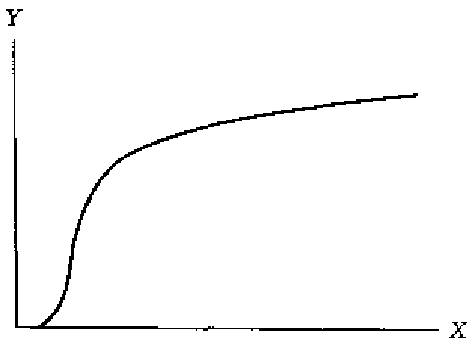


图 6.10　对数倒数模型

§6.8 函数形式的选择

我们在本章讨论了经验模型可以利用的几种函数形式（它们都是线性于参数的回归模型）。在双变量的情形中，由于通过对变量描点就能基本上知道哪个模型合适，所以特定函数形式的选择就相对容易。当我们考虑涉及不止一个回归元的多元回归模型时，这种选择将困难得多，我们在下面两章中讨论这个问题时将会认识到这一点。不可否认，在对经验估计选择适当模型时，需要大量的技巧和经验，但仍有一些指导原则可以参考：

1. 模型背后的理论（如菲利普斯曲线）可能给出了一个特定的函数形式。

2. 最好能求出回归子相对回归元的变化率（即斜率）和回归子对回归元的弹性。我们在表 6.6 中针对本章考虑的各种模型列出了其斜率和弹性系数的公式。了解这些公式将有助于我们比较各种不同的模型。

191 3. 所选模型的系数应该满足一定的先验预期。比如，如果我们考虑对汽车的需求是价格和其他变量的函数，那我们应该预期价格变量的系数为负。

4. 有时不止一个模型能相当不错地拟合一个给定的数据集。在修正的菲利普斯曲线中，我们对同样的数据拟合了一个线性模型和一个倒数模型。在这两种情况下，系数都与先验预期相一致，也都是统计显著的。一个重要的区别在于，线性模型的 r^2 值比倒数模型的 r^2 大。因此人们略微倾向于使用线性模型一些。但一定要注意，在比较两个 r^2 值时，两个模型的因变量或回归子必须相同；回归元则可采用任何形式。我们在下一章将解释其原因。

5. 通常不应该过分强调 r^2 这一度量，也就是说，并非模型的 r^2 越大就越好。如我们在下一章中将讨论的那样，当我们在模型中添加更多的回归元时，r^2 会不断地提高。更重要的地方在于所选模型的理论基础、估计系数的符号及其统计显著性。如果一个模型从这些准则来看不错，那么较低的 r^2 值也是完全可以接受的。我们将在第 13 章更深入地讨论这个重要问题。

表 6.6

模型	方程	斜率$\left(=\frac{dY}{dX}\right)$	弹性$\left(=\frac{dY}{dX}\cdot\frac{X}{Y}\right)$
线性	$Y=\beta_1+\beta_2 X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
对数线性（对数—对数）	$\ln Y=\beta_1+\beta_2\ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2

线性到对数	$\ln Y=\beta_1+\beta_2 X$	$\beta_2(Y)$	$\beta_2(X)^*$
对数到线性	$Y=\beta_1+\beta_2\ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
倒数	$Y=\beta_1+\beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
对数倒数	$\ln Y=\beta_1-\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

注：* 表示弹性系数是可变的，它依赖于 X 或 Y 或二者的取值。在 X 和 Y 未给定时，实践中常常在均值 $\bar{X}$ 和 $\bar{Y}$ 处测度这些弹性。

*§6.9 关于随机误差项的性质的一个注记：相加性与相乘性随机误差项

考虑如下回归模型：

$$Y_i=\beta_1 X^{\beta_2} \tag{6.9.1}$$

这是一个与（6.5.1）相同但没有误差项的模型。为了估计的目的，可把此模型表达成三种不同的形式：

$$Y_i=\beta_1 X_i^{\beta_2} u_i \tag{6.9.2}$$

$$Y_i=\beta_1 X_i^{\beta_2} e^{u_i} \tag{6.9.3}$$

$$Y_i=\beta_1 X_i^{\beta_2}+u_i \tag{6.9.4}$$

对这些方程两边取对数得：

$$\ln Y_i=\alpha+\beta_2\ln X_i+\ln u_i \tag{6.9.2a}$$

$$\ln Y_i=\alpha+\beta_2\ln X_i+u_i \tag{6.9.3a}$$

$$\ln Y_i=\ln(\beta_1 X_i^{\beta_2}+u_i) \tag{6.9.4a}$$

其中 $\alpha=\ln\beta_1$。

192 像（6.9.2）这样的模型是**本质上线性**（对参数而言）回归模型，因为通过适当的（对数）变换即可将该模型变成对参数 α 和 β_2 是线性的。（注：这些模型对 β_1 却是非线性的。）但模型（6.9.4）是本质上对参数非线性的。因为 $\ln(A+B)\neq\ln A+\ln B$。所以没有对（6.9.4）取对数的简单方法。

虽然(6.9.2)和(6.9.3)同是线性回归模型，并且都可用 OLS 或 ML 法加以估计，但我们必须注意进入模型的随机误差项的性质。记得 OLS 的 BLUE 性质要求 u_i 有零均值、恒定方差和零自相关。为了假设检验，我们还假定 u_i 是正态分布的，并有方才说的均值和方差。简言之，我们假定了 $u_i\sim N(0,\sigma^2)$。

* 供选读用。

现在来考虑模型（6.9.2）及其统计匹配（6.9.2a）。为了利用经典正态线性回归模型，我们必须假定：

$$\ln u_i \sim N(0, \sigma^2) \tag{6.9.5}$$

因此，当我们做回归(6.9.2a)时，我们必须把第 5 章讨论的正态性检验，应用到从这个回归得到的残差中。顺便指出，如果 $\ln u_i$ 遵循零均值和恒定方差的正态分布，则统计学理论证明了(6.9.2)中的 u_i 必然遵循均值为$e^{\sigma^2/2}$且方差为$e^{\sigma^2}\cdot(e^{\sigma^2}-1)$的**对数正态分布**（log-normal distribution）。

如上面的分析所表明的，为了做回归分析，我们必须十分留意变换一个模型时的误差项。至于（6.9.4）这个本质上对参数非线性的回归模型，还必须用某种替代性计算机程序来求解。模型（6.9.3）应该没有什么估计上的问题。

总之，当你为回归分析而变换一个模型时，要对干扰项非常仔细的注意，不然的话，对变换后的模型盲目地应用 OLS，将不会产生一个有良好统计性质的模型。

§6.10 要点与结论

本章介绍了经典线性回归模型中的若干更为细致的问题。

1. 有时一个回归模型并不明显含有截距项。这样的模型被称为**过原点**
193 **回归**。虽然估计这种模型的代数方法是简单的，但应小心使用这些模型。对于这种模型，残差和$\sum \hat{u}_i$是非零的；此外，通常计算的 r^2 不一定有意义。除非有很强的理论原因，否则还是在模型中明显地引入一个截距为好。

2. 因为单位和尺度是回归系数赖以解释的关键，所以用什么单位和尺度来表达回归子和回归元是很重要的。在经验研究中，研究者不仅要注明数据的来源，还要声明变量是怎样度量的。

3. 同样重要的是，回归子与回归元之间关系的函数形式。本章讨论的一些重要的函数形式是：(a) 对数线性或恒定弹性模型；(b) 半对数回归模型；(c) 倒数模型。

4. 在对数线性模型中，回归子和回归元都用对数形式来表达。附着于对数回归元的回归系数被解释为回归子对回归元的弹性。

5. 在半对数模型中，或者是回归子或者是回归元以对数形式出现。在回归子为对数形式且回归元为时间的半对数模型中，所估计的斜率系数（乘以 100）度量着回归子的（瞬时）增长率。这样的模型常被用来度量许多经济现象的增长率。在半对数模型中，如果回归元是对数形式，它的系数就测出回归元取值的给定百分比变化所引起的回归子的绝对变化率。

6. 在倒数模型中，或者将回归子或者将回归元表达为倒数或反比形式，以刻画经济变量之间的非线性关系，如同著名的菲利普斯曲线那样。

7. 在选择各种函数形式时，须对随机干扰项 u_i 给予高度的关注。如在第 5 章中所指出的 CLRM 明确地假定了干扰项有零均值和恒定方差（同方差性），并且它与（诸）回归元不相关。在这种假定下 OLS 估计量才是 BLUE。此外，在 CNLRM 下，OLS 估计量还是正态分布的。因此，在选择函数形式做经验分析时，需要查对一下这些假定是否成立。在算完一个回归之后，还应做些诊断性检验，如第 5 章中讨论的正态性检验。这一点再三强调也不为过分，因为经典的假设检验，诸如 t、F 和 χ^2 检验都是以干扰的正态分布为依据的。这对小样本情形尤为重要。

8. 虽然直到现在为止的讨论都限于双变量回归模型，随后各章将表明在许多情况下，把我们的讨论推广到多元回归模型上时也不过是涉及更多的代数，而无须引进更多的基本概念。这说明读者牢牢掌握双变量回归模型是非常重要的。

习　题

194 **问答题**

6.1 考虑回归模型：

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

其中 $y_i = (Y_i - \bar{Y})$，$x_i = (X_i - \bar{X})$。这时回归线必定经过原点。此结论正确或错误？给出你的计算。

6.2 根据 1978 年 1 月至 1987 年 12 月每月数据获得以下回归结果：

$$\begin{aligned}
\hat{Y}_t &= 0.006\,81 \quad + \quad 0.758\,15X_t \\
\text{se} &= (0.025\,96) \qquad (0.270\,09) \\
t &= (0.262\,29) \qquad (2.807\,00) \\
p\text{ 值} &= (0.798\,4) \qquad (0.018\,6) \qquad r^2 = 0.440\,6 \\
\hat{Y}_t &= 0.762\,14X_t \\
\text{se} &= (0.265\,799) \\
t &= (2.954\,08) \\
p\text{ 值} &= (0.013\,1) \qquad\qquad\qquad r^2 = 0.436\,84
\end{aligned}$$

其中 Y = 德士古（Texaco）普通股的月回报率（%）

X = 市场回报率（%）[1]

a. 这两个回归模型有什么区别？

b. 给定上述结果，你会在第一个模型中保留截距项吗？为什么？

c. 你怎样解释这两个模型的斜率系数？

d. 两个模型所依据的理论是什么？

e. 你能不能比较两模型的 r^2 项？为什么？

f. 在此问题中第一个模型的雅克-贝拉正态性统计量是1.116 7，而第二个模型是1.117 0。你能从这些统计量中得出什么结论？

g. 在零截距的模型中斜率系数的 t 值约为 2.95，而在有截距的模型中则约为 2.81。你能对这一结果做些合理的解释吗？

6.3 考虑如下回归模型：

$$\frac{1}{Y_i}=\beta_1+\beta_2\left(\frac{1}{X_i}\right)+u_i$$

注：Y 和 X 都不为零。

a. 这是一个线性回归模型吗？

195 b. 你怎样估计这个模型？

c. 随着 X 趋向无穷大，Y 有怎样的行为？

d. 你能给出一个该模型可能适用的例子吗？

6.4 考虑对数线性模型

$$\ln Y_i=\beta_1+\beta_2\ln X_i+u_i$$

把 Y 画在纵轴上，并把 X 画在横轴上。分别描绘出 $\beta_2=1$，$\beta_2>1$ 和 $\beta_2<1$ 时表现 Y 与 X 之间关系的曲线。

6.5 考虑下列模型：

模型Ⅰ： $Y_i=\beta_1+\beta_2X_i+u_i$

模型Ⅱ： $Y_i^*=\alpha_1+\alpha_2X_i^*+u_i$

其中 Y^* 和 X^* 是习题 6.7 所定义的标准化变量。试说明 $\hat{\alpha}_2=\hat{\beta}_2(S_x/S_y)$，从而建立下面的论断：**虽然回归的斜率系数与原点的变化无关，却与尺度的变化有关。**

6.6 考虑下列模型：

$$\ln Y_i^*=\alpha_1+\alpha_2\ln X_i^*+u_i^*$$
$$\ln Y_i=\beta_1+\beta_2\ln X_i+u_i$$

其中 $Y_i^*=w_1Y_i$，$X_i^*=w_2X_i$，这里的 w 是常数。

a. 建立这两组回归系数以及它们的标准误之间的关系。

b. 两模型的 r^2 有何不同？

6.7 在回归（6.6.8）和（6.6.10）之间，你觉得哪个模型好？为什么？

6.8 对回归（6.6.8）检验假设：斜率系数与 0.005 无显著差异。

6.9 能不能从（6.7.3）所给的菲利普斯曲线估计出自然失业率？怎样估计？

6.10 恩格尔支出曲线（the Engel expenditure curve）把一个消费者在某一商品上的支出同他或她的总收入联系起来。令 Y = 对某一商品的消费支出，X = 消费者收入，考虑下列模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 + \beta_2(1/X_i) + u_i$$
$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + u_i$$
$$\ln Y_i = \ln\beta_1 + \beta_2(1/X_i) + u_i$$
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

你会选择哪个（些）模型作为恩格尔支出曲线，为什么？（提示：解释各种斜率系数，求出支出的收入弹性，等等。）

196 **6.11** 考虑如下模型

$$Y_i = \frac{e^{\beta_1 + \beta_2 X_i}}{1 + e^{\beta_1 + \beta_2 X_i}}$$

它表示一个线性回归模型吗？若否，你能用什么"技巧"使它变成一个线性回归模型？你如何解释由此得到的模型？在什么情况下，这种模型比较适用？

6.12 图示如下模型（为便于说明，我们省略了观测角标 i）：

a. $Y = \beta_1 X^{\beta_2}$，$\beta_2 > 1$，$\beta_2 = 1$，$0 < \beta_2 < 1$，…

b. $Y = \beta_1 e^{\beta_2 X}$，$\beta_2 > 0$，$\beta_2 < 0$

讨论何时这些模型比较适合。

解答题

6.13 给定数据如表 6.7 所示[2]：

表 6.7

Y_i	X_i	Y_i	X_i
86	3	62	35
79	7	52	45
76	12	51	55
69	17	51	70
65	25	48	120

用以下模型去拟合这些数据并求出通常的回归统计量。

$$\left(\frac{100}{100 - Y_i}\right) = \beta_1 + \beta_2\left(\frac{1}{X_i}\right)$$

6.14 为了度量资本投入和劳动力投入之间的替代弹性，当今著名的 CES（恒定替代弹性，constant elasticity of substitution）生产函数的作者阿罗（Arrow）、切纳里（Chenery）、明哈斯（Minhas）和索洛（Solow）用了以下的模型[3]：

$$\log\left(\frac{V}{L}\right)=\log\beta_1+\beta_2\log W+u$$

其中 （V/L） =单位劳动力的附加值

L = 劳动力投入

W = 实际工资率

系数 β_2 度量着劳动力与资本之间的替代弹性（就是，因素的比例变化/因素相对价格的比例变化)。用表 6.8 给出的数据，验证估计的弹性是1.333 8，并且它在统计上与 1 无显著差异。

197 **表 6.8**

工业	$\log(V/L)$	$\log W$
小麦面粉	3.697 3	2.961 7
食糖	3.479 5	2.853 2
涂料与油漆	4.000 4	3.115 8
水泥	3.660 9	3.037 1
玻璃与玻璃器具	3.232 1	2.872 7
陶瓷	3.341 8	2.974 5
三夹板	3.430 8	2.828 7
棉纺品	3.315 8	3.088 8
毛纺品	3.506 2	3.008 6
大麻纺品	3.235 2	2.968 0
化纤	3.882 3	3.090 9
铝制品	3.730 9	3.088 1
铁与钢	3.771 6	3.225 6
自行车	3.660 1	3.102 5
缝纫机	3.755 4	3.135 4

资料来源：Damodar Gujarati，"A Test of ACMS Production Function: Indian Industries, 1958," *Indian Journal of Industrial Relations*, Vol. 2, no. 1, July 1966, pp. 95—97.

6.15 表 6.9 给出 1968—1982 年期间新加坡国内产品的 GDP（国内生产总值）平价因子和进口商品的 GDP 平价因子。GDP 平价因子常用来代替消费者物价指数作为通货膨胀的指标。新加坡是一个小而开放的经济，在很大程度上依赖国际贸易以求得生存。

表 6.9

年份	国内产品的 GDP 平价因子 Y	进口商品的 GDP 平价因子 X
1968	1 000	1 000
1969	1 023	1 042
1970	1 040	1 092
1971	1 087	1 105
1972	1 146	1 110
1973	1 285	1 257
1974	1 485	1 749
1975	1 521	1 770
1976	1 543	1 889
1977	1 567	1 974
1978	1 592	2 015
1979	1 714	2 260
1980	1 841	2 621
1981	1 959	2 777
1982	2 033	2 735

资料来源：Colin Simkin, "Does Money Matter in Singapore?" *The Singapore Economic Review*, Vol. 29, no. 1, April 1984, Table 6, p.8.

为了研究国内物价与世界物价的关系，下面给出两个模型：

1. $Y_t = \alpha_1 + \alpha_2 X_t + u_t$
2. $Y_t = \beta_2 X_t + u_t$

其中 Y = 国内产品的 GDP 平价因子

X = 进口商品的 GDP 平价因子

198 a. 你会怎样先验地在这两个模型之间进行选择？

b. 兼用两个模型去拟合表中数据，然后决定哪个拟合得更好。

c. 还有什么可能适合这些数据的其他模型？

6.16 参照习题 6.15 所给数据。Y 和 X 的均值分别是 1 456 和 1 760，相应的标准差是 346 和 641。估计下面的回归：

$$Y_t^* = \alpha_1 + \alpha_2 X_t^* + u_t$$

其中加星号的变量是标准化变量，并解释所得结果。

6.17 回到表 6.3。求出耐用品支出的增长率。估计的半弹性是多少？解释你的结论。以耐用品支出为回归子、时间为回归元进行双对数回归有意义吗？你如何解释这种情形中的斜率系数？

6.18 从表 6.3 中给出的数据求出非耐用品支出的增长率，并与习题 6.17 中得到的那些结论相比较。

6.19 回到习题 1.7。现在你知道了几种函数形式，哪一个适合用于研究广告印象与广告支出之间的关系？给出必要的计算。

【习题注释】

[1] 背景数据得自 Ernst R. Berndt, *The Practice of Econometrics*: *Classic and Contemporary*, Addison-Wesley, Reading, Mass., 1991。

[2] 资料来源：摘自 J. Johnston, *Econometric Methods*, 3d ed, McGraw-Hill, New York, 1984, p. 87。这实际上取自 1975 年牛津大学的一份计量经济学考试题。

[3] "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics*, August 1961, Vol. 43, no. 5, pp. 225-254.

附录 6A

6A.1 过原点回归的最小二乘估计量的推导

我们求：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2 \tag{1}$$

对 $\hat{\beta}_2$ 最小化。

求 (1) 对 $\hat{\beta}_2$ 的导数得：

$$\frac{d\sum \hat{u}_i^2}{d\hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_2 X_i)(-X_i) \tag{2}$$

令 (2) 等于零并化简得：

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (6.1.6) = (3)$$

将 PRF: $Y_i = \beta_2 X_i + u_i$ 代入此方程得：

$$\hat{\beta}_2 = \frac{\sum X_i(\beta_2 X_i + u_i)}{\sum X_i^2}$$

$$= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} \tag{4}$$

199 [注：$E(\hat{\beta}_2) = \beta_2$。]因此，

$$E(\hat{\beta}_2-\beta_2)^2 = E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \tag{5}$$

将（5）的右端展开，并注意到 X_i 是非随机的以及 u_i 有同方差性且无自相关，故得：

$$\operatorname{var}(\hat{\beta}_2) = E(\hat{\beta}_2-\beta_2)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad (6.1.7)=(6)$$

顺便指出，令（2）等于零即有：

$$\sum \hat{u}_i X_i = 0 \tag{7}$$

从附录 3A，第 3A.1 节我们看到，当截距项在模型中出现时，除了（7）以外，还有条件 $\sum \hat{u}_i = 0$。从刚才的数学推导看，为什么过原点回归模型的误差总和 $\sum \hat{u}_i$ 不一定为零，就很清楚了。

假使我们要增加 $\sum \hat{u}_i = 0$ 这个条件，那么，我们有：

$$\begin{aligned}\sum Y_i &= \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= \hat{\beta}_2 \sum X_i, \quad \text{因为在结构上} \sum \hat{u}_i = 0\end{aligned} \tag{8}$$

这一表达式于是给出：

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum Y_i}{\sum X_i} \\ &= \frac{\bar{Y}}{\bar{X}} = \frac{Y\text{ 的均值}}{X\text{ 的均值}}\end{aligned} \tag{9}$$

但这个估计量和上面的（3）或（6.1.6）并不一样。由于（3）中的 $\hat{\beta}_2$ 是无偏的（为什么？），（9）中的 $\hat{\beta}_2$ 就不可能是无偏的。

要点在于：在过原点回归中，我们不可能像惯用的模型那样，兼令 $\sum \hat{u}_i X_i$ 和 $\sum \hat{u}_i$ 都等于零。惟一得到满足的条件是 $\sum \hat{u}_i X_i$ 等于零。回忆：

$$Y_i = \hat{Y}_i + \hat{u} \tag{2.6.3}$$

两边求和再除以样本大小 N 便得：

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \tag{10}$$

因为对零截距模型来说 $\sum \hat{u}_i$ 从而 $\bar{\hat{u}}$ 不一定是零，所以：

$$\bar{Y} \neq \bar{\hat{Y}} \tag{11}$$

200 就是说，实测的 Y 均值不一定等于估计的 Y 均值；但对出现有截距的模型而言这两个均值是等同的，这可从（3.1.10）看出来。

前面说过，对零截距模型说，r^2 可能是负的，而对惯用的模型说，它永远不会是负的。下面说明这一情况。

利用（3.5.5a），可写：

$$r^2=1-\frac{\text{RSS}}{\text{TSS}}=1-\frac{\sum\hat{u}_i^2}{\sum y_i^2} \tag{12}$$

对于惯用的或出现有截距的模型，方程（3.3.6）表明：

$$\text{RSS}=\sum\hat{u}_i^2=\sum y_i^2-\hat{\beta}_2^2\sum x_i^2\leqslant\sum y_i^2 \tag{13}$$

除非 $\hat{\beta}_2$ 是零（即 X 对 Y 无任何影响）。就是说，对于惯用的模型，RSS≤TSS，或者说，r^2 不可能是负的。

对于无截距模型，可以类似地表明：

$$\text{RSS}=\sum\hat{u}_i^2=\sum Y_i^2-\hat{\beta}_2^2\sum X_i^2 \tag{14}$$

（注：Y 和 X 的平方和并未经过均值调整。）现在并不能保证 RSS 一定小于 $\sum y_i^2=\sum Y_i^2-N\bar{Y}^2$（即 TSS）。这就示意我们，RSS 可能大于 TSS。意谓惯用的定义 r^2 可能是负的。其实，不难看出，如果 $\hat{\beta}_2^2\sum X_i^2<N\bar{Y}^2$，RSS 将大于 TSS。

6A.2 证明标准化变量的均值为零和方差为 1

考虑随机变量 Y，其（样本）均值为 $\bar{Y}$，（样本）标准差为 S_y。

定义

$$Y_i^*=\frac{Y_i-\bar{Y}}{S_y} \tag{15}$$

因此 Y_i^* 就是一个标准化变量。注意标准化涉及双重变化：(1) 原点即 (15) 的分子发生变化；(2) 分母所代表的尺度也发生了变化。因此，标准化同时涉及原点和尺度的变化。

现在，由于一个变量与其均值的离差之和总是等于零，所以

$$\bar{Y}_i^*=\frac{1}{S_y}\cdot\frac{\sum(Y_i-\bar{Y})}{n}=0 \tag{16}$$

201 因此标准化变量的均值为零。（注：我们之所以能把 S_y 项放到求和符号之外，因为它的值是已知的。）

于是

$$\begin{aligned}S_{y^*}^2&=\sum\frac{(Y_i-\bar{Y})^2/(n-1)}{S_y^2}\\&=\frac{1}{(n-1)S_y^2}\sum(Y_i-\bar{Y})^2\\&=\frac{(n-1)S_y^2}{(n-1)S_y^2}=1\end{aligned} \tag{17}$$

注意

$$S_y^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$

是 Y 的样本方差。

【注释】

[1] Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N. J., 1984, Chap. 14.

[2] 例如，参看 Diana R. Harrington, *Modern Portfolio Theory and the Capital Asset Pricing Model: A User's Guide*, Prentice-Hall, Englewood Cliffs, N. J., 1983, p. 71。

[3] 更多的讨论见 Dennis J. Aigner, *Basic Econometrics*, Prentice-Hall, Englewood Cliffs, N. J., 1971, pp. 85–88。

[4] Henri Theil 指出，如果确实没有截距，那么斜率系数要比硬放进一个截距项估计的准确得多。见他的 *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N. J., 1978, p.76. 还可参看他随之给出的数值列子。

[5] 记得在方程（3.1.7）中，截距＝因变量的均值－斜率×回归元的均值。但对标准化变量而言，因变量和回归元的均值都是零，所以截距值为零。

[6] 不要将这些 β 系数与金融理论中的 β 系数相混淆。

[7] 须知对数的这些性质：(1) $\ln AB = \ln A + \ln B$，(2) $\ln(A/B) = \ln A - \ln B$ 和 (3) $\ln(A^k) = k\ln A$，这里假定 A 和 B 是正数而且 k 是常数。

[8] 在实践中可以用普通对数，即以 10 为底的对数。自然对数与普通对数的关系是 $\ln_e X = 2.302\ 6\log_{10} X$。按惯例，ln 指自然对数，而 log 指以 10 为底的对数；从而没有必要写明下标 e 和 10。

[9] 用微积分符号，弹性系数被定义为 $(dY/Y)/(dX/X) = [(dY/dX)(X/Y)]$。熟悉微分学的读者容易看出 β_2 确实是弹性系数。

一个技术性的注解：习惯了微积分表达的读者将看到 $d(\ln X)/dX = 1/X$ 或 $d(\ln X) = dX/X$，即 $\ln X$ 的无穷小变化（注意微分算子 d）等于 X 的相对或比例变化。但在实践中，对于 X 的小的变化，就可将此关系式写成 $\ln X$ 的变化 $\doteq X$ 的相对变化，这里 $\doteq$ 表示近似地等于。如此，对于小的变化，

$$(\ln X_t - \ln X_{t-1}) \doteq (X_t - X_{t-1})/X_{t-1} = X \text{ 的相对变化}$$

顺便指出，读者应留意这些常常出现的名词：(1) **绝对变化**（absolute change），(2) **相对或比例变化**（relative or proportional change），(3) **百分比变化**（percentage change）**或百分数增长率**（percent growth rate）。比如 $(X_t - X_{t-1})$ 表示绝对变化，$(X_t - X_{t-1})/X_{t-1} = (X_t/X_{t-1} - 1)$ 为相对或比例变化，而 $[(X_t - X_{t-1})/X_{t-1}] \cdot 100$ 为百分比变化或增长率。X_t 和 X_{t-1} 分别是变量 X 的现期和前期值。

[10] 对于给定的价格百分比变化，不管价格的绝对水平是什么，一个不变的弹性模型将给出一个不变的总收入变化。读者可将此结果同一个简单的线性需求函数 $Y_i = \beta_1 + \beta_2 X_i + u_i$ 所蕴涵的弹性情形相比较。然而，简单的线性函数却给出价格每单位变化的不变数量变化。再将此情形同一美元的价格变化下的对数线性模型所给出的变化相比较。

[11] 关于偏误的性质以及怎样对付这种偏误，参看 Arthur S. Goldberger, *Topics in Regression Analysis*, Macmillan, New York, 1978, p.120。

[12] 耐用品包括机动车辆及其部件，家具和住房设施；非耐用品包括食物、衣物、汽油、石油、燃油和煤炭；劳务包括家务、电气、交通和医疗等。

[13] 我们增加这个误差项是因为复利公式并不准确地成立。至于为什么要在进行对数变换之后才加进这个误差项，将在第 6.8 节中加以解释。

[14] 由微分学可以推出

$$\beta_2 = \mathrm{d}(\ln Y)/\mathrm{d}X = (1/Y)(\mathrm{d}Y/\mathrm{d}X) = (\mathrm{d}Y/Y)/\mathrm{d}X,$$

而这就是 (6.6.7)。对于 Y 和 X 的微小变化，这个关系式可近似地写为：

$$\frac{(Y_t - Y_{t-1})/Y_{t-1}}{(X_t - X_{t-1})}$$

注：这里 $x = t$。

[15] 仍然利用微分学，我们得到 $\mathrm{d}Y/\mathrm{d}X = \beta_2(1/X)$。因此，$\beta_2 = \mathrm{d}Y/(\mathrm{d}X/X) =$ (6.6.12)。

[16] 见 Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, London, 1998, *p*.158。这里所引用部分归功于 H. Working, "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, vol.38, 1943, pp.43 - 56。

[17] 若置 $X_i^* = (1/X_i)$，则 (6.7.1) 既对参数又对变量 Y 和 X^* 而言是线性的。

[18] (6.7.1) 的斜率是 $\mathrm{d}Y/\mathrm{d}X = -\beta_2(1/X^2)$，其含义是，如果 β_2 是正的，斜率就一直是负的；而如果 β_2 是负的，斜率就总是正的，分别见于图 6.6 (a) 和 6.6 (c)。

[19] A.W. Phillips, "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861—1957," *Economica*, November 1958, vol.15, pp.283 - 299。注意，原始曲线仅拟合 1861—1913 年期间的数据，并没有穿过失业轴，而图 6.8 则代表菲利普斯曲线的现代情形。

[20] 参见 Olivier Blanchard, *Macroeconomics*, Englewood Cliffs, N. J., 1997, Chap.17。

[21] 经济学家相信这个误差项代表某种供给冲击，如 OPEC 1973 和 1979 年的石油禁运。

[22] 利用微积分可以证明

$$\frac{\mathrm{d}}{\mathrm{d}X}(\ln Y) = -\beta_2\left(-\frac{1}{X^2}\right) = \beta_2\left(\frac{1}{X^2}\right)$$

但

$$\frac{\mathrm{d}}{\mathrm{d}X}(\ln Y) = \frac{1}{Y}\frac{\mathrm{d}Y}{\mathrm{d}X}$$

替换后得到

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \beta_2\frac{Y}{X^2}$$

这就是 Y 对 X 的斜率。

第 7 章 多元回归分析：估计问题

202 前面各章所广泛讨论的双变量模型在实践中往往是不适宜的。比如，在我们的消费—收入一例中，我们无形地假定了只有收入影响着消费 Y。但经济理论很少有这般简单的情形。因为，除了收入，还有许多其他的变量会影响着消费支出。一个显然的变量是消费者的财富。作为另一个例子，对某商品的需求很可能不仅依赖于它本身的价格，而且还依赖于其他相互竞争（互替）或相互补充（互补）的产品价格。此外，还有消费者的收入、社会地位，等等。因此，我们需要把这个简单的双变量模型推广到包含有多于两个变量的模型。加进更多的变量，就把我们引到多元（多变量）回归模型的讨论中去。就是说，要讨论因变量或回归子 Y，依赖于两个或更多个解释变量或回归元的模型。

最为简单的多元回归模型，是含有一个因变量和两个解释变量的三变量回归模型。在本章和下章中，我们将研究这种模型，而在第 9 章中，我们将把它推广到多于三个变量的情形。通观全书，我们考虑的是多元线性回归模型，即对参数而言是线性的模型；对变量而言它们可以是或不是线性的。

§7.1 三变量模型：符号与假定

将双变量的总体回归模型（2.4.2）推广，便可写出三变量 PRF 为：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{7.1.1}$$

203 其中 Y 是因变量，X_2 和 X_3 是解释变量（或回归元），u 是随机干扰项，而 i 指第 i 次观测。当数据为时间序列时，下标 t 将用来指第 t 次观测。[1]

在方程（7.1.1）中 β_1 是截距项。虽然机械地去解释，它代表 X_2 和 X_3 均为零时的 Y 的均值，但像通常所描述的，它给出了所有未包含到模型中来的变量对 Y 的平均影响。系数 β_2 和 β_3 被称为**偏回归系数**（partial regression coefficients），其含义稍后即做解释。

我们继续在第 3 章首次引进的经典线性回归模型的框架中运作。具体地说，我们做如下假定：

u_i 有零均值，或：

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \quad \text{对每一个 } i \tag{7.1.2}$$

无序列相关，或：

$$\operatorname{cov}(u_i, u_j) = 0 \quad i \neq j \tag{7.1.3}$$

同方差性，或：

$$\operatorname{var}(u_i) = \sigma^2 \tag{7.1.4}$$

u_i 与每一 X 变量之间都有零协方差，或：

$$\operatorname{cov}(u_i, X_{2i}) = \operatorname{cov}(u_i, X_{3i}) = 0 \tag{7.1.5}$$ [2]

无设定偏误，或：

$$\text{模型被正确地设定} \tag{7.1.6}$$

X 诸变量间无精确的共线性，或：

$$X_2 \text{ 与 } X_3 \text{ 之间无}\textbf{精确的线性关系} \tag{7.1.7}$$

此外，如同第 3 章那样，我们假定多元回归模型是对参数线性的，回归元的值在重复抽样中是被固定的，以及回归元的取值有足够的变异性（variability）。

204 从假定（7.1.2）到假定（7.1.6）的合理性都无异于第 3.2 节所讨论的。假定（7.1.7）是说，X_2 与 X_3 之间无精确的线性关系，专业上名为**无共线性**（no collinearity）或**无多重**（如果涉及不止一个精确线性关系式）**共**

线性（no multicollinearity），则是新的并需要做一些解释。

非正式地说，无共线性是说没有一个解释变量可以写成模型中其余解释变量的线性组合。

正式地说，无共线性的含义是，不存在一组不全为零的数 λ_2 和 λ_3 使得：

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \tag{7.1.8}$$

如果这一关系式存在，则说 X_2 和 X_3 是**共线的**或**线性相关**。另一方面，如果（7.1.8）仅当 $\lambda_2 = \lambda_3 = 0$ 时成立，则说 X_2 和 X_3 线性独立。

因此，如果：

$$X_{2i} = -4X_{3i} \quad \text{或} \quad X_{2i} + 4X_{3i} = 0 \tag{7.1.9}$$

两变量就是线性相关的。如果这样的两个变量都包含在一个回归模型中，我们就遇到一种完全共线性，或两回归之间的一个精确的线性关系。

虽然我们将在第 10 章中详细讨论多重共线性的问题，但不难从直观上去掌握无多重共线性假定的道理。假使（7.1.1）中的 Y、X_2 和 X_3 分别代表消费者的消费支出、收入和财富。在消费支出与收入和财富有线性关系的公设中，经济理论设想财富和收入也许对消费各有一些独立的影响，否则把收入和财富两个变量都包括到模型中来就是没有意义的。在极端的情形中，如果收入与财富之间有准确的线性关系，我们就只有一个独立变量而不是两个，也就无从区分收入和财富对消费的各自影响了。为了看清楚这点，在消费一收入—财富的表达式中，令 $X_{3i} = 2X_{2i}$。于是（7.1.1）变成：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3)X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned} \tag{7.1.10}$$

其中 $\alpha = (\beta_2 + 2\beta_3)$。就是说，事实上我们有一个双变量而不是三变量的回归。而且，如果我们计算回归(7.1.10)并得到 α，那么 α 给出的是$X_2(=\beta_2)$和 $X_3(=\beta_3)$对 Y 的联合影响（combined influence），并且没有什么方法能分别估计出 X_2 的单独影响（β_2）和 X_3 的单独影响（β_3）。[3]

205 总之，无多重共线性假定，要求我们在 PRF 中仅仅把那些不成为模型中其他变量的线性函数的变量包括进来。实际上能不能永远做到这一点则属另一问题。对此我们将在第 10 章中做广泛的探讨。

首先，无多重共线性的假定是对我们的理论（即 PRF）模型而言。实际上，当我们为经验分析搜集数据时，不能保证回归元之间不相关。事实上，我们在本章稍后的说明性例子中将会发现，在分数应用研究中，几乎不可能找到两个或多个在某种程度上不相关的（经济）变量。我们只是要求不存在像方程（7.1.9）中那样精确的线性关系。

其次，记住我们只是在讨论两个或多个变量之间的完全**线性**关系。多重共线性并不排除变量之间的**非线性**关系。假设 $X_{3i} = X_{2i}^2$，这就不违背不完全共线性的假定，因为这里的变量之间的关系不是线性的。

§7.2 对多元回归方程的解释

给定经典回归模型的诸假定，那么，在（7.1.1）的两边对 Y 求条件期望得：

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} \tag{7.2.1}$$

即，（7.2.1）给出以变量 X_2 和 X_3 的固定值为条件的 Y 的条件均值或期望值。因此，如同双变量情形那样，多元回归分析是以多个解释变量的固定值为条件的回归分析，并且我们所获取的，是给定回归元值时 Y 的平均值或 Y 的平均响应。

§7.3 偏回归系数的含义

前面曾指出，回归系数 β_2 和 β_3 被称为**偏回归**（partial regression）**系数**或**偏斜率**（partial slope）**系数**。偏回归系数的含义如下：β_2 度量着在 X_3 保持不变的情况下，X_2 每变化 1 单位时，Y 的均值$E(Y \mid X_2, X_3)$的变化。换
206 句话说，β_2 给出保持 X_3 不变时$E(Y \mid X_2, X_3)$对 X_2 的斜率。[4]再换一种方式说，它给出 X_2 的单位变化对 Y 均值的“直接”或“净”（净在不染有 X_3 的）影响。类似地，β_3 度量着在 X_2 保持不变的情况下，X_3 每变化 1 单位时 Y 均值的变化。即它给出 X_3 的单位变化对 Y 均值的“直接”或“净”影响，净在不沾有 X_2 的影响。[5]

我们实际上应如何理解保持一个回归元的影响不变呢？为了解释这一点，让我们回到儿童死亡率的例子。记得在那个例子中，Y = 儿童死亡率（CM），X_2 = 人均 GNP（PGNP），X_3 = 妇女识字率（FLR）。假设我们想保持 FLR 的影响不变。由于在任何一个给定的具体数据中，FLR 对 CM 和 PGNP 都有些影响，我们所能做的是，通过分别做 CM 对 FLR 和 PGNP 对 FLR 的回归，然后看一下从这些回归中得到的残差，来消除 CM 和 PGNP 中 FLR 的（线性）影响。利用表 6.4 中的数据，我们得到如下回归：

$$\begin{aligned} \text{CM}_i &= 263.863\,5 - 2.390\,5\text{FLR}_i + \hat{u}_{1i} \\ \text{se} &= (12.224\,9)\ (0.213\,3) \quad r^2 = 0.669\,5 \end{aligned} \tag{7.3.1}$$

其中 $\hat{u}_{1i}$表示此回归的残差项。

$$\begin{aligned} \text{PGNP}_i &= -39.303\,3 + 28.142\,7\text{FLR}_i + \hat{u}_{2i} \\ \text{se} &= (734.952\,6) \quad (12.821\,1) \quad r^2 = 0.072\,1 \end{aligned} \tag{7.3.2}$$

其中 $\hat{u}_{2i}$表示此回归的残差项。

现在

$$\hat{u}_{1i} = (CM_i - 263.8635 + 2.3905FLR_i) \tag{7.3.3}$$

表示 CM 中除去 FLR（线性）影响余下的部分。类似地，

$$\hat{u}_{2i} = (PGNP_i + 39.3033 - 28.1427FLR_i) \tag{7.3.4}$$

表示 PGNP 中除去 FLR 的（线性）影响余下的部分。

因此，如果我们现在做 $\hat{u}_{1i}$ 对 $\hat{u}_{2i}$ 的如下回归（即除掉了 FLR 影响后净化的 CM 和 PGNP），我们不就得到 PGNP 对 CM 的净影响了吗？实际上确实如此（见附录 7A，第 7A.2 节）。回归结果如下：

$$\hat{u}_{1i} = -0.0056\hat{u}_{2i}$$
$$\text{se} = (0.0019) \qquad r^2 = 0.1152 \tag{7.3.5}$$

207 注：此回归没有截距项，因为 OLS 残差 $\hat{u}_{1i}$ 和 $\hat{u}_{2i}$ 的均值都是零。（为什么？）

现在的斜率系数 -0.0056 就给出了 PGNP 的单位变化对 CM 的“真实”或净影响，或者说 CM 方程中 PGNP 的真实斜率 β_2。

想得到 CMFLR 的偏回归系数的读者，通过只将 CM 对 PGNP 回归得到残差 $\hat{u}_{2i}$，再将 $\hat{u}_{1i}$ 对 $\hat{u}_{2i}$ 回归，即可得到 FLR 的偏回归系数。我确信读者明白了这一点。

我们每次要求出真实的偏回归系数都必须经过这个多步骤的程序吗？幸运的是，不必如此，通过下一节讨论的 OLS 程序，可以相当迅速而又例行地做到这一切。刚才概括的多步骤程序只是出于教学目的，让读者明白“偏”回归系数的含义。

§7.4 偏回归系数的 OLS 与 ML 估计

为了估计三变量回归模型（7.1.1）的参数，我们先考虑第 3 章介绍的普通最小二乘法，然后再扼要地考虑第 4 章讨论的最大似然法。

OLS 估计量

为了求 OLS 估计量，让我们先写出和（7.1.1）的 PRF 相对应的样本回归函数如下：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{7.4.1}$$

其中 $\hat{u}_i$ 是残差项，为随机干扰项 u_i 的样本对等部分。

如第 3 章中所看到的，OLS 方法是要选择未知参数的值，以使残差平方和（RSS）$\sum \hat{u}_i^2$ 尽可能小，用符号表示：

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \tag{7.4.2}$$

其中 RSS 的表达式得自（7.4.1）的简单代数运算。

208 为了求（7.4.2）的最小化估计量，最直接的方法是将它对未知数求微分，令所得的表达式为零，然后解联立方程。如同附录 7A，第 7A.1 所展示的，此方法给出如下正规方程［比较方程（3.1.4）和（3.1.5）］：

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 \tag{7.4.3}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \tag{7.4.4}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \tag{7.4.5}$$

由方程（7.4.3）我们立即得到：

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \tag{7.4.6}$$

这就是总体截距 β_1 的 OLS 估计量。

按照用小写字母表示对样本均值离差的惯例，我们从正规方程（7.4.3）至（7.4.5）导出以下公式：

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \qquad (7.4.7)^{[6]}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \tag{7.4.8}$$

分别给出总体偏回归系数 β_2 和 β_3 的 OLS 估计量。

顺便指出以下性质：(1) 可以从方程（7.4.7）和（7.4.8）中的一个通过 x_2 和 x_3 的对调而得到另一个，所以它们本质上是对称的；(2) 两个方程的分母完全相同；以及 (3) 三变量情形是双变量情形的自然推广。

OLS 估计量的方差和标准误

得到了偏回归系数的 OLS 估计量，就可按照附录 3A.3 所指示的方法推导出这些估计量的方差和标准误。如同双变量情形，我们计算标准误有两个主要目的：建立置信区间和检验统计假设。有关的公式如下[7]：

209

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2 \bar{X}_3 \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}\right] \cdot \sigma^2 \tag{7.4.9}$$

$$\operatorname{se}(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)} \tag{7.4.10}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}\sigma^2 \tag{7.4.11}$$

或等价地

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{23}^2)} \tag{7.4.12}$$

其中 r_{23} 是在第 3 章中定义的 X_2 和 X_3 的样本相关系数。[8]

$$\operatorname{se}(\hat{\beta}_2) = +\sqrt{\operatorname{var}(\hat{\beta}_2)} \tag{7.4.13}$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}\sigma^2 \tag{7.4.14}$$

或等价地：

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2(1 - r_{23}^2)} \tag{7.4.15}$$

$$\operatorname{se}(\hat{\beta}_3) = +\sqrt{\operatorname{var}(\hat{\beta}_3)} \tag{7.4.16}$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{x_{2i}^2}\sqrt{x_{3i}^2}} \tag{7.4.17}$$

在所有的这些公式中 σ^2 是总体干扰项 u_i 的（同方差性的）方差。

仿照附录 3A，第 3A.5 节的辩解，读者能证实 σ^2 的一个无偏估计量是：

$$\boxed{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - 3}} \tag{7.4.18}$$

210 注意 σ^2 的这一估计量和它的双变量对等部分 $[\hat{\sigma}^2 = (\sum \hat{u}_i^2)/(n-2)]$ 之间的相似性。现在的自由度是 $(n-3)$，这是因为在估计 $\sum \hat{u}_i^2$ 之前，我们必须先估计 β_1，β_2 和 β_3，从而消耗了 3 个自由度。(这种辩解是很有一般性的，例如，在四变量的情形中，自由度就是 $n-4$。)

一旦算出残差 u_i，就能从 (7.4.18) 算出估计量 $\hat{\sigma}^2$。但更容易算得 $\hat{\sigma}^2$ 的方法是利用以下的关系式（证明见附录 7A，第 7A.3 节）：

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

这是 (3.3.6) 所给的关系式的三变量对等部分。

OLS 估计量的性质

多元回归模型的 OLS 估计量和双变量模型的 OLS 估计量有着平行的性质。具体地说：

1. 三变量回归线（面）通过均值 $\bar{Y}$，$\bar{X}_2$ 和 $\bar{X}_3$。(7.4.3) 表明了这点

[比较双变量模型的方程（3.1.7）]，这个性质可以推广到一般情形。例如，在 k 变量线性回归模型［一个回归子和（$k-1$）个回归元］中：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{7.4.20}$$

我们有：

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \bar{X}_3 - \cdots - \beta_k \bar{X}_k \tag{7.4.21}$$

2．估计的 Y_i（$=\hat{Y}_i$）的均值等于真实 Y_i 的均值，这是容易证明的：

$$\begin{aligned}\hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \qquad \text{（为什么？）} \\ &= \bar{Y} + \hat{\beta}_2 (X_{2i} - \bar{X}_2) + \hat{\beta}_3 (X_{3i} - \bar{X}_3) \\ &= \bar{Y} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}\end{aligned} \tag{7.4.22}$$

其中，和平常一样，小写字母表示有关变量对各自均值的离差。

将（7.4.22）两边对所有样本值求和并除以样本大小 n 即得 $\bar{\hat{Y}} = \bar{Y}$。（注：$\sum x_{2i} = \sum x_{3i} = 0$。为什么？）注意，由于（7.4.22），我们可写为：

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \tag{7.4.23}$$

其中 $\hat{y}_i = (\hat{Y}_i - \bar{Y})$。

因此，SRF（7.4.1）可用离差形式表达为：

$$y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{7.4.24}$$

211 3．$\sum \hat{u}_i = \bar{\hat{u}} = 0$，这可从（7.4.24）得到证实。［提示：（7.4.24）两边对样本值求和。］

4．残差 $\hat{u}_i$ 与 X_{2i} 和 X_{3i} 都不相关，就是，$\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$（证明见附录 7A.1）。

5．残差 $\hat{u}_i$ 与 $\hat{Y}_i$ 不相关，即 $\sum \hat{u}_i \hat{Y}_i = 0$。为什么？［提示：（7.4.23）两边同时乘以 $\hat{u}_i$，然后对样本值求和。］

6．由（7.4.12）和（7.4.15）显见，随着 X_2 和 X_3 的相关系数 r_{23} 朝着 1 增大，对给定的 σ^2 和 $\sum x_{2i}^2$ 或 $\sum x_{3i}^2$ 值来说，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差也增大。在 $r_{23}=1$ 即完全共线的极限中，这些方差变成无限大。这种情形的含义将在第 10 章中充分探讨，但直觉上读者能看出，随着 r_{23} 的增大，要知道 β_2 和 β_3 的真值所在将变得越来越困难。［参照方程（7.1.10），但下章有更多的讨论。］

7．由（7.4.12）和（7.4.15）还明显看到，对给定的 r_{23} 和 $\sum x_{2i}^2$ 或 $\sum x_{3i}^2$ 值，OLS 估计值的方差正比于 σ^2，即这些方差随 σ^2 增加而增加。类似地，对给定的 σ^2 和 r_{23} 值，$\hat{\beta}_2$ 的方差反比于 $\sum x_{2i}^2$，即 X_2 的样本值变化越大，$\hat{\beta}_2$ 的方差越小，并从而能更精确地估计 β_2。关于 $\hat{\beta}_3$ 的方差也可做类似的叙述。

8．在第 7.1 节所声称的经典线性模型的假定下，可以证明偏回归系数

的 OLS 估计量不仅是线性和无偏的，而且在所有线性无偏估计量类中有最小方差。简单地说，它们是 BLUE；换一种方式说，它们满足高斯-马尔可夫定理。(其证明完全类似于附录 3A，第 3A.6 节所证的双变量情形，而我们将在第 9 章中用矩阵符号更简洁地给出这个证明。)

最大似然估计量

在第 4 章中我们曾指出，在总体干扰项 u_i 遵循零均值和常数方差 σ^2 的正态分布的假定下，双变量模型的回归系数的最大似然估计量和普通最小二乘估计量是相等的。这一等号可推广到含任意多个变量的模型中去。(证明见附录 7A，第 7A.4 节。) 然而，对 σ^2 的估计量而言则不然。可以证明，不管模型中有多少个变量，σ^2 的 ML 估计量都是 $\sum \hat{u}_i^2/n$，而 σ^2 的 OLS 估计量则对双变量情形为 $\sum \hat{u}_i^2/(n-2)$，对三变量情形为 $\sum \hat{u}_i^2/(n-3)$，对 k 变量模型 (7.4.20) 为 $\sum \hat{u}_i^2/(n-k)$。总之，σ^2 的 OLS 估计量考虑了自由度的个数，而 ML 估计量则无此考虑。当然，如果 n 很大，σ^2 的 ML 和 OLS 估计量将趋于一致，彼此接近。(为什么?)

§7.5 多元判定系数 R^2 与复相关系数 R

212 在双变量的情形中我们曾看到，由 (3.5.5) 定义，r^2 是回归方程拟合优度的一个度量；即它给出在因变量 Y 的总变异中由 (单一个) 解释变量 X 解释了的比例或百分比。容易把 r^2 这个符号推广应用到含有多于 2 个变量的回归模型中去。这样，在三变量模型中，我们也许想知道 Y 的变异由变量 X_2 和 X_3 联合解释的比例。提供这一信息的数量被称为**复判定系数** (multiple coefficient of determination) 并记为 R^2；概念上 R^2 近似于 r^2。

可仿照第 3.5 节推导 r^2 的方法来推导 R^2。回忆：

$$
\begin{aligned}
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\
&= \hat{Y}_i + \hat{u}_i
\end{aligned} \tag{7.5.1}
$$

其中 $\hat{Y}_i$ 是从所拟合的回归线估计的 Y_i 值，它是真实 $E(Y_i \mid X_{2i}, X_{3i})$ 的一个估计量。把它换成小写字母，用以表示对均值的离差，方程 (7.5.1) 就可写为：

$$
\begin{aligned}
y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \\
&= \hat{y}_i + \hat{u}_i
\end{aligned} \tag{7.5.2}
$$

将 (7.5.2) 两边平方，再对样本值求和，得：

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i$$

$$= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \quad (为什么?) \tag{7.5.3}$$

也就是说，方程（7.5.3）是说，总平方和（TSS）等于解释平方和(ESS) + 残差平方和（RSS）。用（7.4.19）代 $\sum \hat{u}_i^2$ 得：

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

整理后得：

$$\text{ESS} = \sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} \tag{7.5.4}$$

213 于是，按定义：

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \tag{7.5.5}$$ [19]

[比较（7.5.5）和（3.5.6）。]

因为进入（7.5.5）的各项通常已按既定程序算出，故 R^2 能容易计算。R^2 和 r^2 一样，落在 0 与 1 之间。如果是 1，则所拟合的回归线 100% 地解释了 Y 的变异；如果是 0，则模型不解释 Y 的任何变异。典型的情形是 R^2 位于这两个极端值之间。R^2 越靠近 1，我们说模型的“拟合”越好。

回顾双变量的情形，我们把 r 这个量定义为相关系数，表示它度量着两个变量之间的（线性）关联程度。类似于 r 的三或多变量度量是**多元相关**(multiple correlation) 系数，记为 R。它度量着 Y 和所有解释变量在一起的关联程度。虽然 r 可正可负，但 R 则永远取正值。实际上，R 没有多大意思，更有意义的量是 R^2。

在继续讨论前，让我们建立下述 R^2 与 k 变量多元回归模型（7.4.20）中的一个偏回归系数的方差var($\hat{\beta}_j$)之间的关系：

$$\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1 - R_j^2}\right) \tag{7.5.6}$$

其中 $\hat{\beta}_j$ 是回归元 X_j 的偏回归系数，而 R_j^2 是 X_j 对其余（$k-2$）个回归元回归的 R^2。[注：k 变量回归模型中有（$k-1$）个回归元。] 虽然方程（7.5.6）的用途要等到第 10 章讲多重共线性时才变得明显，但应看到此方程不过是公式（7.4.12）和（7.4.15）的推广。后面这两个公式适用于三变量回归模型，即一个回归子和两个回归元的模型。

§7.6 例 7.1 儿童死亡率与人均 GNP 和妇女识字率的关系

214 我们在第 6 章中考虑了儿童死亡率（CM）与人均 GNP（PGNP）之间的关系。在那里，我们发现 PGNP 如预期般对 CM 有负影响。现在让我们引

入由妇女识字率（FLR）度量的妇女识字变量。据经验，可预计 FLR 对 CM 也具有负影响。现在在模型中同时引入这两个变量之后，我们就不需要专门净化每个回归元的净影响。即我们需要估计每个回归元的（偏）回归系数。于是我们的模型为：

$$\mathrm{CM}_i = \beta_1 + \beta_2 \mathrm{PGNP}_i + \beta_3 \mathrm{FLR}_i + u_i \tag{7.6.1}$$

所需数据由表 6.4 给出。记住，CM 为每 1 000 名产婴中不足 5 岁便死亡的人数，PGNP 为 1980 年的人均 GNP，而 FLR 以百分比度量。我们的样本由 64 个国家构成。

使用 Eviews 3 统计软件，我们得到如下结果：

$$\widehat{\mathrm{CM}}_i = 263.641\,6 - 0.005\,6\mathrm{PGNP}_i - 2.231\,6\mathrm{FLR}_i \tag{7.6.2}$$

$$\mathrm{se} = \quad (11.593\,2) \qquad (0.001\,9) \qquad (0.209\,9) \qquad R^2 = 0.707\,7$$

$$\bar{R}^2 = 0.698\,1^*$$

括号中的数字为估计的标准误。在我们解释这个回归之前，先观察 PGNP 的偏回归系数 $-0.005\,6$。它不是与我们在上一节中用三步法所得到的结果［见方程（7.3.5）］完全一样吗？你对此感到吃惊吗？不仅如此，两个标准误也完全相同，这些都无足为奇。我们没有使用麻烦的三步法也得到了这些结果。

现在让我们来解释这些回归系数：$-0.005\,6$ 是 PGNP 的偏回归系数，它告诉我们，保持 FLR 的影响不变，PGNP 提高 1 美元，儿童死亡率平均下降 0.005 6 个单位。为了在经济上更容易解释，若人均 GNP 提高 1 000 美元，则每 1 000 名产婴中不足 5 岁便死亡的儿童数平均下降约 5.6%。系数 $-2.231\,6$ 告诉我们，保持 PGNP 的影响不变，妇女识字率每提高 1 个百分点，每 4 名产婴中不足 5 岁便死亡的儿童数平均减少约 2.23 人。约等于 263 的截距值，机械地解释就是，若 PGNP 和 FLR 固定为零，则每 4 名产婴中儿童死亡人数的均值为 263。当然，对这种解释应该有所保留。从实际情况可以推断出来的是，若这两个回归元都固定为零，儿童死亡率则应该相当高。约为 0.71 的 R^2 值意味着，儿童死亡率变异中约有 70% 可由 PGNP 和 FLR 来解释，考虑到 R^2 的最大值充其量为 1，这个值已相当高了。所有这些都说明回归结果讲得通。

215 估计系数的统计显著性如何呢？我们将在第 8 章讨论这个问题。我们在那里会看到，本章在许多方面都是对讨论双变量模型的第 5 章的一个推广。我们还将证明，双变量和多变量回归模型在统计推断（即假设检验）方面有一些重要的差别。

标准化变量的回归

在上一章中，我们介绍了对标准化变量进行回归的问题，并说过这种分

析可推广到多元回归的情况。记住，如果将一个变量用它与其均值的离差除以其标准差的形式来表示，就称之为标准化变量或以标准差单位表示的变量。

对儿童死亡率一例，结果如下：

$$\mathrm{CM}^* = -0.202\,6\mathrm{PGNP}_i^* - 0.763\,9\mathrm{FLR}_i^*$$

$$\mathrm{se} = (0.071\,3) \qquad (0.071\,3) \qquad r^2 = 0.707\,7 \tag{7.6.3}$$

注：加星号的变量都是标准化变量。还要注意，出于上一章已经讨论过的原因，此模型中没有截距项。

你从这个回归中可以看到，保持 FLR 不变，PGNP 提高一个标准差，导致 CM 平均下降0.202 6个标准差。类似地，保持 PGNP 不变，FLR 提高一个标准差，导致 CM 平均下降 0.763 9 个标准差。相对而言，妇女识字率比人均 GNP 对儿童死亡率的影响更大。这里你将看到使用标准化变量的好处，由于所有的标准化变量的均值都是零、方差都是 1，所以标准化使所有变量都处在同一个标准下。

§7.7 从多元回归的角度看简单回归：设定偏误初探

记得经典线性回归模型的假定（7.1.6）声称，分析中所用的回归模型是正确设定的，就是说，没有设定上的偏误或误差（对此第 3 章曾有过一些初步的评论）。虽然第 13 章将对设定的分析问题做更透彻的讨论，但上一节的说明性例子给了我们机会去深入认识假定（7.1.6）的重要性，不仅如此，还启发了我们更好地去认识偏回归系数的含义，从而比较正式地把我们引导到设定偏误的问题上来。

216 假定（7.6.1）是解释儿童死亡率行为与人均 GNP 和妇女识字率（FLR）之关系的“真实”模型。但假设我们去掉 FLR 而估计如下简单回归：

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \tag{7.7.1}$$

其中 $Y = \mathrm{CM}$，$X_2 = \mathrm{PGNP}$。

既然（7.6.1）是真实模型，所以估计（7.7.1）将构成设定误差；这里的误差因省略妇女识字率变量 X_3 而形成。注意，我们在（7.7.1）中使用了不同的参数符号（α），以有别于真实模型（7.6.1）中的参数（β）。

现在 α_2 能给出模型（7.6.1）中 β_2 所表示的 PGNP 的真实影响的一个无偏估计吗？换言之，$E(\hat{\alpha}_2) = \beta_2$ 吗？其中 $\hat{\alpha}_2$ 是 α_2 的估计值。再言之，知道我们从模型中省略了变量X_3(FLR)后，（7.7.1）中 PGNP 的系数是 PGNP 对 CM 真实影响的无偏估计吗？恰如所料，α_2 通常不是 β_2 的无偏估计。为粗略地了解一下偏误，让我们做回归（7.7.1），并得到如下结论：

$$\widehat{\mathrm{CM}}_i = 157.424\,4 - 0.011\,4\ \mathrm{PGNP}_i$$

$$\text{se} = \quad (9.8455) \quad (0.0032) \qquad r^2 = 0.1662 \tag{7.7.2}$$

与“真实”多元回归（7.6.1）相比，从此回归中可观察到如下几点：

1. 从绝对值看（即忽略符号），PGNP 系数从 0.005 6 增加到 0.011 4，几乎扩大一倍。

2. 标准误不同。

3. 截距值不同。

4. r^2 值明显不同，尽管随着模型中回归元个数的增加，r^2 值通常都会提高。

现在假设你将儿童死亡率对妇女识字率回归，而无视 PGNP 的影响，并得到如下结果：

$$\widehat{\text{CM}}_i = 263.8635 - 2.3905\text{FLR}_i$$

$$\text{se} = \quad (21.2249) \quad (0.2133) \qquad r^2 = 0.6696 \tag{7.7.3}$$

同样，你如果将此（误设）回归结果与“真实”多元回归的结论相比较，你会看出其差别，只是这里的差别没有回归（7.7.2）中那么显著。

217 要指出的重要一点是，错误拟合一个模型会导致严重后果。我们在有关设定误差的第 13 章中将更全面地讨论这个专题。

§7.8 R^2 及校正 R^2

R^2 的一个重要性质是，它是出现在模型中的解释变量或回归元的个数的非减函数；随着回归元个数的增大，R^2 几乎必然增大并永不减小。换一种说法，多加一个 X 变量必不会减少 R^2。比如，将回归（7.7.2）或（7.7.3）与（7.6.2）相比较。为了要看清楚这点，回忆一下判定系数的定义：

$$\begin{aligned} R^2 &= \frac{\text{ESS}}{\text{TSS}} \\ &= 1 - \frac{\text{RSS}}{\text{TSS}} \\ &= 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \end{aligned} \tag{7.8.1}$$

这里，$\sum y_i^2$ 就是 $\sum(Y_i - \bar{Y})^2$，与模型中 X 变量的个数无关。但 RSS 即 $\sum \hat{u}_i^2$ 却与模型中出现的回归元个数相关。直观上，显而易见，随着 X 变量个数的增加，$\sum \hat{u}_i^2$ 很可能减小（至少不会增大）；随之（7.8.1）所定义的 R^2 也将增大。有鉴于此，在比较有同一因变量（same dependent variable）但有不同个数的 X 变量的两个回归时，选择有最高 R^2 值的模型必须当心。

要比较两个 R^2 项，必须考虑到模型中出现的 X 变量的个数。如果我们

考虑如下的另一种判定系数，这个问题就能容易解决：

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} \tag{7.8.2}$$

式中 k = 包括截距项在内的模型中的参数个数。(在三变量回归中 $k=3$。为什么?) 如此定义的 R^2，称**校正 R^2** (adjusted R^2)，记为 $\bar{R}^2$。“校正”一词指对 (7.8.1) 中的平方和所涉及的自由度的校正：在一个涉及 k 个参数 (包括截距项) 的模型中 $\sum \hat{u}_i^2$ 有 $n-k$ 个自由度，而 $\sum y_i^2$ 有 $n-1$ 个自由度。(为什么?) 我们知道，对于三变量情形 $\sum \hat{u}_i^2$ 有 $n-3$ 个自由度。

218 方程 (7.8.2) 又可写为：

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \tag{7.8.3}$$

其中 $\hat{\sigma}^2$ 是残余方差，是真实 σ^2 的一个无偏估计，而 S_Y^2 是 Y 的样本方差。

容易看出 $\bar{R}^2$ 和 R^2 有一定关系；将 (7.8.1) 代入 (7.8.2) 即得：

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{7.8.4}$$

从 (7.8.4) 立即看出：(1) 对于 $k>1$，$\bar{R}^2 < R^2$。这意味着，随着 X 变量的个数增加，校正的 R^2 比未校正的 R^2 增加得慢些；(2) 虽然 R^2 必定是非负的，但 $\bar{R}^2$ 可以是负的。[10]在应用中，如果遇到 $\bar{R}^2$ 为负值的情形，就把它的值取为零。

实践中应选用哪一个 R^2? 如泰尔所注释的：

> ……R^2 对回归拟合的描述，特别是当解释变量的个数相对于观测次数来说不算很少的时候，明显地偏向于乐观，因此，用 $\bar{R}^2$ 而不用 R^2 是一种好的实践。[11]

但泰尔的观点并没有被人们一致地接受，因为他未曾为 $\bar{R}^2$ 的“优越性”提出任何一般理论性的论点。例如，戈德伯格称辩，如下的 R^2，称修正的 (modified) R^2 也能做同样好的工作[12]：

$$\text{校正的 } R^2 = (1 - k/n)\, R^2 \tag{7.8.5}$$

他建议把 R^2，n 和 k 一起报道出来，让读者去决定怎样为说明 n 和 k 的作用而调整 R^2。

且不管这个建议怎样，大多数统计软件包都是把 (7.8.4) 所给的校正 R^2 连同惯用的 R^2 一起报告的。读者完全可以把 $\bar{R}^2$ 当作另一摘要统计量来看待。

219 顺便提一句，对于儿童死亡率回归 (7.6.2)，读者可以验证，$\bar{R}^2$ 为 0.698 1，记住此例中 $(n-1)=63$，$(n-k)=60$。恰如所料，$\bar{R}^2=0.698\ 1$小

于 $R^2=0.7077$。

除了用 R^2 和校正 R^2 作为拟合优度的度量外，还有其他准则（或判据），也常被用来判断一个回归模型的适用性。其中的两个是**赤池**（Akaike）**的信息准则**和**雨宫**（Amemiya）**的预测准则**，用以挑选相互媲美的模型。当我们在以后的一章（第 13 章）考虑模型选择的问题时，我们将详细讨论这些准则。

比较两个 R^2 值

根据判定系数，不管是用校正的还是未经校正的系数来评比两个模型，一定要注意**样本大小 n 和因变量**都必须相同，而解释变量则可取任何形式。因此，对模型：

$$\ln Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i \tag{7.8.6}$$

$$Y_i=\alpha_1+\alpha_2X_{2i}+\alpha_3X_{3i}+u_i \tag{7.8.7}$$

计算的两个 R^2 项是不可比较的。理由如下：按定义，R^2 度量着因变量的变异被（诸）解释变量解释了的部分（占 Y 总变异的比例）。因此，在(7.8.6) 中，R^2 度量着由 X_2 和 X_3 解释的 $\ln Y$ 的变异部分，而在（7.8.7）中，则度量着被解释的 Y 的变异部分，两者不是同一回事。如第 6 章中所指出的，$\ln Y$ 的变化给出 Y 的相对或比例变化，而 Y 的变化则指 Y 的绝对变化。所以 $\mathrm{var}\hat{Y}_i/\mathrm{var}Y_i$ 不等于 $\mathrm{var}(\widehat{\ln Y_i})/\mathrm{var}(\ln Y_i)$。就是说，两个判定系数是不同尺度的。[13]

在回归子的形式不同的两个模型中，如何比较其 R^2 呢？为回答这个问题，让我们首先考虑一个数字例子。

例 7.2：美国 1970—1980 年的咖啡消费

220 考虑表 7.1 中的数据。表中是有关美国 1970—1980 年日均咖啡消费量（Y）和真实零售价格（X）的数据。对这些数据应用 OLS 法，我们得到如下回归结果：

$$\begin{aligned}\hat{Y}_t &= 2.6911-0.4795X_t\\ \mathrm{se}&=(0.1216)(0.1140)\\ \mathrm{RSS}&=0.1491; \qquad r^2=0.6628\end{aligned} \tag{7.8.8}$$

这个结果的经济含义是：随着咖啡价格的上涨，日均咖啡消费量平均下降约半杯。约等于 0.66 的 r^2 值意味着，咖啡价格大约能解释咖啡消费量变化的 66%。读者很容易验证，这个方程的斜率系数是统计上显著的。

表 7.1 美国 1970—1980 年咖啡消费（Y）与平均真实零售价格（X）的关系*

年份	Y（人均日消费杯数）	X（美元/磅）
1970	2.57	0.77
1971	2.50	0.74
1972	2.35	0.72
1973	2.30	0.73
1974	2.25	0.76
1975	2.20	0.75
1976	2.11	1.08
1977	1.94	1.81
1978	1.97	1.39
1979	2.06	1.20
1980	2.02	1.17

*注：名义价格除以食品与饮料的消费者价格指数，1967 年 = 100。

资料来源：Y 数据取自 *Summary of National Coffee Drinking Study*, Data Group, Elkins Park, Penn., 1981；名义 X 数据（即 X 的当时价格）取自 *Nielsen Food Index*, A.C.Nielsen, New York, 1981。

感谢斯科特·桑德伯格（Scott E. Sandberg）为我搜集到这些数据。

利用同样的数据，也可以估计出如下双对数或常弹性模型：

$$
\begin{aligned}
\ln\widehat{Y}_t &= 0.7774 - 0.2530\ln X_t \\
\text{se} &= (0.0152)(0.0494) \qquad\qquad (7.8.9)\\
\text{RSS} &= 0.0226; \qquad\qquad r^2 = 0.7448
\end{aligned}
$$

由于这是一个双对数模型，所以斜率系数就直接给出了价格弹性系数的一个估计值。在目前的例子中，它告诉我们，若每磅咖啡的价格上涨 1%，则日咖啡消费量平均下降约 0.25 个百分点。记住，在线性模型（7.8.8）中，斜率系数只给出了咖啡消费量相对价格的变化率。（你如何在线性模型中估计价格弹性呢？）约等于 0.74 的 r^2 值意味着，咖啡价格的对数变化大约能解释咖啡消费量的对数变化的 74%。读者很容易验证，这个方程的斜率系数是统计上显著的。

既然线性模型的 r^2 值 0.6628 比对数线性模型的 r^2 值 0.7448 小，那你可能禁不住要选择 r^2 值高的后面一个模型。但出于前面提到的原因，我们不能这么做。若你非要对这两个 r^2 值进行比较，那你可以采取如下步骤：

221 1. 从（7.8.9）中计算出每个观测的 $\widehat{\ln Y_t}$；即从此模型中得到每个观测值的对数估计值。取其反对数，然后按照方程（3.5.14）所指明的方法计算

$\widehat{\ln Y_t}$的反对数与实际 Y_t 值之间的 r^2。这个 r^2 值和得自（7.8.8）的 r^2 值才是可比的。

2. 或者，假设所有的 Y 值都为正，则取其对数便得到 $\ln Y$。从线性模型（7.8.8）中得到 Y_t 的估计值（即 $\hat{Y}_t$），然后取这些估计值的对数（即得到 $\ln \hat{Y}_t$），并按照方程（3.5.14）所指明的方法计算（$\ln Y_t$）与（$\ln \hat{Y}_t$之间的 r^2 值。这个 r^2 值与从（7.8.9）得到的 r^2 值是可比的。

对我们的咖啡需求一例而言，我们在表 7.2 中给出了计算这些可比较的 r^2 值所必须的原始数据。为了将线性模型（7.8.8）中的 r^2 值与（7.8.9）的 r^2 值相比较，我们首先得到 $\hat{Y}_t$的对数［在表 7.2 中第（6）列给出］，然后得到实际 Y 值的对数［表中第（5）列给出］，再利用方程（3.5.14）计算这两组数值之间的 r^2。计算出来的 r^2 值为 0.731 8，现在便可与对数线性模型的 r^2 值 0.744 8 相比较了。就此看来，这两个 r^2 值之间的差别很小。

表 7.2　　用于比较两个 R^2 值的原始数据

年份	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	$\widehat{\ln Y_t}$的反对数 (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321 887	0.843 555	2.324 616	0.943 906	0.842 380
1971	2.50	2.336 272	0.853 611	2.348 111	0.916 291	0.848 557
1972	2.35	2.345 863	0.860 544	2.364 447	0.854 415	0.852 653
1973	2.30	2.341 068	0.857 054	2.356 209	0.832 909	0.850 607
1974	2.25	2.326 682	0.846 863	2.332 318	0.810 930	0.844 443
1975	2.20	2.331 477	0.850 214	2.340 149	0.788 457	0.846 502
1976	2.11	2.173 233	0.757 943	2.133 882	0.746 688	0.776 216
1977	1.94	1.823 176	0.627 279	1.872 508	0.662 688	0.600 580
1978	1.97	2.024 579	0.694 089	2.001 884	0.678 034	0.705 362
1979	2.06	2.115 689	0.731 282	2.077 742	0.722 706	0.749 381
1980	2.02	2.130 075	0.737 688	2.091 096	0.703 098	0.756 157

注：第(1)列：来自表 7.1 的实际 Y 值；
第(2)列：来自线性模型(7.8.8)的估计 Y 值；
第(3)列：来自双对数模型(7.8.9)的估计 Y 值；
第(4)列：列(3)中的值的反对数；
第(5)列：列(1)中的 Y 的对数值；
第(6)列：列(2)中的 $\hat{Y}_t$ 的对数值。

另一方面，如果我们想将对数线性模型的 r^2 值与线性模型中的 r^2 值相

比较，那么，我们首先从（7.8.9）计算出每个观测的$\widehat{\ln Y_i}$［表中第（3）列给出］，再求出其反对数值［表中第（4）列给出］，然后按照公式（3.5.14）计算这些反对数值与实际 Y 值之间的 r^2。计算出来的 r^2 值为 0.718 7，比从线性模型（7.8.8）中所得到的 r^2 值0.662 8略高一些。

如此看来，无论用哪一种方法，对数线性模型拟合得总是略好一些。

在回归元之间分配 R^2

222 让我们回到儿童死亡率一例。我们在（7.6.2）中看到，PGNP 和 FLR 两个回归元解释了儿童死亡率变异中的 0.707 7 或 70.77%。但现在我们再考虑去掉 FLR 变量的回归（7.7.2），r^2 值下降到 0.166 2。这是否意味着 r^2 值的差值 0.541 5（0.707 7 − 0.166 2）都是因为去掉的变量 FLR 呢？另一方面，如果考虑去掉 PGNP 变量的回归（7.7.3），r^2 值下降到 0.669 6。这是否又意味着 r^2 值的差值 0.038 1（0.707 7 − 0.669 6）是源于省略变量 PGNP 呢？

于是问题就来了：我们是否能够如此将多元回归的 R^2 值 0.707 7 在 PGNP 和 FLR 这两个回归元之间分配？不幸的是，我们不能这么做，因为正如我们刚才说明的那样，这种分配取决于我们引入回归元的顺序。这里的部分问题在于这两个回归元的相关关系，其相关系数为 0.268 5（用表 6.4 中的数据来验证）。在大多数含有多个回归元的应用研究中，回归元之间的相关都是一个常见问题。当然，若回归元之间存在完全共线性，则问题就很严重了。

最好的实践忠告就是，试图将 R^2 值在其包含的回归元中进行分配会有些问题。

关于 R^2 最大化的"游戏"

在结束本节讨论时，提出一个警告是合适的：有时一些研究者玩弄 $\bar{R}^2$ 最大化的游戏。这是说，要选择有最高 $\bar{R}^2$ 值的模型。但这样做可能是危险的。因为，在回归分析中，我们的目的并不是为了得到一个高的 R^2，而是要得到真实总体回归系数的可靠估计并做出有关的统计推断。在经验分析中，得到一个很高的 $\bar{R}^2$，然而发现某些回归系数或者统计上不显著，或者取与先验预期相反的符号的情形屡见不鲜。故研究者应更关心解释变量对因变量的逻辑或理论关系及统计显著性。如果在这样的研究过程中，我们得到了一个高的 $\bar{R}^2$，自然很好；另一方面，如果 $\bar{R}^2$ 偏低，模型也未必是坏的。[14]

223 实际上，戈德伯格曾对 R^2 的作用有过告诫：

> 我们看来，在回归分析中，R^2 作为从一堆数据里拟合一个样本最小二乘回归的优度，只起到一种很平常的作用。在经典回归（经典线性回归）模型中没有哪一点要求 R^2 须是高的。因而，一个高的 R^2 不是肯定模型的证据，而一个低的 R^2 也不是否定模型的证据。
>
> 事实上，关于 R^2 的最重要的事情是，它在经典回归模型中是不重要的。CR 模型是用来研究一个总体中的参数的，它不问在一个样本中拟合的好坏，……如果人们坚持要有对预测成功（而不是失败）有一个度量，那么有了 σ^2 也许就足够了。毕竟，如果用总体回归函数作为预测元的话，参数 σ^2 就是期望预测误差的平方。换句话说，预测的标准误差的平方……也许对适当取定的 x（回归元）值来说，是富有信息(informative) 的。[15]

§7.9 例 7.3 柯布 道格拉斯生产函数：函数形式再议

在 6.4 节中，我们说明怎样能通过适当的变量代换把非线性关系式转换为线性，以便在经典线性回归模型的框架内考虑问题。当时相对于双变量情形而讨论的各种变换，是能够容易地推广应用到多元回归模型上来的。本节拟通过双变量对数线性模型的多变量推广，来讲解变量代换；其他的函数形式则散见于各章习题以及本书其余部分所讨论的说明性例子中。我们即将讨论的特殊例子，是生产理论中著名的**柯布-道格拉斯**（Cobb-Douglas）**生产函数**。

随机形式的柯布-道格拉斯生产函数可表达为：

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \tag{7.9.1}$$

其中 Y = 产出

X_2 = 劳动投入

X_3 = 资本投入

u = 随机干扰项

e = 自然对数的底

显然方程（7.9.1）给出的产出与两种投入之间的关系式是非线性的。然而，通过模型的对数变换，可得：

224

$$\begin{aligned} \ln Y_i &= \ln\beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \end{aligned} \tag{7.9.2}$$

其中 $\beta_0 = \ln\beta_1$。

这样书写的模型对参数 β_0，β_2 和 β_3 是线性的，因而是一个线性回归模型。注意，虽然如此，它对变量 Y 和 X 为非线性而对这些变量的对数为线性。简言之，(7.9.2) 是**一个对数—对数，双对数或对数线性模型**。它是双

变量对数线性模型（6.5.3）的多元回归对等物。

柯布-道格拉斯生产函数的性质是为人熟知的：

1. β_2 是产出对劳动投入的（偏）弹性，即它度量着在资本投入保持不变下劳动投入变化 1% 时的产出百分比变化（见习题 7.9）。

2. 同样，β_3 是在劳动投入保持不变下产出对资本投入的（偏）弹性。

3. 总和（$\beta_2+\beta_3$）给出关于规模报酬（returns to scale）的信息，就是产出对投入的比例变化的反应。如果此总和为 1，则规模报酬不变（constant returns to scale），即 2 倍的投入将带来 2 倍的产出，3 倍的投入将带来 3 倍的产出，等等。如果总和小于 1，则规模报酬递减——2 倍的投入将带来少于 2 倍的产出。最后，如果总和大于 1，则规模报酬递增——2 倍的投入将带来多于 2 倍的产出。

在继续往下讲之前，应看到无论你的对数线性回归模型涉及多少个 X 变量，每个 X 变量的系数代表因变量对该变量的（偏）弹性。例如，如果你有一个 k 变量对数线性模型：

$$\ln Y_i=\beta_0+\beta_2\ln X_{2i}+\beta_3\ln X_{3i}+\cdots+\beta_k\ln X_{ki}+u_i \tag{7.9.3}$$

则从 β_2 到 β_k 的每个（偏）回归系数，都是 Y 对从 X_2 到 X_k 变量的（偏）弹性。[16]

为了说明柯布-道格拉斯生产函数，我们收集了表 7.3 中的数据；这些数据反映了 1958—1972 年的台湾地区农业部门经济。

225 **表 7.3　1958—1972 年台湾地区农业部门的实际总产值、劳动日和实际资本投入**

年份	实际总产值 Y（百万元新台币）	劳动日 X_2（百万日）	实际资本投入 X_3（百万元新台币）
1958	16 607.7	275.5	17 803.7
1959	17 511.3	274.4	18 096.8
1960	20 171.2	269.7	18 271.8
1961	20 932.9	267.0	19 167.3
1962	20 406.0	267.8	19 647.6
1963	20 831.6	275.0	20 803.5
1964	24 806.3	283.0	22 076.6
1965	26 465.8	300.7	23 445.2
1966	27 403.0	307.5	24 939.0
1967	28 628.7	303.7	26 713.7
1968	29 904.5	304.7	29 957.8
1969	27 508.2	298.6	31 585.9

1970	29 035.5	295.5	33 474.5
1971	29 281.5	299.0	34 821.8
1972	31 535.8	288.1	41 794.3

资料来源：Thomas Pei-Fan Chen, "Economic Growth and Structural Change in Taiwan—1952—1972. A Production Function Approach," unpublished Ph.D.thesis, Dept. of Economics, Graduate Center, City University of New York, June 1976, Table Ⅱ.

假定模型(7.9.2)满足经典线性回归模型。[17]用OLS法得到如下回归（计算机打印结果见附录7A,第7A.5节）：

$$
\begin{aligned}
\widehat{\ln Y_i} &= -3.338\,4 + 1.498\,8\ln X_{2i} + 0.489\,9\ln X_{3i} \\
&\quad\ (2.449\,5) \quad (0.539\,8) \qquad (0.102\,0) \\
t &= (-1.362\,9) \quad (2.776\,5) \qquad (4.800\,5) \\
R^2 &= 0.889\,0 \qquad\qquad \mathrm{df}=12 \\
\bar{R}^2 &= 0.870\,5
\end{aligned}
\tag{7.9.4}
$$

从方程（7.9.4）我们看到1958—1972年台湾地区农业部门产出的劳动和资本弹性分别是1.498 8和0.489 9。换言之，在研究时期，保持资本投入不变，劳动投入增加1%，导致产出平均增加约1.5%。类似地，保持劳动投入不变，资本投入增加1%导致产出平均增加约0.5%。把两个产出弹性相加得1.988 7，就是规模报酬参数的取值。看得出来，在此研究期间，台湾地区农业部门经历着递增的规模报酬。[18]

226 从纯粹的统计观点看，所估计的回归线对数据的拟合相当良好。R^2取值0.889 0，表示产出（的对数）的变异的89%都可由劳动和资本（的对数）来解释。在第8章中，我们将看到怎样能用估计的标准误去检验有关台湾地区经济的柯布-道格拉斯生产函数中的参数"真"值的假设。

§7.10 多项式回归模型

作为本章的最后一部分，我们现在考虑一类多元回归模型，就是**多项式回归模型**（polynomial regression models）。这类模型在有关成本和生产函数的计量经济研究中有广泛的用途。在这些模型的引进当中，我们进一步扩大经典线性回归模型的适用范围。

为了明确概念，考虑图7.1。该图描述生产一种商品的生产（Y）的短期边际成本（MC）和它的产出水平（X）的关系。图中随手画出的、教科书般的U形MC曲线，表明了MC和产出之间的关系是非线性的。如果我们要把这种得自给定散点图的关系加以量化，怎么办？换句话说，什么类型的计量经济模型能抓住边际成本先降后升的性质？

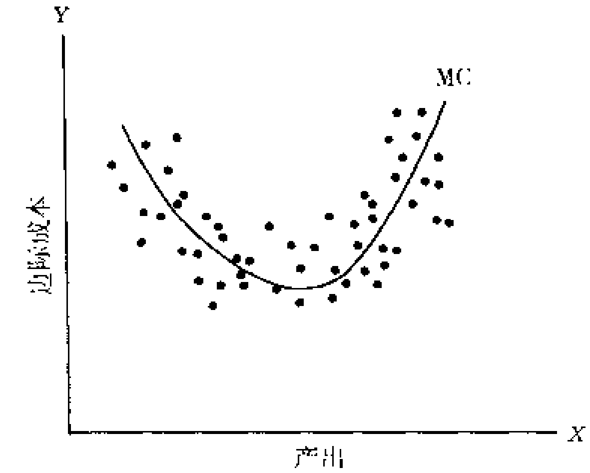


图 7.1　U型边际成本曲线

从几何上看，图 7.1 描绘的 MC 曲线代表一条抛物线。在数学上，抛物线的表达式是

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \tag{7.10.1}$$

它被称为**二次函数**或更一般地称为变量 X 的**二次多项式**——X 的最高乘方代表多项式的次数（如果在上述函数中加进 X^3 项，就称为三次多项式，如此类推）。

227 （7.10.1）的随机形式可写为：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \tag{7.10.2}$$

此即二阶多项式回归。

k 阶多项式回归可写成

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i \tag{7.10.3}$$

注意，在这类多项式回归中，方程右边只有一个解释变量，但以不同乘方出现，从而使方程成为多元回归模型。顺便提一下，如果 X 被假定为固定的或非随机的，那么带有乘方的各 X_i 项也都变成固定的或非随机的。

这种模型会提出什么特殊的估计问题吗？由于二次多项式（7.10.2）或 k 次多项式（7.10.3）对参数 β 而言都是线性的，故可用普通最小二乘法或最大似然法去估计。但会有什么共线性的问题吗？既然各个 X 项都是 X 的幂函数，它们会不会高度相关？不错，但应记住，像 X^2、X^3、X^4 等项都是 X 的非线性函数，所以严格地说，并不违反无多重共线性假定。总之，多项式回归模型没有提出任何新的估计问题，可用本章讲过的方法去估计它们。

例 7.4　估计总成本函数

作为多项式回归的一个例子。考虑表 7.4 给出的短期内某商品产出及其

生产总成本数据。什么类型的回归模型能拟合这些数据呢？为此，我们先做出散点图，如图7.2所示。

表7.4　　总成本（Y）与产出（X）

产出	总成本（美元）
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

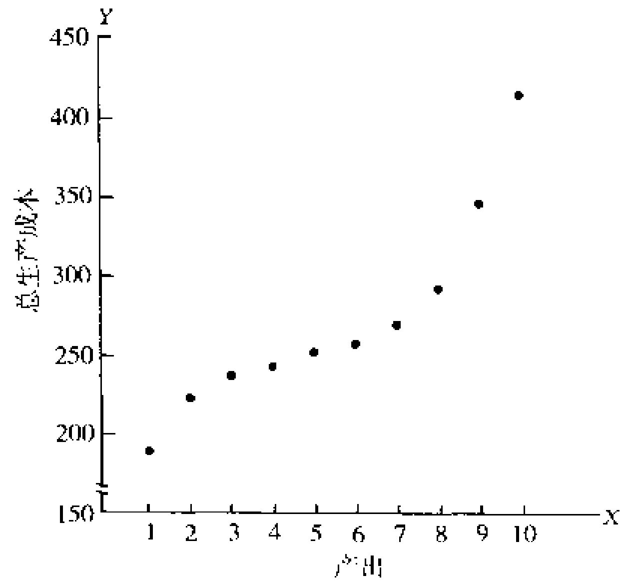


图7.2　总成本曲线

由图显见，总成本与产出之间的关系像一条拉长的S曲线；注意这条总成本曲线是怎样如同著名的**递减报酬律**（law of diminishing returns）所描述的，先是缓慢，然后急剧上升。总成本曲线的S形状可以由下面的立方或**三次多项式**来刻画：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \tag{7.10.4}$$

其中 Y = 总成本，而 X = 产出。

228 给定表 7.4 的数据，可用 OLS 法估计（7.10.4）中的参数。在做估计前，先看看经济理论是怎样描述短期立方成本函数（7.10.4）的。基本价格理论表明，在短期里，典型地说，生产的边际成本（MC）和平均成本（AC）都是 U 型的——开始时，随着产出增加，MC 和 AC 都下降，但到了一定产出水平之后，两者均转而升高，再度显示递减报酬律的后果。这可以从图 7.3（还有图 7.1）看出。而由于 MC 和 AC 曲线可以从总成本曲线导出，故这些曲线的 U 形性质给总成本曲线（7.10.4）的参数添加了一些约束。事实上，可以证明，如果短期边际和平均成本曲线遵循 U 形的话，（7.10.4）的参数必须满足如下约束条件[19]：

$$\begin{aligned} &1. \beta_0,\ \beta_1, \text{ 和 } \beta_3 > 0 \\ &2. \beta_2 < 0 \\ &3. \beta_2^2 < 3\beta_1\beta_3 \end{aligned} \tag{7.10.5}$$

所有这些理论探讨也许看来有些令人厌烦，但这种知识在我们分析经验结果时却非常有用，如果经验结果与事先的理论预期不符，那么，且假定我们的模型没有设定误差（即没有选用错误的模型），我们就必须修改我们的理论或寻求新的理论，然后重新开始我们的经验研究。但如同引言中所提的，这是任何经验调查的共性。

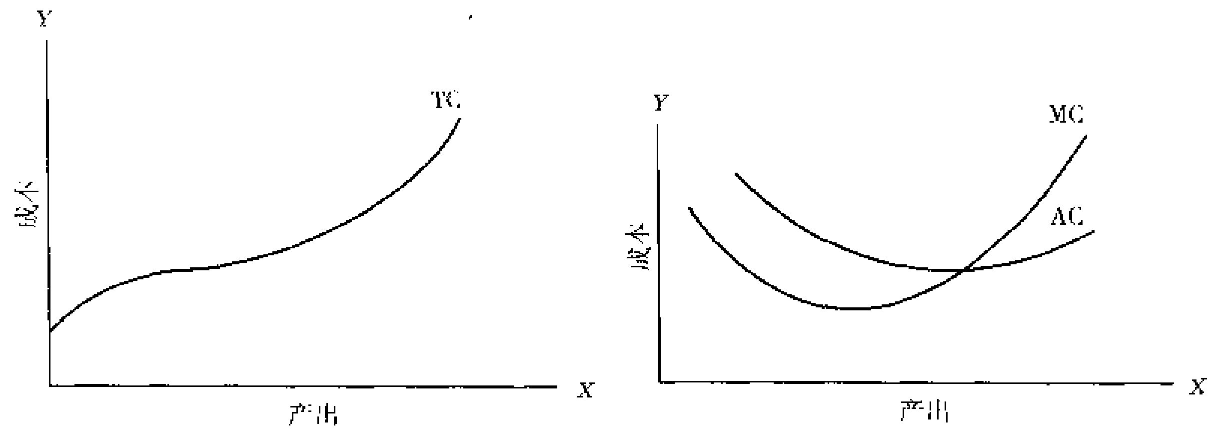


图 7.3　短期成本函数

经验结果

用三次多项式拟合表 7.4 的数据，得到如下结果：

$$\begin{aligned} \hat{Y}_i = 141.766\,7 &+ 63.477\,6X_i - 12.961\,5X_i^2 + 0.939\,6X_i^3 \\ (6.375\,3)\quad &(4.778\,6)\quad (0.985\,7)\quad (0.059\,1) \\ &R^2 = 0.998\,3 \end{aligned} \tag{7.10.6}$$

（注：括号中的数字是估计的标准误。）虽然在下章里我们将分析这些结果的统计显著性，但读者能证实它们和（7.10.5）所列举的理论预期是一致的。至于怎样解释回归（7.10.6），我们把它作为习题留给读者。

例 7.5　119 个发展中国家 1960—1985 年的 GDP 增长率与相对人均 GDP

作为对多项式回归模型附加的一个经济例子，考虑如下回归结果[20]：

$$
\begin{aligned}
\widehat{\text{GDPG}}_i &= 0.013 + 0.062\text{RGDP} - 0.061\text{RGDP}^2 \\
\text{se} &= (0.004) \quad (0.027) \qquad (0.033) \qquad\qquad (7.10.7) \\
& R^2 = 0.053 \quad 校正\ R^2 = 0.036
\end{aligned}
$$

其中 GDPG 表示 GDP 的百分比增长率（1960—1985 年间的平均增长率），RGDP 表示 1960 年的相对人均 GDP（即占美国 1960 年人均 GDP 的百分比）。校正后的 R^2 告诉我们，在考虑了回归元的个数之后，该模型也只解释了 GDPG 变动的 3.6%。即便从未经校正的 R^2 来看，0.053 也很小。这个值听起来让人气馁，但我们在下一章将会看到，在一个含有大量观测的横截面数据中，经常会遇到这么低的 R^2 值。而且，我们在下一章将会证明，即便一个明显很低的 R^2 值也可能是统计上显著的（即异于零）。

这个回归表明，发展中国家的 GDPG 随着 RGDP 的提高而递增，但增加的速度递减；发展中国家无法赶上发达国家。[21]此例表明了如何用相当简单的计量模型来解释重要的经济现象。

*§7.11　偏相关系数

简单与偏相关系数的释义

230 在第 3 章中，我们介绍了作为度量两个变量之间的线性关联程度的相关系数 r。对于三变量回归模型，我们可以算出三个相关系数：r_{12}（Y 与 X_2 之间的相关），r_{13}（Y 与 X_3 的相关）和 r_{23}（X_2 与 X_3 的相关）；注意，为了记号上的便利，我们让下标 1 代表 Y。这些相关系数可称毛或**简单相关系数**，或称**零阶相关系数**（correlation coefficients of zero order）。这些系数都能按（3.5.13）所给相关系数的定义计算出来。

但考虑下述问题：比方说，如果有第三个变量 X_3 与 Y 和 X_2 都可能有关联的话，r_{12} 果真度量了 Y 与 X_2 之间的"真实"（线性）关联度吗？这个问题又可类比于下述问题：假使真实的回归模型是（7.1.1），但我们从模型中略去了变量 X_3，仅做了 Y 对 X_2 的回归，并得到斜率系数 b_{12}（比方说）。问这个系数是否等于我们一开始就要估计的模型（7.1.1）的真实系数 β_2

呢？从我们在 7.7 节中的讨论看，答案是明显的。一般来说，r_{12}在 X_3 出现的情形下不大可能反映 Y 和 X_2 之间的真实关联度。事实上，我们即将看到，它容易给出 Y 与 X_2 之间的关联性质的一个错误印象。因此，我们还需要有一个不依赖于 X_3 的对 X_2 和 Y 的影响（如果有这个影响的话）的一种相关系数。这种系数就是我们所称的**偏相关系数**（partial correlation coefficient），并且可以计算出来。它类似于偏回归系数。让我们定义：

$r_{12.3}$ = X_3 保持不变下的 Y 和 X_2 的偏相关系数

$r_{13.2}$ = X_2 保持不变下的 Y 和 X_3 的偏相关系数

$r_{23.1}$ = Y 保持不变下的 X_2 和 X_3 的偏相关系数

从简单或零阶相关系数很容易就能计算出这些偏回归系数如下（证明见习题）[12]：

$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \tag{7.11.1}$$

$$r_{13.2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} \tag{7.11.2}$$

$$r_{23.1}=\frac{r_{23}-r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} \tag{7.11.3}$$

231 由方程（7.11.1）到（7.11.3）给出的偏相关称**一阶相关系数**（first-order correlation coefficients），这里说的阶是指次段下标（secondary subscripts）的个数。例如，$r_{12.34}$是二阶相关系数，$r_{12.345}$是三阶相关系数，等等。如前面所指出的，r_{12}，r_{13}等等称为简单或零阶相关。$r_{12.34}$（比方说）的含义是在 X_3 和 X_4 保持不变下的 Y 和 X_2 之间的相关系数。

简单与偏相关系数的解说

在双变量情形中，简单相关 r 有明显的含义：它度量着因变量 Y 与单一解释变量 X 之间的线性关联（而非因果关系）程度。但一旦我们越出双变量情形，我们对简单相关系数的解释就需要细心一些。例如，从（7.11.1）我们观察到：

1. 即使 $r_{12}=0$，$r_{12.3}$并不为零，除非 r_{13}或 r_{23}或两者为零。

2. 如果 $r_{12}=0$，而 r_{13}和 r_{23}均不为零且有相同符号，则 $r_{12.3}$为负，而如果 r_{13}和 r_{23}异号，则 $r_{12.3}$为正。举一个例子说明这点。令 Y = 农作物收成，X_2 = 雨量，及 X_3 = 气温。假定 $r_{12}=0$，即作物收成和雨量没有关联。再假定 r_{13}是正的，而 r_{23}是负的。这时（7.11.1）表明 $r_{12.3}$将是正的；就是说，在气温保持不变的情况下，收成和雨量有正的关联。然而这种看起来似乎矛盾的结论并不令人惊奇。因为气温 X_3 兼影响收成 Y 和雨量 X_2。要找出作物收成与雨量之间的净关系，需要除掉“讨厌”变量气温的影响。本例表明了人们怎样会被简单相关系数误导。

3. $r_{12.3}$项和 r_{12}项（及类似项的比较）不一定同号。

4. 在双变量情形中，r^2 落在 0 与 1 之间。偏相关系数的平方也有同样性质。利用这一事实，读者应能证实从（7.11.1）可推出下列表达式：

$$0\leqslant r_{12}^2+r_{13}^2+r_{23}^2-2r_{12}r_{13}r_{23}\leqslant 1 \qquad (7.11.4)$$

这给出了三个零阶相关系数的相互关系。同样的表达式可从方程（7.9.3）和（7.9.4）导出。

5. 假使 $r_{13}=r_{23}=0$，这是否意味着 r_{12}也等于零？答案可从（7.11.4）明显看出。Y 与 X_3 以及 X_2 与 X_3 不相关，并不意味着 Y 与 X_2 不相关。

232 顺便提一下，表达式 $r_{12.3}^2$可称为**偏判定系数**（coefficient of partial determination），并可解释为未被 X_3 解释的 Y 的变异部分由于 X_2 被引进到模型中来而得到解释的比例（参看习题 7.5）。它在概念上类似于 R^2。

在继续往下讲之前，请注意 R^2、简单相关系数以及偏相关系数之间有如下关系：

$$R^2=\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2} \qquad (7.11.5)$$

$$R^2=r_{12}^2+(1-r_{12}^2)r_{13.2}^2 \qquad (7.11.6)$$

$$R^2=r_{13}^2+(1-r_{13}^2)r_{12.3}^2 \qquad (7.11.7)$$

在结束本节讨论之际，请考虑：前面我们说过，如果在模型中多引进一个解释变量，R^2 必不会减小。这点可由方程（7.11.6）明显看出。该方程说，由 X_2 和 X_3 联合解释的 Y 的变异部分（比例）是两个部分之和：由 X_2 单独解释的部分（$=r_{12}^2$），以及未被 X_2 解释的部分（$=1-r_{12}^2$）乘以在 X_2 的影响保持不变下由 X_3 解释的比例。现在，只要 $r_{13.2}^2>0$，就有 $R^2>r_{12}^2$。$r_{13.2}^2$最小也不过是零，这时 $R^2=r_{12}^2$。

§7.12 要点与结论

1. 本章介绍最可能简单的多元（多变量）线性回归模型，即三变量回归模型。我们默认，“线性”一词指对参数为线性，对变量不一定为线性。

2. 虽然三变量回归模型在多个方面都是双变量模型的推广，却涉及一些新的概念，诸如偏回归系数，偏相关系数，多元相关系数，校正与未校正（对自由度）R^2，多重共线性和设定偏误。

3. 本章还考虑多元回归模型的函数形式，诸如柯布-道格拉斯生产函数和多项式回归模型。

4. 虽然 R^2 和校正 R^2 是所选模型对给定数据集拟合好坏的总度量，但它们的重要性不可过分夸大。最为关键的，是对进入模型的变量的系数、应带有什么先验性的符号，从而对这个模型，有一个基本的理论预期，以及下

章要讲的关于这些系数的统计显著性。

233 5. 本章所给出的结果，能够容易地推广到涉及任意多个回归元的复线性回归模型中，但代数运算会变得非常繁琐。如果用矩阵代数就可避免这种繁琐性。为了有兴趣的读者，在作为选读用的附录C中，我们用矩阵代数做出对 k 变量回归模型的推广。一般读者可越过它，继续阅读本书的其他部分。

习 题

问答题

7.1 考虑如下数据

Y	X_2	X_3
1	1	2
3	2	1
8	3	-3

根据这些数据估计以下回归：

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \tag{1}$$

$$Y_i = \lambda_1 + \lambda_3 X_{3i} + u_{2i} \tag{2}$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{3}$$

注：只估计系数，不用估计标准误。

a. $\alpha_2 = \beta_2$？为什么？

b. $\lambda_3 = \beta_3$？为什么？

你能从这道题引出什么重要的结论？

7.2 利用以下数据，估计偏回归系数、它们的标准误，以及校正与未校正的 R^2 值：

$$\bar{Y} = 367.693 \qquad X_2 = 402.760 \qquad \bar{X}_3 = 8.0$$

$$\sum (Y_i - \bar{Y})^2 = 66\ 042.269$$

$$\sum (X_{2i} - \bar{X}_2)^2 = 84\ 855.096$$

$$\sum (X_{3i} - \bar{X}_3)^2 = 280.000$$

$$\sum (Y_i - \bar{Y})(X_{2i} - \bar{X}_2) = 74\ 778.346$$

$$\sum (Y_i - \bar{Y})(X_{3i} - X_3) = 4\ 250.900$$

$$\sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) = 4\ 796.000$$

$$n = 15$$

7.3 说明方程（7.4.7）还可表达为：

$$\hat{\beta}_2 = \frac{\sum y_i(x_{2i} - b_{23}x_{3i})}{\sum (x_{2i} - b_{23}x_{3i})^2}$$

$$= \frac{y \text{ 与 } x_2 \text{ 的净(除去 } x_3\text{)协变}}{x_2 \text{ 的净(除去 } x_3\text{)变异}}$$

其中 b_{23}是 X_2 对 X_3 回归的斜率系数，见（7.3.2）。（提示：回忆 $b_{23} = \sum x_{2i}x_{3i} / \sum x_{3i}^2$。）

234 **7.4** 在一个多元回归模型中，告诉你误差项 u_i 服从如下概率分布，即 $u_i \sim N(0,4)$。你将如何构造一个蒙特卡罗实验来验证真实方差为 4?

7.5 证明 $r_{12.3}^2 = (R^2 - r_{13}^2)/(1 - r_{13}^2)$。并解释该方程。

7.6 如果关系式 $\alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 = 0$ 对 X_1，X_2 和 X_3 的所有值都成立，试求三个偏相关系数的值。

7.7 能从一组数据得到以下结果吗?

a. $r_{23} = 0.9$，　$r_{13} = -0.2$，　$r_{12} = 0.8$

b. $r_{12} = 0.6$，　$r_{23} = -0.9$，　$r_{31} = -0.5$

c. $r_{21} = 0.01$，　$r_{13} = 0.66$，　$r_{23} = -0.7$

7.8 考虑如下模型：

$$Y_i = \beta_1 + \beta_2 \text{ 教育}_i + \beta_2 \text{ 工作年限}_i + u_i$$

假设你漏掉了工作年限变量。我预计会出现什么类型的问题或偏误？并口头加以解释。

7.9 说明（7.9.2）中的 β_2 和 β_3 确实给出产出的劳动和资本弹性。(不用微积分也能回答此问题；只要回忆弹性系数的定义，并记住一个变量的对数的变化是一种相对变化，如果变化是比较小的话。)

7.10 考虑本章讨论的三变量线性回归模型。

a. 假设你将所有的 X_2 值都乘以 2。这种度量单位的改变对参数估计值及其标准误有什么影响（如果有的话)?

b. 现在假设与 a 不同，你将所有 Y 的值都乘以 2，对所估计的参数及其标准误又将有何影响（如果有的话)?

7.11 一般地说 $R^2 \neq r_{12}^2 + r_{13}^2$，但等式成立仅当 $r_{23} = 0$。试加评论并指出这一发现的意义。[提示：参看方程（7.11.5)。]

7.12 考虑以下模型[1]：

模型 A：$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_{1t}$

模型 B：$(Y_t - X_{2t}) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_{2t}$

a. α_1 和 β_1 的 OLS 估计会不会是一样的？为什么?

b. α_3 和 β_3 的 OLS 估计会不会是一样的？为什么？

c. α_2 和 β_2 有什么关系？

d. 你能比较两个模型的 R^2 项吗？为什么？

7.13 假使你估计消费函数[2]

$$Y_i = \alpha_1 + \alpha_2 X_i + u_{1i}$$

和储蓄函数

$$Z_i = \beta_1 + \beta_2 X_i + u_{2i}$$

其中 Y = 消费，Z = 储蓄，X = 收入，并且 $X = Y + Z$，即收入等于消费加储蓄。

235 a. α_2 和 β_2 有什么关系（如果有的话）？说明你的计算。

b. 两个模型的残差平方和 RSS 会不会是一样的？做出解释。

c. 你能比较两模型的 R^2 项吗？为什么或为什么不？

7.14 假使你把(7.9.1)中给出的柯布-道格拉斯模型表达成如下形式：

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i$$

如果你做这个模型的对数变换，你将在等式右边得到 $\ln u_i$ 作为干扰项。

a. 为了能应用经典正态线性回归模型的性质，你需要对 $\ln u_i$ 做些什么概率假设？你会怎样利用表 7.3 中的数据去检验这个假设。

b. 同样的假设也适用于 u_i 吗？为什么？

7.15 **过原点回归。**考虑以下过原点回归

$$Y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

a. 你打算怎样去估计这些未知数？

b. 对这个模型而言 $\sum \hat{u}_i$ 会是零吗？为什么？

c. 对这个模型会不会有 $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$？

d. 什么时候你会使用这样的模型？

e. 你能把你的结果推广到 k 变量模型吗？

（提示：仿照第 6 章对双变量情形的讨论。）

解答题

7.16 对玫瑰的需求。[3]表 7.6 给出如下变量的季度数据：

表 7.6

年与季	Y	X_2	X_3	X_4	X_5
1971 - Ⅲ	11 484	2.26	3.49	158.11	1
Ⅳ	9 348	2.54	2.85	173.36	2
1972 - Ⅰ	8 429	3.07	4.06	165.26	3

－Ⅱ	10 079	2.91	3.64	172.92	4
－Ⅲ	9 240	2.73	3.21	178.46	5
－Ⅳ	8 862	2.77	3.66	198.62	6
1973－Ⅰ	6 216	3.59	3.76	186.28	7
－Ⅱ	8 253	3.23	3.49	188.98	8
－Ⅲ	8 038	2.60	3.13	180.49	9
－Ⅳ	7 476	2.89	3.20	183.33	10
1974－Ⅰ	5 911	3.77	3.65	181.87	11
－Ⅱ	7 950	3.64	3.60	185.00	12
－Ⅲ	6 134	2.82	2.94	184.00	13
－Ⅳ	5 868	2.96	3.12	188.20	14
1975－Ⅰ	3 160	4.24	3.58	175.67	15
－Ⅱ	5 872	3.69	3.53	188.00	16

Y＝售出的玫瑰数量，单位：打

X_2＝玫瑰的平均批发价格，美元/打

X_3＝石竹的平均批发价格，美元/打

X_4＝每周家庭平均可支配收入，美元/周

X_5＝底特律（Detroit）市区从 1971 年第 3 季度到 1975 年第 2 季度的趋势变量，取值 1，2，等等。

请你考虑以下的需求函数：

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t$$
$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_t$$

a. 估计线性模型的参数并解释所得结果。

b. 估计对数线性模型的参数并解释计算结果。

c. β_2，β_3 和 β_4 分别给出需求的**自价格**（own-price）**弹性**，**交叉价格**（cross-price）**弹性和收入弹性**。它们的先验符号是什么？你的结果同先验预期相符吗？

236 d. 你怎样对线性模型计算自价格，交叉价格和收入三弹性？

e. 根据你的分析，你会选择哪个模型（如有可选的话）呢？为什么？

7.17 “野猫”活动（wildcat activity）。“野猫”是指为了在一个开发区发现和生产石油或天然气，或者为了在一个已发现的油田里寻找新的产油或产气池，或者为了提高一个已知的产油或产气池的最大产量而冒险钻打的井口。表 7.7 给出以下变量的数据。*

表 7.7

以千计的野猫数 (Y)	每桶价格（不变美元）(X_2)	国内产量（每天百万桶）(X_3)	GNP（10亿不变美元）(X_4)	时间 (X_5)
8.01	4.89	5.52	487.67	1948 = 1
9.06	4.83	5.05	490.59	1949 = 2
10.31	4.68	5.41	533.55	1950 = 3
11.76	4.42	6.16	576.57	1951 = 4
12.43	4.36	6.26	598.62	1952 = 5
13.31	4.55	6.34	621.77	1953 = 6
13.10	4.66	6.81	613.67	1954 = 7
14.94	4.54	7.15	654.80	1955 = 8
16.17	4.44	7.17	668.84	1956 = 9
14.71	4.75	6.71	681.02	1957 = 10
13.20	4.56	7.05	679.53	1958 = 11
13.19	4.29	7.04	720.53	1959 = 12
11.70	4.19	7.18	736.86	1960 = 13
10.99	4.17	7.33	755.34	1961 = 14
10.80	4.11	7.54	799.15	1962 = 15
10.66	4.04	7.61	830.70	1963 = 16
10.75	3.96	7.80	874.29	1964 = 17
9.47	3.85	8.30	925.86	1965 = 18
10.31	3.75	8.81	980.98	1966 = 19
8.88	3.69	8.66	1 007.72	1967 = 20
8.88	3.56	8.78	1 051.83	1968 = 21
9.70	3.56	9.18	1 078.76	1969 = 22
7.69	3.48	9.03	1 075.31	1970 = 23
6.92	3.53	9.00	1 107.48	1971 = 24
7.54	3.39	8.78	1 171.10	1972 = 25
7.47	3.68	8.38	1 234.97	1973 = 26
8.63	5.92	8.01	1 217.81	1974 = 27
9.21	6.03	7.78	1 202.36	1975 = 28
9.23	6.12	7.88	1 271.01	1976 = 29
9.96	6.05	7.88	1 332.67	1977 = 30
10.78	5.89	8.67	1 385.10	1978 = 31

资料来源：Energy Information Administration，1978 Report to Congress.

Y = 钻打（探）的井口（野猫）数

X_2 = 前期井地价格（1972 = 100，不变美元）

X_3 = 国内产量

X_4 = GNP，不变美元（1972 = 100）

X_5 = 趋势变量，1948 = 1，1949 = 2，…，1978 = 31

看看下面的模型对数据拟合得怎样：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

a. 你能对此模型做些合理的先验性预期吗？

b. 假定模型可以接受，估计模型的参数及其标准误，并求 R^2 和 R^2。

c. 按照你的先验预期的观点评论你的结果。

d. 为了解释野猫活动，你能提出其他的模型吗？理由是什么？

7.18 1962—1981 年美国国防预算支出。为了说明美国国防预算，请你考虑如下模型：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

其中 Y_t = 年度 t 的国防预算支出，10 亿美元计

X_{2t} = 年度 t 的 GNP，10 亿美元计

X_{3t} = 年度 t 的美国军事销售或（与）援助，10 亿美元计

X_{4t} = 太空工业销售，10 亿美元计

X_{5t} = 涉及多于 10 万军队的军事冲突。当军队为100 000或多于100 000人时，此变量取值 1；当军队人数小于100 000人时，它取值零

238 为了检验此模型，现在为你提供如下表 7.8 中的数据：

表 7.8

年份	国防预算支出 Y	GNP，X_2	美国军事销售/援助 X_3	太空工业销售 X_4	100 000 人及以上的冲突 X_5
1962	51.1	560.3	0.6	16.0	0
1963	52.3	590.5	0.9	16.4	0
1964	53.6	632.4	1.1	16.7	0
1965	49.6	684.9	1.4	17.0	1
1966	56.8	749.9	1.6	20.2	1
1967	70.1	793.9	1.0	23.4	1
1968	80.5	865.0	0.8	25.6	1
1969	81.2	931.4	1.5	24.6	1

1970	80.3	992.7	1.0	24.8	1
1971	77.7	1077.6	1.5	21.7	1
1972	78.3	1185.9	2.95	21.5	1
1973	74.5	1326.4	4.8	24.3	0
1974	77.8	1434.2	10.3	26.8	0
1975	85.6	1549.2	16.0	29.5	0
1976	89.4	1718.0	14.7	30.4	0
1977	97.5	1918.3	8.3	33.3	0
1978	105.2	2163.9	11.0	38.0	0
1979	117.7	2417.8	13.0	46.2	0
1980	135.9	2633.1	15.3	57.6	0
1981	162.1	2937.7	18.0	68.9	0

资料来源：数据是由 Albert Lucchino 从各种政府出版物收集到的。

a. 估计此模型的参数及其标准误并求 R^2，修正的 R^2 和 $\bar{R}^2$。

b. 评论所得结果，同时参照你对 Y 与各 X 变量之间的关系的任何先验预期。

c. 你还想把什么别的变量包括在这个模型中吗？什么理由？

7.19 1960—1982 年美国对子鸡的要求。为了研究美国每人的子鸡消费量，我们提供表 7.9 如下的数据：

239

表 7.9

年份	Y	X_2	X_3	X_4	X_5	X_6
1960	27.8	397.5	42.2	50.7	78.3	65.8
1961	29.9	413.3	38.1	52.0	79.2	66.9
1962	29.8	439.2	40.3	54.0	79.2	67.8
1963	30.8	459.7	39.5	55.3	79.2	69.6
1964	31.2	492.9	37.3	54.7	77.4	68.7
1965	33.3	528.6	38.1	63.7	80.2	73.6
1966	35.6	560.3	39.3	69.8	80.4	76.3
1967	36.4	624.6	37.8	65.9	83.9	77.2
1968	36.7	666.4	38.4	64.5	85.5	78.1
1969	38.4	717.8	40.1	70.0	93.7	84.7
1970	40.4	768.2	38.6	73.2	106.1	93.3
1971	40.3	843.3	39.8	67.8	104.8	89.7

1972	41.8	911.6	39.7	79.1	114.0	100.7
1973	40.4	931.1	52.1	95.4	124.1	113.5
1974	40.7	1 021.5	48.9	94.2	127.6	115.3
1975	40.1	1 165.9	58.3	123.5	142.9	136.7
1976	42.7	1 349.6	57.9	129.9	143.6	139.2
1977	44.1	1 449.4	56.5	117.6	139.2	132.0
1978	46.7	1 575.5	63.7	130.9	165.5	132.1
1979	50.6	1 759.1	61.6	129.8	203.3	154.4
1980	50.1	1 994.2	58.9	128.0	219.6	174.9
1981	51.7	2 258.1	66.4	141.0	221.6	180.8
1982	52.9	2 478.7	70.4	168.2	232.6	189.4

注：实际价格是用食品的消费者价格指数去除名义价格得到的。

资料来源：Y 数据来自城市数据库（Citibase）；从 X_2 到 X_6 的数据来自美国农业部。感谢罗伯特·费希尔对这些数据的收集和统计分析。

其中 Y = 每人的子鸡消费量，磅

X_2 = 每人实际可支配收入，美元

X_3 = 每磅子鸡实际零售价格，美分

X_4 = 每磅猪肉实际零售价格，美分

X_5 = 每磅牛肉实际零售价格，美分

X_6 = 子鸡替代品每磅综合实际价格，美分。这是猪肉和牛肉每磅实际零售价格的加权平均。其权数是在猪肉和牛肉的总消费量中两者各占的相对消费量。

现考虑下面的需求函数：

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \tag{1}$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \tag{2}$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \tag{3}$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \tag{4}$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \tag{5}$$

由微观经济学得知，对一种商品的需求通常较依赖于消费者的实际收入，该商品的实际价格，以及互替或互补商品的实际价格。按照这些思路，回答以下问题。

a. 从这里所列举的需求函数中你会选择哪一个并且为什么？

b. 你怎样解释这些模型中的 $\ln X_2$ 和 $\ln X_3$ 的系数？

c. 模型的设定（2）和（4）有什么不同？

d. 如果你采用设定（4），你会预见到什么问题？（提示：猪肉和牛肉价格都同子鸡价格一道被引进。）

240 e. 因为设定（5）含有牛肉和猪肉的综合价格，你会认为需求函数（5）优于函数（4）吗？为什么？

f. 猪肉和（或）牛肉是子鸡的竞争或替代产品吗？你怎样知道？

g. 假定函数(5)是"正确"的需求函数。估计此模型的参数。算出它们的标准误，以及 R^2，$\bar{R}^2$ 和修正的 R^2。解释你的结果。

h. 现假如你用了"不正确"的模型（2）。通过分别考虑 r_2 值和 r_3 值同 β_2 和 β_3 的关系，来评估这一错误设定的后果。（提示：注意第 7.7 节的讨论。）

7.20 在一项劳动市场的人员周转研究中，J.F. 拉根（James F. Ragan）对 1950 年第 1 季度至 1979 年第 4 季度期间的美国经济获得了如下结果。[15]（括号中的数字是估计的 t 统计量。）

$$\widehat{\ln Y_t} = 4.47 - 0.34\ln X_{2t} + 1.22\ln X_{3t} + 1.22\ln X_{4t}$$
$$\qquad (4.28) \qquad (-5.31) \qquad (3.64) \qquad (3.10)$$
$$+ 0.80\ln X_{5t} - 0.0055X_{6t} \quad \bar{R}^2 = 0.5370$$
$$(1.10) \qquad (-3.09)$$

注：我们将在下章中讨论 t 统计量。

其中 Y = 制造业中的辞职率，定义为每 100 个雇员中自愿离职的人数

X_2 = 成年男性失业率的工具变量或替代变量

X_3 = 年龄小于 25 岁的雇员百分比

$X_4 = N_{t-1}/N_{t-4}$ = 季度$_{t-1}$与季度$_{t-4}$的制造业就业比率

X_5 = 妇女雇员百分比

X_6 = 时间趋势（1950 - I = 1）

a. 解释上面列出的结果。

b. 所观测的对数 Y 和对数 X_2 之间的负的关系，在先验上是否说得过去？

c. 为什么 $\ln X_3$ 的系数是正的？

d. 既然趋势系数是负的，那么辞职率的百分比就有一长期下降趋势，而为什么有这样一种趋势？

e. $\bar{R}^2$ 是否"太"低？

f. 你能从所给数据估计回归系数的标准误吗？为什么或为什么不？

7.21 考虑美国1980—1998 年间的货币需求函数：

$$M_t = \beta_1 Y_t^{\beta_2} r_t^{\beta_3} e^{u_t}$$

241 其中 M = 真实货币需求，利用货币的 M_2 定义

Y = 真实 GDP

r = 利率

为估计上述货币需求函数，为你提供了表 7.10 中的数据。注意，为了把名义变量转换成真实变量，将 M 和 GDP 除以

CPI。利率变量则不必除以 CPI。另外还要注意，这里给出了两个利率，一个是以 3 个月国库券利率度量的短期利率，一个是以 30 年期国债的收益率度量的长期利率，前面的经验研究已经使用过这两个利率。

a. 给定这些数据，估计货币需求函数。货币需求对收入和利率的弹性是多少？

b. 如你不去拟合上述需求函数，而代之以对模型 $(M/Y)_t = \alpha_1 r_t^{\alpha_2} e^{u_t}$ 的拟合，你会怎样解释所得的结果？说明必需的计算。

c. 你如何决定哪个设定更好？（注：在第 8 章将给出一个规范的统计检验。）

表 7.10　　1980—1998 年美国的货币需求

观测	GDP	M2	CPI	LTRATE	TBRATE
1980	2 795.6	1 600.4	82.4	11.27	11.506
1981	3 131.3	1 756.1	90.9	13.45	14.029
1982	3 259.2	1 911.2	96.5	12.76	10.686
1983	3 534.9	2 127.8	99.6	11.18	8.630
1984	3 932.7	2 311.7	103.9	12.41	9.580
1985	4 213.0	2 497.4	107.6	10.79	7.480
1986	4 452.9	2 734.0	109.6	7.78	5.980
1987	4 742.5	2 832.8	113.6	8.59	5.820
1988	5 108.3	2 995.8	118.3	8.96	6.690
1989	5 489.1	3 159.9	124.0	8.45	8.120
1990	5 803.2	3 279.1	130.7	8.61	7.510
1991	5 986.2	3 379.8	136.2	8.14	5.420
1992	6 318.9	3 434.1	140.3	7.67	3.450
1993	6 642.3	3 478.5	144.5	6.59	3.020
1994	7 054.3	3 502.2	148.2	7.37	4.290
1995	7 400.5	3 649.3	152.4	6.88	5.510
1996	7 813.2	3 824.2	156.9	6.71	5.020
1997	8 300.8	4 046.7	160.5	6.61	5.070
1998	8 759.9	4 401.4	163.0	5.58	4.810

注：GDP＝国内生产总值（10 亿美元计）；
M2＝M2 货币供给；
CPI＝消费者价格指数（1982—1984＝100）；
LTRATE＝长期利率（30 年国债）；
TBRATE＝3 月期国库券利率（年百分率）。

资料来源：*Economic Report of the President*，2000，Tables B-1，B-58，B-67，B-71.

解答题

7.22 表 7.11 给出希腊 1961—1987 年制造业的数据。

表 7.11 希腊工业部门

观测	产出*	资本	劳动†	资本劳动比
1961	35.858	59.600	637.0	0.093 6
1962	37.504	64.200	643.2	0.099 8
1963	40.378	68.800	651.0	0.105 7
1964	46.147	75.500	685.7	0.110 1
1965	51.047	84.400	710.7	0.118 8
1966	53.871	91.800	724.3	0.126 7
1967	56.834	99.900	735.2	0.135 9
1968	65.439	109.100	760.3	0.143 5
1969	74.939	120.700	777.6	0.155 2
1970	80.976	132.000	780.8	0.169 1
1971	90.802	146.600	825.8	0.177 5
1972	101.955	162.700	864.1	0.188 3
1973	114.367	180.600	894.2	0.202 0
1974	101.823	197.100	891.2	0.221 2
1975	107.572	209.600	887.5	0.236 2
1976	117.600	221.900	892.3	0.248 7
1977	123.224	232.500	930.1	0.250 0
1978	130.971	243.500	969.9	0.251 1
1979	138.842	257.700	1 006.9	0.255 9
1980	135.486	274.400	1 020.9	0.268 8
1981	133.441	289.500	1 017.1	0.284 6
1982	130.388	301.900	1 016.1	0.297 1
1983	130.615	314.900	1 008.1	0.312 4
1984	132.244	327.700	985.1	0.332 7
1985	137.318	339.400	977.1	0.347 4
1986	137.468	349.492	1 007.2	0.347 0
1987	135.750	358.231	1 000.0	0.358 2

* 以 1970 年不变价格的 10 亿德拉克马计。

† 以每千人计。

资料来源：感谢弗吉尼亚州 Christopher Newport 大学的 George K. Zestos。

a. 看柯布-道格拉斯生产函数是否拟合表中给出的数据，并解释得到的结果。你能得到什么一般性结论？

b. 现考虑如下模型：产出/劳动 $= A(K/L)^{\beta}e^{u}$，其中回归子代表劳动生产率，回归元代表资本劳动比，这种关系有没有什么经济意义？估计此模型并解释其结果。

7.23 参考习题3.3和表2.6中给出的数据，现考虑如下模型：

a. $\ln(\text{hwage})_i = \beta_1 + \beta_2 \ln(\text{education})_i + \beta_3 (\ln \text{education})^2 + u_i$，
其中 ln = 自然对数。你如何解释此模型？估计这个模型，得出常见的统计量，并评论你的结果。

243 b. $\ln(\text{hwage}) = \beta_1 + \beta_2 \ln(\text{education}) + \beta_3 \ln(\text{education}^2) + u_i$，
如果你试图估计这个模型，你会遇到什么问题？尝试估计这个模型，并看你的软件包能否估计此模型。

7.24 蒙特卡罗实验：考虑如下模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

若告诉你 $\beta_1 = 262$，$\beta_2 = -0.006$，$\beta_3 = -2.4$，$\sigma^2 = 42$，$u_i \sim N(0, 42)$。从给定正态分布中生成10组64个观测 u_i 的集合，并利用表6.4中给出的64次观测（其中 $Y = \text{CM}$，$X_2 = \text{PGNP}$，$X_3 = \text{FLR}$）生成10组估计的 β 系数集（每个集都有三个估计系数）。求出每个估计 β 系数的平均值，并将它们与上面给出这些系数的真实值相联系，你能得到什么总体性结论？

【习题注释】

[1] 改定自 Wojciech W. Charemza and Derek F. Deadman, *Econometric Practice: General to Specific Modelling, Cointegration and Vector Autogression*, Edward Elgar, Brookfield, Vermont, 1992, p.18。

[2] 改定自 Peter Kennedy, *A Guide To Econometrics*, 3d ed., The MIT Press, Cambridge, Massachusetts, 1992, p.308, Question #9。

[3] 我感谢乔·沃尔什(Joe Walsh)从底特律市区的一个大批发商收集到这些数据并随后做了加工。

[4] 我感谢雷蒙德·萨文诺(Raymond Savino)收集并加工了这些数据。

[5] 资料来源：见拉根的论文："Turnover in the Labor Market: A Study of Quit and Layoff Rates," *Economic Review*, Federal Reserve Bank of Kansas City, May 1981, pp.13–22。

附录 7A

7A.1 方程(7.4.3)至(7.4.5)所给OLS估计量的推导

将方程

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \tag{7.4.2}$$

对三个未知数求偏导数，并令所得结果为零，得：

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-1) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0$$

简化后即得方程（7.4.3）至（7.4.5）。

顺便指出，以上三个方程又可写为：

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = 0 \qquad \text{（为什么？）}$$

$$\sum \hat{u}_i X_{3i} = 0$$

从而表明最小二乘拟合的一些性质，即残差和为零，以及残差与解释变量 X_2 和 X_3 均不相关。

244 还容易看到，按照类似方法可以得到 k 变量线性回归模型（7.4.20）的 OLS 估计量。为此，先写出：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

将此表达式对 k 个未知数的每一个求偏微分并令其结果等式为零，经适当整理后，即得 k 个未知数的 k 个正规方程如下：

$$\begin{aligned}
\sum Y_i &= n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki} \\
\sum Y_i X_{2i} &= \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i} X_{ki} \\
\sum Y_i X_{3i} &= \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i} X_{ki} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum Y_i X_{ki} &= \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{2i} X_{ki} + \hat{\beta}_3 \sum X_{3i} X_{ki} + \cdots + \hat{\beta}_k \sum X_{ki}^2
\end{aligned}$$

或者，转换到小写字母，将这些方程表达为：

$$\begin{aligned}
\sum y_i x_{2i} &= \hat{\beta}_2 \sum x_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} + \cdots + \hat{\beta}_k \sum x_{2i} x_{ki} \\
\sum y_i x_{3i} &= \hat{\beta}_2 \sum x_{2i} x_{3i} + \hat{\beta}_3 \sum x_{3i}^2 + \cdots + \hat{\beta}_k \sum x_{3i} x_{ki} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum y_i x_{ki} &= \hat{\beta}_2 \sum x_{2i} x_{ki} + \hat{\beta}_3 \sum x_{3i} x_{ki} + \cdots + \hat{\beta}_k \sum x_{ki}^2
\end{aligned}$$

还应看到，k 变量模型也满足如下方程：

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = \cdots = \sum \hat{u}_i X_{ki} = 0$$

7A.2 （7.3.5）和（7.6.2）中 PGNP 系数的相等性质

令 $Y = \text{CM}$，$X_2 = \text{PGNP}$，$X_3 = \text{FLR}$，并用离差形式写成

$$y_i = b_{13} x_{3i} + \hat{u}_{1i} \tag{1}$$

$$x_{2i} = b_{23} x_{3i} + \hat{u}_{2i} \tag{2}$$

现在将 $\hat{u}_1$ 对 $\hat{u}_2$ 回归得到

$$a_1 = \frac{\sum \hat{u}_{1i}\hat{u}_{2i}}{\hat{u}_{2i}^2} = -0.005\,6\ (\text{对本例}) \tag{3}$$

245 注意，由于 $\hat{u}$ 都是残差，所以它们的均值为零。利用（1）和（2），我们可以把（3）写成

$$a_1 = \frac{\sum (y_i - b_{13}x_{3i})(x_{2i} - b_{23}x_{3i})}{\sum (x_{2i} - b_{23}x_{3i})^2} \tag{4}$$

展开上述表达式，并注意到

$$b_{23} = \frac{\sum x_{2i}x_{3i}}{\sum x_{3i}^2} \tag{5}$$

和
$$b_{13} = \frac{\sum y_i x_{3i}}{\sum x_{3i}^2} \tag{6}$$

把它们代入(4)即得到

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i}x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i}x_{3i})^2}$$
$$= -0.005\,6\ (\text{对本例}) \tag{7.4.7}$$

7A.3 方程（7.4.19）的推导

回忆：

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

又可写为：

$$\hat{u}_i = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$$

其中小写字母照例指对均值的离差。

但

$$\sum \hat{u}_i^2 = \sum (\hat{u}_i\hat{u}_i)$$
$$= \sum \hat{u}_i(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})$$
$$= \sum \hat{u}_i y_i$$

这里利用了关系式$\sum \hat{u}_i x_{2i} = \sum \hat{u}_i x_{3i} = 0$。(为什么?) 又因,

$$\sum \hat{u}_i y_i = \sum y_i \hat{u}_i = \sum y_i(y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})$$

即有:

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

这就是所要的结果。

7A.4 多元回归模型的最大似然估计法

246 将第 4 章附录 4A 中的概念加以推广，即可写出 k 变量线性回归模型 (7.4.20) 的对数似然函数为:

$$\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2}{\sigma^2}$$

将此函数对 $\beta_1, \beta_2, \cdots, \beta_k$ 和 σ^2 求偏微分,便得以下的($k+1$) 个方程:

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-1) \tag{1}$$

$$\frac{\partial \ln L}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{2i}) \tag{2}$$

…………………………………………………………………

$$\frac{\partial \ln L}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{ki}) \tag{K}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2 \tag{K+1}$$

令这些偏导数为零(最优的一阶条件),并用 $\tilde{\beta}_1, \tilde{\beta}_2, \cdots, \tilde{\beta}_k$ 和 $\tilde{\sigma}^2$ 表示 ML 估计量,经代数简化后可得(前 k 个方程):

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_{2i} + \cdots + \tilde{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \tilde{\beta}_1 \sum X_{2i} + \tilde{\beta}_2 \sum X_{2i}^2 + \cdots + \tilde{\beta}_k \sum X_{2i} X_{ki}$$

…………………………………………………………………

$$\sum Y_i X_{ki} = \tilde{\beta}_1 \sum X_{ki} + \tilde{\beta}_2 \sum X_{2i} X_{ki} + \cdots + \tilde{\beta}_k \sum X_{ki}^2$$

如同我们在附录 7A，第 7A.1 节所看到的，这正是最小二乘理论所导出的正规方程。因此，ML 估计量 $\tilde{\beta}_i$ 就是前面已给出的 OLS 估计量 $\hat{\beta}_i$。如第 4 章附录 4A 所指出的，这两者的相等并非偶然。

把 ML（= OLS）估计量代入刚才推出的第（$K+1$）个方程，经过简

化，便得到 σ^2 的 ML 估计量为：

$$
\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{n}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} - \cdots - \tilde{\beta}_k X_{ki})^2 \\
&= \frac{1}{n}\sum \hat{u}_i^2
\end{aligned}
$$

如本书已指出的，此估计量不同于 OLS 估计量 $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-k)$，而由于后者是 σ^2 的无偏估计，故引出结论：ML 估计量 $\tilde{\sigma}^2$ 是一个有偏误的估计量。然而，容易验证，$\tilde{\sigma}^2$ 是渐近无偏的。

7A.5 柯布-道格拉斯生产函数 (7.9.4) 的 SAS 打印结果

247 因变量：Y1

来源	自由度	平方和	均方	F 值	$>F$ 的概率
模型	2	0.538 038	0.269 019	48.069	0.000 1
误差	12	0.067 153	0.005 596 531		
总计	14	0.605 196			
根方误		0.074 810	R^2	0.889 0	
应变均值		10.096 535	校正 R^2	0.870 5	
变异系数		0.740 946 9			

变量	自由度	参数估计值	标准误(差)	对 H_0：参数 = 0 的 t 的值	$>\|T\|$ 的概率
截距	1	−3.338 455	2.449 508	−1.363	0.197 9
Y2	1	1.498 767	0.539 803	2.777	0.016 8
Y3	1	0.489 858	0.102 043	4.800	0.000 4

估计的协方差

系数 β_i 的协方差	截距	Y2	Y3
截距	6.000 091	−1.260 56	0.112 195 1
Y2	−1.260 56	0.291 386 8	−0.038 427 2
Y3	0.011 219 51	−0.038 427 2	0.010 412 88

Y	X2	X3	Y1	Y2	Y3	Y1HAT	Y1RESID
16 607.7	275.5	17 803.7	9.717 6	5.618 59	9.787 2	9.876 8	−0.159 20
17 511.3	274.4	18 096.8	9.770 6	5.614 59	9.803 5	9.878 8	−0.108 22

20 171.2	269.7	18 271.8	9.912 0	5.597 31	9.813 1	9.857 6	0.054 37
20 932.9	267.0	19 167.3	9.949 1	5.587 25	9.861 0	9.866 0	0.083 07
20 406.0	267.8	19 647.6	9.923 6	5.590 24	9.885 7	9.882 6	0.040 97
20 831.6	275.0	20 803.5	9.944 2	5.616 77	9.942 9	9.950 4	-0.006 15
24 806.3	283.0	22 076.6	10.118 9	5.645 45	10.002 3	10.022 5	0.096 40
26 465.8	300.7	23 445.2	10.183 6	5.706 11	10.062 4	10.142 8	0.040 77
27 403.0	307.5	24 939.0	10.218 4	5.728 48	10.124 2	10.206 6	0.011 80
28 628.7	303.7	26 713.7	10.262 2	5.716 04	10.192 9	10.221 7	0.040 51
29 904.5	304.7	29 957.8	10.305 8	5.719 33	10.307 5	10.282 7	0.023 04
27 508.2	298.6	31 585.9	10.222 2	5.699 10	10.360 5	10.278 3	-0.056 10
29 035.5	295.5	33 474.5	10.276 3	5.688 67	10.418 5	10.291 1	-0.014 87
29 281.5	299.0	34 821.8	10.284 7	5.700 44	10.458 0	10.328 1	-0.043 41
31 535.8	288.1	41 794.3	10.358 9	5.66 331	10.640 5	10.361 9	-0.002 99

	共线性诊断		方差比例		
		分配	分配	分配	
编号	条件本征值	指数	截距	Y2	Y3
1	3.000	1.000	0.000 0	0.000 0	0.000 0
2	.000 375 451	89.383	0.049 1	0.006 9	0.595 9
3	.000 024 219	351.925	0.950 9	0.003 1	0.404 0
	德宾-沃森 d 值			0.891	
	1 阶自相关			0.366	

注：Y1 = ln Y；Y2 = ln X_2；Y3 = ln X_3。标题 PROB > |T| 下的数代表 p 值。关于共线性诊断的讨论，见第 10 章。

【注释】

[1] 为了符号上的对称，方程（7.1.1）也可写成：

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

但是，对一切 i 都有 $X_{1i} = 1$。

[2] 如果 X_2 和 X_3 是非随机的，并且（7.1.2）成立，这个假定就自动得到满足。

[3] 从数学的角度说，$\alpha = (\beta_2 + 2\beta_3)$ 是一个有两个未知数的方程，因此没有从 α 估计 β_2 和 β_3 的惟一方法。

[4] 习惯于用微积分思考的读者会立即看出，β_2 和 β_3 是 $E(Y|X_2, X_3)$ 对 X_2 和 X_3 的偏导数。

[5] 控制、保持不变、考虑到或解释……的影响，修正……的影响，和去掉……的影响等词句是同义语，并将在本书中交替地使用。

[6] 这个估计量等于（7.3.5）中的 a_1，参看附录7A，第7A.2节。

[7] 这些公式的推导利用矩阵符号较为容易，有基础的读者可参见附录C。

[8] 由第3章给出的 r 的定义，我们有：$r_{23}^2=\dfrac{(\sum x_{2i}x_{3i})^2}{\sum x_{2i}^2\sum x_{3i}^2}$。

[9] 注意，R^2 还可计算为 $R^2=1-\dfrac{\text{RSS}}{\text{TSS}}=1-\sum\hat{u}_i^2/\sum y_i^2=1-\dfrac{(n-3)\hat{\sigma}^2}{(n-1)S_y^2}$。

[10] 然而，要注意，如 $R^2=1$，则 $\bar{R}^2=R^2=1$。当 $R^2=0$ 时，$\bar{R}^2=(1-k)/(n-k)$。这时，如果 $k>1$，$\bar{R}^2$ 就可能是负的。

[11] Henri Theil, *Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1978, p.135.

[12] Arthur S.Goldberger, *A Course in Econometrics*, Harvard University Press, Cambridge, Mass., 1991, p. 178. 关于 R^2 更多的评判意见，见S.Cameron,"Why is the R Squared Adjusted Reported?", *Journal of Quantitative Economics*, vol.9, no.1, January 1993, pp.183－186。作者称辩："R^2 不是一个检验统计量而且没有任何清晰直观的理由把它用做描述统计量。最后，我们应该明白，它并非预防数据开采的有效手段。"(第186页)

[13] 由 R^2 的定义知，对于线性模型：

$$1-R^2=\frac{\text{RSS}}{\text{TSS}}=\frac{\sum\hat{u}_i^2}{\sum(Y_i-\bar{Y})^2}$$

对于对数模型：

$$1-R^2=\frac{\sum\hat{u}_i^2}{\sum(\ln Y_i-\overline{\ln Y}\,\widehat{\ln Y_i})^2}$$

由于这两个表达式右端的分母各不相同，我们不能直接比较其 R^2 项。

如例7.2所示，对于线性形式，RSS＝0.149 1（咖啡消费的残差平方和）。而对于对数线性形式，RSS＝0.022 6（对数咖啡消费的残差平方和）。这两种残差属于不同的数量级，因而不可直接比较。

[14] 一些作者意在抵消对 R^2 用于衡量拟合优度以及比较两个或多个 R^2 值的过分重视。参看 Christopher H.Achen, *Interpreting and Using Regression*, Sage Publications, Beverly Hills, Calif., 1982, pp.58－67, 以及 C.Granger and P.Newbold, "R^2 and the Transformation of Regression Variables," *Journal of Econometrics*, vol.4, 1976, pp.205－210。值得一提的是，根据最高的 $\bar{R}^2$ 挑选模型的做法，属于一种数据开采，导致了所谓的**预检偏误**（pretest bias）。这种偏误可能破坏了经典回归模型的OLS估计量的某些性质。关于这一论题，读者可参考George G. Judge, Carter R. Hill, William E. Griffiths, Helmut Lütkepohl and Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley, New York, 1982, Chap.21。

[15] Arther S.Goldberger, *op.cit.*, *pp*.177－178.

[16] 为了看到这点，将(7.9.3)对每一 X 变量的对数求偏微分。于是 $\partial\ln Y/\partial\ln X_2=(\partial Y/\partial X_2)(X_2/Y)=\beta_2$，根据定义，这就是 Y 对 X_2 的弹性；$\partial\ln Y/\partial\ln X_3=(\partial Y/\partial X_3)(X_3/Y)=\beta_3$，这就是 Y 对 X_3 的弹性，等等。

[17] 注意在柯布-道格拉斯生产函数(7.9.1)中，我们以特殊方式引进随机误差项，以使对数变换后得到通常的线性形式。参看6.9节。

[18] 我们避开从理论观点看模型的适宜性问题，以及能否从时间序列测得规模报酬的问题。

[19] 参看 Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, 3d ed., McGraw-Hill, New York, 1984, pp. 250－252。

[20] 资料来源：*The East Asian Economic Miracle: Economic Growth and Public Policy*, A World Bank Policy Research Report, Oxford University Press, U.K., 1993, p. 29。

[21] 如果对(7.10.7)求导，你将得到

$$\frac{dGDPG}{dRGDP} = 0.062 - 0.122RGDP$$

表明 GDPG 相对 RGDP 的变化率在下降。若令此导数为零，你将得到 RGDP≈0.508 2。因此，如果一个国家的 GDP 接近美国 GDP 的一半，那么 GDPG 的增长率将降至 0。

第 8 章　多元回归分析：推断问题

248　本章为第 5 章的续篇。它把该章所讲的区间估计与假设检验的思想延伸到涉及三个或多个变量的模型上来。虽然第 5 章所论述的概念在许多方面都可直截了当地应用于多元回归，但有少数特点则为多元回归模型所特有。而正是这些特点在本章中受到了更多的关注。

§8.1　再一次正态性假定

至此我们知道，如果我们惟一的目的是对回归模型的参数作点估计，则普通最小二乘法将足够使用，并不需要对干扰项 u_i 的概率分布做任何假定。但若我们的目的兼在于估计和推断，则如同第 4 章和第 5 章所称辩的，我们还需要假定 u_i 遵循某个概率分布。

由于已明确申述的一些理由，我们曾假定 u_i 遵循均值为零、方差 σ^2 为常数的正态分布。我们对多元回归模型继续作同样的假定。有了正态性假定，并仿照第 4 章和第 7 章的讨论，我们发现，偏回归系数的 OLS 估计量，
249　无异于最大似然估计量，是最优线性无偏估计量。[1] 此外，估计量 $\hat{\beta}_2$、$\hat{\beta}_3$ 和 $\hat{\beta}_1$ 本身也是正态分布的，其均值等于 β_2，β_3 和 β_1。而方差由第 7 章

[(7.4.9)至(7.4.16)。——译者注]给出。而且，$(n-3)\hat{\sigma}^2/\sigma^2$ 遵循自由度为 $n-3$ 的 χ^2 分布，并且三个 OLS 估计量均独立于 $\hat{\sigma}^2$ 而分布，其证明类似于附录 C 所讨论的双变量情形。因此，再仿照第 5 章，即可表明，在标准误的计算中，σ^2 由它的无偏估计 $\hat{\sigma}^2$ 代替时便知：

$$t=\frac{\hat{\beta}_1-\beta_1}{\operatorname{se}(\hat{\beta}_1)} \tag{8.1.1}$$

$$t=\frac{\hat{\beta}_2-\beta_2}{\operatorname{se}(\hat{\beta}_2)} \tag{8.1.2}$$

$$t=\frac{\hat{\beta}_3-\beta_3}{\operatorname{se}(\hat{\beta}_3)} \tag{8.1.3}$$

均遵循自由度为 $n-3$ 的 t 分布。

注意自由度为 $n-3$ 是因为在计算 $\sum\hat{u}_i^2$ 和 $\hat{\sigma}^2$ 之前，我们要先估计三个偏回归系数，从而给残差平方和（RSS）的计算加上了三个约束。（按照这个逻辑，在四变量情形中将有 $n-4$ 个自由度，依此类推。）于是，t 分布可用于建立关于真实总体偏回归系数的置信区间并检验统计假设。同理，χ^2 分布可用于检验关于真实 σ^2 的假设。我们用下面的说明性例子来阐明具体操作步骤。

§8.2 例 8.1 修正儿童死亡率例子

在第 7 章，我们用一个 64 个国家构成的样本将儿童死亡率对人均 GNP 和妇女识字率进行回归。(7.6.2) 中给出的回归结果增加某些信息后复制如下：

$$\begin{aligned}
\widehat{\mathrm{CM}}_i &= 263.641\,6 \quad - 0.005\,6\mathrm{PGNP}_i - 2.231\,6\mathrm{FLR}_i\\
\mathrm{se} &= (11.593\,2) \quad (0.001\,9) \qquad (0.209\,9)\\
t &= (22.741\,1) \quad (-2.818\,7) \qquad (-10.629\,3)\\
p\text{ 值} &= (0.000\,0)^{*} \quad (0.006\,5) \qquad (0.000\,0)^{*}\\
&R^2=0.707\,7,\ \bar{R}^2=0.698\,1
\end{aligned} \tag{8.2.1}$$

其中 * 表示极低值。

250 在方程 (8.2.1) 中，我们沿袭了方程 (5.11.1) 中首次引入的格式，其中第一行括号中的数字是估计标准误，第二行为相关总体系数为零的虚拟假设下的 t 值，第三行为估计的 p 值。还给出 R^2 和校正 R^2 值。我们已在例 7.1 中解释过这个回归。

所观察到这些结论的统计显著性如何呢？比方考虑 PGNP 的系数 $-0.005\,6$。这个系数是统计显著的吗？即它统计显著地异于零吗？类似地，

FLR 的系数 −2.231 6是统计显著的吗？这两个系数都是统计显著的吗？为回答这个及与此相关的问题，让我们首先考虑在多元回归模型中可能会遇到的假设检验类型。

§8.3 多元回归中的假设检验：总评

一旦我们走出简单的双变量线性回归模型的范围，假设检验就会以多种有趣的形式出现，诸如：

1. 检验关于个别偏回归系数的假设（8.4 节）。

2. 检验所估计的多元回归模型的总显著性，也就是要判明是否全部偏斜率系数同时为零（8.5 节）。

3. 检验两个或多个系数是否相等（8.6 节）。

4. 检验诸偏回归系数是否满足某种约束条件（8.7 节）。

5. 检验所估计的回归模型在时间上或在不同横截面单元上的稳定性（8.8 节）。

6. 检验回归模型的函数形式（8.9 节）。

因为在经验分析中常常出现这些类型的一种或多种检验，我们将分节讨论每一种类型的检验。

§8.4 检验关于个别偏回归系数的假设

如在 8.1 节中指出的，引用假定 $u_i \sim N(0,\ \sigma^2)$，便可用 t 检验统计量对任一个别的偏回归系数的假设进行检验。为了说明操作步骤，考虑我们的数值例子，我们公设：

$$H_0:\ \beta_2=0 \quad 及 \quad H_1:\ \beta_2\neq 0$$

251 虚拟假设是说，保持 X_3（妇女识字率）不变，X_2（PGNP）对 Y（儿童死亡率）无（线性）影响。[2] 为了检验这个虚拟假设，我们利用（8.1.2）中给出的 t 检验。仿照第 5 章的作法，如果计算的 t 值超过选定显著性水平的临界 t 值，就可拒绝假设；否则就不拒绝它。对于我们的例子，利用（8.1.2）并注意到在虚拟假设下 $\beta_2=0$，我们得到：

$$t=\frac{-0.005\ 6}{0.002\ 0}=-2.818\ 7 \tag{8.4.1}$$

如方程（8.2.1）所示。

注意我们有 64 个观测。因此本例中的自由度为 61（为什么?），你若查

阅附录 D 中的 t 表，则没有与自由度 61 相对应的数据。与之最接近的自由度是 60。若使用自由度 60，并假设显著性水平（即犯第 I 类错误的概率）α 为 5%，则双尾检验的临界 t 值为 2.0（查自由度为 60 的 $t_{\alpha/2}$），而单尾检验的临界 t 值为 1.671（查自由度为 60 的 t_{α}）。

对于本例，对立假设是双侧的。因此我们用双尾 t 值。既然计算出来的 t 值2.818 7（绝对值）超过了临界 t 值 2，那我们就可以拒绝 PGNP 对儿童死亡率没有影响的虚拟假设。更明确地讲，保持妇女识字率不变，人均 GNP 对儿童死亡率具有显著的（负面）影响，这与先验预期完全一致。图 8.1 从图解上展示了这一情形。

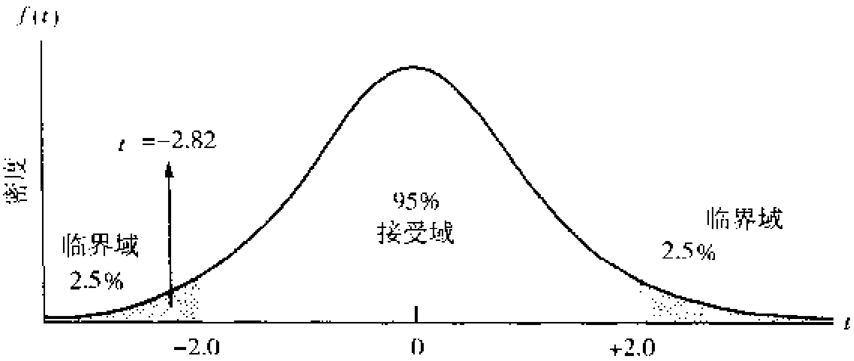


图 8.1　t 的 95%置信区间（自由度为 60）

252 实际上，人们不必假定一个特定的 α 值来进行假设检验，仅使用(8.2.2）中的 p 值即可。本例中的 p 值就是0.006 5。对这个 p 值（即精确的显著性水平）的解释是，如果虚拟假设正确，得到一个大于等于2.818 7（在绝对值上）的 t 值的概率仅为0.006 5或 0.65%，这确实是一个相当小的概率，比人为选定的 $\alpha=5\%$ 小得多。

本例还为我们提供了一个决定是用单尾还是用双尾 t 检验的机会。既然推测儿童死亡率与人均 GNP 负相关（为什么?），那我们就应该使用单尾检验。即虚拟和对立假设应该是：

$$H_0:\ \beta_2<0 \text{ 和 } H_1:\ \beta_2\geqslant 0$$

读者已经知道，我们在本例中能基于单尾 t 检验而拒绝虚拟假设。

在第 5 章中，我们曾看到假设检验和置信区间估计之间的密切关系。拿我们的例子说，β_2 的 95% 置信区间是：

$$\hat{\beta}_2-t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)\leqslant\beta_2\leqslant\hat{\beta}_2+t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)$$

具体到我们的数值就变成：

$$-0.005\,6-2(0.002\,0)\leqslant\beta_2\leqslant-0.005\,6+2(0.002\,0)$$

即是：

$$-0.009\,6\leqslant\beta_2\leqslant-0.001\,6 \qquad (8.4.2)$$

就是说，β_2 以 95% 的置信系数落在 -0.0096 与 -0.0016 之间。这样，如果选取了大小同为 64 的 100 个样本并构造像（8.4.2）这样的 100 个置信区间，则我们预期其中的 95 个包含着真实总体参数 β_2。由于虚拟假设的零值不落在区间（8.4.2）上，故能以 95% 的置信系数拒绝虚拟假设 $\beta_2=0$。

由此，无论我们用（8.4.1）的 t 显著性检验或用（8.4.2）的置信区间估计，我们得到同样的结论。但鉴于置信区间估计与假设检验之间的密切关系，这是不足为奇的。

按照刚才讲的方法，我们可以利用方程（8.2.1）提供的数据，检验关于儿童死亡率回归模型中的其他参数假设。例如，假使我们想检验的假设是保持 PGNP 的影响不变，妇女识字率对儿童死亡率没有什么影响。我们确信能拒绝这个假设，因为在此假设下，得到一个绝对值大于等于 10.6 的 t 值的 p 值实际上是 0。

253 在继续下一步之前，记住 t 检验的程序是基于误差项 u_i 服从正态分布的假定。尽管我们不能直接观测 u_i，但我们能够观测到它们的代理变量 $\hat{u}_i$，即残差。对儿童死亡率一例而言，残差直方图如图 8.2 所示。

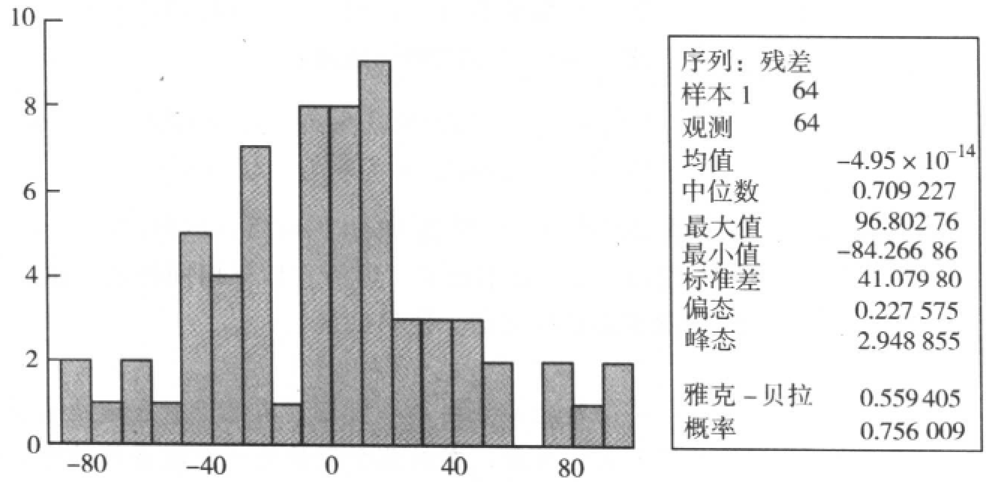


图 8.2 回归（8.2.1）的残差直方图

从直方图来看，残差是正态分布的。我们也可以计算方程（5.12.1）所示的雅克-贝拉（JB）正态性检验。这里的 JB 值为0.559 4，p 值为 0.76。[3] 因此，本例中的误差项看上去服从正态分布。当然也要记住，JB 检验是一个大样本检验，我们在这个例子中只有 64 次观测可能不够多。

§8.5 检验样本回归的总显著性

上节全节我们都仅讨论**个别的**，也就是在每一真实偏回归系数为零的**单个**（separate）假设下的，检验所估偏回归系数的显著性的问题，而现在我

们考虑如下的假设：

$$H_0: \ \beta_2 = \beta_3 = 0 \tag{8.5.1}$$

这个虚拟假设是关于 β_2 和 β_3 联合地或同时地等于零的一个联合假设（joint hypothesis）。对这样一个假设的检验被称作对所测或所估回归线的**总显著性**（overall significance）检验，也就是，检验 Y 是否与 X_2 和 X_3 两者有线性关系。

254 能不能用 8.4 节逐一检验 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的显著性的方法来检验（8.5.1）中的联合假设呢？答案是否定的，理由如下。

在第 8.4 节中检验一个所测偏回归系数的个别显著性时，我们隐含地假定每一个显著性检验都是根据一个不同的（即独立的）样本进行的。这样，在假设 $\beta_2=0$ 下检验 $\hat{\beta}_2$ 的显著性时，我们无形地假定了用于这一检验的样本不同于用来检验在假设 $\beta_3=0$ 下 $\hat{\beta}_3$ 的显著性的那个样本。但是，如果我们用同一样本数据去检验（8.5.1）的联合假设，我们就违反了检验方法所依据的基本假定。[4] 这一问题可另行表述如下：在（8.4.2）中我们对 β_2 构造了一个 95% 置信区间。如果我们仍按 95% 的置信系数用同样的样本数据构造 β_3 的一个置信区间，我们就不能断言 β_2 和 β_3 同时落在其各自的置信区间上的（置信）概率是 $(1-\alpha)(1-\alpha)=(0.95)(0.95)$。

换言之，虽然下面的两个陈述：

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\text{se}(\hat{\beta}_2) \leqslant \beta_2 \leqslant \hat{\beta}_2 + t_{\alpha/2}\text{se}(\hat{\beta}_2)] = 1-\alpha$$
$$\Pr[\hat{\beta}_3 - t_{\alpha/2}\text{se}(\hat{\beta}_3) \leqslant \beta_3 \leqslant \hat{\beta}_3 + t_{\alpha/2}\text{se}(\hat{\beta}_3)] = 1-\alpha$$

个别地看是正确的，但说 β_2 和 β_3 同时落入区间 $[\hat{\beta}_2 \pm t_{\alpha/2}\text{se}(\hat{\beta}_2), \hat{\beta}_3 \pm t_{\alpha/2}\text{se}(\hat{\beta}_3)]$ 的概率是 $(1-\alpha)^2$ 就不对了。因为，如果用同样的数据去导出这些区间，这些区间就不会是独立的。换句话说，

> ……检验一系列单个假设，不等于联合地检验同样的这些假设。其直观上的理由是，在对几个假设的一个联合检验中，任一个单一假设都受其他假设所含信息的影响。[5]

以上辩论的要点在于：对于一个给定的实例（样本），只能得到一个置信区间或者只能做出一个显著性检验。那么，我们怎样来检验联立的虚拟假设 $\beta_2=\beta_3=0$ 呢？答案如下文。

检验所测多元回归的总显著性的方差分析法：F 检验

这个理由刚才解释过了，我们不能用通常的 t 检验去检验多个真实偏斜率系数同时为零的联合假设。然而，这个联合假设可以用最先在 5.9 节介绍的方差分析技术去检验。现说明如下：

回忆恒等式：

$$\sum y_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum \hat{u}_i^2 \tag{8.5.2}$$
$$\text{TSS} = \qquad \text{ESS} \qquad + \text{RSS}$$

和平常一样，TSS 有 $n-1$ 个自由度，而 RSS 有 $n-3$ 个自由度，它的理由已经讨论过了。ESS 因为是 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的函数所以有 2 个自由度。因此，按照 5.9 节讨论的 ANOVA 程序，我们可以列出表 8.1。

255

表 8.1　　三变量回归的 ANOVA 表

变异来源	SS	df	MSS
来自回归(ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\dfrac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
来自残差(RSS)	$\sum \hat{u}_i^2$	$n-3$	$\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-3}$
总计	$\sum y_i^2$	$n-1$	

现在可以证明[6]，在 u_i 的正态分布假定下以及在虚拟假设 $\beta_2 = \beta_3 = 0$ 下，变量：

$$\begin{aligned} F &= \frac{\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)/2}{\sum \hat{u}_i^2/(n-3)} \\ &= \frac{\text{ESS/df}}{\text{RSS/df}} \end{aligned} \tag{8.5.3}$$

遵循自由度为 2 和 $n-3$ 的 F 分布。

上述 F 能有什么用？可以证明[7]，在 $u_i \sim N(0, \sigma^2)$ 的假定下，

$$E\frac{\sum \hat{u}_i^2}{n-3} = E(\hat{\sigma}^2) = \sigma^2 \tag{8.5.4}$$

表 8.2　　F 统计量小结

虚拟假设 H_0	对立假设 H_1	拒绝 H_0 的临界域
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$S_1^2/S_2^2 > F_{\alpha, ndf, ddf}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$S_1^2/S_2^2 > F_{\alpha/2, ndf, ddf}$ 或 $< F_{(1-\alpha/2), ndf, ddf}$

注：1. σ_1^2 和 σ_2^2 为两个总体方差。
2. S_1^2 和 S_2^2 为两个样本方差。
3. $n$df 和 $d$df 分别为分子和分母自由度。
4. 在计算 F 比率时，将较大的 S^2 值放在分子上。
5. 临界 F 值在最后一列给出，F 的第一个下标是显著性水平，第二个下标为分子和分母自由度。
6. 注意 $F_{(1-\alpha/2), ndf, ddf} = 1/F_{\alpha/2, ndf, ddf}$。

再加一个假定 $\beta_2=\beta_3=0$，便能证明：

$$\frac{E(\hat{\beta}_2\sum y_i x_{2i}+\hat{\beta}_3\sum y_i x_{3i})}{2}=\sigma^2 \tag{8.5.5}$$

因此，如果虚拟假设是真实的，(8.5.4) 和 (8.5.5) 都将对真实 σ^2 给出同样的估计。这一声称并不足以为奇。因为，如果 Y 与 X_2 和 X_3 的关系微不足道，则 Y 变异的惟一来源是来自 u_i 所代表的随机因素。然而，如果
256 虚拟假设错误，即 X_2 和 X_3 确实影响 Y，则不能在 (8.5.4) 和 (8.5.5) 之间划等号。这时，在适当考虑自由度之后，ESS 要相对大于 RSS。从而 (8.5.3) 的 F 值对真实斜率系数同时为零的这一虚拟假设，提供了一种检验。如果从 (8.5.3) 算出的 F 值大于 2% 显著水平的 F 表中的临界 F 值，我们就拒绝 H_0；否则就不拒绝它。另一种方法是，如果所测 F 的 p 值足够低，则可拒绝 H_0。

表 8.2 对 F 检验进行了总结。回到我们的例子，我们算得表 8.3。利用 (8.5.3) 即得：

$$F=\frac{128\ 681.2}{1\ 742.88}=73.832\ 5 \tag{8.5.6}$$

表 8.3　　　　儿童死亡率一例的 ANOVA 表

变异来源	SS	df	MSS
来自回归	257 362.4	2	128 681.2
来自残差	106 315.6	61	1 742.88
总计	363 678	63	

得到一个大于或等于 73.832 5 的 F 值的 p 值几乎是 0，从而拒绝 PGNP 和 FLR 同时对儿童死亡率都没有影响的假设。如果使用惯常的 5% 的显著性水平，分子自由度为 2 和分母自由度为 60（实际自由度为 61）的临界 F 值
257 约为 3.15，若用 1% 的显著性水平，临界 F 值则约为 4.98。显然，观察到约为 74 的 F 值比这些临界 F 值中的任何一个都大得多。

我们可将上述 F 检验方法推广到一般情况，如下所述。

检验多元回归的总显著性：F 检验

决策规则。给定 k 变量回归模型：

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i$$

检验假设：

$$H_0:\ \beta_2=\beta_3=\cdots=\beta_k=0$$

（就是，全部斜率系数同时为零。）

相对于 H_1：非全部斜率系数同时为零。

计算：
$$F=\frac{\text{ESS/df}}{\text{RSS/df}}=\frac{\text{ESS}/(k-1)}{\text{RSS}/(n-k)} \tag{8.5.7}$$

如果 $F>F_{\alpha}(k-1,n-k)$，则拒绝 H_0；否则不拒绝它，其中 $F_{\alpha}(k-1,n-k)$ 是 α 显著水平、$(k-1)$ 个分子自由度和 $(n-k)$ 个分母自由度的临界 F 值。另法，如果由（8.5.7）所得 F 的 p 值足够低，即可拒绝 H_0。

不言而喻，在三变量情形（Y 和 X_2、X_3）中 $k=3$，在四变量情形中 $k=4$，依此类推。

顺便一提，大多数回归软件包都照例把（方差分析表中给出的）F 值连同通常的回归结果，诸如估计的系数，它们的标准误（差），t 值等等一起算出。在 t 的计算中，通常都把虚拟假设取为 $\beta_i=0$。

个别与联合假设检验对比。在 8.4 节中我们讨论了单一回归系数的显著性检验，并在 8.5 节中又讨论了整个回归（即全部系数同时为零）的总显著性检验。**我们要重申这两类检验是不相同的。**因此，有可能根据 t 检验或（8.4 节的）置信区间而接受某一系数 β_k 为零的假设，但另一方面却拒绝全部系数为零的联合假设。

> 应吸取的经验教训是，个别的（多个）置信区间所提供的联合“信息”，不能代替假设的联合检验和联合置信陈述中所蕴涵的联合置信区域（蕴涵于 F 检验中）。[8]

R^2 和 F 之间的一个重要关系式

258 判定系数 R^2 与方差分析中所用的 F 检验之间有一紧密关系。假定干扰项 u_i 为正态分布，并且虚拟假设 $\beta_2=\beta_3=0$ 成立，我们曾看到：

$$F=\frac{\text{ESS}/2}{\text{RSS}/(n-3)} \tag{8.5.8}$$

遵循 2 和 $n-3$ 个自由度的 F 分布。

推广到 k 变量情形（包括截距），如果假定干扰项是正态分布的，而且有虚拟假设：

$$H_0:\ \beta_2=\beta_3=\cdots=\beta_k=0 \tag{8.5.9}$$

则随之有：

$$F=\frac{\mathrm{ESS}/(k-1)}{\mathrm{RSS}/(n-k)} \qquad (8.5.7)=(8.5.10)$$

遵循 $k-1$ 和 $n-k$ 个自由度的 F 分布。(注：待估参数的总个数是 k，其中之一为截距项。)

对（8.5.10）做如下演算：

$$\begin{aligned}F&=\frac{n-k}{k-1}\cdot\frac{\mathrm{ESS}}{\mathrm{RSS}}\\&=\frac{n-k}{k-1}\cdot\frac{\mathrm{ESS}}{\mathrm{TSS}-\mathrm{ESS}}\\&=\frac{n-k}{k-1}\cdot\frac{\mathrm{ESS}/\mathrm{TSS}}{1-(\mathrm{ESS}/\mathrm{TSS})}\\&=\frac{n-k}{k-1}\cdot\frac{R^2}{1-R^2}\\&=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}\end{aligned} \qquad (8.5.11)$$

其中我们用到了定义 $R^2=\mathrm{ESS}/\mathrm{TSS}$。方程（8.5.11）表明了 F 和 R^2 是怎样一种关系——两者是同向变化的。当 $R^2=0$ 时，F 随之等于零。R^2 越大，F 值也越大。终其极限，当 $R=1$ 时，F 变为无限大。因此，F 检验既是所估回归的总显著性的一个度量，也是 R^2 的一个显著性检验。换句话说，检验虚拟假设（8.5.9）等价于检验（总体）R^2 等于零的虚拟假设。

259 将三变量情形（8.5.11）变为：

$$F=\frac{R^2/2}{(1-R^2)/(n-3)} \qquad (8.5.12)$$

利用 F 与 R^2 之间的紧密联系，可把 ANOVA 表 8.2 重新设计成表 8.4。

对于我们的说明性例子，利用（8.5.12）我们得到：

$$F=\frac{0.707\ 7/2}{(1-0.707\ 7)/61}=73.872\ 6$$

除四舍五入的误差外，这个 F 值与前面得到的 F 值相同。

表 8.4　　由 R^2 表示的 ANOVA 表

变异来源	SS	df	MSS*
来自回归	$R^2(\sum y_i^2)$	2	$R^2(\sum y_i^2)/2$
来自残差	$(1-R^2)(\sum y_i^2)$	$n-3$	$(1-R^2)(\sum y_i^2)/(n-3)$
总计	$\sum y_i^2$	$n-1$	

* 注意，在计算 F 值时无须用 $\sum y_i^2$ 乘 R^2 和 $(1-R^2)$。因为，如(8.5.12)所示，它将被消掉。

用 R^2 来表示 F 检验的一个优点在于计算上的便易：需要知道的仅仅是 R^2 值而已。因此，(8.5.7) 所给的总 F 显著性检验可重新用 R^2 表达，如表 8.4 所示。

检验一个用 R^2 表示的多元回归的总显著性

决策规则。检验一个用 R^2 表示的多元回归的总显著性：另一个等价于 (8.5.7) 的检验。

给定 k 变量回归模型：

$$Y_i = \beta_i + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

为了检验假设：

$$H_0:\ \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

相对于

H_1：非全部斜率系数都同时为零

计算：

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{8.5.13}$$

260

如果 $F > F_{\alpha(k-1,n-k)}$，则拒绝 H_0；否则可接受 H_0，其中 $F_{\alpha(k-1,n-k)}$ 是 α 显著性水平、$k-1$ 个分子自由度和 $(n-k)$ 个分母自由度的临界 F 值。另法：如果得自 (8.5.13) 的 F 的 p **值足够小，则拒绝** H_0。

在继续讲下去之前，回头看一下第 7 章的例 7.5。我们从 (7.10.7) 看到，RGDP (相对人均 GDP) 和 RGDP 的平方只解释了 119 个国家构成的样本中 GDPG (GDP 增长率) 变异的 5.3%。$R^2 = 0.053$ 看起来是个 "低" 值。它真的就在统计上异于零吗？我们如何说明？

回想我们前面在 "R^2 和 F 之间的一个最要关系式" 专题中对 (8.5.11) 或两个回归元特殊情形下的 (8.5.12) 给出的 R^2 和 F 值之间关系的讨论。前面曾指出，若 R^2 为零，则 F 因此为零，若回归元对回归子没什么影响便是如此。因此，若我们将 $R^2 = 0.053$ 代入公式 (8.5.12)，我们可得到

$$F = \frac{0.053/2}{(1-0.053)/116} = 3.247\ 5 \tag{8.5.13}$$

在虚拟假设 $R^2 = 0$ 下，上述 F 值服从自由度为 2 和 116 的 F 分布（注：有 119 个观测值和 2 个回归元）。我们从 F 表看出，这个 F 值在 5% 的显著性水平上仍是显著的，p 值实际上是 0.042 5。因此，尽管 R^2 只有 0.053，我

们仍能拒绝这两个回归元对回归子没有影响的虚拟假设。

本例引申出一个重要的经验心得，在涉及几个变量观测值的横截面数据中，由于横截面单元的多样性，所以得到的 R^2 一般都很低。所以，对横截面回归中得到低 R^2 值不用感到吃惊或着急。重要的是，正确地设定模型，回归元具有正确（即理论预期）的符号，以及统计上显著（希望如此）的回归系数。读者应该在 5% 或更好（即低于 5%）的显著性水平上验证 (7.10.7) 中两个回归元都是个别统计显著的。

一个解释变量的“增量”或“边际”贡献

我们在第 7 章中说过，我们一般不能将 R^2 值在各个回归元之间分配。在我们儿童死亡率的例子中，我们发现 R^2 为 0.707 7，但由于这两个回归元在我们手头的样本中可能相关，所以我们不知道这个值中哪些属于回归元 PGNP 的功劳，哪些又属于妇女识字率的功劳。利用协方差分析的方法，我们可以对此有更深入的了解。

261 对我们的说明性例子而言，我们发现 X_2(PGNP)和 X_3(FLR)基于（分别进行的）t 检验都是个别统计显著的。我们还发现，基于 F 检验，这两个回归元一起对回归子 Y（儿童死亡率）具有明显影响。

现在假设我们顺序引入 PGNP 和 FLR；即首先将儿童死亡率对 PGNP 回归并评价其显著性，然后在模型中增加 FLR，以判明它是否有任何贡献（当然，可以对调 X_2 和 X_3 进入的次序）。所谓贡献，我们意指增加一个变量到模型中来，是否相对于 RSS 说“显著地”增加了 ESS（从而影响 R^2）。把这一贡献称作一个解释变量的**增量**（incremental）或**边际**（marginal）贡献，也许是适当的。

在实践中，增量贡献的问题，是一个重要的论题。在大多数的经验研究工作中，研究者对一个已含有若干个变量的模型是否值得再添加一个新的 X 变量，是没有足够把握的。他也不愿意放进那些对 ESS 贡献很少的变量。同样的道理，他也不愿意排除一些能实质上增加 ESS 的变量。但怎样决定一个 X 变量的引进是否能显著地减少 RSS 呢？容易通过方差分析的应用来回答此问题。

假设我们先做儿童死亡率对 PGNP 的回归，并得到下述结果：

$$\begin{aligned}\widehat{\mathrm{CM}}_i &= 157.424\,4 - 0.011\,4\mathrm{PGNP} \\ t &= (15.989\,4) \quad (-3.515\,6) \\ p\text{ 值} &= (0.000\,0) \quad (0.000\,8) \quad r^2 = 0.166\,2 \\ &\qquad\qquad \text{校正 } r^2 = 0.152\,8 \end{aligned} \tag{8.5.14}$$

如这些结论所示，PGNP 对 CM 具有明显影响。表 8.5 给出了上述回归的 ANOVA 表。

表 8.5 回归（8.5.14）的 ANOVA 表

变异来源	SS	df	MSS
ESS（来自 PGNP）	60 449.5	1	604 495.1
RSS	303 228.5	62	4 890.782 2
总计	363 678.00	63	

假定干扰项 u_i 是正态分布的，并且虚拟假设 $\beta_{12}=0$。我们知道：

$$F=\frac{60\ 449.5}{4\ 890.782\ 2}=12.359\ 8 \tag{8.5.15}$$

262 遵循自由度为 1 和 62 的 F 分布。因为计算出来的 p 值为 0.000 8，所以这个 F 值是高度显著的。因此，和前面一样，我们拒绝 PGNP 对 CM 没有影响的假设。顺带指出，$t^2=(\quad 3.515\ 6)^2=12.359\ 4$。这近似等于用（8.5.14）中 t 值计算出来的 F 值（8.5.15）。但这个结果是没有什么可惊奇的。因为，如第 5 章第一次证明的，在同样虚拟假设和同样显著性水平下，自由度为 n 的 t 平方等于自由度为 1 和 n 的 F。本例中 $n=64$。

假如在做完回归（8.5.14）之后，我们决定把 X_3 加进到模型中来，并得到多元回归（8.2.1）。我们需要回答的问题是：

（1）知道 PGNP 在模型中并且和 CM 有显著关系，FLR 的边际或增量贡献为何？

（2）FLR 的增量贡献在统计上显著吗？

(3)根据什么准则把变量加进模型？可通过 ANOVA 技术来回答上述问题。

为了说明这点，让我们构造表 8.6。此表中 X_2 表示 PGNP，X_3 表示 FLR。

表 8.6 用于评估变量的增量贡献的 ANOVA 表

变异来源	SS	df	MSS
ESS 仅由于 X_2	$Q_1=\hat\beta_{12}^2\sum x_2^2$	1	$\frac{Q_1}{1}$
ESS 由于 X_3 的加入	$Q_2=Q_3-Q_1$	1	$\frac{Q_2}{1}$
ESS 由于 X_2 和 X_3	$Q_3=\hat\beta_2\sum y_i x_{2i}+\hat\beta_3\sum y_i x_{3i}$	2	$\frac{Q_3}{2}$
RSS	$Q_4=Q_5-Q_3$	$n-3$	$\frac{Q_4}{n-3}$
总计	$Q_5=\sum y_i^2$	$n-1$	

为了评估在扣除 X_2 的贡献后 X_3 的增量贡献，我们构造：

$$
\begin{aligned}
F &= \frac{Q_2 / \text{自由度}}{Q_4 / \text{自由度}} \\
&= \frac{(\mathrm{ESS}_{新} - \mathrm{ESS}_{老}) / \text{新回归元个数}}{\mathrm{RSS}_{新} / \text{自由度}(= n - \text{新模型中的参数个数})} \\
&= \frac{Q_2 / 1}{Q_4 / 12} \quad \text{对于本例}
\end{aligned} \tag{8.5.16}
$$

其中 $\mathrm{ESS}_{新}$ = 新模型的 ESS·(指加进新回归元后的 Q_3)，$\mathrm{ESS}_{老}$ = 老模型的 ESS（$=Q_1$），$\mathrm{RSS}_{新}$ = 新模型的 RSS（指扣除所有回归元的贡献后的 Q_4）。对于我们的说明性例子，结果如表 8.7 所示。

263 **表 8.7　　说明性例子的 ANOVA 表：增量分析**

变异来源	SS	df	MSS
ESS，仅由 PGNP	$Q_1 = 60\ 449.5$	1	60 449.5
ESS，由 FLR 的添加	$Q_2 = 196\ 912.9$	1	196 912.9
ESS，由 PGNP 和 FLR	$Q_3 = 257\ 362.4$	2	128 681.2
RSS	$Q_4 = 106\ 315.6$	61	1 742.878 6
总计	$Q_5 = 363\ 678$	63	

现在应用（8.5.16），我们算得：

$$
\begin{aligned}
F &= \frac{196\ 912.9}{1\ 742.878\ 6} \\
&= 112.981\ 4
\end{aligned} \tag{8.5.17}
$$

在 u_i 的通常的假定下，这个 F 值服从自由度为 1 和 62 个自由度的 F 分布。读者应该验证，这个 F 值是高度显著的，表明模型中增加 FLR 明显提高了 ESS 并因此提高 R^2 值。因此，模型中应该增加 FLR。同样注意到，如果你将多元回归（8.2.1）中 FLR 系数值平方，即（$-10.629\ 3)^2$，在允许存在四舍五入误差的情况下，你将得到（8.5.17）中的 F 值。

另外，如同我们在（8.5.13）中所做的那样，（8.5.16）的 F 比率还可仅用 R^2 值重新再表达出来。如习题 8.2 所表明的，（8.5.16）的 F 比率等价于以下的 F 比率[9]：

$$F=\frac{(R^2_{新}-R^2_{老})/自由度}{(1-R^2_{新})/自由度}=\frac{(R^2_{新}-R^2_{老})/新回归元个数}{(1-R^2_{新})/自由度(=n-新模型中的参数个数)} \tag{8.5.18}$$

此 F 比率遵循有适当分子和分母自由度（在我们的例子中分别是 1 和 61）的 F 分布。

对我们的例子，$R^2_{新}=0.707\ 7$ [由 (8.2.1)] 和 $R^2_{老}=0.166\ 2$ [由 (8.5.14)]。因此，

$$\begin{aligned}F&=\frac{(0.707\ 7-0.166\ 2)/1}{(1-0.707\ 7)/61}\\&=113.05\end{aligned} \tag{8.5.19}$$

除近似计算中的误差外，和 (8.5.17) 的 F 值差不多一样。这个 F 值高度显著，强化了变量 FLR 应当在此模型中的早期结论。

一个忠告性注释：你若使用 (8.5.11) 中给出的 F 检验的 R^2，要保证新老模型中因变量是相同的。如果它们不同，则使用 (8.5.16) 中给出的 F 检验。

何时加进一个新变量。 刚才描述的 F 检验程序，为决定是否增加一个变量到回归模型中来，提供了一个程式化方法。研究者常常遇到从几个互相媲美的模型中挑选一个的问题，这些模型**有同一因变量**但有不同的解释变量。作为一种见机行事的选择（因为分析的理论基础薄弱），研究者常常选择有最高校正 R^2 值的模型。这样一来，如果只要增加一个变量能增加 $\bar{R}^2$ 的值，就把它保留在模型中，即使它并不在统计意义上显著地减少 RSS。问题于是变为：校正 R^2 值在什么情况下增加？可以证明，如果新增变量的系数的 t 值在绝对值上大于 1，$\bar{R}^2$ 就会增加。这里的 t 值，是在所指系数的总体值为零的假设下计算的 [即在真实 β 值为零的假设下由 (5.3.2) 计算出来的 t 值]。[10] 上述准则又可用不同的方式叙述为：仅当一个新增解释变量的 F $(=t^2)$ 值大于 1 时，它的引进才使 $\bar{R}^2$ 增大。

不管用哪一准则，儿童死亡率一例中的 FLR 变量的 t 值为 $-10.629\ 3$ 或 F 值为 $112.981\ 4$，均表明 $\bar{R}^2$ 将增大。的确，当 FLR 加进模型时，$\bar{R}^2$ 从 0.152 8 增大到 0.698 1。

何时加进一组变量。 能不能找到类似的规则，以决定是否值得把一组变量加进到模型中来，或从模型中排除出去？从 (8.5.18) 应能看到答案：如果一组变量的加入（排除）给出一个大（小）于 1 的 F 值，$\bar{R}^2$ 将增加（减小）。当然，从 (8.5.18) 容易看出，一组变量的加入（删除）是否显著地增加（减少）了回归模型的解释能力。

§8.6 检验两个回归系数是否相等

假如在多元回归中

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \tag{8.6.1}$$

265 我们要检验假设

$$H_0: \beta_3 = \beta_4 \quad \text{或} \quad (\beta_3 - \beta_4) = 0$$
$$H_1: \beta_3 \neq \beta_4 \quad \text{或} \quad (\beta_3 - \beta_4) \neq 0 \tag{8.6.2}$$

即两个斜率系数 β_3 和 β_4 相等。

这样的虚拟假设没有实际意义。例如，令（8.6.1）代表对某商品的需求函数，其中 Y = 某商品的需求量，X_2 = 该商品的价格，X_3 = 消费者的收入，和 X_4 = 消费者的财富。这时，虚拟假设是说，收入系数与财富系数相等。或者，如果把 Y 和诸 X 表达为对数形式，则（8.6.2）中的虚拟假设意味着消费的收入弹性与财富弹性相同。(为什么?)

怎样检验这样一种虚拟假设？在经典假设下，可以证明：

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{se}(\hat{\beta}_3 - \hat{\beta}_4)} \tag{8.6.3}$$

遵循自由度为（$n-4$）的 t 分布，因为（8.6.1）是一个四变量模型或写得更一般，自由度为（$n-k$），其中 k 为包括截距项的待估参数的总个数。标准误 se（$\hat{\beta}_3 - \hat{\beta}_4$）则可从以下熟知的公式中得到（详见本书统计附录）：

$$\text{se}(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)} \tag{8.6.4}$$

如果将虚拟假设和se($\beta_3 - \beta_4$)的表达式代入(8.6.3),我们的检验统计量就变为：

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \tag{8.6.5}$$

于是，检验方法包括如下步骤：

1. 估计 $\hat{\beta}_3$ 和 $\hat{\beta}_4$。任何标准计算机软件包都能做到。

2. 大多数标准计算机软件包都照常算出所估参数的方差与协方差。[11] 从这些（方差与协方差）的估计值就容易算得（8.6.5）的分母中的标准误。

3. 从（8.6.5）算出 t 比率。注意本例中的虚拟假设是（$\beta_3 - \beta_4$）$=0$。

266 4. 如果从（8.6.5）计算出来的 t 变量超过给定自由度的在指定显著性水平上的临界 t 值，则可拒绝虚拟假设；否则不拒绝。另法，如得自（8.6.5）的 t 统计量的 p 值合理的低，就可拒绝虚拟假设。注意，p 值越

低，拒绝虚拟假设的证据就越强。因此，当我们说一个 p 值低或合理的低，那就意味着它低于显著性水平 10%，5% 或 1%。在此决策中有些个人的判断在其中。

例 8.2　立方成本函数再议

回顾 7.10 节所估的立方总成本函数，为方便起见，将其重写如下：

$$\hat{Y}_i = 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3$$
$$\text{se} = (6.3753)\quad(4.7786)\quad(0.9857)\quad(0.0591)\qquad(7.10.6)$$
$$\text{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.0576;\quad R^2 = 0.9983$$

其中 Y 为总成本，而 X 为产出。括号中的数字代表估计的标准误。

假如我们要检验假设：立方成本函数中的 X^2 和 X^3 项的系数相等，即 $\beta_3 = \beta_4$ 或 $(\beta_3 - \beta_4) = 0$。在回归 (7.10.6) 中，我们已具备用于 (8.6.5) 的 t 检验的全部必需数字结果。具体的演算步骤如下：

$$\begin{aligned} t &= \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}} \\ &= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} \\ &= \frac{-13.9011}{1.0442} \\ &= -13.3130 \end{aligned} \qquad (8.6.6)$$

读者可以验证，对 6 个自由度（为什么？），即使在 0.002（或 0.2%）显著性水平上（双尾检验），所测 t 值也超过了临界 t 值；p 值极端地小，仅 0.000 006。从而我们拒绝假设：立方成本函数中的 X^2 和 X^3 有相同的系数。

§8.7　受约束的最小二乘法：检验线性等式约束条件

经济理论有时会提出某一回归模型中的系数满足一些线性等式约束条件。例如，考虑柯布-道格拉斯生产函数：

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \qquad (7.9.1) = (8.7.1)$$

其中 Y = 产出，X_2 = 劳力投入，X_3 = 资本投入。写成对数形式，方程就变为：

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \qquad (8.7.2)$$

其中 $\beta_0 = \ln\beta_1$。

267 现在如果是规模报酬不变（每一同比例的投入变化有同比例的产出变化），经济理论将提出：

$$\beta_2+\beta_3=1 \tag{8.7.3}$$

这就是线性等式约束之一例。[12]

怎样判知规模报酬是否不变，即约束条件（8.7.3）是否正确？有两种方法。

t 检验方法

最简单的程序，是先不明显考虑约束条件（8.7.3），而按照通常的方式估计（8.7.2）。即做所谓**无约束或无限制的回归**（unrestricted or unconstrained regression）。一旦估计了 β_2 和 β_3（比方说，用 OLS 法），就可通过（8.6.3）的 t 检验来检验假设或约束（8.7.3），即：

$$\begin{aligned} t &= \frac{(\hat\beta_2+\hat\beta_3)-(\beta_2+\beta_3)}{\mathrm{se}(\hat\beta_2+\hat\beta_3)} \\ &= \frac{(\hat\beta_2+\hat\beta_3)-1}{\sqrt{\mathrm{var}(\hat\beta_2)+\mathrm{var}(\hat\beta_3)+2\mathrm{cov}(\hat\beta_2\hat\beta_3)}} \end{aligned} \tag{8.7.4}$$

其中在虚拟假设下，$(\beta_2+\beta_3)=1$，而分母是$(\hat\beta_2+\hat\beta_3)$的标准误。然后仿照8.6节，如果从（8.7.4）计算的 t 值超过在选定显著性水平上的临界 t 值，则拒绝不变规模报酬假设；否则不拒绝它。

F 检验法：受约束最小二乘法

前述 t 检验是一种静观后效法（postmortem examination）。因为我们是在估计“无约束”回归之后再看线性约束是否被满足。一种直接方法则是一开始便把约束条件（8.7.3）纳入估计过程中。在本例中，这种过程是不难实现的。由（8.7.3）可得：

$$\beta_2=1-\beta_3 \tag{8.7.5}$$

或：

$$\beta_3=1-\beta_2 \tag{8.7.6}$$

因此，利用两等式之一便可消去（8.7.2）中的一个 β 系数，然后估计所得的方程。例如，我们利用（8.7.5）把柯布-道格拉斯生产函数写为：

$$\begin{aligned} \ln Y_i &= \beta_0+(1-\beta_3)\ln X_{2i}+\beta_3\ln X_{3i}+u_i \\ &= \beta_0+\ln X_{2i}+\beta_3(\ln X_{3i}-\ln X_{2i})+u_i \end{aligned}$$

268 或：

$$(\ln Y_i-\ln X_{2i})=\beta_0+\beta_3(\ln X_{3i}-\ln X_{2i})+u_i \tag{8.7.7}$$

或：

$$\ln(Y_i/X_{2i}) = \beta_0 + \beta_3\ln(X_{3i}/X_{2i}) + u_i \tag{8.7.8}$$

其中（Y_i/X_{2i}）＝产出/劳动力比率，（X_{3i}/X_{2i}）＝资本/劳动力比率，两者都是有重大经济意义的数量。

留意原始方程经过了怎样的变换。一旦我们从（8.7.7）或（8.7.8）估计出 β_3，β_2 就容易从关系式（8.7.5）算出。不言而喻，这种估计程序保证了所估计的两投入系数之和必然等于 1。（8.7.7）或（8.7.8）所描述的程序被称为**受约束的最小二乘法**（restricted least squares，RLS）。此程序可推广到含有任意多个解释变量，以及有多于一个线性等式约束的模型。推广方法见于泰尔（Theil，1971）。[13]（还可参见下面的一般 F 检验法。）

怎样比较无约束的和受约束的两个最小二乘回归呢？换句话说，我们怎样知道，比方说，约束（8.7.3）是否真实？这个问题可通过以下的 F 检验的应用加以检验。令：

$$\sum \hat{u}_{\mathrm{UR}}^2 = \text{无约束回归(8.7.2)的 RSS}$$
$$\sum \hat{u}_{\mathrm{R}}^2 = \text{受约束回归(8.7.7)的 RSS}$$
$$m = \text{线性约束个数（本例中是 1）}$$
$$k = \text{无约束回归中的参数个数}$$
$$n = \text{观测次（个）数}$$

于是，

$$\begin{aligned} F &= \frac{(\mathrm{RSS_R} - \mathrm{RSS_{UR}})/m}{\mathrm{RSS_{UR}}/(n-k)} \\ &= \frac{(\sum \hat{u}_{\mathrm{R}}^2 - \sum \hat{u}_{\mathrm{UR}}^2)/m}{\sum \hat{u}_{\mathrm{UR}}^2/(n-k)} \end{aligned} \tag{8.7.9}$$

遵循 m，$(n-k)$ 个自由度的 F 分布。（注：UR 和 R 分别表示无约束和受约束。）

上述 F 检验还可通过 R^2 表示如下：

$$F = \frac{(R_{\mathrm{UR}}^2 - R_{\mathrm{R}}^2)/m}{(1 - R_{\mathrm{UR}}^2)/(n-k)} \tag{8.7.10}$$

269 其中 R_{UR}^2 和 R_{R}^2 分别是得自无约束和受约束回归的 R^2 值。即得自（8.7.2）和（8.7.7）的 R^2 值。应注意到：

$$R_{\mathrm{UR}}^2 \geqslant R_{\mathrm{R}}^2 \tag{8.7.11}$$

及：

$$\sum \hat{u}_{\mathrm{UR}}^2 \leqslant \sum \hat{u}_{\mathrm{R}}^2 \tag{8.7.12}$$

习题 8.4 要求你对这些陈述做出解释。

提醒注意：在使用（8.7.10）时，要记住，如果在受约束和无约束的两个模型中因变量不相同，则 R^2_{UR}和 R^2_R 不可直接相比，这时，可用第 7 章所述的程序把两个 R^2 值转化为可比的（参看下面的例 8.3），或使用（8.7.9）给出的 F 检验。

例 8.3　1955—1974 年墨西哥经济的柯布-道格拉斯生产函数

借助说明上述讨论来考虑表 8.8 中给出的数据。尝试对这些数据拟合柯布-道格拉斯生产函数得到如下结果：

$$
\begin{aligned}
&\widehat{\ln\text{GDP}}_t = -1.652\,4 + 0.339\,7\ln\text{Labor}_t + 0.846\,0\ln\text{Capital}_t\\
&t = (-2.725\,9)\quad(1.829\,5)\quad(9.062\,5)\\
&p\ \text{值} = (0.014\,4)\quad(0.084\,9)\quad(0.000\,0)\\
&R^2 = 0.995\,1\qquad \text{RSS}_{\text{UR}} = 0.013\,6
\end{aligned}
\tag{8.7.13}
$$

其中 RSS_{UR}因我们在估计（8.7.13）时没有施加限制而成为无约束的 RSS。

我们在第 7 章已经看到，如何解释柯布-道格拉斯生产函数的系数。如你所见，产出/劳动的弹性约为 0.34，而产出/资本的弹性约为 0.85。如果我们把这些系数相加则得到 1.19，表明考察期内墨西哥经济可能正经历着规模报酬递增的阶段。当然我们不知道 1.19 是否显著异于 1。

为看出是否如此，让我们施加规模报酬不变的约束，并给出如下回归：

$$
\begin{aligned}
&\widehat{\ln(\text{GDP}/\text{Labor})}_t = -0.494\,7 + 1.015\,3\ln(\text{Capital}/\text{Labor})_t\\
&t = (-4.061\,2)\qquad(28.105\,6)\\
&p\ \text{值} = (0.000\,7)\qquad(0.000\,0)\\
&R^2_R = 0.977\,7\qquad \text{RSS}_{\text{R}} = 0.016\,6
\end{aligned}
\tag{8.7.14}
$$

270　其中 RSS_{R} 为约束 RSS，因为我们已经施加了规模报酬不变的限制。

由于上面两个回归的因变量不同，所以我们必须使用（8.7.9）中给出的 F 检验：我们有得到 F 值所需要的数据。

$$
\begin{aligned}
F &= \frac{(\text{RSS}_{\text{R}} - \text{RSS}_{\text{UR}})/m}{\text{RSS}_{\text{UR}}/(n-k)}\\
&= \frac{(0.016\,6 - 0.013\,6)/1}{(1-0.013\,6)/(20-3)}\\
&= 0.051\,7
\end{aligned}
$$

注：因为我们只施加了一个约束，所以 $m=1$；而因为我们有 20 个观测，且在无约束回归中有 3 个参数，所以 $n-k=17$。

此 F 值服从分子自由度为 1 和分母自由度为 17 的 F 分布。读者很容易验证，即使在 25% 的显著性水平上，这个 F 值仍不显著。（见附录 D 的表 D.3。）

于是结论就是，墨西哥经济在样本期内可能仍具有规模报酬不变的特征，因此采用（8.7.14）中给出的约束回归没有害处。此回归表明，若资本劳动比提高 1%，则劳动生产率也平均上升 1%。

表 8.8　　墨西哥的真实 GDP、就业和真实固定资本

年份	GDP*	就业†	固定资本‡
1955	114 043	8 310	182 113
1956	120 410	8 529	193 749
1957	129 187	8 738	205 192
1958	134 705	8 952	215 130
1959	139 960	9 171	225 021
1960	150 511	9 569	237 026
1961	157 897	9 527	248 897
1962	165 286	9 662	260 661
1963	178 491	10 334	275 466
1964	199 457	10 981	295 378
1965	212 323	11 746	315 715
1966	226 977	11 521	337 642
1967	241 194	11 540	363 599
1968	260 881	12 066	391 847
1969	277 498	12 297	422 382
1970	296 530	12 955	455 049
1971	306 712	13 338	484 677
1972	329 030	13 738	520 553
1973	354 057	15 924	561 531
1974	374 977	14 154	609 825

* 1960 年的百万比索。

† 千人。

‡ 1960 年的百万比索。

资料来源：Victor J. Elias, *Sources of Growth: A Study of Seven Latin American Economies*, International Center for Economic Growth, ICS Press, San Francisco, 1992. Data from Tables E5, E12 and E14.

一般的 F 检验方法[14]

271 （8.7.10）中的 F 检验或与它等效的（8.7.9），为检验有关 k 变量回归模型

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i \tag{8.7.15}$$

中的一或多个参数的假设，提供了一般的方法。（8.5.16）的 F 检验或（8.6.3）的 t 检验不过是（8.7.10）的一个应用特例。例如，如同：

$$H_0:\ \beta_2=\beta_3 \tag{8.7.16}$$

$$H_0:\ \beta_3+\beta_4+\beta_5=3 \tag{8.7.17}$$

这样的涉及 k 变量模型的参数的一些线性约束的假设，或者如同：

$$H_0:\ \beta_3=\beta_4=\beta_5=\beta_6=0 \tag{8.7.18}$$

这样意味着某些回归元在模型中并不出现的假设，却可通过（8.7.10）的 F 检验方法来检验。

从 8.5 节和 8.7 节的讨论，读者一定已察知 F 检验方法的一般策略是：先是有一个较大的模型，如无约束模型（8.7.15），然后，通过从中删除某些变量如（8.7.18），或通过对较大模型中的一或多个参数加以某种线性约束如（8.7.16）或（8.7.17），而有一个较小的受约束或较大的受限制模型。

然后，分别用无约束和受约束模型去拟合数据，以获得判定系数 R^2_{UR} 和 R^2_{R}。注意，无约束模型的自由度为 $n-k$，而对受约束模型，自由度为 m，这里 m 为线性约束［如（8.7.16）或（8.7.17）］个数，或为从模型中略去的回归元个数。［例如，$m=4$，如果（8.7.18）成立，则因为后者假定了 4 个回归元不在模型中出现。］然后按照（8.7.10）计算 F，并使用决策规则：如果计算的 F 超过 $F_\alpha(m,n-k)$，这里 $F_\alpha(m,n-k)$ 是显著性水平为 α 的临界 F 值，我们就拒绝虚拟假设；否则不拒绝它。

272 让我们来说明。

例 8.4 1960—1982 年美国子鸡需求

习题 7.19 中的一问是要求读者考虑以下对子鸡的需求函数：

$$\ln Y_t=\beta_1+\beta_2\ln X_{2t}+\beta_3\ln X_{3t}+\beta_4\ln X_{4t}+\beta_5\ln X_{5t}+u_t \tag{8.7.19}$$

其中 Y＝每人子鸡消费量（磅），X_2＝每人实际可支配收入（美元），X_3＝每磅子鸡实际零售价格（美分），X_4＝每磅猪肉实际零售价格（美分），X_5＝每磅牛肉实际零售价格（美分）。

在此模型中 β_2，β_3，β_4 和 β_5 分别是收入、自价格、交叉价格（猪肉）

和交叉价格（牛肉）弹性。（为什么？）根据经济理论，

$$\beta_2 > 0$$

$$\beta_3 < 0$$

$\beta_4 > 0$，如果鸡和猪肉是**替代**（competing）产品

< 0，如果鸡和猪肉是**互补**（complementary）产品　　(8.7.20)

$= 0$，如果鸡和猪肉是无关产品

$\beta_5 > 0$，如果鸡和牛肉是替代产品

< 0，如果鸡和牛肉是互补产品

$= 0$，如果鸡和牛肉是无关产品

假使某人执意认为鸡与猪肉和牛肉为无关产品，即鸡的消费不受猪肉和牛肉价格的影响。简单地表示，即：

$$H_0:\ \beta_4 = \beta_5 = 0 \tag{8.7.21}$$

从而有受约束回归：

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t \tag{8.7.22}$$

当然，方程（8.7.19）是无约束回归。

利用习题 7.19 中的数据，我们得到以下：

无约束回归：

$$\begin{aligned}\widehat{\ln Y_t} = {} & 2.189\,8 + 0.342\,5 \ln X_{2t} - 0.504\,6 \ln X_{3t} + 0.148\,5 \ln X_{4t} + 0.091\,1 \ln X_{5t} \\ & (0.155\,7) \qquad (0.083\,3) \qquad (0.110\,9) \qquad (0.099\,7) \qquad (0.100\,7) \\ & R^2_{\mathrm{UR}} = 0.982\,3\end{aligned} \tag{8.7.23}$$

受约束回归：

$$\begin{aligned}\widehat{\ln Y_t} = {} & 2.032\,8 + 0.451\,5 \ln X_{2t} - 0.377\,2 \ln X_{3t} \\ & (0.116\,2) \qquad (0.024\,7) \qquad (0.063\,5) \\ & R^2_{\mathrm{R}} = 0.980\,1\end{aligned} \tag{8.7.24}$$

273 其中括号内的数字是估计的标准误。注：因两模型有相同的因变量，故（8.7.23）和（8.7.24）的两个 R^2 值是可比的。

现在检验假设（8.7.21）的 F 比率是：

$$F = \frac{(R^2_{\mathrm{UR}} - R^2_{\mathrm{R}})/m}{(1 - R^2_{\mathrm{UR}})/(n-k)} \tag{8.7.10}$$

因为在本例中涉及两个约束：$\beta_4 = 0$ 和 $\beta_5 = 0$，故 m 的值为 2，而分母自由度 $(n-k) = 23 - 5 = 18$（5 个 β 系数），因此 F 比率是：

$$\begin{aligned}F &= \frac{(0.982\,3 - 0.980\,1)/2}{(1 - 0.982\,3)/18} \\ &= 1.122\,4\end{aligned} \tag{8.7.25}$$

它遵从自由度为 2 和 18 的 F 分布。

显然，在 5% 水平上这个 F 值是统计上不显著的 $[F_{0.5}(2,18) = 3.55]$。

p 值是 0.347 2。因此没有理由拒绝虚拟假设——对子鸡的需求不依赖于猪肉和牛肉价格。简言之，我们可以接受受约束回归（8.7.24）作为子鸡需求函数的表达式。

注意，在自价弹性为负和收入弹性为正的意义上，需求函数符合先验的经济预期。然而，估计的价格弹性在绝对值上统计地小于 1，这意味着子鸡需求是缺乏价格弹性的（price inelastic）。（为什么？）而且，收入弹性虽是正数，仍然统计地小于 1。这表明子鸡不是奢侈品；按照惯例，如果一种商品的收入弹性大于 1，它就被称作奢侈品。

§8.8 检验回归模型的结构或参数稳定性：邹至庄检验

在我们使用一个涉及时间序列数据的回归时，回归子 Y 和回归元之间的关系可能会出现**结构变动**（structural change）。结构变动意味着，模型中的参数值在整个期间内不能保持相同。结构变动有时源于外部力量（如 1973 年和 1979 年 OPEC 石油卡特尔提出的石油涨价或 1990—1991 年的海湾战争），或源于政策变化（如 1973 年期间从固定汇率制向浮动汇率制的转换），或国会所采取的行动（如里根总统在其两任任期内的税收变化或最小工资率的变化），或一系列其他原因。

我们如何发现结构变动确实存在？具体而言，考虑表 8.9 中给出的数据。此表给出了美国 1970—1995 年个人可支配收入和个人储蓄的数据（以 10 亿美元计）。假使我们想估计储蓄（Y）与个人可支配收入 DPI（X）之间
274
的简单储蓄函数。既然我们有数据，那就能得到 Y 对 X 的 OLS 回归。但如果我们那么做，那我们就认为储蓄和 DPI 的关系在 26 年间没多大变化。这是一个难以置信的假定。比如，众所周知，美国 1982 年遭受了其和平时期最大的衰退，即城市失业率当年达到了自 1948 年以来最高水平 9.7%。这种事件可能会破坏储蓄和 DPI 之间的关系。为看出是否如此，我们把样本数据分为两个时期：1970—1981 年的衰退前时期和 1982—1995 年的衰退后时期。

我们现在有三个可能的回归：

时期 1970—1981： $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$ （8.8.1）

时期 1982—1995： $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$ （8.8.2）

时期 1970—1995： $Y_t = \alpha_1 + \alpha_2 X_t + u_t \quad n = (n_1 + n_2) = 26$ （8.8.3）

回归（8.8.3）假定这两个时期之间没有区别，因此对 26 个观测构成的整个时期估计储蓄和 DPI 之间的关系。换言之，此回归假定截距和斜率系数在整个期间保持不变；即不存在结构变动。若确实如此，则 $\alpha_1 = \lambda_1 = \gamma_1$，

$\alpha_2 = \lambda_2 = \gamma_2$。

表 8.9　美国 1970—1995 年储蓄和个人可支配收入（10 亿美元计）

观测	储蓄	收入	观测	储蓄	收入
1970	61.0	727.1	1983	167.0	2 522.4
1971	68.6	790.2	1984	235.7	2 810.0
1972	63.6	855.3	1985	206.2	3 002.0
1973	89.6	965.0	1986	196.5	3 187.6
1974	97.6	1 054.2	1987	168.4	3 363.1
1975	104.4	1 159.2	1988	189.1	3 640.8
1976	96.4	1 273.0	1989	187.8	3 894.5
1977	92.5	1 401.4	1990	208.7	4 166.8
1978	112.6	1 580.1	1991	246.4	4 343.7
1979	130.1	1 769.5	1992	272.6	4 613.7
1980	161.8	1 973.3	1993	214.4	4 790.2
1981	199.1	2 200.2	1994	189.4	5 021.7
1982	205.5	2 347.3	1995	249.3	5 320.8

资料来源：*Economic Report of the President*，1997，Table B-28，p.332.

回归（8.8.1）和（8.8.2）假定这两个时期的回归不同，即截距和斜率参数如带下标的参数所示都不相同。在上述回归中，u 都表示误差项，n 都表示观测次数。

275 针对表 8.9 中给出的数据，上述三个回归的经验结果如下：

$$\hat{Y}_t = 1.016\,1 + 0.080\,3X_t$$
$$t = \quad (0.087\,3)(9.601\,5)$$
$$R^2 = 0.902\,1 \quad \mathrm{RSS}_1 = 1\,785.032 \quad \mathrm{df} = 10 \tag{8.8.1a}$$
$$\hat{Y}_t = 153.494\,7 + 0.014\,8X_t$$
$$t = \quad (4.692\,2) \quad (1.770\,7)$$
$$R^2 = 0.297\,1 \quad \mathrm{RSS}_2 = 10\,005.22,\ \mathrm{df} = 12 \tag{8.8.2a}$$
$$\hat{Y}_t = 62.422\,6 + 0.037\,6X_t + \cdots$$
$$t = \quad (4.8917) \quad (8.893\,7) + \cdots$$
$$R^2 = 0.767\,2 \quad \mathrm{RSS}_3 = 23\,248.30,\ \mathrm{df} = 24 \tag{8.8.3a}$$

在上述回归中，RSS 表示残差平方和，括号中的数字都是估计的 t 值。

粗看之下，所估计的回归表明储蓄和 DPI 之间的关系在这两个子时期并不相同。上述储蓄—收入回归中的斜率表示**边际储蓄倾向**（marginal propensity to save, MPS)，即个人可支配收入增加一美元导致储蓄的（平均）变化。在1970—1981年期间，MPS约为0.08，但在1982—1995年期间，MPS约为0.02。这种变化是否源于里根总统所追求的经济政策很难说。这进一步表明，无视两个时期的差异而将26个观测放在一起做一个通常的回归，**即混合回归**（pooled regression）（8.8.3a）可能不适当。当然，上述判断仍需要由适当的统计检验来支持。顺便指出，散点图和估计的回归线如图8.3所示。

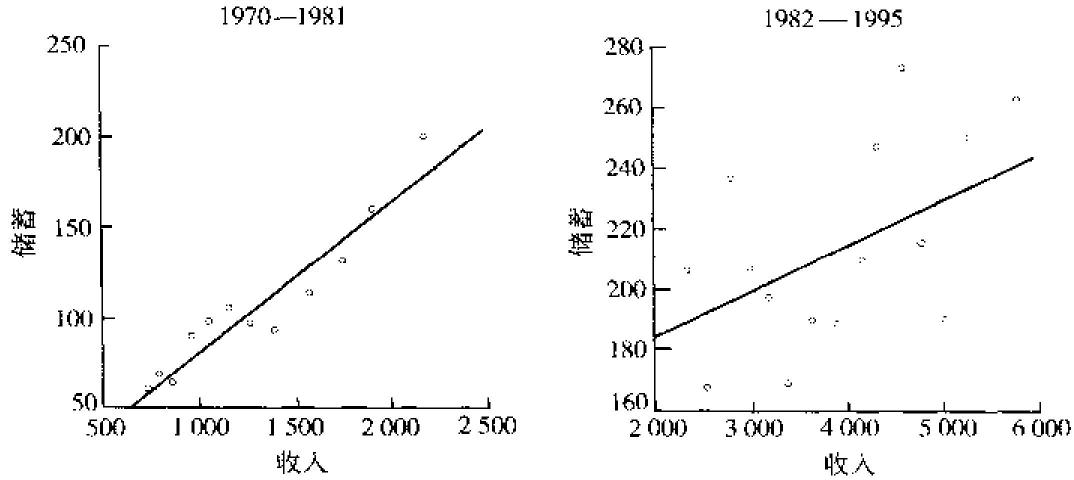


图 8.3

现在可能的区别即结构变动可能因截距或斜率或二者共同所致。我们如何找出其差别呢？从图8.3中可对此得到直觉上的结论，但规范的检验会更有帮助。

邹至庄检验（Chow test）适逢所需。[15] 该检验假定：

1. $u_{1t} \sim N(0, \sigma^2)$和$u_{2t} \sim N(0, \sigma^2)$，就是说，两个子期间回归的误差项是有相同方差$\sigma^2$的（同方差性）正态分布变量。

276 2. 两个误差项u_{1t}和u_{2t}是独立分布的。[18]

邹至庄检验的机制如下：

1. 估计回归（8.8.3）（若无参数不稳定性，则为适当）并得到RSS_3，其自由度为(n_1+n_2-k)，其中k为所估参数的个数，在本例中为2。对我们的例子而言，$RSS_3 = 23\,248.30$，我们称RSS_3为**约束残差平方和**（restricted residual sum of squares, RSS_R），因为它是施加了$\lambda_1=\gamma_1$和$\lambda_2=\gamma_2$（即子期间回归没有不同）的约束后得到。

2. 估计（8.8.1）并在$df=(n_1-k)$下得到其残差平方和RSS_1。本例中$RSS_1 = 1\,785.032$，$df=10$。

3. 估计（8.8.2）并在$df=(n_2-k)$下得到其残差平方和RSS_2。本例中$RSS_2 = 10\,005.22$，$df=12$。

4. 既然这两个样本集被视为独立，所以我们就能把RSS_1和RSS_2相加得到所谓**无约束残差平方和**（unrestricted residual sum of squares, RSS_{UR}），即

$$RSS_{UR} = RSS_1 + RSS_2 \quad df=(n_1+n_2-2k)$$

在本例中，$RSS_{UR}=(1\ 785.032+10\ 005.22)=11\ 790.252$

277 5. 可见藏在邹至庄检验背后的思想是，若不存在结构变动［即回归(8.8.1)和(8.8.2)实质相同］，则 RSS_R 和 RSS_{UR} 在统计上不应该不同。因此，若我们构造如下比率：

$$F=\frac{(RSS_R-RSS_{UR})/k}{(RSS_{UR})/(n_1+n_2-2k)}\sim F_{[k,(n_1+n_2-2k)]} \tag{8.8.4}$$

邹至庄已经证明，在回归(8.8.1)和(8.8.2)(在统计上)相同(即没有结构变动或转折)的虚拟假设下，以上给出的 F 比率服从分子和分母自由度分别为 k 和 (n_1+n_2-2k) 的 F 分布。

6. 因此，若在应用中计算出的 F 值没有超过 F 表中在选定显著性水平(或 p 值)上的临界 F 值，则不能拒绝参数稳定(parameter stability)(即没有结构变动)的虚拟假设。此时用混合(约束)回归(8.8.3)就是合理的。相反，若计算出的 F 值超过了临界 F 值，则拒绝参数稳定的假设，并断定回归(8.8.1)和(8.8.2)是不同的，此时混合回归(8.8.3)至少是没有把握的。

回到我们的例子中，我们发现

$$F=\frac{(23\ 248.30-11\ 790.252)/2}{(11\ 790.252)/22}=10.69 \tag{8.8.5}$$

我们在 F 表中发现，自由度为2和22的1%临界 F 值为7.72。因此，得到一个大于等于10.69的 F 值的概率远小于1%；实际的 p 值只有0.000 57。

因此，邹至庄检验看起来支持我们前面的预感，假设该检验背后的假定都满足，美国1970—1995年期间的储蓄—收入关系已经历了一次结构变动。稍后我们对此还有补充。

顺便一提，注意邹至庄检验可轻而易举地推广到不止一次结构变动的情形。比如，若我们相信，储蓄—收入关系在克林顿总统1992年1月入主白宫后发生了变化，那我们就会把样本分为三个时期：1970—1981，1982—1991，1992—1995，并进行邹至庄检验。当然，我们将有4个RSS项，每个子期间一个，混合数据一个。但检验的逻辑仍然是一样的。直至2001年的数据现在都可以利用，所以最后一个期间可延伸到2001年。

必须牢记关于邹至庄检验的一些警告：

1. 必须满足该检验背后的假定。比如，必须弄清楚回归(8.8.1)和(8.8.2)中的误差方差是否相同。我们稍后再谈。

2. 邹至庄检验只告诉我们回归(8.8.1)和(8.8.2)是否有差别，并没有告诉我们差别是来自截距、斜率、还是二者都有。但在讨论虚拟变量的第9章，我们将看到如何回答这个问题。

278 3. 邹至庄检验假定我们知道结构转折点。在我们的例子中，我们假定是1982年。但若不能确定结构变动何时发生，我们就必须用其他方法。[16]

在我们离开邹至庄检验和储蓄—收入回归之前，让我们考察邹至庄检验背后的一个假定，即两个时期的误差方差相同。因为我们不能观测到真实的

误差方差，所以我们能从回归（8.8.1a）和（8.8.2a）中给出的 RSS 来得到它们的估计值，即

$$\hat{\sigma}_1^2=\frac{\mathrm{RSS}_1}{n_1-2}=\frac{1\,785.032}{10}=178.503\,2 \tag{8.8.6}$$

$$\hat{\sigma}_2^2=\frac{\mathrm{RSS}_2}{n_2-2}=\frac{10\,005.22}{14-2}=833.768\,3 \tag{8.8.7}$$

注意，由于每个方程中都有两个估计参数，所以我们从观测数中减去 2 得到 df。给定邹至庄检验背后的假定，则 $\hat{\sigma}_1^2$ 和 $\hat{\sigma}_2^2$ 为两个子期间真实方差的无偏估计量。因此，可以证明，若 $\sigma_1^2=\sigma_2^2$（即两个子总体的方差相同，如邹至庄检验所假定的那样），则

$$\frac{(\hat{\sigma}_1^2/\sigma_1^2)}{(\hat{\sigma}_2^2/\sigma_2^2)}\sim F_{(n_1-k),(n_2-k)} \tag{8.8.8}$$

服从分子和分母自由度分别为（n_1-k）和（n_2-k）的 F 分布。在我们的例子中，由于在每个子回归中都只有两个参数，所以 $k=2$。

当然，$\sigma_1^2=\sigma_2^2$ 使上述 F 检验简化为计算

$$F=\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \tag{8.8.9}$$

注：按惯例，我们把两个估计方差中较大的一个放在分子中。（关于 F 和其他概率分布的详细知识，可参见附录 A。）

在应用中计算这个 F 值并与适当 df 的临界 F 值相比较，就能决定是否拒绝两个子总体的方差相同的虚拟假设。若虚拟假设未被拒绝，则可以使用邹至庄检验。

回到我们的储蓄—收入回归，我们得到如下结果：

$$F=\frac{833.768\,3}{178.503\,2}=4.670\,1 \tag{8.8.10}$$

279 在两个子总体的方差相等的虚拟假设之下，此 F 值服从分子和分母自由度分别是 12 和 10 的 F 分布。（注：我们已经将较大的估计方差放在分子上。）从附录 D 中的 F 表看到，自由度为 12 和 10 的 5% 和 1% 临界 F 值分别是 2.91 和 4.71。计算出来的 F 值在 5% 的显著性水平上是显著的，在 1% 的水平上也几乎是显著的。于是，我们的结论将是两个子总体方差并不相同，因此，严格地讲，我们不应该使用邹至庄检验。

我们在这里的目的是要说明应用研究中经常用到的邹至庄检验的机制。若两个子总体的误差方差不同，则邹至庄检验可进行修正。但这种程序超出了本书的范围。[17]

我们前面提到的另外一点是，邹至庄检验对回归参数可能发生变化的时间的选择十分敏感。我们在例子中假定变化可能发生在出现衰退的 1982 年。如果我们当时假定它是罗纳德·里根开始执政的 1981 年，那我们可能会发现，计算出来的 F 值不同。事实上，习题 8.34 要求读者来验证这一点。

如果我们不想选择结构关系可能出现转折的时点，那我们可以选用其他方法，如**递归残差检验**（recursive residual test）。在有关模型设定分析的第13章，我们将讨论这个问题。

§8.9 用多元回归做预测

在5.10节我们曾说明怎样能用双变量回归模型（1）作均值预测（mean prediction），即预测总体回归函数（或总体回归线）上的点，以及（2）作个值预测（individual prediction），即对回归元 X 的特定值 $X = X_0$ 预测 Y 的个值。

多元回归的估计结果也可用于同样目的，并且预测程序明显地是双变量情形的一个直接推广，只不过用于估计预测值的方差或标准误的公式［与（5.10.2）和（5.10.6）比较］更为复杂，故宜于放到讲矩阵方法的附录C中处理。当然，大多数标准回归软件包可例行做到这一点，所以没有必要查找矩阵表述。附录C对喜欢数学的学生有好处，并给出了一个完整的示例。

* §8.10 假设检验三联体：似然比，瓦尔德与拉格朗日乘数检验[21]

280 大体上说，在本章和前些章里，我们曾用 t、F 和 χ^2 检验对线性（对参数言）回归范围内的各种假设进行了检验。但是我们一旦越出线性回归模型的这个多少是如意的世界，我们还需要有检验关于线性或非线性回归模型的假设检验方法。

著名的三位一体：**似然比**（likelihood ratio）、**瓦尔德**（Wald）**和拉格朗日乘数**（Lagrange multipler）**检验的使用能达到此目的**。值得注意的是，这三种检验在渐近（即大样本）意义下都是等价的。因为每一种检验的检验统计量都遵循 χ^2 分布。

虽然我们将在本章附录中讨论**似然比检验**，但以实用为由，本书将不使用这些检验。因为，大多数研究者遇到的样本，不幸都是小样本或有限样本，而对于小样本，我们至今一直在使用的 F 检验已经够用了。且看戴维森（Davidson）和麦金农（MacKinnon）是怎样说的：

> 对于线性回归模型，不管它的误差是或不是正态分布的，当然都不需要过问LM，W和LR。因为我们不能从这些统计量中得到任何不为 F 所含有的信息。[19]

§8.11 检验回归的函数形式：在线性与对数线性回归模型之间进行选择

选择线性回归模型（回归子是诸回归元的线性函数）或对数线性回归模型（回归子的对数是诸回归元的对数的线性函数）是经验分析中由来已久的一个问题。我们可用麦金农，怀特和戴维森提出的一种检验，简称它为 **MWD 检验**（MWD Test），在上述两个模型之间进行选择。[20]

为说明这种检验，假定：

H_0：线性模型：Y 是诸回归元 X 的线性函数。

H_1：对数线性模型：$\ln Y$ 是诸回归元的对数（即诸 X 的对数）的线性函数。

其中，和平常一样，H_0 和 H_1 指虚拟和对立假设。

281 MWD 检验可分以下几个步骤[21]：

步骤 1： 估计线性模型并获得 Y 的估计值，且计为 Yf（即 $\hat{Y}$）。

步骤 2： 估计对数线性模型并获得 $\ln Y$ 的估计值，且记为 $\ln f$（即 $\widehat{\ln Y}$）。

步骤 3： 算出 $Z_1 = (\ln Yf - \ln f)$。

步骤 4： 做 Y 对诸 X 和得自步骤 3 的 Z_1 的回归。如果按通常的 t 检验 Z_1 的系数是统计上显著的，就拒绝 H_0。

步骤 5： 算出 $Z_2 =$（$\ln f$ 的反对数 $- Yf$）。

步骤 6： 做 Y 的对数对诸 X 的对数和 Z_2 的回归。如果按通常的检验 Z_2 的系数是统计上显著的，就拒绝 H_1。

MWD 检验虽然看似复杂，其实这个检验的逻辑很简单，如果线性模型是事实上正确的模型，第 4 步的构造变量就不会是统计上显著的，因为这时从线性模型估计出的 Y 值和从对数线性模型估计出来的（为了比较而取反对数之后的）就不会有什么差别，同样的评语也适用于对立假设 H_1。

例 8.5 玫瑰需求

参照习题 7.16 所给的 1971 年第 2 季度至 1975 年第 2 季度底特律市区对玫瑰的需求季度数据，为了说明上的便利，我们将把对玫瑰的需求看作仅仅是玫瑰和荷兰石竹两种价格的函数，而暂不考虑收入变量。现在考虑以下模型：

线性模型： $$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_t \tag{8.11.1}$$

对数线性模型： $\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_t$ (8.11.2)

其中 Y 是玫瑰数量（打），X_2 是玫瑰平均批发价格（美元/打），而 X_3 是荷兰石竹平均批发价格（美元/打）。先验地，预期 α_2 和 β_2 为负（为什么?），而 α_3 和 β_3 为正。（为什么?）我们知道，对数线性模型中的斜率系数是弹性系数。

回归结果如下：

$$
\begin{aligned}
\hat{Y}_t &= 973\,4.217\,6 - 378\,2.195\,6X_{2t} + 2\,815.251\,5X_{3t} \\
t &= \quad (3.370\,5) \qquad (-6.606\,9) \qquad (2.971\,2) \qquad\qquad (8.11.3) \\
&\qquad F = 21.84; \quad R^2 = 0.770\,96
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\ln Y_t} &= 9.227\,8 - 1.760\,7\ln X_{2t} + 1.339\,8\ln X_{3t} \\
t &= \quad (16.234\,9) \quad (-5.904\,4) \quad (2.540\,7) \qquad\qquad (8.11.4) \\
&\qquad F = 17.50; \quad R^2 = 0.729\,2
\end{aligned}
$$

282 这些结果表明，似乎线性和对数线性两模型均对数据拟合良好：参数有预期符号且 t 和 R^2 值均在统计上显著。

为要根据 **MWD 检验**在两个模型之间做出选择，我们先检验真实模型是线性的假设。然后按照检验的步骤 4，算得以下回归：

$$
\begin{aligned}
\hat{Y}_t &= 9\,727.568\,5 - 3\,783.062\,3X_{2t} + 2\,817.715\,7X_{3t} + 85.231\,9Z_{1t} \\
t &= \quad (3.217\,8) \quad (-6.333\,7) \qquad (2.836\,6) \qquad (0.020\,7) \\
&\qquad F = 13.44; \qquad R^2 = 0.770\,7 \qquad\qquad (8.11.5)
\end{aligned}
$$

因 Z_1 的系数在统计上不显著（p 值是 0.98），故我们不拒绝真模型是线性的假设。

假使我们调换一下假设。假设真模型是对数线性的。按照 MWD 检验的步骤 6，得到以下回归结果：

$$
\begin{aligned}
\widehat{\ln Y_t} &= 9.148\,6 - 1.969\,9\ln X_t + 1.589\,1\ln X_{2t} - 0.001\,3Z_{2t} \qquad (8.11.6) \\
t &= \quad (17.082\,5) \quad (-6.418\,9) \qquad (3.072\,8) \quad (-1.661\,2) \\
&\qquad F = 14.17; \quad R^2 = 0.779\,8
\end{aligned}
$$

Z_2 的系数约在 12% 的水平上统计上显著（p 值是 0.122 5），因此可在这一显著性水平上拒绝真模型是对数线性的假设。当然，如果我们要墨守成规地引用 1% 或 5% 显著性水平，则还不能拒绝真实模型是对数线性的假设。本例表明，在一定情况下，有可能对任一模型都不能拒绝。

§8.12 要点与结论

1. 本章推广并细致地分析了最先在第 5 章中对双变量线性回归模型引进的区间估计与假设检验的思想。

2. 在一个多元回归中，检验一个偏回归系数的个别显著性（用 t 检验）和

检验回归的总显著性（即 H_0：全部偏斜率系数为零或 $R^2=0$）是不相同的。

3. 特别是，在个别 t 检验的基础上发现了一或多个偏回归系数在统计上不显著，并不意味着全部多个偏回归系数也是（集体地）在统计上不显著的。后一假设只能用 F 统计量加以检验。

4. F 检验是丰富多彩的。它可用于检验各种各样的假设。诸如：(1) 个别的回归系数是否统计上显著，(2) 是否全部偏斜率系数为零，(3) 两或多个系数是否统计上相等，(4) 一些系数是否满足某些线性约束条件，和 (5) 回归模型是否有结构稳定性。

5. 和双变量情形一样，多元回归模型可用于均值或个值预测的目的。

习 题

283

问答题

8.1 假如你要研究某产品，比如说汽车，在一些年里的销售情况，有人建议你试用下面的模型：

$$Y_t = \beta_0 + \beta_1 t$$

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

其中 Y_t = 时间 t 的销售量，而 t = 时间（以年计）。第一个模型假设销售量是时间的线性函数，而第二个模型把它表述为时间的二次函数。

a. 讨论这些模型的性质。

b. 你会怎样在两个模型之间做出选择？

c. 在什么情况下二次模型会是有用的？

d. 试图找到美国在过去 20 年里的汽车销售量数据，并看哪个模型对数据的拟合较好。

8.2 证明 (8.5.16) 的 F 比率等于 (8.5.18) 的 F 比率。（提示：ESS/TSS = R^2。）

8.3 证明 (8.5.18) 和 (8.7.10) 的 F 检验是等价的。

8.4 证明 (8.7.11) 和 (8.7.12) 的陈述。

8.5 考虑柯布-道格拉斯生产函数：

$$Y = \beta_1 L^{\beta_2} K^{\beta_3} \tag{1}$$

其中 Y = 产出，L = 劳动力投入，和 K = 资本投入。用 K 除 (1) 得：

$$(Y/K) = \beta_1 (L/K)^{\beta_2} K^{\beta_2+\beta_3-1} \tag{2}$$

取 (2) 的自然对数得：

$$\ln(Y/K) = \beta_0 + \beta_2 \ln(L/K) + (\beta_2+\beta_3-1)\ln K + u_i \tag{3}$$

其中 $\beta_0 = \ln\beta_1$。

a. 假如你有做回归（3）的数据，你会怎样检验规模报酬不变即（$\beta_2 + \beta_3$）$=1$ 这个假设？

b. 如果有规模报酬不变情形，你会怎样解释回归（3）？

c. 用 L 而不用 K 遍除（1），会有什么不同吗？

8.6 **当 $R^2=0$ 时的临界 R^2 值。**方程（8.5.11）给出在全部偏斜率系数同时为零（即 $R^2=0$）的假设下 F 与 R^2 的关系。正如我们能从 F 表求出在 α 显著性水平上的临界 F 值，我们能通过以下的关系式求出临界 R^2 值：

$$R^2 = \frac{(k-1)F}{(k-1)F + (n-k)}$$

284 其中 k 是回归模型中包括截距在内的参数个数，而 F 是在 α 显著性水平上的临界 F 值。如果所测的 R^2 超过由上述公式计算出来的临界 R^2，就可拒绝真 R^2 为零的假设。

建立上述公式并求出（在 $\alpha = 5\%$ 处）回归（8.2.1）的临界 R^2 值。

8.7 根据 1968—1987 年年度数据得到如下回归结果：

$$\hat{Y}_t = -859.92 + 0.647\,0X_{2t} - 23.195X_{3t} \quad R^2 = 0.977\,6 \quad (1)$$

$$\hat{Y}_t = -261.09 + 0.245\,2X_{2t} \quad R^2 = 0.938\,8 \quad (2)$$

其中 Y = 美国进口商品支出（1982 年 10 亿美元计），X_2 = 个人可支配收入（1982 年 10 亿美元计），而 X_3 = 趋势变量。（1）中的 X_3 的标准误是 4.275 0？真或不真？说明你的计算。（提示：利用 R^2，F 与 t 的关系。）

8.8 假使回归：

$$\ln(Y_i/X_{2i}) = \alpha_1 + \alpha_2\ln X_{2i} + \alpha_3\ln X_{3i} + u_i$$

中的回归系数及其标准误均已知[1]，你会怎样估计以下回归模型参数及其标准误？

$$\ln Y_i = \beta_1 + \beta_2\ln X_{2i} + \beta_3\ln X_{3i} + u_i$$

8.9 假定：

$$Y_i = \beta_1 + \beta_2X_{2i} + \beta_3X_{3i} + \beta_4X_{2i}X_{3i} + u_i$$

其中 Y 是个人消费支出，X_2 是个人收入，X_3 是个人财富。[2]（$X_{2i}X_{3i}$）被称为**交互作用项**（interaction term）。此表达式的含义是什么？你会怎样检验边际消费倾向（即 β_2）独立于消费者财富的假设？

8.10 给定如下回归结果：

$$\hat{Y}_t = 16\,899 \qquad -2\,978.5X_{2t} \qquad R^2 = 0.614\,9$$
$$t = \qquad (8.515\,2) \qquad (-4.728\,0)$$

$$\hat{Y}_t = 9\,734.2 \quad - 3\,782.2X_{2t} + 2\,815X_{3t} \quad R^2 = 0.770\,6$$
$$t = \quad (3.370\,5) \quad (-6.607\,0) \quad (2.971\,2)$$

你能发现这些结果所依据的样本大小吗？（提示：回顾 R^2、F 与 t 值的关系。）

285 **8.11** 根据我们对分别以 t 和 F 检验为基础的个别和联合假设检验的讨论，以下哪些情况看来较近于可能？

a. 拒绝基于 F 统计量的联合虚拟假设，但不拒绝基于个别 t 检验的每一分开的虚拟假设；

b. 拒绝基于 F 统计量的联合虚拟假设，拒绝基于 t 检验的一个个别的假设，而不拒绝基于 t 检验的其他个别的假设；

c. 拒绝基于 F 统计量的联合虚拟假设；并且拒绝基于个别 t 检验的每一分开的虚拟假设；

d. 不拒绝基于 F 统计量的联合虚拟假设，并且不拒绝基于个别 t 检验的每一分开的虚拟假设。

e. 不拒绝基于 F 统计量的联合虚拟假设，拒绝基于 t 检验的一个个别的假设，而不拒绝基于 t 检验的其他个别的假设；

f. 不拒绝基于 F 统计量的联合虚拟假设，但拒绝基于个别 t 检验的每一分开的虚拟假设。[3]

解答题

8.12 参照习题 7.21。

a. 实际现金结余的实际收入弹性和利率弹性是什么？

b. 上述弹性是个别地在统计上显著的吗？

c. 检验所估回归的总显著性。

d. 对实际现金结余的需求的收入弹性显著地异于 1 吗？

e. 应该把利率变量留在模型中吗？为什么？

8.13 根据美国1992 年 46 个州的数据，Baltagi 得到如下回归结果[4]：

$$\log C = 4.30 \quad - \quad 1.34\log P \quad + \quad 0.17\log Y$$
$$\text{se} = (0.91) \quad (0.32) \quad (0.20) \quad \bar{R}^2 = 0.27$$

其中 C = 香烟消费（每年的包数计）

P = 每包香烟的真实价格

Y = 真实人均可支配收入

a. 香烟需求的价格弹性是多少？它统计显著吗？若显著，它在统计上异于 1 吗？

b. 香烟需求的收入弹性是多少？它显著吗？若不显著，其原因是什么？

c. 你如何根据上面给出的校正 R^2 来得到 R^2？

286 **8.14** Wooldridge 从 209 个企业的样本得到如下回归结果[5]：

$$\log(\text{salary}) = 4.32 + 0.280\log(\text{sales}) + 0.017\,4\text{roe} + 0.000\,24\text{ros}$$
$$\text{se} = (0.32) \quad (0.035) \quad (0.004\,1) \quad (0.000\,54)$$

$$R^2 = 0.283$$

其中 salary = CEO 薪水，sales = 企业年销售额，roe = 股权每百分比收益，ros = 企业股票回报。

括号中的数字为估计的标准误。

a. 根据你对各个系数符号的先验预期，解释上述回归。

b. 哪个系数在 5% 的显著性水平上是个别统计显著的？

c. 回归的总显著性如何？你用哪个检验？为什么？

d. 你能把 roe 和 ros 的系数解释成弹性系数吗？为什么？

8.15 假定 Y 和 X_1，X_2，…，X_k 是联合正态分布的，并假定虚拟假设为：总体偏相关个别地等于零，R.A. 费希尔曾证明：

$$t = \frac{r_{1\,2.3\,4\cdots k}\sqrt{n-k-2}}{\sqrt{1-r^2_{1\,2.3\,4\cdots k}}}$$

遵循 $n-k-2$ 个自由度的 t 分布，其中 k 指第 k 阶偏相关系数，而 n 是观测值的总个数。（注：$r_{1\,2.3}$ 是第一阶偏相关系数，$r_{1\,2.3\,4}$ 是第二阶偏相关系数，依此类推。）参照习题 7.2。假定 Y，X_2 和 X_3 是联合正态分布的，计算三个偏相关 $r_{1\,2.3}$，$r_{1\,3.2}$，和 $r_{2\,3.1}$，并在相应总体相关个别地为零的假设下检验它们的显著性。

8.16 在研究 1921—1941 年和 1948—1957 年两时期美国对农用拖拉机的需求中格里利谢斯（Griliches）[6]得到如下结果：

$$\widehat{\log Y_t} = \text{常数} - \underset{(0.231)}{0.519}\log X_{2t} - \underset{(0.477)}{4.933}\log X_{3t} \qquad R^2 = 0.793$$

287 其中 Y_t = 以 1935—1939 年美元计的。每年 1 月 1 日农场拥有拖拉机存量的价值，X_2 = 时间 t 的拖拉机支付价格指数除以全部农作物收取价格指数，X_3 = 第 $t-1$ 年流行利率，而括号中的数字是估计的标准误。

a. 解释上述回归。

b. 所估斜率系数个别地在统计上显著？它们显著地异于 1 吗？

c. 使用方差分析技术检验整个回归的显著性。提示：利用 ANOVA 技术的 R^2 形式。

d. 怎样计算农用拖拉机需求的利率弹性？

e. 怎样检验所估 R^2 的显著性？

8.17 考虑如下的 1950—1969 年时期英国经济[7]的工资—决定方程：

$$\hat{W}_t = \underset{(1.129)}{8.582} + \underset{(0.080)}{0.364}(\mathrm{PF})_t + \underset{(0.072)}{0.004}(\mathrm{PF})_{t-1} - \underset{(0.658)}{2.560}U_t$$

$$R^2 = 0.873; \qquad \mathrm{df} = 15$$

其中 W = 每个雇员的工资和薪金，PF = 最终产品的要素成本价格，U = 表示为英国总劳工人数的百分比的英国失业率，t = 时

间。(括号内的数字是估计标准误。)

a. 解释上述方程。

b. 所估系数个别地显著吗?

c. 引进 $(PF)_{t-1}$的合理性何在?

d. 是否应把变量 $(PF)_{t-1}$从模型中删去? 为什么?

e. 怎样计算每个雇员的工薪对失业率的弹性?

8.18 习题 8.17 所给的工资决定方程的一个变异有如下式[8]:

$$\hat{W}_t = 1.073 + 5.288V_t - 0.116X_t + 0.054M_t + 0.046M_{t-1}$$
$$(0.797)(0.812) \quad (0.111) \quad (0.022) \quad (0.019)$$
$$R^2 = 0.934; \qquad \mathrm{df} = 14$$

其中 W 同前, V = 表示为占英国总劳工人数百分比的英国未填补的职位(工作)空缺率, X = 每就业人员的国内总产值, M = 进口价格及 M_{t-1} = 上(或滞后)年的进口价格。(估计的标准误放在括号中。)

288 a. 解释上述方程。

b. 哪些估计系数是个别地在统计上显著的?

c. 引进 X 变量的合理性何在? 先验上, X 的预期符号应为负吗?

d. 在模型中兼引进 M_t 和 M_{t-1}, 用意何在?

e. 哪些变量可从模型中删去?

f. 检验所测回归的总显著性。

8.19 对于(8.7.24)中估计的子鸡需求函数: 所估计的收入弹性等于 1 吗? 价格弹性等于 −1 吗?

8.20 对于需求函数(8.7.24)你怎样检验假设: 收入弹性与价格弹性数值相同而符号相反? 说明必要的计算。[注: $\mathrm{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.001\,42$。]

8.21 参照习题 7.16 的玫瑰需求函数。限于考虑对数设定形式。

a. 所估计的需求的自价格弹性(即对玫瑰价格的弹性)是什么?

b. 它在统计意义上显著吗?

c. 如果是, 它是不是在统计意义上异于 1?

d. 先验上, X_3(石竹价格)和 X_4(收入)的预期符号为何? 经验结果和这些预期相符吗?

e. 如果 X_3 和 X_4 的系数在统计意义上不显著, 可能是什么理由?

8.22 参照关于野猫活动的习题 7.17。

a. 所估计的每一个斜率系数个别地说在 5% 的显著性水平上都是统计显著的吗?

b. 你会拒绝假设 $R^2 = 0$ 吗?

c. 1948—1978 年期间野猫活动的瞬时增长率是什么? 相应的复合增长率呢?

8.23 参照习题 7.18 所估计的美国国防预算支出回归。

a. 对所估计的回归结果作一般性评论。

b. 建立 ANOVA 表并检验全部斜率系数为零的假设。

8.24 下面列出所谓的**超越生产函数**（transcendental production function，TPF），是著名的柯布-道格拉斯生产函数的一个推广：

$$Y_i = \beta_1 L^{\beta_2} k^{\beta_3} e^{\beta_4 L + \beta_5 K}$$

其中 Y = 产出，L = 劳力投入，K = 资本投入。

取对数并加入随机干扰项便得到随机的 TPF：

$$\ln Y_i = \beta_0 + \beta_2 \ln L_i + \beta_3 \ln K_i + \beta_4 L_i + \beta_5 K_i + u_i$$

其中 $\beta_0 = \ln\beta_1$。

a. 此函数有些什么性质？

b. 要使 TPF 化为柯布-道格拉斯生产函数，β_4 和 β_5 的值必须是什么？

c. 如果你拥有数据，你会怎样判明 TPF 是否可简化为柯布-道格拉斯生产函数？你会用什么检验方法？

d. TPF 对习题 7.18 的数据拟合得怎样？说明你的计算。

8.25 1948—1978 年美国能源价格与资本形成。为了检验假设：相对于产出的能源价格上升导致现有资本与劳动力资源的生产力下降，J.A. 塔托姆（John A. Tatom）估计了 1948 年第 1 季度至 1978 年第 2 季度时期的美国生产函数。[9]

$$\begin{aligned}\ln(y/k) = {} & 1.549\,2 + 0.713\,5\ln(h/k) - 0.108\,1\ln(P_e/P)\\ & (16.33) \qquad (21.69) \qquad\qquad (-6.42)\\ & + 0.004\,5t \qquad R^2 = 0.98\\ & (15.86)\end{aligned}$$

其中 y = 私有企业部门的实际产出，k = 资本（提供）服务流量的一种度量，h = 私有企业部门的工时（人员小时），P_e = 燃料及相关产品的生产者价格，P = 私有企业部门价格平缩因子，t = 时间。括号中的数字是 t 统计量。

289 a. 这些结果是否支持了作者的假设？

b. 在 1972 年和 1977 年之间，能源相对价格 P_e/P 增加了 60%。按照估算的回归，生产力损失了多少？

c. 除去 (h/k) 和 (P_e/P) 的变化后，在样本期间里的生产力的趋势增长率如何？

d. 你会怎样解释系数 0.713 5？

e. 估计的每一个偏斜率系数都是个别地在统计意义上显著的这一事实（为什么？），意味着我们可以拒绝假设：$R^2 = 0$？为什么或为什么不？

8.26 **电线电缆需求**。表 8.10 给出一个电话线制造商用来预测 1968—1983 年期间对一主要用户的销售量的数据。[10]

表 8.10 回归变量

年份	X_2 GNP	X_3 新房动工数	X_4 失业率%	X_5 滞后 6 个月的最惠利率	X_6 用户用线增量%	Y 总塑料购买(MPE)
1968	1 051.8	1 503.6	3.6	5.8	5.9	5 873
1969	1 078.8	1 486.7	3.5	6.7	4.5	7 852
1970	1 075.3	1 434.8	5.0	8.4	4.2	8 189
1971	1 107.5	2 035.6	6.0	6.2	4.2	7 497
1972	1 171.1	2 360.8	5.6	5.4	4.9	8 534
1973	1 235.0	2 043.9	4.9	5.9	5.0	8 688
1974	1 217.8	1 331.9	5.6	9.4	4.1	7 270
1975	1 202.3	1 160.0	8.5	9.4	3.4	5 020
1976	1 271.0	1 535.0	7.7	7.2	4.2	6 035
1977	1 332.7	1 961.8	7.0	6.6	4.5	7 425
1978	1 399.2	2 009.3	6.0	7.6	3.9	9 400
1979	1 431.6	1 721.9	6.0	10.6	4.4	9 350
1980	1 480.7	1 298.0	7.2	14.9	3.9	6 540
1981	1 510.3	1 100.0	7.6	16.6	3.1	7 675
1982	1 492.2	1 039.0	9.2	17.5	0.6	7 419
1983	1 535.4	1 200.0	8.8	16.0	1.5	7 923

表中变量定义如下：

Y = 年销售量（百万尺双线，记 MPF）

X_2 = 国民生产总值（10 亿美元）

X_3 = 新房动工数（单位：1000）

X_4 = 失业率（%）

X_5 = 滞后 6 个月的最惠利率

X_6 = 用户用线增量%

290 考虑以下模型：

$$Y_i = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t$$

a. 估计以上回归。

b. 此模型中各系数的预期符号是什么？

c. 经验结果与事前预期一致吗？

d. 这些估计的偏回归系数个别地看在 5% 的显著性水平上都是统计显著的吗？

e. 假使你先做 Y 对 X_2、X_3 和 X_4 的回归，然后决定是否再加进变量 X_5 和 X_6。你会怎样知道值不值得把 X_5 和 X_6 加进来呢？你用哪一种检验？说明必要的计算。

8.27 M. 纳洛夫（Marc Nerlove）曾估计如下的电力产生的成本函数[11]：

$$Y = AX^{\beta}P_1^{\alpha_1}P_2^{\alpha_2}P_3^{\alpha_3}u \tag{1}$$

291 其中 Y = 总生产成本，X = 小时产出，P_1 = 劳动投入价格，P_2 = 资本投入价格，P_3 = 燃料价格，u = 干扰项。理论上，预期价格弹性之和为 1，即 $(\alpha_1+\alpha_2+\alpha_3)=1$。引进这一约束，上述成本函数就可写为：

$$(Y/P_3) = AX^{\beta}(P_1/P_3)^{\alpha_1}(P_2/P_3)^{\alpha_2}u \tag{2}$$

换言之，(1) 是无约束而 (2) 是受约束成本函数。

根据 29 个中等厂家的一个样本并通过对数变换，纳洛夫得到如下回归结果：

$$\begin{aligned}
\widehat{\ln Y_i} &= -4.93 \quad + \quad 0.94\ln X_i + 0.31\ln P_1 \\
\text{se} &= (1.96) \quad (0.11) \qquad (0.23) \\
&\quad -0.26\ln P_2 + 0.44\ln P_3 \\
&\qquad (0.29) \quad (0.07) \qquad \text{RSS} = 0.336
\end{aligned} \tag{3}$$

$$\begin{aligned}
\widehat{\ln(Y/P_3)} &= -6.55 + 0.91\ln X + 0.51\ln(P_1/P_3) + 0.09\ln(P_2/P_3) \\
\text{se} &= (0.16) \quad (0.11) \quad (0.19) \qquad\qquad (0.16) \\
&\qquad\qquad\qquad\qquad \text{RSS} = 0.364
\end{aligned} \tag{4}$$

a. 解释方程 (3) 和 (4)。

b. 你怎样判明约束 $(\alpha_1+\alpha_2+\alpha_3)=1$ 是否正确？说明你的计算。

8.28 估计资本资产定价模型（CAPM）。在第 6.1 节中我们简要地叙述了现代证券组合理论中著名的资本资产定价模型。在经验分析中，CAPM 的估计分为两阶段：

阶段Ⅰ（时间序列回归）。对样本所含 N 种证券的每一种，做如下的一个时间上的回归：

$$R_{it} = \hat{\alpha}_i + \hat{\beta}_i R_{mt} + e_{it} \tag{1}$$

其中 R_{it} 和 R_{mt} 是年度 t 第 i 种证券和市场组合证券（比如，S & P 500）的回报率，β_i，如本书其他地方说过的，是第 i 种证券的 β 系数或市场波动系数，而 e_{it} 是干扰项。一共有 N 个这样的回归（每一种证券有一个），从而给出 β_i 的 N 个估计值。

阶段Ⅱ（横截面回归）。在这一阶段中，我们在 N 种证券上做以下的回归：

$$\bar{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2\hat{\beta}_i + u_i \tag{2}$$

其中 $\bar{R}_i$ 是在阶段Ⅰ的样本时期内算出的第 i 种证券的平均回报率，$\hat{\beta}_i$ 是从第一阶段回归估计出来的 β 系数，而 u_i 是干扰项。

拿第二阶段的回归 (2) 和：

$$\mathrm{ER}_i = r_f + \beta_i(\mathrm{ER}_m - r_f) \tag{3}$$

292 （其中 r_f 代表无风险回报率）这样的 CAPM 方程（6.1.2）相比，我们即看到 $\hat{\gamma}_1$ 是 r_f 的一个估计值，以及 $\hat{\gamma}_2$ 是市场风险溢价（$\mathrm{ER}_m - r_f$）的一个估计值。

因此，在 CAPM 的经验检验中，R_i 和 $\hat{\beta}_i$ 被分别地用作 ER_i 和 β_i 的估计量。现在，如果 CAPM 成立，则在统计意义上：

$$\hat{\gamma}_1 = r_f$$

$$\hat{\gamma}_2 = R_m - r_f, (\mathrm{ER}_m - r_f)\text{的估计量}$$

其次，考虑另一模型：

$$\bar{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2\hat{\beta}_i + \hat{\gamma}_3 s_{e_i}^2 + u_i \tag{4}$$

其中 $s_{e_i}^2$ 是得自第一阶段回归的第 i 种证券的残差方差。那么，如果 CAPM 正确，$\hat{\gamma}_3$ 就不会显著地异于零。

为了检验 CAPM，利维（Levy）根据 1948—1968 年时期的 101 种股票的一个样本，做回归(2) 和 (4)，并得到以下结果[12]：

$$\begin{aligned} \bar{R}_i &= 0.109 + 0.037\beta_i \\ &\quad (0.009)\quad (0.008) \\ t &= (12.0)\quad (5.1)\quad R^2 = 0.21 \end{aligned} \tag{2)'$$

$$\begin{aligned} \bar{R}_i &= 0.106 + 0.002\,4\hat{\beta}_i + 0.201 s_{e_i}^2 \\ &\quad (0.008)\quad (0.007)\quad (0.038) \\ t &= (13.2)\quad (3.3)\quad (5.3)\quad R^2 = 0.39 \end{aligned} \tag{4)'$$

a. 这些结果支持了 CAPM 吗？

b. 值不值得把变量 $s_{e_i}^2$ 加进模型中来？你怎样知道？

c. 如果 CAPM 成立，（2）中的 $\hat{\gamma}_1$ 应接近无风险利率 r_f 的均值。这个估计值是 10.9%。此值像不像是观测期间（1948—1968 年）的无风险回报率的一个合理的估计呢？（不妨考虑国库券或类似的较无风险的资产回报率。）

d. 如果 CAPM 成立，由（2）得到的市场风险溢价（$\bar{R}_m - r_f$）将是 3.7%。如果假定 r_f 为 10.9%，则意味着在样本期间 $\bar{R}_m$ 约为 14.6%。这像不像是一个合理的估计呢？

e. 你对 CAPM 能做些什么一般性的评论？

8.29 参照习题7.21c。现在，你有了必要的工具，你将用哪种检验来在两个模型之间做出选择。给出必要的计算。注意，两个模型的因变量不同。

8.30 参照习题 8.3。利用（8.7.4）中所示的 t 检验，说明墨西哥经济在研究期内是否存在规模报酬不变的情况。

293 **8.31** 回到我们曾几次讨论的儿童死亡率一例。在回归（7.6.2），我们将儿童死亡率对人均 GNP 和妇女识字率回归。现在我们通过增

加总生育率（TFR）来扩展模型。表 6.4 中有所有这些数据可用。重做回归（7.6.2）并给出扩展模型的回归结果如下：

1. $\widehat{\mathrm{CM}}_i = 263.641\,6 - 0.005\,6\mathrm{PGNP}_i - 2.231\,6\mathrm{FLR}_i$

$$\mathrm{se} = (11.593\,2)\ (0.001\,9) \qquad (0.209\,9) \quad R^2 = 0.707\,7 \tag{7.6.2}$$

2. $\widehat{\mathrm{CM}}_i = 168.306\,7 - 0.005\,5\mathrm{PGNP}_i - 1.768\,0\mathrm{FLR}_i + 12.868\,6\mathrm{TFR}_i$

$$\mathrm{se} = (32.891\,6)(0.001\,8) \qquad (0.248\,0) \qquad (?)$$

$$R^2 = 0.747\,4$$

a. 你如何解释 TFR 的系数？据经验，预期 CM 和 TFR 之间的关系是正还是负？判明你的答案。

b. 这两个方程之间 PGNP 和 FLR 的系数有变化吗？若有，这种变化的原因是什么？所观测的差别统计上显著吗？你用哪个检验，为什么？

c. 你如何在模型 1 和 2 之间做选择？你用哪个检验来回答这个问题？给出必要的计算。

d. 我们没有给出 TFR 系数的标准误。你能求出它吗？（提示：回忆 t 和 F 分布之间的关系。）

8.32 回到习题 1.7，它给出 21 个企业在广告印象和广告支出方面的数据。习题 5.11 要你对这些数据描点，并决定二者关系的适当模型。令 Y 表示保留的印象数，X 为广告支出，则得到如下回归：

模型Ⅰ：$\hat{Y}_i = 22.163 + 0.363\,1X_i$

$$\mathrm{se} = (7.089)\ (0.097\,1) \quad r^2 = 0.424$$

模型Ⅱ：$\hat{Y}_i = 7.059 + 1.084\,7X_i - 0.004\,0X_i^2$

$$\mathrm{se} = (9.986)\ (0.369\,9) \quad (0.001\,9) \quad R^2 = 0.53$$

a. 解释这两个模型。

b. 哪个模型更好？为什么？

c. 你用哪个检验来选择模型？

d. 广告支出存在“收益递减”吗，即在一定的广告支出水平（饱和水平）后就不再支出广告费吗？你能求出这个支出水平吗？给出必要的计算。

294

8.33 在回归(7.9.4)中，我们给出用柯布-道格拉斯生产函数来拟合台湾农业部门 1958—1972 年间数据的结果。基于此回归，说明是否存在规模报酬不变的情况。

a. 使用(8.7.4)中给出的 t 检验来说明。并告诉你两个斜率估计量间的协方差为 $-0.038\,43$。

b. 用(8.7.9)中给出的 F 检验来说明。

c. 这两种检验结果是否不同？对于台湾的农业部门在样本期内

的规模报酬，你有什么结论？

8.34 重新考虑第8.8节中的储蓄—收入回归。假设我们把样本分为1970—1982年和1983—1995年两个时期，利用邹至庄检验判断储蓄—收入回归在两个时期是否有结构变动。将你的结论与8.8节中给出的结论相比较，对邹至庄检验对样本分成两（或多）期的转折点的敏感性，你有什么总体结论？

[习题注释]

[1] 取材自 Peter Kennedy, *A Guide to Econometrics*, the MIT Press, 3d ed., Cambridge, Mass., 1992, p.301。

[2] Ibid., p.327.

[3] 取自 Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, Reading, Mass., 1991, p.79。

[4] 参见 Badi H. Baltagi, *Econometrics*, Springer-Verlag, New York, 1998, p.111。

[5] 参见 Jeffrey M. Wooldridge, *Introductory Econometrics*, South-Western Publishing Co., 2000, pp.154－155。

[6] Z. Griliches, "The Demand for a Durable Input: Farm Tractors in the United States, 1921—1957," *The Demand for Durable Goods*, Arnold C. Harberger (ed.), The University of Chicago Press, Chicago, 1960, Table 1, p.192.

[7] 取自 *Prices and Earnings in 1951—1969: An Econometric Assessment*, Dept. of Employment, HMSO, 1971, Eq.(19), p.35。

[8] 同上，第37页，方程（67）。

[9] 见 "Energy Prices and Capital Formation: 1972—1977," *Review*, Federal Reserve Bank of St. Louis, vol.61, no. 5, May 1979, p.4。

[10] 我感谢 Daniel J. Reardon 收集并加工了这些数据。

[11] Marc Nerlove, "Returns to Scale in Electric Supply," in Carl Christ, ed., *Measurement in Economics*, Stanford University Press, Palo Alto, Calif., 1963.

[12] H. Levy, "Equilibrium in an Imperfect Market: A constraint on the number of securities in the portfolio," *American Economic Review*, vol.68, no.4, September 1978, pp.643－658.

*附录 8A

似然比检验

在附录4A中我们讨论最大似然原理时，说明了怎样获得双变量回归模

* 供选读用。

型的 ML 估计量；LR 检验是以此 ML 原理为根据的。该原理可直接推广应用到多元回归模型中。在干扰项 u_i 为正态分布的假定下，我们证明对双变量模型而言，回归系数的 OLS 和 ML 估计量是相同的。但所估计的误差方差却不相同。σ^2 的 OLS 估计量是 $\sum\hat{u}_i^2/(n-2)$，而 ML 估计量是 $\sum\hat{u}_i^2/n$。前者是无偏的。但后者有偏误，尽管这种偏误在大样本中将消失。

这些结论对多元回归而言也是对的，为了说明，且考虑三变量回归模型

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{1}$$

对应于附录 4A 中的方程（5），模型（1）的对数似然函数可写成：

$$\ln\mathrm{LF} = -\frac{n}{2}\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \tag{2}$$

295 如附录 4A 所示，将此函数对 α_1，α_2，α_3 和 σ^2 微分并令其结果为零，解方程组即得到这些参数的 ML 估计量；α_1，α_2 和 α_3 的 ML 估计量将无异于 OLS 估计量［后者已见于方程（7.4.6）到（7.4.8）］，而误差方差则有所不同，因为残差平方和将被除以 n，而不是 OLS 情形中的（$n-3$）。

现假定虚拟假设 H_0 是：变量 X_3 的系数 β_3 为零。这时，由（2）给出的对数似然函数变为：

$$\ln\mathrm{LF} = -\frac{n}{2}\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum(Y_i - \beta_1 - \beta_2 X_{2i})^2 \tag{3}$$

因方程（3）是在先验约束 $\beta_3=0$ 下估计的，故称**受约束对数似然函数**（restricted log-likelihood function，RLLF），而方程（1）则无参数方面的先验约束，故可称**无约束对数似然函数**（unrestricted log LF，ULLF）。为了检验先验约束 β_3 为零的真实性，LR 检验使用如下的检验统计量：

$$\lambda = 2(\mathrm{ULLF} - \mathrm{RLLF}) \tag{4}$$[1]

其中 ULLF 和 RLLF 分别是无约束对数似然函数［方程（1）］和受约束对数似然函数［方程（3）］。可以证明，在大样本中，由（4）给出的检验统计量 λ 遵循自由度等于虚拟假设中所加约束的个数的 χ^2 分布。本例中此个数为 1。

LR 检验的基本思想是简单的：如果先验约束真实，则受约束与无约束（对数）LF 不应有差异。这时（4）中的 λ 将是零。但如果先验约束不真，则两个 LF 必定相异。而我们知道，在大样本中 λ 遵循 χ^2 分布，于是能找出这个差异在（比方说）1% 或 5% 显著性水平上是否是统计上显著的。此外，我们还能找出估计值 λ 的 p 值。

现在让我们用儿童死亡率的例子来说明 LR 检验。如果我们像在（8.2.1）中那样将儿童死亡率对人均 GNP 和妇女识字率回归，我们就得到 ULLF 为 -328.1012，但如果我们只将 CM 对 PGNP 回归，则得到 RLLF 为 -361.6396。从绝对值看（即不考虑符号），前者较小，由于我们在前面一个模型中多一个变量，所以这也讲得通。

296 现在的问题是，是否值得增加 FLR 变量。若不值得，则约束和无约束

的 LLF 就不应该有大的差别，但若值得，二者就会有所差别。为了看出这个差别在统计上是否显著，我们现在利用（4）式给出的 LR 检验：

$$\lambda = 2[-328.1012 - (-361.6396)] = 67.0768$$

此量在渐近意义上遵循 1 个自由度（因从完整模型中去掉变量 FLR 而只有 1 个约束）的 χ^2 分布。得到这样一个 χ^2 值的 p 值几乎为 0，从而得到变量 FLR 不应该从模型中去掉的结论。换言之，约束回归在目前的情况下不能成立。

由于 W 检验和 LM 检验数学的复杂性，这里将不予讨论。但如在正文中提到的，LR，W 和 LM 检验在渐近意义上给出相同的答案，如何选择检验方法，就要看计算上的便利了。

【附录注释】

[1] 此式又可表达为 $-2(\mathrm{RLLF}-\mathrm{ULLF})$或 $-2\ln(\mathrm{RLF}/\mathrm{ULF})$。

【注释】

[1] 有了正态性假定，OLS 估计量 $\hat{\beta}_2$、$\hat{\beta}_3$，和 $\hat{\beta}_1$ 就是在整个无偏估计类中的最小方差估计量，不管它是不是线性估计量。简言之，它们是 BUE（最优无偏估计量）。参看 C.R.Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, 1965, p.258。

[2] 在大多数的经验研究中，虚拟假设都被叙述成这种形式，即采取如下的极端立场（一种打倒草人的策略）：因变量与所考虑的解释变量之间无任何关系。用意是要从判明两变量之间是否存在一个无关重要的关系开始。

[3] 对我们的例子而言，偏态值为0.2276，峰态值为2.9488。而对一个正态分布变量而言，偏态值和峰态值分别为 0 和 3。

[4] 在任一给定的样本中，协方差$\mathrm{cov}(\hat{\beta}_2,\hat{\beta}_3)$未必是零；即 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 可能相关。见（7.4.17）。

[5] Thomas B.Fomby, R.Carter Hill, and Stanley R.Johnson, *Advanced Econometric Methods*, Springer-Verlag, New York, 1984, p.37.

[6] 参看 K.A.Brownlee, *Statistical Theory and Methodology in Science and Engineering*, John Wiley & Sons, New York, 1960, pp.278－280。

[7] 同上。

[8] Fomby et al., op.cit., p.42.

[9] 以下的 F 检验是8.7节中（8.7.9）或（8.7.10）所给的更为一般的 F 检验的一个特殊情形。

[10] 证明见 Dennis J.Aigner, *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1971, pp.91－92。

[11] 协方差公式的代数表达式颇为复杂。用矩阵符号简洁地描述这个表达式，则在附录 C 中给出。

[12] 如果 $\beta_2+\beta_3<1$，此关系式将构成线性不等式约束。为处理这种约束，需要用到数学规划技术。

[13] Henri Theil, *Principles of Econometrics*, John Wiley & Sons, New York, 1971, pp.43－45.

[14] 如果是用最大似然法进行估计，则一个类似于即将讨论的方法是似然比检验，因它较复杂，故放在本章附录中讨论。进一步的讨论，见于泰尔：所引书的第179－184页。

[15] Gregory C.Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, vol.28, no.3, 1960, pp.591－605.

[16] 至于详尽讨论，参见 William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp.293－297。

[17] 在异方差条件下对邹至庄检验的讨论，参见 William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp.292－293, and Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, U.K., 1994, p.51。

[18] 一篇易读的论文，见于 A.Buse, "The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note," *American Statistician*, vol. 36, 1982, pp.153－157。

[19] Russell Davidson and James G.MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, p.456.

[20] J.MacKinnon, H.White, and R.Davidson, "Tests for Model Specification in the Presence of Alternative Hypothesis; Some Further Results", *Journal of Econometrics*, Vol.21, 1983, pp.53－70. A Similar test is proposed in A.K.Bera and C.M.Jarque,"Model Specification Tests: A Simultaneous Approach," *Journal of Econometrics*, vol.20, 1982, pp.59－82.

[21] 根据 William H.Greene, *ET.The Econometrics Toolkit Version 3*, Econometric Software, Bellport, New York, 1992, pp.245－246。

第 9 章 虚拟变量回归模型

297 我们在第 1 章简单讨论了经验分析中通常会遇到的四种变量类型：比率尺度、区间尺度、序数尺度和名义尺度。我们在前面几章曾遇到的变量类型基本上都是比率尺度。但这不应该给我们留下回归模型只能处理比率尺度变量的印象。回归模型也可以处理前面提到的其他几种数据类型。我们在本章要考虑不仅涉及比率尺度变量而且涉及名义尺度变量的模型。这种变量也被称为**指标变量**（indicator variables）、**分类变量**（categorical variables）、**定性变量**（qualitative variables）或**虚拟变量**（dummy variables）。[1]

§9.1 虚拟变量的性质

在回归分析中，因变量或回归子不仅经常受到比率尺度变量（如收入、产出、价格、成本、身高、温度）的影响，还会受到定性性质的变量或名义尺度变量的影响，如性别、种族、肤色、宗教、国籍、地区、政治动乱和党派等。例如，保持所有其他因素不变，我们发现女性工人比相应男性工人挣得少，而白人比非白人挣得多。[2]这种情况可能是性别或种族歧视所致，但

不论如何，诸如性别和种族之类的定性变量看来都能影响回归子，而且明显应该包含在解释变量或回归元之中。

298 由于这种变量通常都标志着出现或不出现某种“品质”或属性，如男性或女性、黑人或白人、天主教或非天主教、民主党或共和党等，所以它们基本上都是名义尺度变量。我们能量化这种属性的途径之一，就是构造一个取值为 1 或 0 的人为变量，1 表示出现（或具备）那种属性，0 表示没有那种属性。比如，1 可能标志着一个人是女性，而 0 则标志着男性；或者 1 标志着一个人是大学毕业生，而 0 标志着不是，等等。假定这种取值 0 和 1 的变量称为**虚拟变量**。[3]这种变量实质上就是一个将数据区分为相互排斥的类别（如男性或女性）的工具。

虚拟变量也可以像定量变量那样轻而易举地放到回归模型中。事实上，一个回归模型所包含的回归元可以都是虚拟或定性变量。这种模型被称为方差分析（ANOVA）模型。[4]

§9.2 ANOVA 模型

为说明 ANOVA 模型，考虑如下例子。

例 9.1 不同地理区域公立学校教师的薪水

表 9.1 给出了 1985 年 50 个州和哥伦比亚特区公立学校教师平均薪水（美元）的数据。这 51 个地区被分为三个地理区域：(1) 东北和中北部（共 21 个州），(2) 南部（共 17 个州），及 (3) 西部（共 13 个州）。目前，暂不考虑表的格式及其中的其他数据。

假设我们想知道，公立教师平均年薪（AAS）在这个国家的三个地理区域之间是否有所不同。如果你仅对这三个地区中教师的平均年薪进行简单的算术平均，那么这三个地区的平均值分别是 24 424.14 美元（东北和中北
299 部），22 894 美元（南部）和 26 158.62 美元（西部）。这些数字看起来不同，但它们在统计上彼此相异吗？有各种统计方法来比较两个和多个均值，通常是做**方差分析**（analysis of variance）。[5]但在回归分析的框架下也能做到这一点。

为看出这一点，考虑如下模型：

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.1}$$

其中 Y_i = 第 i 个州公立学校教师的平均薪水

D_{2i} = 1 若该州位于东北和中北部

= 0 否则（即位于美国其他地区）

$D_{3i}=1$ 若该州位于南部

$=0$ 否则（即位于美国其他地区）

300 应看到，除了不是定量回归元而是定性和虚拟回归元（若观测属于某特定的组则取值为 1，若它不属于那一组则取值 0）外，(9.2.1) 与前面考虑的任何一个多元回归模型都一样。此后，我们所有的虚拟变量都用字母 D 表示。表 9.1 中的虚拟变量就是这样构造的。

模型 (9.2.1) 告诉了我们什么呢？假定误差项满足通常的 OLS 假定，通过对 (9.2.1) 的两边同时取期望，我们得到：

东北和中北地区公立学校教师薪水的均值为：

$$E(Y_i \mid D_{2i}=1, D_{3i}=0)=\beta_1+\beta_2 \tag{9.2.2}$$

南部公立学校教师薪水的均值为：

$$E(Y_i \mid D_{2i}=0, D_{3i}=1)=\beta_1+\beta_3 \tag{9.2.3}$$

你可能想知道，我们如何求西部教师薪水的均值。若你猜测它等于 β_1，那就完全正确，因为

$$E(Y_i \mid D_{2i}=0, D_{3i}=0)=\beta_1 \tag{9.2.4}$$

换句话说，西部公立学校教师薪水的均值由多元回归 (9.2.1) 中的截距 β_1 给出，而“斜率”系数 β_2 和 β_3 则告诉我们，中东北地区和西部地区教师薪水的均值与西部相比有多大的差别。但我们如何才能知道这种差别在统计上是否显著呢？在回答这个问题之前，让我们先给出回归 (9.2.1) 的结果。利用表 9.1 中给出的数据，我们得到如下结果：

$$\begin{aligned}
\hat{Y}_i &= 26\,158.62 - 1\,734.473D_{2i} - 3\,264.615D_{3i} \\
\text{se} &= (1\,128.523)\quad(1\,435.953)\quad(1\,499.615) \\
t &= (23.175\,9)\quad(-1.207\,8)\quad(-2.177\,6) \\
&\quad\ (0.000\,0)^*\quad(0.233\,0)^*\quad(0.034\,9)^*\qquad R^2=0.090\,1
\end{aligned} \tag{9.2.5}$$

其中 * 表示 p 值。如这些回归结果所示，西部教师薪水的均值约为26 158美元，而东北和中北教师薪水的均值则约低1 734美元，南部教师薪水均值则约低3 265美元。如方程 (9.2.3) 和 (9.2.4) 所示，将西部教师薪水加上这些地区差薪，很容易就能得到后两个地区实际的薪水均值。我们如此可以得到后两个地区的薪水均值约为24 424美元和22 894美元。

表 9.1　　1986 年公立学校教师的州平均薪水

薪水	支出	D_2	D_3	薪水	支出	D_2	D_3
19 583	3 346	1	0	22 795	3 366	0	1
20 263	3 114	1	0	21 570	2 920	0	1
20 325	3 554	1	0	22 080	2 980	0	1
26 800	4 642	1	0	22 250	3 731	0	1

29 470	4 669	1	0	20 940	2 853	0	1
26 610	4 888	1	0	21 800	2 533	0	1
30 678	5 710	1	0	22 934	2 729	0	1
27 170	5 536	1	0	18 443	2 305	0	1
25 853	4 168	1	0	19 538	2 642	0	1
24 500	3 547	1	0	20 460	3 124	0	1
24 274	3 159	1	0	21 419	2 752	0	1
27 170	3 621	1	0	25 160	3 429	0	1
30 168	3 782	1	0	22 482	3 947	0	0
26 525	4 247	1	0	20 969	2 509	0	0
27 360	3 982	1	0	27 224	5 440	0	0
21 690	3 568	1	0	25 892	4 042	0	0
21 974	3 155	1	0	22 644	3 402	0	0
20 816	3 059	1	0	24 640	2 829	0	0
18 095	2 967	1	0	22 341	2 297	0	0
20 939	3 285	1	0	25 610	2 932	0	0
22 644	3 914	1	0	26 015	3 705	0	0
24 624	4 517	0	1	25 788	4 123	0	0
27 186	4 349	0	1	29 132	3 608	0	0
33 990	5 020	0	1	41 480	8 349	0	0
23 382	3 594	0	1	25 845	3 766	0	0
20 627	2 821	0	1				

注：$D_2=1$，若该州位于东北和中北部；0，其他地区。

$D_3=1$，若该州位于南部；0，其他地区。

资料来源：National Educational Association, as reported by *Albuquerque Tribune*, Nov.7, 1986.

但我们如何才能知道，这些薪水均值是否与参照组西部地区教师薪水的均值在统计上有所差异？这相当容易，我们所需要做的只是辨别（9.2.5）中的每一个斜率系数是否统计显著。从这个回归中可以看出，东北和中北地区的估计系数在统计上是不显著的，因为其 p 值为 23%，而南部地区的估计系数则是统计上显著的，因为其 p 值约为 3.5%。因此总体结论便是，西部和东中部地区公立学校教师薪水的均值大致相同，而南部教师薪水的均值则统计上显著地略低 3 265 美元。从图上来看，这种情形如图 9.1 所示。

301 在解释这些差异时，我们必须给予警告。这些虚拟变量只是简单地指出了可能存在的这些差异，但不能给出导致这些差异的原因。受教育水平、生活成本指数、性别和种族上的差异都可能对所观测到的差异具有某种影响。因此，除非我们能考虑所有其他可能影响教师薪水的变量，否则我们就不能确定导致这些差异的原因。

从上述讨论明显可见，我们所要做的，只是看附属于各个虚拟变量的系

数是不是个别统计显著的。此例还表明，在回归模型中包含定性或虚拟回归元是多么容易。

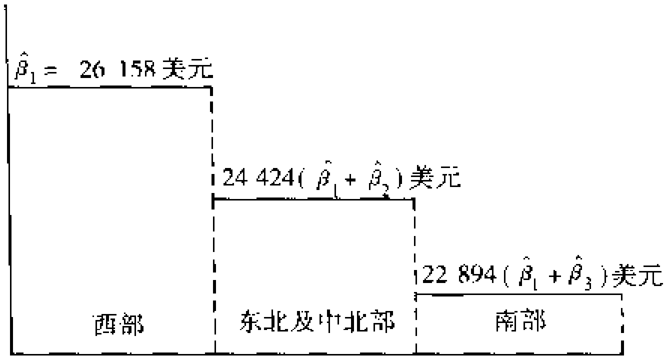


图 9.1　三个地区公立学校教师的平均薪水（以美元计）

对使用虚拟变量的告诫

尽管在回归模型中包含虚拟变量很容易，但在使用它们时仍必须小心。具体而言，需考虑如下方面：

1. 我们在例 9.1 中为区分三个区域而只使用了两个虚拟变量 D_2 和 D_3。我们为什么不用三个虚拟变量来区分这三个区域呢？假设我们这样做并将模型（9.2.1）写成：

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.6}$$

302 其中 D_{1i}对西部的观测取值 1，对其他观测取值 0。于是我们现在对每个地理区域都有了一个虚拟变量。利用表 9.1 中的数据，如果你做回归（9.2.6），计算机将“拒绝”做这个回归（不妨试试看）。[6] 为什么呢？原因在于，在（9.2.6）的背景下，你既有每个类别或组的虚拟变量，又有一个截距，即你遇到了完全共线性的情况，即变量之间出现完全线性的关系。为什么？回到表 9.1，设想我们现在增加了 D_1 列：如果该州处于西部就取值 1，否则取值 0。现在，如果你将这三个 D 列水平相加就得到由 51 个 1 构成的一列。但由于截距 α 对每个观测都（隐含地）为 1，所以你又得到一个由 51 个 1 构成的一列。换言之，这三个 D 列之和再次生成了截距列，由此导致了完全共线性。在这种情况下，估计模型（9.2.6）是不可能的。

这里的信息是：**若定性变量有 m 个类别，则只需引入（$m-1$）个虚拟变量**。在我们的例子中，定性变量“区域”有三类，所以我们只需引入两个虚拟变量。如果你不遵守这个规则，那你就会陷入所谓的**虚拟变量陷阱**，即完全共线性或完全多重共线性（若变量之间存在不止一个精确的关系）的情形。在模型中有不止一个定性变量时这个规则依然适用，稍后会给一个例子。因此我们应该将前述规则重新表述为：**对每个定性回归元而言，所引入的虚拟变量个数必须比该变量的类别数少一个**。因此，如果我们在例 9.1 中

有教师性别的信息，那我们就应该再使用一个（而非两个）虚拟变量，对女性取值 1 和对男性取值 0 或反之。

2. 不指定其虚拟变量的那一组被称为**基**（base）**组**、**基准**（benchmark）**组**、**控制**（control）**组**、**比较**（comparison）**组**、**参照**（reference）**组**或**省略**（omitted）**组**。所有其他的组都与基准组进行比较。

3. 截距值（β_1）代表了基准组的均值。在例 9.1 中基准组为西部地区。因此，在回归（9.2.5）中约为 26 159 的截距值代表着西部各州教师薪水的均值。

4. 附属于（9.2.1）中虚拟变量的系数被称为**级差截距系数**（differential intercept coefficients），因为它们告诉我们的是，取值为 1 的地区的截距值与基准组的截距系数之间的差别。比如在（9.2.5）中，约为 −1 734 的系数值意味着，与作为基准组的西部地区的薪水均值 26 159 美元相比，东北和中北地区教师的薪水均值约低 1 734 美元。

5. 如果像在我们的说明性的例子那样定性变量不止一类，那么，基准组的选择完全取决于研究者。基准组的选择有时候受所研究的特殊问题的支配。在我们的说明性例子中，我们可以选择南部作为基准组。在那种情况下，（9.2.5）中的回归结果将有所变化，因为现在都是与南部做比较。当然，这不会改变此例中的总体结论。（为什么？）此时，截距值将是南部教师薪水的均值，约为22 894美元。

303 6. 我们前面对虚拟变量陷阱做过警告。如果我们在这种模型中不使用截距项，那么引入与变量的类别相同数量的虚拟变量就能够回避虚拟变量陷阱的问题。因此，如果我们从（9.2.6）中去掉截距项，并考虑如下模型：

$$Y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \tag{9.2.7}$$

由于此时没有完全共线性，就不会陷入虚拟变量的陷阱。但要确定当你做这个回归时，一定要使用回归软件包中的无截距选项。

我们如何解释回归(9.2.7)呢？如果你对(9.2.7)求期望，你将会发现：

$\beta_1 =$ 西部教师薪水的均值

$\beta_2 =$ 东北和中北地区教师薪水的均值

$\beta_3 =$ 南部教师薪水的均值

换言之，去掉截距项并容许每一类别都有一个虚拟变量，我们就直接得到不同群组的均值。在我们说明性的例子中，（9.2.7）的结果如下：

$$\begin{aligned} \hat{Y}_i &= 26\,158.62\ D_{1i} + 24\,424.14\ D_{2i} + 22\,894\ D_{3i} \\ \text{se} &= (1\,128.523) \qquad (887.917\,0) \qquad (986.864\,5) \\ t &= \quad (23.179\,5)^* \qquad (27.507\,2)^* \qquad (23.198\,7)^* \qquad R^2 = 0.090\,1 \end{aligned} \tag{9.2.8}$$

其中 * 表示这些 t 比率的 p 值很小。如你所见，虚拟变量的系数直接给出了西部、东中部和南部三个区域的（薪水）均值。

7. 如下引入虚拟变量的方法中哪种更好呢：（1）为每个类别都引入一个虚拟变量并省略截距项；或（2）引入截距项和（$m-1$）个虚拟变量，其

中 m 为虚拟变量的类别数吗？如肯尼迪（kennedy）所指出：

大多数研究者认为，在一个含有截距的方程中，他们能更容易地处理他们通常最感兴趣的问题，是否有某个组与基准组有所不同以及有多大的不同，所以在方程中包括截距更方便。为了检查分组是否得当，也可通过将虚拟变量的系数相对 0 做 t 检验（或者更一般地，对适当的虚拟变量系数集做一个 F 检验），就可以检验分类是否适当（因为分组的人可能预料在基准组与其他组之间存在统计上的显著差异。——译者注）。[7]

§9.3 含有两个定性变量的 ANOVA 模型

304 我们在上一节考虑了含有一个三类别定性变量的 ANOVA 模型。我们在本节将考虑另外一个含两个定性变量的 ANOVA 模型，并揭示虚拟变量的某些其他特点。

例 9.2 小时工资与婚姻状况和居住地的关系

从 1985 年 5 月的一个 528 人的样本中得到如下结论[8]：

$$
\begin{aligned}
\hat{Y}_i &= \quad 8.8148 \quad + \quad 1.0997\,D_{2i} - 1.6729 D_{3i} \\
\text{se} &= \ (0.4015) \qquad (0.4642) \qquad (0.4854) \\
t &= (21.9528) \qquad (2.3688) \quad (-3.4462) \\
& \quad\ (0.0000)^* \qquad (0.0182)^* \quad (0.0006)^* \qquad R^2 = 0.0322
\end{aligned}
\tag{9.3.1}
$$

其中 Y = 小时工资(美元)

D_2 = 婚姻状况；1 = 已婚，0 = 其他

D_3 = 居住地；1 = 南部，0 = 其他

* 表示 p 值。

在此例中，我们有两个定性回归元，且每个回归元分为两个类别。因此，我们为每个定性回归元都指定一个虚拟变量。

这里的基准组是什么呢？显然是未婚的非南部居民组。换句话说，不住在南部的未婚人士就属于被省略组。于是所有的组都与这个组进行比较。此基准组的小时工资均值约为 8.81 美元。与其相比，已婚者的平均小时工资约高 1.10 美元，即实际的平均工资为 9.91(=8.81+1.10)美元。对比之下，那些住在南部的人的平均小时工资约低 1.67 美元，实际小时工资为 7.14 美元。

上述小时工资与基组相比在统计上有差异吗？是的，因为所有的级差截距都是统计上显著的，且 p 值都相当小。

在此例中值得注意的一点是：一旦遇到多于一个定性变量，你就必须密切

注意基组，因为所有其他组都是与基组进行比较。在有几个定性回归元，而且每个回归元又有几个类别时这一点就特别重要。但现在怎样操作几个定性变量应该清楚了。

§9.4 同时含有定性和定量回归元的回归：ANCOVA 模型

前两节讨论的那种 ANOVA 模型，尽管在社会学、心理学、教育学和市场研究等领域很常见，但在经济学中并不普遍。在多数典型的经济研究中，回归模型的解释变量既有一些定性的，又有一些定量的。同时包含定性和定量变量的回归模型被称为**协方差分析**（analysis of covariance, ANCOVA）**模型**。ANCOVA 模型是对 ANOVA 模型的推广，在一个同时包括定量和定性（或虚拟）回归元的模型中，这种模型提供了一种方法，能在统计上控制定量回归元
305 ［被称为**协变量**（covariates）或**控制变量**（control variables）］的影响。我们现在就来说明 ANCOVA 模型。

为了说明为什么要进行这种分析，让我们重新考虑例 9.1，并设想三个区域的公立学校教师的平均薪水可能本来就没有什么不同，因为我们应考虑到有些变量无法在不同地区使之标准化，比如考虑地方政府对公立学校的支出这种变量（因为公共教育基本上是地方和州政府的事情）。为了看出是否如此，我们提出如下模型：

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i \tag{9.4.1}$$

其中 Y_i = 公立学校教师的平均薪水（美元）

X_i = 对公立学校每个学生的支出（美元）

D_{2i} = 1，若该州位于东北和中北部

= 0，其他

D_{3i} = 1，若该州位于南部

= 0，其他

X 的数据在表 9.1 中给出。记住我们把西部作为基准组。除两个定性回归元之外，我们还有一个定量变量 X，前面曾提到，在 ANCOVA 模型的背景下，X 被称为**协变量**。

例 9.3 教师薪水与区域和对公立学校每个学生的支出之间的关系

从表 9.1 中的数据得到模型(9.4.1)的结论如下：

$$\begin{aligned}
\hat{Y} &= 13\,269.11 - 1\,673.514 D_{2i} - 1\,144.157 D_{3i} + 3.288\,9\, X_i \\
\text{se} &= (1\,395.056) \quad (801.170\,3) \quad (861.118\,2) \quad (0.317\,6) \qquad (9.4.2)\\
t &= (9.511\,5)^{*} \quad (-2.088\,9)^{*} \quad (-1.328\,6)^{**} \quad (10.353\,9)^{*}\\
& \qquad\qquad\qquad\qquad R^2 = 0.726\,6
\end{aligned}$$

其中，*表示 p 值低于5%，**表示 p 值高于5%。

如这些结论所示，在其他条件不变的情况下：公共支出每增加1美元，公立学校教师的薪水约上升3.29美元。控制教育支出后，我们现在看到，东北和中北地区的级差截距系数依然显著，但南部的系数则不显著。这些结论与（9.2.5）中的结论不同。但这无足为奇，因为我们在（9.2.5）中没有解释对每个学生的公共教育支出的差别这个协变量。在图上看，如图9.2所示。

注：尽管我们对三个区域给出三个回归线，但从统计上看，西部和南部的回归线是一样的。还可以注意到，这三个回归线是平行的。（为什么？）

306

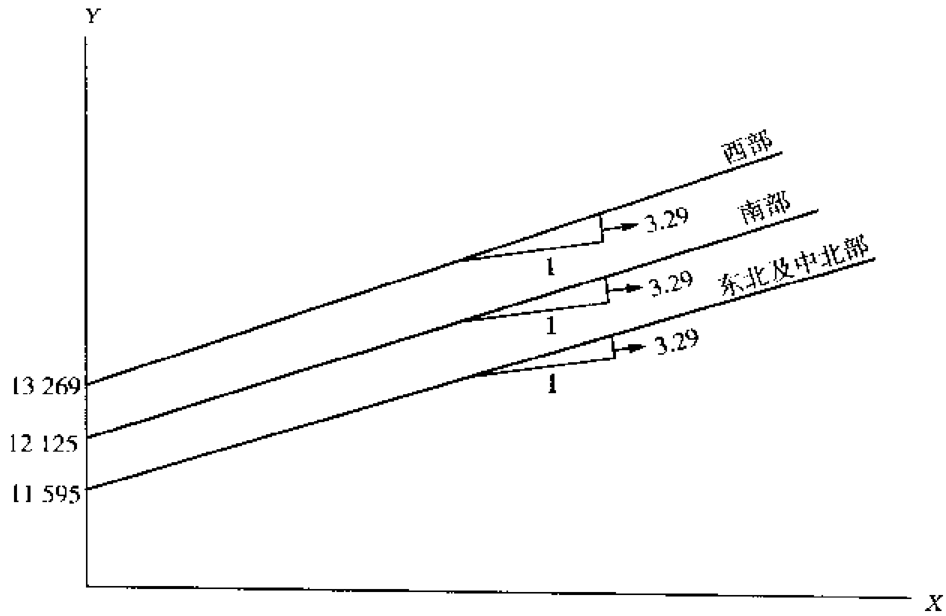


图9.2 公立学校教师薪水（Y）与对每个学生教育支出（X）的关系

§9.5 邹至庄检验的虚拟变量方法[9]

我们在第8.8节讨论过邹至庄检验，以考察一个回归模型的结构稳定性。我们在那里讨论的例子涉及美国1970—1995年间储蓄与收入之间的关系。我们将样本期间一分为二，1970—1981年和1982—1995年，邹至庄检验表明，储蓄对收入的回归在这两个区间存在着差异。

然而，我们不知道这两个回归的差异是源于截距项、斜率系数还是二者兼而有之。这种知识本身常常是十分有用的。

参照方程（8.8.1）和（8.8.2），我们看到四种可能性，如图9.3所示。

1. 两个回归的截距和斜率都相同。这种**重合回归**(coincident regressions)的情形如图9.3a所示。

307 2. 两个回归的斜率相同但截距不同。这种**平行回归**（parallel regressions）的情形如图 9.3b 所示。

3. 两个回归的截距相同但斜率不同。这是**同截距回归**（concurrent regressions）的情形（图 9.3c）。

4. 两个回归的截距和斜率都不同。这种**非相似回归**（dissimilar regressions）如图 9.3d 所示。

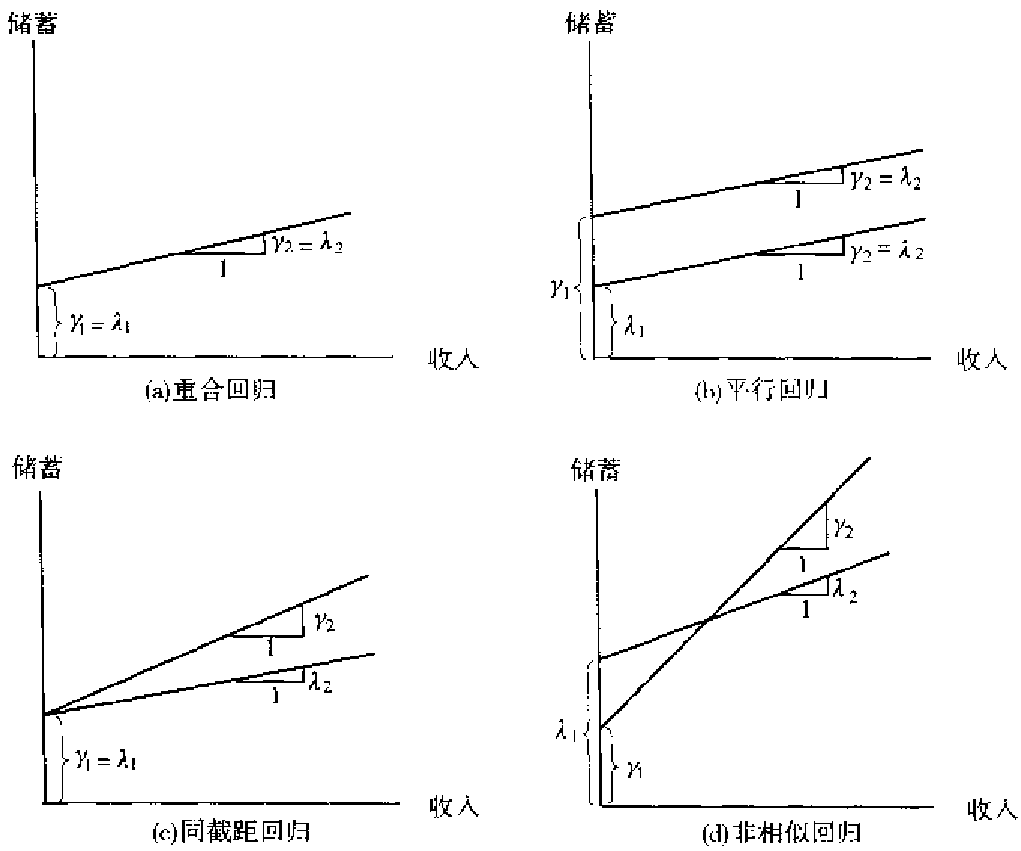


图 9.3 合理的储蓄—收入回归

前面提到，第 8.8 节中所讨论的多步骤邹至庄检验程序只告诉我们两个（或多个）回归是否不同，但没有告诉我们这种不同来自哪里。通过将所有观测（共 26 个）混合起来，如果存在着差异，只须做如下多元回归，便能探明这种差异的根源[10]：

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \tag{9.5.1}$$

其中 Y = 储蓄

X = 收入

t = 时间

D = 1 1982—1995 年之间的观测

= 0 其他(即 1970—1981 年之间的观测)

308 表 9.2 说明了数据矩阵的结构。

为了看出（9.5.1）的含义，像通常一样假定$E(u_i)=0$，我们得到：

1970—1981 年的均值储蓄函数：

$$E(Y_t \mid D_t=0, X_t)=\alpha_1+\beta_1 X_t \tag{9.5.2}$$

1982—1995 的均值储蓄函数：

$$E(Y_t \mid D_t=1, X_t)=(\alpha_1+\alpha_2)+(\beta_1+\beta_2)X_t \tag{9.5.3}$$

读者将会注意到，它们是与（8.8.1）和（8.8.2）相同的函数，其中$\lambda_1=\alpha_1$，$\lambda_2=\beta_1$，$\gamma_1=(\alpha_1+\alpha_2)$和$\gamma_2=(\beta_1+\beta_2)$。因此，估计（9.5.1）就等同于逐一估计两个储蓄函数（8.8.1）和（8.8.2）。

在（9.5.1）中，和前面一样，α_2是级差截距，而级差斜率系数（也被称为斜率漂移因子）。β_2表示的是，第二个期间储蓄函数（虚拟变量取值为 1 的那一组）的斜率与第一个期间相比有多大的不同。注意，以**交互或相乘**
309 **形式**（interactive or multiplicative form）引入虚拟变量 D（D 乘以 X），如何使我们区别两个期间的斜率系数，与我们引入**相加形式**（additive form）虚拟变量来区分两个期间的截距殊途同归。

表 9.2　　美国 1970—1995 年间的储蓄与收入数据

观测	储蓄	收入	虚拟变量
1970	61	727.1	0
1971	68.6	790.2	0
1972	63.6	855.3	0
1973	89.6	965	0
1974	97.6	1 054.2	0
1975	104.4	1 159.2	0
1976	96.4	1 273	0
1977	92.5	1 401.4	0
1978	112.6	1 580.1	0
1979	130.1	1 769.5	0
1980	161.8	1 973.3	0
1981	199.1	2 200.2	0
1982	205.5	2 347.3	1
1983	167	2 522.4	1
1984	235.7	2 810	1
1985	206.2	3 002	1
1986	196.5	3 187.6	1
1987	168.4	3 363.1	1
1988	189.1	3 640.8	1
1989	187.8	3 894.5	1

1990	208.7	4 166.8	1
1991	246.4	4 343.7	1
1992	272.6	4 613.7	1
1993	214.4	4 790.2	1
1994	189.4	5 021.7	1
1995	249.3	5 320.8	1

注：虚拟变量取值 1 表示始于 1982 年的观测；取值 0 表示其他观测。收入和储蓄数据都以 10 亿美元计。

资料来源：*Economic Report of the President*，Table B－28，p.332.

例 9.4 美国储蓄—收入回归中的结构差异：虚拟变量方法

在继续下去之前，让我们先给出应用美国储蓄—收入数据估计模型 (9.5.1) 得到的回归结果：

$$\begin{aligned} \hat{Y}_t &= 1.016\,1 + 152.478\,6\,D_t + 0.080\,3X_t - 0.065\,5(D_tX_t) \\ \text{se} &= (20.164\,8) \quad (33.082\,4) \quad (0.014\,4) \quad (0.015\,9) \qquad (9.5.4) \\ t &= (0.050\,4)^{**} \quad (4.609\,0)^{*} \quad (5.541\,3)^{*} \quad (-4.096\,3)^{*} \\ R^2 &= 0.881\,9 \end{aligned}$$

其中，*表示 p 值低于 5%，而**表示 p 值高于 5%。

如这些回归结果所示，级差截距和斜率系数都是统计上显著的，这强烈地表明，两个期间的储蓄—收入回归如图 9.3 (d) 那样是不同的。

从 (9.5.4) 我们可以推导出方程 (9.5.2) 和 (9.5.3) 如下：

1970—1981 年的储蓄收入回归函数：

$$\hat{Y}_t = 1.016\,1 + 0.080\,3X_t \tag{9.5.5}$$

1982—1995 年的储蓄收入回归函数：

$$\begin{aligned} \hat{Y}_t &= (1.016\,1 + 152.478\,6) + (0.080\,3 - 0.065\,5)X_t \\ &= 153.494\,7 + 0.014\,8X_t \end{aligned} \tag{9.5.6}$$

它们与我们在 (8.8.1a) 和 (8.8.2a) 中所得到的结果完全相同，这无足为奇。这些回归已如图 8.3 所示。

现在很容易看出，虚拟变量方法［即估计 (9.5.1)］相对邹至庄检验［即估计 (8.8.1)、(8.8.2) 和 (8.8.3) 三个回归］有如下优势：

1. 我们只需要做一个回归，因为个别回归很容易就能以 (9.5.2) 和 (9.5.3) 所指明的方式推导出来。

2. 单一回归 (9.5.1) 可用于检验各种假设。因此，如果级差截距系数 α_2 是统计非显著的，那我们或许可以接受这两个回归具有相同截距的同截

距假设［见图9.3（c）］。类似地，如果级差斜率系数β_2统计不显著而α_2是显著的，那我们或许就不能拒绝这两个回归具有相同斜率的平行回归假设［比较图9.3（b）］。用通常的F检验（回想受约束最小二乘F检验）能对整个回归的稳定性（即α_2和β_2同时为零）进行检验。如果不能拒绝这个假设，那么回归线将像图9.3（a）那样重合。

310 3．邹至庄检验不能明确告诉我们截距和斜率系数中到底是哪个不同，还是都不相同。也就是说，只是截距不同、只是斜率不同或二者都不同皆可得到一个显著的邹至庄检验。换言之，我们不能从邹至庄检验中得知，在给定情形中，图9.3中的四种可能性到底存在哪一种。就此看来，虚拟变量方法具有明显的优势，因为它不仅能告诉我们两个回归是否不同，而且还能确定这种差别的来源——是源于截距、斜率还是二者皆有。实践中，了解两个回归到底是哪个系数不同，其重要性绝不亚于对这两个回归是否不同的了解。

4．最后，由于数据混合（即在一个回归中包括所有的观测）增加了自由度，这可能会提高估计系数的相对精度。当然，要记住的是，每增加一个虚拟变量都要求增加一个自由度。

§9.6 使用虚拟变量的交互效应

虚拟变量是一个能处理一系列有趣问题的灵活工具。为看出这一点，考虑如下模型：

$$Y_i=\alpha_1+\alpha_2D_{2i}+\alpha_3D_{3i}+\beta X_i+u_i \tag{9.6.1}$$

其中 Y＝以美元计的小时工资

X＝受教育水平（读书年数）

$D_2=1$，若为女性；0，其他

$D_3=1$，若既非白人又非西班牙人；0，其他

在这个模型中，性别和种族为定性回归元，而受教育水平为定量变量。[11]这个模型中的暗含假定是，性别虚拟变量D_2的差别影响对两个种族类别而言是一样的，而种族虚拟变量D_3的差别影响对两个性别而言也是一样的。这就是说，若男性工资的均值比女性高，则不论是对哪一个种族来讲都是如此。同理，若既非白人又非西班牙人的工资均值较低，则不论他们是男性还是女性都是如此。

在许多应用中，这种假定是不能成立的。在既非白人又非西班牙人的种族中，女性可能比男性挣得少（原书为“少”，疑误。——译者注）。换句话说，两个定性变量D_2和D_3之间可能会相互影响。因此，它们对Y均值的影响可能不像（9.6.1）那样单纯是相加形式的，还有可能像如下模型一样是乘积形式的。

$$\hat{Y}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \qquad (9.6.2)$$

其中变量定义与模型（9.6.1）相同（原书中误将公式中的 Y 写成其拟合值。——译者注）。

我们从（9.6.2）得到：

$$E(Y_i \mid D_{2i}=1,\ D_{3i}=1,\ X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i \qquad (9.6.3)$$

311 这就是非白人和非西班牙人中女工人的小时工资均值。观察到：

α_2 = 作为女性的级差效应

α_3 = 作为非白人/非西班牙人的级差效应

α_4 = 作为非白人/非西班牙人女性的级差效应

它表明，女性非白人/非西班牙人的小时工资均值与女性或非白人/非西班牙人有所不同（差别为 α_4）。例如，如果所有三个级差系数都为负，那就意味着，在与基组（在本例中即男性白人或男性西班牙人组）相比时，非白人/非西班牙人女性比女性或男性非白人/非西班牙人挣得更少。

读者现在可以看出，交互虚拟变量（即两个定性或虚拟变量之积）如何能改变两个单独考虑的影响因素的作用（即将它们的影响相加）。

例 9.5　平均小时工资与受教育水平、性别和种族的关系

让我们首先给出基于模型（9.6.1）的回归结论。利用用于估计（9.3.1）的数据，我们得到如下结果：

$$\begin{aligned} \hat{Y}_i &= -0.261\,0 \quad - \quad 2.360\,6D_{2i} - 1.732\,7D_{3i} + 0.802\,8X_i \\ t &= (-0.235\,7)^{**} \quad (-5.487\,3)^{*} \; (-2.180\,3)^{*} \; (9.909\,4)^{*} \qquad (9.6.4) \\ & \qquad R^2 = 0.203\,2,\ n = 528 \end{aligned}$$

其中 * 表示 p 值低于 5%，** 表示 p 值高于 5%。

读者可以验证，级差截距系数都是统计上显著的，并具有预期的符号（为什么?），而且作为一个无足为奇的结论，受教育水平对小时工资有很大的影响。

如（9.6.4）所示，在其他条件不变的情况下，女性的平均小时工资约低 2.36 美元，非白人/非西班牙人的平均小时工资也约低 1.73 美元。

我们现在来考虑包括交互虚拟变量的模型（9.6.2）的结论。

$$\begin{aligned} \hat{Y}_i &= -0.261\,00 \quad - 2.360\,6D_{2i} - 1.732\,7D_{3i} + 2.128\,9D_{2i}D_{3i} + 0.802\,8X_i \\ t &= (-0.235\,7)^{**} \quad (-5.487\,3)^{*} \; (-2.180\,3)^{*} \; (1.742\,0)^{**} \qquad (9.909\,5)^{**} \\ & \qquad (9.6.5) \\ & \qquad R^2 = 0.203\,2,\ n = 528 \end{aligned}$$

其中 * 表示 p 值低于 5%，** 表示 p 值高于 5%。

如你所见，相加的两个虚拟变量仍是统计上显著的，但交互虚拟变量在

通常5%的显著性水平上不是统计显著的；其实际的 p 值为8%。若你认为这是一个足够低的概率，(9.6.5) 中的结论则可做如下解释：保持受教育水平不变，若将三个虚拟系数相加则得到 -1.964（$=-2.3605-1.7327+2.1289$），这意味着，女性非白人/非西班牙人的小时工资均值约低1.96美元，这个数字介于 -2.3605（单纯性别差异）和 -1.7327（单纯种族不同）之间。

312 上例明显地揭示了模型中包含有两个或多个虚拟变量时交互虚拟变量的作用。重要的是要注意到，我们在模型（9.6.5）中假定，小时工资相对受教育水平的增长率（多受一年教育小时工资约增加80美分）对不同性别和不同种族而言没有什么不同。但情况可能并非如此，若你想对此做检验，则必须引入级差斜率系数（见习题9.25）。

§9.7 季节分析中虚拟变量的使用

许多基于月度或季度数据的经济时间序列都表现出季节特征（规则地摆动）。比如在圣诞或其他重大的节假日期间百货公司的销售额、节假日家庭对货币（或现金余额）的需求、夏天对冰淇淋和软饮料的需求、谷物在收割季节过后的价格、搭乘飞机旅行的需求，等等。从一个时间序列中去掉季节因素或成分，使我们能专注于诸如趋势之类的其他因素，通常都是好事情。[12]从时间序列中去除季节成分的过程被称为**除季节性**（deseasonalization）或**季节调整**（seasonal adjustment），由此得到的时间序列被称为除季节性或季节调整后的时间序列。诸如失业率、消费者价格指数、生产者价格指数和工业生产指数等重要的经济时间序列通常都以季节调整后的形式公布。

虽然有几种方法能去除一个时间序列的季节性，但我们只考虑这些方法中的一种，即虚拟变量方法。[13]为了说明如何用虚拟变量来去除经济时间序列中的季节性，考虑表9.3中给出的数据。此表给出1978—1985年四种主要厨具销售数量的季度数据，这四种厨具是洗碗机、污物粉碎机、冰箱和洗衣机，均以千台计。此表还给出以1982年10亿美元计的耐用品支出数据。

为了说明虚拟变量方法，我们将只考虑样本期内冰箱的销售数据。我们首先看一下数据，如图9.4所示。这个数据表明，数据中可能存在与各个季节相联系的季节性。为了解是否如此，考虑如下模型：

$$Y_t=\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\alpha_4 D_{4t}+u_t \tag{9.7.1}$$

313 其中 $Y_t=$（以千台计）冰箱销售数量，四个 D 都是虚拟变量，分别在相应季节取值1，在其他季节取值0。注意，为避免虚拟变量陷阱，我们为一年中的每个季度都指定一个虚拟变量，但不要截距项。若某给定季度中存在某种季节效应，将会由该季度虚拟变量系数的统计显著的 t 值表现出来。[14]

表 9.3　厨具销售和耐用品支出的季度数据（1978 年第 1 季度到 1985 年第 4 季度）（以千台计）

DISH	DISP	FRIG	WASH	DUR	DISH	DISP	FRIG	WASH	DUR
841	798	1 317	1 271	252.6	480	706	943	1 036	247.7
957	837	1 615	1 295	272.4	530	582	1 175	1 019	249.1
999	821	1 662	1 313	270.9	557	659	1 269	1 047	251.8
960	858	1 295	1 150	273.9	602	837	973	918	262
894	837	1 271	1 289	268.9	658	867	1 102	1 137	263.3
851	838	1 555	1 245	262.9	749	860	1 344	1 167	280
863	832	1 639	1 270	270.9	827	918	1 641	1 230	288.5
878	818	1 238	1 103	263.4	858	1 017	1 225	1 081	300.5
792	868	1 277	1 273	260.6	808	1 063	1 429	1 326	312.6
589	623	1 258	1 031	231.9	840	955	1 699	1 228	322.5
657	662	1 417	1 143	242.7	893	973	1 749	1 297	324.3
699	822	1 185	1 101	248.6	950	1 096	1 117	1 198	333.1
675	871	1 196	1 181	258.7	838	1 086	1 242	1 292	344.8
652	791	1 410	1 116	248.4	884	990	1 684	1 342	350.3
628	759	1 417	1 190	255.5	905	1 028	1 764	1 323	369.1
529	734	919	1 125	240.4	909	1 003	1 328	1 274	356.4

注：DISH＝洗碗机；DISP＝污物粉碎机；FRIG＝冰箱；WASH＝洗衣机；DUR＝以 1982 年 10 亿美元计耐用品支出。

资料来源：*Business Statistics and Survey of Current Business*, Department of Commerce carious issues).

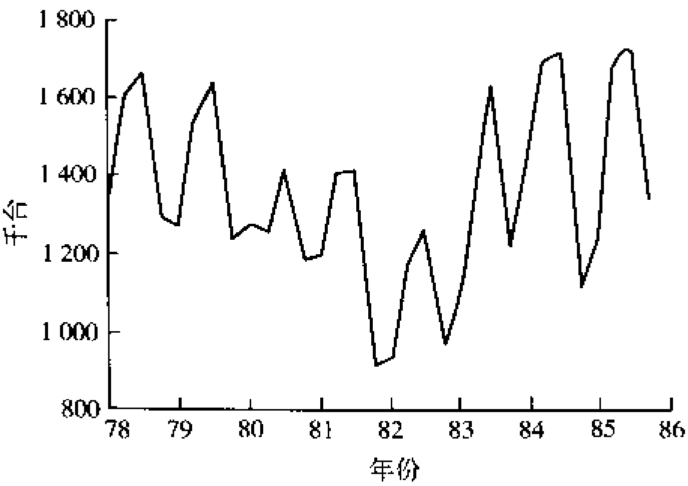


图 9.4　1978—1985 年冰箱销售数量（季度数据）

314 注意，我们在（9.7.1）中实际上将 Y 对一个截距进行回归，不同之处在于，我们容许不同的季节（即季度）有不同的截距。结果，每个虚拟变量的系数都将给出相应季度或季节里冰箱销售的均值。（为什么？）

例 9.6　冰箱销售中的季节性

从表 9.3 中给出的冰箱销售数据，我们得到如下回归结果：

$$\hat{Y}_t = 1\,222.125D_{1t} + 1\,467.500D_{2t} + 1\,569.750D_{3t} + 1\,160.000D_{4t}$$
$$t = \quad (20.372\,0) \qquad (24.462\,2) \qquad (26.166\,6) \qquad (19.336\,4) \tag{9.7.2}$$
$$R^2 = 0.531\,7$$

注：我们没有给出估计系数的标准误，因为所有的虚拟变量都只取值 1 或 0，所以每个标准误都等于 59.990 4。

（9.7.2）中估计的 α 系数表示了每个季节（即季度）冰箱销售（以千台计）的平均数量或均值。于是，第 1 季度冰箱的平均销售量约为1 222千台，第 2 季度约1 468千台，第 3 季度约1 570千台，第 4 季度约1 160千台。

315 顺便指出，为避免虚拟变量陷阱，我们曾为每个季度指定一个虚拟变量并略去截距项，我们也可以在包括截距项的同时只指定 3 个虚拟变量。假设我们把第 1 季度作为参照季度，并为第 2、3 和 4 季度指定虚拟变量。这就得到如下回归结果（见表 9.4 中的数据背景）：

$$\hat{Y}_t = 1\,222.125\,0 + 245.375\,0D_{2t} + 347.625\,0D_{3t} - 62.125\,0D_{4t}$$
$$t = \quad (20.372\,0)^{*} \qquad (2.892\,2)^{*} \qquad (4.097\,4)^{*} \qquad (-0.732\,2)^{**} \tag{9.7.3}$$
$$R^2 = 0.531\,8$$

其中 * 表示 p 值低于 5%，** 表示 p 值高于 5%。

表 9.4　美国 1978—1985 年各季度冰箱销售数据（以千台计）

FRIG	DUR	D_2	D_3	D_4	FRIG	DUR	D_2	D_3	D_4
1 317	252.6	0	0	0	943	247.7	0	0	0
1 615	272.4	1	0	0	1 175	249.1	1	0	0
1 662	270.9	0	1	0	1 269	251.8	0	1	0
1 295	273.9	0	0	1	973	262.0	0	0	1
1 271	268.9	0	0	0	1 102	263.3	0	0	0
1 555	262.9	1	0	0	1 344	280.0	1	0	0
1 639	270.9	0	1	0	1 641	288.5	0	1	0
1 238	263.4	0	0	1	1 225	300.5	0	0	1
1 277	260.6	0	0	0	1 429	312.6	0	0	0

1 258	231.9	1	0	0	1 699	322.5	1	0	0
1 417	242.7	0	1	0	1 749	324.3	0	1	0
1 185	248.6	0	0	1	1 117	333.1	0	0	1
1 196	258.7	0	0	0	1 242	344.8	0	0	0
1 410	248.4	1	0	0	1 684	350.3	1	0	0
1 417	255.5	0	1	0	1 764	369.1	0	1	0
919	240.4	0	0	1	1 328	356.4	0	0	1

注：FRIG＝以千计的冰箱销售数量；DUR＝以 1982 年 10 亿美元计的耐用品支出；D_2 取值 1 表示第 2 季度，否则取值 0；D_3 取值 1 表示第 3 季度，否则取值 0；D_4 取值 1 表示第 4 季度，否则取值 0。

资料来源：*Business Statistics and Survey of Current Business*, Department of Commerce (Various issues).

因为我们把第 1 季度视为基准组，所以各季度虚拟变量的系数现在就是级差截距，表示虚拟变量取值 1 的那个季度 Y 的平均值与基准季度相比有多大的不同。换言之，季节虚拟变量的系数将给出 Y 的平均值相对基准季节的增加或减少。如果你将各个级差截距值与基准平均值 1 222.125 相加，你就会得到各个季度的平均值。如此以来，除四舍五入的误差外，你将恰好重新得到方程 (9.7.2)。

但现在你将看到，将一个季度作为基准季度的价值，(9.7.3) 表明第 4 季度 **Y** 的平均值并非统计上异于第 1 季度的平均值，因为第 4 季度的虚拟变量系数并非统计显著。当然，你的结论将随着哪一组作为基准组而变化，但总体结论将不会改变。

我们如何能得到冰箱销售的除季节性时间序列呢？这很容易做到。用模型 (9.7.2) [或 (9.7.3)] 对每个观测估计 Y 值，然后用每个实际 Y 值减去估计的 Y 值便得到回归 (9.7.2) 的残差 ($Y_t-\hat{Y}_t$)。我们在表 9.5 中给出这些残差。[15]

这些残差代表着什么呢？它们代表着冰箱销售时间序列中除去季节因素后剩余的成分，即趋势、周期和随机几种成分（但参见章末注 15 中的忠告）。

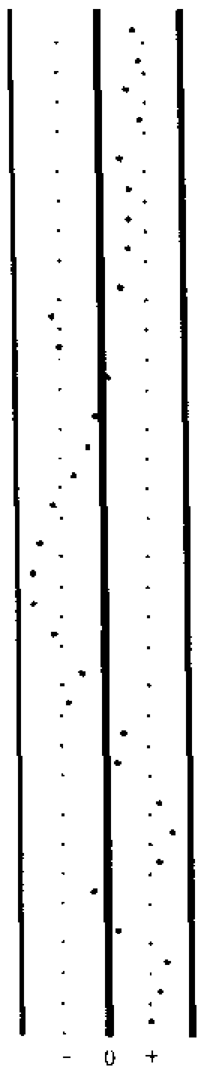
既然模型 (9.7.2) 和 (9.7.3) 都没有包含任何协变量，那我们在模型中引入定量回归元会改变这种情况吗？由于耐用品支出对冰箱需求有重要影响，所以让我们通过引入这个变量来扩展我们的模型 (9.7.3)。以 1982 年 10 亿美元计的耐用品支出数据已经在表 9.3 中给出。这就是我们模型中的（定量）X 变量。回归结果如下：

$$\hat{Y}_t=456.244\,0+242.497\,6D_{2t}+325.264\,3D_{3t}-86.080\,4D_{4t}+2.773\,4X_t$$
$$t=\ (2.559\,3)^{*}\quad(3.695\,1)^{*}\quad(4.942\,1)^{*}\quad(-1.307\,3)^{**}\quad(4.449\,6)^{*}\tag{9.7.4}$$
$$R^2=0.729\,8$$

316 其中 * 表示 p 值低于 5%，** 表示 p 值高于 5%。

表 9.5　　冰箱销售回归：实际值、拟合值和残差[方程(9.7.3)]

	实际值	拟合值	残差	残差图
1978－I	1 317	1 222.12	94.875	
1978－II	1 615	1 467.50	147.500	
1978－III	1 662	1 569.75	92.250	
1978－IV	1 295	1 160.00	135.000	
1979－I	1 271	1 222.12	48.875	
1979－II	1 555	1 467.50	87.500	
1979－III	1 639	1 569.75	69.250	
1979－IV	1 238	1 160.00	78.000	
1980－I	1 277	1 222.12	54.875	
1980－II	1 258	1 467.50	－209.500	
1980－III	1 417	1 569.75	－152.750	
1980－IV	1 185	1 160.00	25.000	
1981－I	1 196	1 222.12	－26.125	
1981－II	1 410	1 467.50	57.500	
1981－III	1 417	1 569.75	－152.750	
1981－IV	919	1 160.00	－241.000	
1982－I	943	1 222.12	－279.125	
1982－II	1 175	1 467.50	－292.500	
1982－III	1 269	1 569.75	－300.750	
1982－IV	973	1 160.00	－187.000	
1983－I	1 102	1 222.12	－120.125	
1983－II	1 344	1 467.50	－123.500	
1983－III	1 641	1 569.75	71.250	
1983－IV	1 225	1 160.00	65.000	
1984－I	1 429	1 222.12	206.875	
1984－II	1 699	1 467.50	231.500	
1984－III	1 749	1 569.75	179.250	
1984－IV	1 117	1 160.00	－43.000	
1985－I	1 242	1 222.12	19.875	
1985－II	1 684	1 467.50	216.500	
1985－III	1 764	1 569.75	194.250	
1985－IV	1 328	1 160.00	168.000	



同样记住，我们仍把第 1 季度作为基组。与 (9.7.3) 中一样，我们看到第 2 和第 3 季度的级差截距系数都统计显著地异于第 1 季度，但第 4 季度的截距与第 1 季度的截距在统计上大致相同。约为 2.77 的 X（耐用品支出）系数告诉我们，容许季节效应，若耐用品支出增加一个单位，则冰箱销售平均上升约 2.77 个单位，即近似为 3 个单位；记住冰箱销售以千台计，而 X 则以 1982 年 10 亿美元计。

317 这里一个有趣的问题是：正如冰箱销售所表现出来的季节性类型一样，耐用品支出不也表现出季节类型吗？那我们如何考虑 X 的季节性呢？(9.7.4) 中有趣的是，模型中的虚拟变量不仅去掉了 Y 中的季节性，还去

掉了 X 中的季节性（若存在的话）。（这来自于统计学中的一个著名定理，即**弗里希-沃夫**（Frisch-Waugh）**定理**。[16]）真可谓是一石（虚拟变量方法）二鸟（除去两个时间序列中的季节性）！

若你想得到上述命题的非正式证明，则只须按如下步骤：（1）如（9.7.2）或（9.7.3）那样将 Y 对虚拟变量回归并存储残差 S_1，这些残差代表着除季节性趋势后的 Y。（2）对 X 做一个类似的回归并得到此回归的残差 S_2；这些残差代表着除趋势后的 X。（3）将 S_1 对 S_2 回归。你将发现此回归中的斜率系数恰好就是回归（9.7.4）中 X 的系数。

§9.8 分段线性回归

为了说明虚拟变量的另一种用处，考虑图 9.5，它表明一个假想的公司是如何酬劳其销售代表的。其支付佣金的方式取决于销售量的一个目标或临界水平 X^*，低于那个水平，就使用一种（随机）佣金结构，高于那个水平就使用另一种佣金结构。（注：除销售量外，还有其他因素影响销售佣金。假定所有这些其他因素都由随机干扰项表示。）更具体而言，假定销售佣金在临界 X^* 之前随销售量线性地增加，在这个临界水平之后仍线性地增加，只是斜率更大。于是，我们得到一个由两段或两部分构成的**分段线性回归**（piecewise linear regression），在图 9.5 中用 Ⅰ 和 Ⅱ 表示这两段，而且销售佣金是在临界值处改变斜率的。给定佣金、销售额和临界值 X^* 的数据，就能用虚拟变量的方法估计图 9.5 中所示的分段线性回归的两个线段的（不同）斜率。我们可以如下进行：

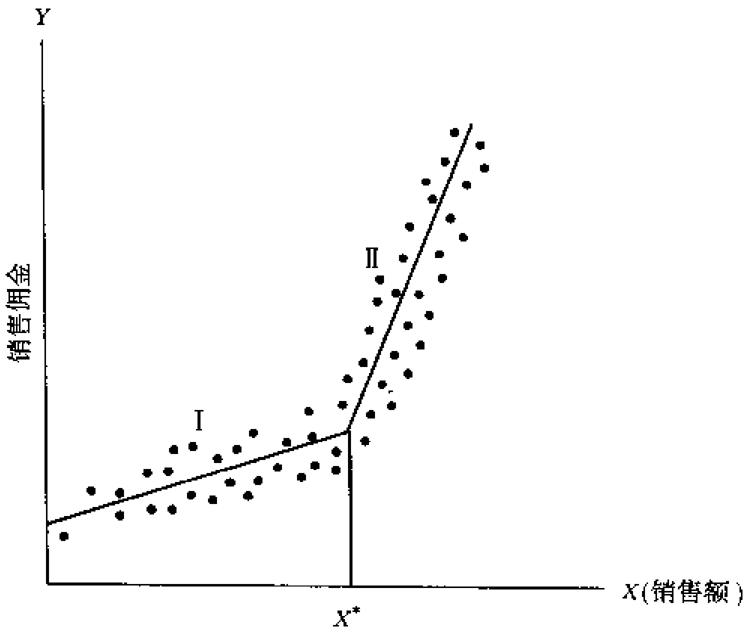


图 9.5 假想销售佣金与销售量之间的关系

注：Y 轴上的截距表示确保的最低佣金。

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i \tag{9.8.1}$$

其中 Y_i = 销售佣金

X_i = 销售员带来的销售量

X^* = 销售临界值，也被称为**结点**（knot）（事先已知）[17]

$D = 1$，若 $X_i > X^*$

$= 0$，若 $X_i < X^*$

假定 $E(u_i) = 0$，我们就立即看到：

$$E(Y_i \mid D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i \tag{9.8.2}$$

它给出目标水平 X^* 之前销售佣金的均值，而：

$$E(Y_i \mid D_i = 1, X_i, X^*) = \alpha_1 - \beta_2 X^* + (\beta_1 + \beta_2) X_i \tag{9.8.3}$$

则给出目标水平 X^* 之后销售佣金的均值。

于是，对图 9.5 中的分段线性回归而言，β_1 就给出了线段 Ⅰ 中回归线的斜率，而（$\beta_1 + \beta_2$）则给出了线段 Ⅱ 中回归线的斜率。要检验回归在临界值 X^* 处没有转折的假设，通过了解所估计的级差斜率系数 $\hat{\beta}_2$ 的统计显著性很容易就能做到（见图 9.6）。

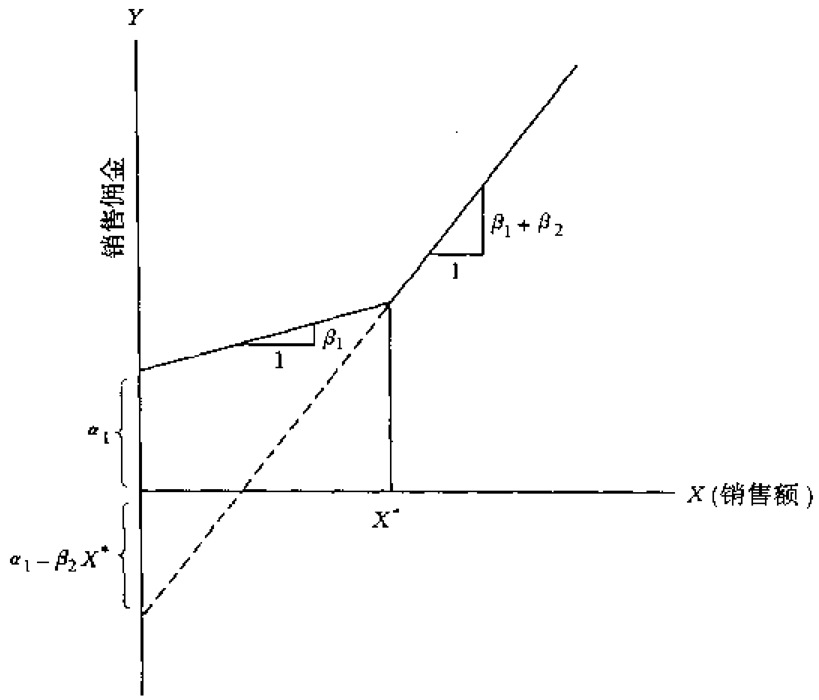


图 9.6　分段线性回归的参数

顺便指出，我们刚才所讨论的分段线性回归，只是所谓间断函数（spline functions）这种更一般函数类型中的一个例子。[18]

例 9.7 总成本与产出之间的关系

319 作为分段线性回归的一个应用例子，考虑表 9.6 中所给出的假想的总成本—总产出数据。我们还被告知，总成本在产出为 5 500 单位时可能会改变其斜率。

以 Y 表示总成本，X 表示总产出，我们得到如下结果：

$$\begin{aligned}\hat{Y}_i &= -145.72 \quad + \quad 0.279\,1X_i + 0.094\,5(X_i - X_i^*)D_i \\ t &= \ (-0.824\,5) \ (6.066\,9) \quad (1.144\,7) \\ & \qquad\qquad R^2 = 0.973\,7,\ X^* = 5\,500\end{aligned} \tag{9.8.4}$$

如这些结论所示，生产的边际成本约为每单位产出 28 美分，而当产出高于 5 500 单位后，边际成本则约为 37（=28+9）美分，由于虚拟变量在 5% 的显著性水平上仍不显著，所以这两个斜率之间的差别不是统计上显著的。于是，在所有的实践中，都可以去掉虚拟变量而将总成本直接对总产出回归。

表 9.6　　产出与总成本的假想数据

总成本（美元）	产出（单位）
256	1 000
414	2 000
634	3 000
778	4 000
1 003	5 000
1 839	6 000
2 081	7 000
2 423	8 000
2 734	9 000
2 914	10 000

§9.9 综列数据回归模型

320 记得我们在第 1 章曾讨论过一系列可用于经验分析的数据，如横截面数据、时间序列数据、混合数据（时间序列数据与横截面数据的综合）和综列数据等。虚拟变量方法很容易就能扩展应用到混合数据和综列数据。由于综列数据的使用在应用研究中日益常见，所以我们将在第 16 章更详尽地讨论

这个专题。

§9.10　虚拟变量方法的某些技术问题

在半对数回归中对虚拟变量的解释

我们在第 6 章讨论过对数线性模型，其中的回归子为对数形式，而回归元为线性形式。在这样的模型中，回归元的斜率系数给出了半弹性的解释，即回归元的单位变化导致回归子的百分比变化。这只在回归元为定量变量时如此。如果回归元是一个虚拟变量又会怎么样？为明确起见，考虑如下模型：

$$\ln Y_i = \beta_1 + \beta_2 D_i + u_i \tag{9.10.1}$$

其中，Y 表示小时工资率，D 取值 1 表示女性，取值 0 表示男性。

我们如何解释这样的一个模型呢？假定$E(u_i)=0$，我们得到：

男工人的工资函数：

$$E(\ln Y_i \mid D_i = 0) = \beta_1 \tag{9.10.2}$$

女工人的工资函数：

$$E(\ln Y_i \mid D_i = 1) = \beta_1 + \beta_2 \tag{9.10.3}$$

因此，截距 β_1 给出了小时工资对数的均值，而“斜率”系数则给出了男女之间小时工资对数均值的差别。这是一种相当笨拙的表述方法。但如果我们对 β_1 取反对数，所得到的就不是男工人小时工资的均值，而是他们工资的**中位数**（median）。如你所知，均值、中位数和众数是度量一个随机变量中心趋势的三种方法。如果我们对（$\beta_1+\beta_2$）取反对数，则得到女工人小时工资的中位数。

例 9.8　小时工资的对数与性别的关系

321　为了说明（9.10.1），我们使用例 9.2 背后的数据，基于 528 个观测所得到的回归结果如下：

$$\begin{aligned}\ln \hat{Y}_i &= \quad 2.176\,3 \quad - \quad 0.243\,7D_i \\ t &= (72.294\,3)^* \quad (-5.504\,8)^* \qquad R^2 = 0.054\,4\end{aligned} \tag{9.10.4}$$

其中，* 表示 p 值实际上是零。

取 2.176 3 的反对数便得到男工人小时工资的中位数8.813 6美元，而取[(2.176 3 − 0.243 7) = 1.928 6]的反对数则得到女工人小时工资的中位数6.879 6美元。因此，与男工人相比，女工人小时工资的中位数约低 21.94%[(8.813 6 − 6.879 6)/8.813 6]。

有趣的是，我们按照霍尔沃森（Halvorsen）和帕姆奎斯特（Palmquist）所建议的方法，我们可直接得到一个虚拟回归元的半弹性。[19]对所估计的虚拟系数取反对数（以 e 为底）并减去 1，再将差值乘以 100 即可。（至于其背后的原因，可参见附录 9A。）因此，取 −0.243 7 的反对数得到 0.783 66，减去 1 后得到 −0.216 3，再乘以 100 便得到 21.63%。这就表明一个女工人（$D = 1$）薪水的中位数比同等男工人薪水的中位数低约 21.63%，除四舍五入的误差，与我们前面所得到的结论相同。

虚拟变量与异方差性

让我们重新回到美国 1970—1981 年期间和 1982—1995 年期间及整个 1970—1995 年期间的储蓄—收入回归。在用虚拟变量方法检验结构稳定性时，我们假定了误差$\mathrm{var}(u_{1i}) = \mathrm{var}(u_{2i}) = \sigma^2$，即两个期间的误差方差相同。这也是邹至庄检验所需要的一个假定。如果这个假定不成立——即两个子期间的误差方差不同，那就很有可能得到具有误导性的结论。所以，必须首先利用合适的统计方法来验证两个子期间方差的相等性。尽管我们在异方差性那一章中将更全面地讨论这个专题，但我们在第 8 章还是看到了，如何用 F 检验来解决这个问题。[20]（参见我们在那一章中对邹至庄检验的讨论。）我们在那里曾指出，两个期间的误差方差看来有所不同。因此，前面给出的邹至庄检验和虚拟变量方法的结论可能都不是那么可靠。当然，我们这里只是说明可以用于解决一个问题（如结构性稳定的问题）的各种方法。在某个特定的研究中，这些方法可能站不住脚，但大多数统计方法都是如此。当然，我们在稍后有关异方差性的一章中将会看到，也可以采用一些适当的补救措施来解决这个问题（参见习题 9.28）。

虚拟变量与自相关

322 除异方差性外，经典线性回归模型还假定回归模型中的误差项不相关。但若这些误差项（特别是涉及虚拟回归元的模型）相关，结果会怎么样呢？由于我们在关于自相关的一章中将深入讨论自相关这个专题，所以我们把对这个问题的回答留到那个时候。

若因变量是一个虚拟变量会怎么样?

在我们到目前为止所讨论过的模型中，回归子都是定量的，而回归元可以是定性的，也可以是定量的，也可以二者都有。但回归子偶尔也可能是定性或虚拟变量。比如，考虑一个工人是否参加劳动力市场的决策。参与的决策是个是否命题，如果决定参与则回答是，决定不参与则回答否。当然，参与劳动力市场的决策还取决于其他几个因素，如起始工资率、受教育水平和劳动力市场状况（由失业率度量）等。

我们还能用 OLS 估计回归子是虚拟变量的回归模型吗？是的，我们仍可机械地这么做。但在这种模型中会遇到几个统计问题。而且既然还有一些不同于 OLS 估计的方法不存在这些问题，那我们就把这个专题留到稍后的一章中讨论（见关于 logit 和 probit 模型的第 15 章）。我们在那一章中还将讨论回归子不止两种类型的模型；比如，是自己开车、乘公汽还是坐地铁去上班的决策，或者选择兼职工作、全职工作还是根本就不工作的决策。与因变量只有两种类型的**二值因变量模型**（dichotomous dependent variable models）相比，这种模型被称为**多值因变量**（polytomous dependent variable）模型。

§9.11 进一步研究的专题

文献中讨论的几个与虚拟变量相关的专题相当高深，包括（1）**随机或变参数模型**（random or varying parameters models），（2）**转换回归模型**（switching regression models），及（3）**非均衡模型**（disequilibrium models）。

在本书所讨论的回归模型中，假定参数 β 都是未知但固定的。随机系数模型及其几个变形则假定 β 也可以是随机的。此领域一个重要的参考文献就是斯瓦迷（Swamy）的一本书。[21]

323 在同时使用级差截距与级差斜率的虚拟变量模型中，暗含地假定了我们知道转折点。于是，在我们 1970—1995 年期间的储蓄—收入例子中，深信 1982 年的衰退改变了储蓄与收入之间的关系，我们将期间分为 1970—1981 年和 1982—1995 年两个（衰退前和衰退后）子区间。有时候不容易确定转折什么时候发生。**转换回归模型**（switching regression models，SRM）就是在这种情况下提出的方法。SRM 将转折点视为一个随机变量，并通过一个迭代过程来决定转折实际上在什么时候发生。此领域的开山之作当属戈德菲尔德（Goldfeld）和匡特（Quandt）的那本书。[22]

欲处理市场未出清（即需求不等于供给）的**非均衡状态**（disequilibrium situations），则需要一些特殊的估计方法。经典的例子是一种商品的供给与需求。对一种商品的需求是其价格和其他一些变量的函数，而对一种商品的

供给也是其价格和其他一些变量的函数，但这些变量与进入需求函数的变量不完全相同。现在，实际购买和卖出的商品数量并不一定等于供求相等时的数量，因此导致非均衡。对非均衡模型的全面讨论，读者可参见匡特的著作。[23]

§9.12 要点与结论

1. 取值 1 或 0（或其线性变形）的虚拟变量是在回归模型中引入定性回归元的一种手段。

2. 虚拟变量是一种基于性质或属性（性别、婚姻状况、种族和宗教信仰等）而将一个样本分为不同子群的数据分类方法，并暗含地容许对每个子群做个别回归。如果回归子对各子群中定性变量的变化有不同的响应，那就由各子群回归的截距或斜率或二者的差别反映出来。

3. 尽管虚拟变量方法是一种通用方法，但仍需仔细处理。首先，如果回归中包含了截距项，那么，虚拟变量的个数就必须比每个定性变量的类别数少一个。其次，附属于虚拟变量的系数必须总是相对基组或参照组（即虚拟变量取值为 0 的组）来解释。基组的选择取决于所进行研究的目的。最后，如果一个模型含有几个均可分为几类的定性变量，那么，引入虚拟变量就会占用很大的自由度。因此，总是要根据分析中可以供使用的观测数来决定要引入的虚拟变量数。

324 4. 在虚拟变量的诸多应用中，本章仅考虑了如下为数不多的几种：(1) 将两个或多个回归进行比较，(2) 除去时间序列数据中的季节性，(3) 交互虚拟变量，(4) 在半对数模型中解释虚拟变量，及 (5) 分段线性回归。

5. 我们对在异方差性和自相关情形下使用虚拟变量提出了郑重警告。但由于我们在以后的章节中将会详尽讨论这些专题，所以还是留待以后再探个究竟。

习 题

问答题

9.1 如果你有连续几年的月度数据，为检验如下假设，需要引入多少个虚拟变量：

a. 一年中 12 个月份都表现出季节类型。

b. 只有 2、4、6、8、10、12 月表现出季节类型。

9.2 考虑如下回归结果（t 比率放在括号内）[1]：

$$\hat{Y}_t = 1\,286 \quad + \quad 104.97X_{2i} - \quad 0.026X_{3i} + \quad 1.20X_{4i} + \quad 0.69X_{5i}$$

$$t = \quad (4.67) \qquad (3.70) \qquad (-3.80) \qquad (0.24) \qquad (0.08)$$

$$-19.47X_{6i} + 266.06X_{7i} - 118.64X_{8i} - 110.61X_{9i}$$

$$(-0.40) \qquad (6.94) \qquad (-3.04) \qquad (-6.14)$$

$$R^2 = 0.383 \quad n = 1\ 543$$

其中FY＝妻子希望每年花在工作上的小时数，以每年工作的小时数加上花在找工作上的时间之和计

X_2＝妻子税后真实时薪

X_3＝丈夫在上一年度税后真实收入

X_4＝妻子的年龄

X_5＝妻子的受教育年数

X_6＝态度变量，若被调查者愿意工作而且她丈夫也同意她工作则取值1，否则取值0

X_7＝态度变量，若被调查者的丈夫支持她工作则取值1，否则取值0

X_8＝年龄低于6岁的子女数

X_9＝年龄在6～13岁之间的子女数

a. 各非虚拟回归元系数的符号有经济含义吗？说明你的观点。

b. 你如何解释虚拟变量 X_6 和 X_7？这些虚拟变量统计显著吗？由于样本相当大，你可以用"$2-t$"经验法则来回答问题。

325 c. 在这项研究中，一位妇女的年龄和受教育程度不是影响其劳动力参与决策的显著因素，你认为这是为什么？

9.3 考虑如下回归结果。[2] (实际数据在表9.7中。)

$$\widehat{\mathrm{UN}}_t = \quad 2.749\ 1 + 1.150\ 7D_t \ - \ 1.529\ 4V_t \ - \ 0.851\ 1\ (D_tV_t)$$

$$t = (26.896) \quad (3.628\ 8) \quad (-12.555\ 2) \quad (-1.981\ 9)$$

$$R^2 = 0.912\ 8$$

326 其中 UN＝失业率，%

V＝岗位空缺率，%

D＝1，对1966年第4季度之后的期间

＝0，对1966年第4季度之前的期间

t ＝时间，以季度度量

注：1966年的第4季度，当时英国的劳工党政府放松了国民保险法案的限制，以统一收费率和救济金与先前收入相关相结合的混合体系取代原来短期失业救济金的统一收费率体系，从而提高了失业救济金水平。

a. 你对失业率和空岗率之间的关系有何先验预期？

b. 保持岗位空缺率不变，在从1966年第4季度开始的期间内，平均失业率为多少？它与1966年第4季度之前的期间有显著不同吗？你如何知道？

c. 1966年第4季度之前和之后的斜率在统计上不同吗？你如何

知道？

d. 从这项研究能安全地断定慷慨的失业救济金导致更高的失业率吗？这在经济上讲得通吗？

表 9.7　　习题 9.3 中所做回归的数据矩阵

年份和季度	失业率 UN,%	岗位空缺率 V,%	D	DV	年份和季度	失业率 UN,%	岗位空缺率 V,%	D	DV
1958 - Ⅳ	1.915	0.510	0	0	1965 - Ⅰ	1.201	0.997	0	0
1959 - Ⅰ	1.876	0.541	0	0	- Ⅱ	1.192	1.035	0	0
- Ⅱ	1.842	0.541	0	0	- Ⅲ	1.259	1.040	0	0
- Ⅲ	1.750	0.690	0	0	- Ⅳ	1.192	1.086	0	0
- Ⅳ	1.648	0.771	0	0	1966 - Ⅰ	1.089	1.101	0	0
1960 - Ⅰ	1.450	0.836	0	0	- Ⅱ	1.101	1.058	0	0
- Ⅱ	1.393	0.908	0	0	- Ⅲ	1.243	0.987	0	0
- Ⅲ	1.322	0.968	0	0	- Ⅳ	1.623	0.819	1	0.819
- Ⅳ	1.260	0.998	0	0	1967 - Ⅰ	1.821	0.740	1	0.740
1961 - Ⅰ	1.171	0.968	0	0	- Ⅱ	1.990	0.661	1	0.661
- Ⅱ	1.182	0.964	0	0	- Ⅲ	2.114	0.660	1	0.660
- Ⅲ	1.221	0.952	0	0	- Ⅳ	2.115	0.698	1	0.698
- Ⅳ	1.340	0.849	0	0	1968 - Ⅰ	2.150	0.695	1	0.695
1962 - Ⅰ	1.411	0.748	0	0	- Ⅱ	2.141	0.732	1	0.732
- Ⅱ	1.600	0.658	0	0	- Ⅲ	2.167	0.749	1	0.749
- Ⅲ	1.780	0.562	0	0	- Ⅳ	2.107	0.800	1	0.800
- Ⅳ	1.941	0.510	0	0	1969 - Ⅰ	2.104	0.783	1	0.783
1963 - Ⅰ	2.178	0.510	0	0	- Ⅱ	2.056	0.800	1	0.800
- Ⅱ	2.067	0.544	0	0	- Ⅲ	2.170	0.794	1	0.794
- Ⅲ	1.942	0.568	0	0	- Ⅳ	2.161	0.790	1	0.790
- Ⅳ	1.764	0.677	0	0	1970 - Ⅰ	2.225	0.757	1	0.757
1964 - Ⅰ	1.532	0.794	0	0	- Ⅱ	2.241	0.746	1	0.746
- Ⅱ	1.455	0.838	0	0	- Ⅲ	2.366	0.739	1	0.739
- Ⅲ	1.409	0.885	0	0	- Ⅳ	2.324	0.707	1	0.707
- Ⅳ	1.296	0.978	0	0	1971 - Ⅰ	2.516*	0.583*	1	0.583*
					- Ⅱ	2.909*	0.524*	1	0.524*

*初步估计值。

资料来源：Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958－1971," *The Economic Journal*, vol. 82, March 1972, p. 202.

9.4 威廉·诺德豪斯（William Nordhaus）从 1972—1979 年度数据估计了如下模型来解释 OPEC 的石油价格行为（括号中为标准误）。[3]

$$y_t = 0.3\ x_{1t} + 5.22\ x_{2t}$$
$$\text{se} = (0.03) \quad (0.50)$$

其中 y = 当年与前一年的价差（美元/桶）

x_1 = 当年现货价格与前一年 OPEC 价格之差

x_2 = 为虚拟变量，在 1974 年取值 1，否则取值 0

解释这个结果并用图说明结论。这些结论表明 OPEC 什么样的垄断势力？

9.5 考虑如下模型：

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + u_i$$

其中 Y = 一位大学教授的年薪

X = 从教年限

D – 性别虚拟变量

考虑定义虚拟变量的三种方式：

a. D 对男性取值 1，对女性取值 0

b. D 对女性取值 1，对男性取值 2

c. D 对女性取值 1，对男性取值 −1

对每种虚拟变量定义解释上述回归模型。是否有某个方法比另外一个方法更好？说明你的理由。

327 **9.6** 参考回归（9.7.3）。你如何检验 D_2 和 D_3 的系数相同的假设？D_2 和 D_4 的系数相同的假设？如果 D_3 的系数在统计上与 D_2 的系数不同，而且 D_4 的系数与 D_2 的系数也不同，那是否意味着 D_3 和 D_4 的系数也不同呢？

提示：$\text{var}(A \pm B) = \text{var}(A) + \text{var}(B) \pm 2\text{cov}(A, B)$。

9.7 参考本章中讨论的美国储蓄—收入一例。

a. 你如何得到从混合回归（9.5.4）中推导出的（9.5.5）和（9.5.6）的回归系数的标准误？

b. 为了得到数值结果，可能的话，还需要什么附加信息？

9.8 米勒（R. J. Miller）在研究联邦储蓄保险公司（FDIC）为准备 91 项银行检查所花费的时间中，估计了如下函数[4]：

$$\widehat{\ln Y_i} = 2.41 + 0.367\,4\ln X_1 + 0.221\,7\ln X_2 + 0.080\,3\ln X_3$$
$$(0.047\,7) \qquad (0.062\,8) \qquad (0.028\,7)$$
$$-0.175\,5D_1 + 0.279\,9D_2 + 0.563\,4D_3 - 0.257\,2D_4$$
$$(0.290\,5) \qquad (0.104\,4) \qquad (0.165\,7) \qquad (0.078\,7)$$
$$R^2 = 0.766$$

其中 Y = FDIC 检查者的劳动时间

X_1 = 银行的总资产

X_2 = 银行的办公室总数

X_3 = 银行分类存款占总存款的比例

D_1 = 若管理评级为“优”则取值 1

D_2 = 若管理评级为“普通”则取值 1

D_3 = 若管理评级为“满意”则取值 1

D_4 = 若检查与评级同时进行则取值 1

括号中的数字为估计的标准误。

a. 解释这些结论。

b. 在解释模型中的虚拟变量时因 Y 以对数形式出现而带来什么问题了吗?

c. 你如何解释虚拟变量的系数?

9.9 为了评价美联储自1979 年 7 月以来放松利率管制的政策效果，我们的一个学生兰格 (Sidney Langer) 利用 1975 - Ⅲ至 1983 - Ⅱ期间的季度数据估计了如下模型[5]：

$$\begin{aligned}
\hat{Y} = &\ 8.5871 - 0.1328P_t - 0.7102\mathrm{Un}_t - 0.2389M_t \\
\text{se} = &\ (1.9563)\quad (0.0992)\quad (0.1909)\quad (0.0727) \\
&+ 0.6592Y_{t-1} + 2.5831\mathrm{Dum}_t \\
&\quad (0.1036)\qquad (0.7549) \\
&R^2 = 0.9156
\end{aligned}$$

328 其中 Y = 3 月期国库券利率

P = 预期通货膨胀率

Un = 进行季节调整后的失业率

M = 货币基的变化

Dum = 虚拟变量，对 1979 年 7 月 1 号之后的观测都取值 1

a. 解释这些结论。

b. 放松利率管制有什么样的影响? 这些结论在经济上讲得通吗?

c. P_t、Un_t 和 M_t 的系数都为负，你能给出经济学上的根据吗?

9.10 参考文中讨论的分段回归。假设在 X^* 处不仅斜率系数发生了变化，而且回归线还发生了跳跃，如图 9.7 所示。你将如何修正 (9.8.1)，以考虑回归线在 X^* 处的跳跃?

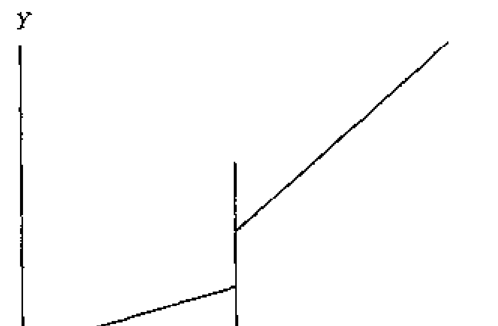


图 9.7 不连续的分段线性回归

9.11 可乐每盎司价格的决定因素。我的一个学生谢弗（Cathy Schaefer）对 77 个横截面观测数据估计了如下回归[6]：

$$P_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \mu_i$$

其中 P_i = 可乐的每盎司价格

D_{1i} = 001，若是折扣商店

= 010，若是连锁商店

= 100，若是方便商店

D_{2i} = 10，若是名牌可乐

= 01，若是无牌可乐

D_{3i} = 0 001，若 67.6 盎司（2 升）瓶装

= 0 010，若 28～33.8 盎司瓶装

= 0 100，若 16 盎司瓶装

= 1 000，若 12 盎司听装

329 估计的结果如下：

$$\hat{P}_i = 0.014\,3 - 0.000\,004 D_{1i} + 0.009\,0 D_{2i} + 0.000\,01 D_{3i}$$
$$\text{se} = \quad (0.000\,01) \quad (0.000\,11) \quad (0.000\,00)$$
$$t = \quad (-0.383\,7) \quad (8.392\,7) \quad (5.812\,5)$$
$$R^2 = 0.603\,3$$

注：标准误只保留到小数点后 5 位。

a. 对模型中引入虚拟变量的方式进行评论。

b. 假定虚拟变量的分类是可以接受的，你将如何对上述结果进行解释？

c. D_3 的系数为正且显著，你如何理性化这个结论？

9.12 根据20 世纪 70 年代早期 101 个国家以美元计人均收入和预期寿命的数据，森（Sen）和斯里瓦斯特瓦（Srivastava）得到如下回归结果[7]：

$$\hat{Y}_i = -2.40 + 9.39\ln X_i - 3.36[D_i(\ln X_i - 7)]$$
$$\text{se} = \quad (4.73) \quad (0.859) \quad (2.42) \qquad R^2 = 0.752$$

其中，D_i在 $\ln X_i$ 大于 7 时取值 1，否则取值 0。注：当 $\ln X_i = 7$ 时，X = 109 7 美元（近似）。

a. 以对数形式引入收入变量的原因是什么？

b. 你如何解释 $\ln X_i$ 的系数 9.39？

c. 引入回归元$D_i(\ln X_i - 7)$的理由是什么？你如何口头上解释这个回归元？你又如何解释这个回归元的系数 −3.36？（提示：分段线性回归。）

d. 假定穷国与富国之间的分界线为人均收入1 097美元，你如何推导出收入低于1 097美元国家的回归线和收入高于1 097美元国家的回归线？

e. 你从这个问题给出的结果中能得到什么一般结论？

9.13 考虑如下模型：

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

其中，D_i对前20个观测取值0，而对后30个观测取值1。并告诉你$\operatorname{var}(u_i^2) = 300$。

a. 你如何解释 β_1 和 β_2？

b. 这两组的均值分别是多少？

c. 你如何计算（$\hat{\beta}_1 + \hat{\beta}_2$）的方差？注：已知$\operatorname{cov}(\hat{\beta}_1 + \hat{\beta}_2) = -15$。

330 **9.14** 为了评价工作权法案（不把工会会员作为就业的先决条件）对工会关系的影响，利用美国1982年50个州的数据估计出如下回归结果[8]：

$$\widehat{\text{PVT}}_i = 19.8066 - 9.3917\text{RTW}_i$$
$$t = (17.0352) \quad (-5.1086) \qquad R^2 = 0.3522$$

其中PVT＝1982年私人部门雇员加入工会的百分比，若工作权法案生效则RTW取值1，否则取值0。注：1982年美国有20个州颁布工作权法案。

a. 据经验，预期PVT和RTW之间有什么样的关系？

b. 回归结果支持先验预期吗？

c. 解释回归结果。

d. 在那些没有颁布工作权法案的州，私人部门雇员加入工会的平均百分比是多少？

9.15 在如下回归模型中：

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

Y表示以美元度量的小时工资；D为虚拟变量，对大学毕业生取值1，对高中毕业生取值0。利用第3章中的OLS公式，证明$\hat{\beta}_1 = Y_{hg}$和$\hat{\beta}_2 = Y_{cg} - \bar{Y}_{hg}$，其中下角标有如下含义：hg表示高中毕业，cg表示大学毕业。总之，有n_1个高中毕业生和n_2个大学毕业生，总样本为$n = n_1 + n_2$。

9.16 为了研究伯利兹（Belize）在1970—1992年间的人口增长率，马克杰（Mukherjee）等人估计了如下模型[9]：

模型Ⅰ：

$$\ln(\text{Pop})_t = 4.73 + 0.024\,t$$
$$t = (781.25) \quad (54.71)$$

模型Ⅱ：

$$\ln(\text{Pop})_t = 4.77 + 0.015t - 0.075D_t + 0.011\,(D_t t)$$
$$t = (2\,477.92) \quad (34.01) \quad (-17.03) \quad (25.54)$$

其中，Pop＝以百万计的人口数量，t ＝趋势变量，D_t 对 1978 年开始的观测取值 1，对此前的观测取值 0，ln 表示自然对数。

a. 在模型Ⅰ中，伯利兹人口在样本期的增长率是多少？

b. 1978 年之前和之后的人口增长率在统计上不同吗？你如何知道？若不同，1972—1977 年和 1978—1992 年期间的增长率各为多少？

解答题

331 **9.17** 利用表 9.7 中给出的数据，检验 1958－Ⅳ至 1966－Ⅲ和 1966－Ⅳ至 1971－Ⅱ两个子区间误差方差相同的假设。

9.18 利用第8章所讨论的方法，比较无约束和有约束回归（9.7.3）和（9.7.4）；即检验所施加约束的有效性。

9.19 在本章讨论的美国储蓄—收入回归（9.5.4）中，假设虚拟变量 D_i 的取值不再是 1 和 0，而是令 $Z_i = a + bD_i$，其中 $D_i = 1$ 和 0，$a = 2$，$b = 3$。比较你的结论。

9.20 继续储蓄—收入回归（9.5.4），假设你让 D_i 对第二区间的观测取值 0，而对第一区间的观测取值 1。（9.5.4）所示的结论有何变化？

9.21 利用表 9.2 中给出的数据，考虑如下模型：

$$\ln \text{Savings}_i = \beta_1 + \beta_2 \ln \text{Income}_i + \beta_3 \ln D_i + u_i$$

其中，ln 表示自然对数；D_i 对 1970—1981 年取值 1，对 1982—1995 取值 10。

a. 如此确定虚拟变量的根据是什么？

b. 估计上述模型并解释你的结论。

c. 两个子期间储蓄函数的截距值是多少？你如何解释它们？

9.22 参考表 9.3 给出的厨具销售季节数据，并考虑如下模型：

$$\text{Sales}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + u_i$$

其中 D 都是虚拟变量，对第 2 至 4 季度取值 1 和 0。

a. 分别对洗碗机、污物粉碎机和洗衣机估计上述模型。

b. 你如何解释估计的斜率系数？

c. 你如何应用所估计的 α 对个别厨具销售数据除去季节变化？

9.23 增加耐用品支出这个回归元后重新估计习题 9.22 中的模型。

a. 这两道题的回归结果有所不同吗？如果有，什么因素能解释这个差别？

b. 如果耐用品支出的数据中也有季节性因素，你如何做出说明？

9.24 表 9.8 给出了美国 1916—1996 年四年一次的总统选举数据。[10]

a. 利用表 9.8 中给出的数据，提出一个合适模型来预测民主党在两党总统投票中所占的份额。

b. 你如何用这个模型去预测总统选举的结果？

332 c. 查特吉（Chatterjee）等人建议考虑如下尝试模型预测总统选举：

$$V=\beta_0+\beta_1 I+\beta_2 D+\beta_3 W+\beta_4(GI)+\beta_5 P+\beta_6 N+u$$

估计这个模型，并评论其结果与你所选用模型得到的结果之间的关系。

表 9.8　　美国总统选举：1916—1996

年份	V	W	D	G	I	N	P
1916	0.516 8	0	1	2.229	1	3	4.252
1920	0.361 2	1	0	−11.463	1	5	16.535
1924	0.417 6	0	−1	−3.872	−1	10	5.161
1928	0.411 8	0	0	4.623	−1	7	0.183
1932	0.591 6	0	−1	−14.901	1	4	7.069
1936	0.624 6	0	1	11.921	1	9	2.362
1940	0.550 0	0	1	3.708	1	8	0.028
1944	0.537 7	1	1	4.119	1	14	5.678
1948	0.523 7	1	1	1.849	1	5	8.722
1952	0.446 0	0	0	0.627	1	6	2.288
1956	0.422 4	0	−1	−1.527	−1	5	1.936
1960	0.500 9	0	0	0.114	−1	5	1.932
1964	0.613 4	0	1	5.054	1	10	1.247
1968	0.496 0	0	0	4.836	1	7	3.215
1972	0.382 1	0	−1	6.278	−1	4	4.766
1976	0.510 5	0	0	3.663	−1	4	7.657
1980	0.447 0	0	1	−3.789	1	5	8.093
1984	0.408 3	0	−1	5.387	−1	7	5.403
1988	0.461 0	0	0	2.068	−1	6	3.272
1992	0.534 5	0	−1	2.293	1	1	3.692
1996	0.547 4	0	1	2.918	1	3	2.268

注：

V　两党总统投票中民主党所占的份额；

I　指标变量，1 表示选举时在位总统为民主党，−1 表示选举时在位总统为共和党；

D　指标变量，1 表示民主党在任总统参加竞选，−1 表示共和党在任总统参加竞选，其他情况下为 0；

W　指标变量，对 1920、1944 和 1948 年的选举取值 1，其他取值 0；

G　选举年份前三个季度真实人均 GDP 的增长率；

P　现任政府前 15 个季度 GDP 缩减指数增长率的绝对值；

N　现任政府在前 15 个季度中真实人均 GDP 增长率超过 3.2% 的季度数。

9.25 参考回归（9.6.4）。检验平均时薪随受教育水平的增长率因性别和种族不同而不同。（提示：使用多个虚拟变量。）

9.26 参考回归（9.3.1）。为了发现性别和居住地这两个虚拟变量之间是否有某种交互作用，你该如何修正此模型？给出基于此模型的结果，并与（9.3.1）中给出的结果相比较。

9.27 在模型 $Y_i = \beta_1 + \beta_2 D_i + u_i$ 中，令 D_i 对前 40 个观测取值 0，而对其余 60 个观测取值 1。已知 u_i 的均值为 0，方差为 100。这两个观测集的均值和方差各为多少？

9.28 参考本章讨论的美国储蓄—收入回归。与（9.5.1）不同，考虑如下模型：

$$\ln Y_t = \beta_1 + \beta_2 D_t + \beta_3 X_t + \beta_4 (D_t X_t) + u_t$$

其中，Y 为储蓄，X 为收入。

a. 估计上述模型，并与（9.5.4）的结论相比较。哪个模型更好？

b. 你如何解释此模型中虚拟变量的系数？

c. 如我们在有关异方差性的章节中将看到的那样，对因变量取对数常常会减小数据中的异方差性。分两个期间将 Y 的对数对 X 做回归，看本例中是否如此？并看一下两个期间的误差方差在统计上是否相同。若相同，则可以按照本章中给出的方法将数据混合，再用邹至庄检验。

[习题注释]

[1] Jane Leuthold, "The Effect of Taxation on the Hours Worked by Married Women," *Industrial and Labor Relations Review*, no. 4, July 1978, pp. 520－526.（为适合我们的格式符号有所变化。）

[2] Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britiain, 1958－1971," *The Economic Journal*, vol. 82, March 1972, pp. 195－202.

[3] "Oil and Economic Performance in Industrial Countries," *Brookings Papers on Economic Activity*, 1980, pp. 341－388.

[4] "Examination of Man-Hour Cost for Independent, Joint, and Divided Examination Programs," *Journal of Bank Research*, vol. 11, 1980, pp. 28－35. 注：为了与我们的符号相一致，符号有所变化。

[5] Sidney Langer, "Interest Rate Deregulation and Short-Term Interest Rates," 未发表的学期论文。

[6] Cathy Schaefer, "Price Per Ounce of Cola Beverage as a Function of Place of Purchase, Size of Container, and Branded or Unbranded Product," 未发表的学期论文。

[7] Ashish Sen and Muni Srivastava, *Regression Analysis: Theory, Methods, and Applications*, Springer-Verlag, New York, 1990, p. 92. 符号有所变化。

[8] 回归结果中所用数据得自于 N. M. Meltz, "Interstate and interprovincial Differ-

ences in Union Density," *Industrial Relations*, vol. 28, no. 2, 1989, pp. 142－158.

[9] Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, London, 1998, pp. 372－375. 符号有所变化。

[10] 这些数据最初由耶鲁大学已预测总统选举结果多年的 Ray Fair 编制。我们只是复制于 Samprit Chatterjee, Ali S. Hadi, and Petram Price, *Regression Analysis by Example*, 3d ed., John Wiley & Sons, New York, 2000, pp. 150－151。

附录 9A　含虚拟回归元的半对数回归

我们在第 9.10 节注意到，在如下形式的模型中：

$$\ln Y_i = \beta_1 + \beta_2 D_i \tag{1}$$

Y 相对取值 1 或 0 的虚拟变量的变化（即半弹性）可用如下方法得到：(β_2 估计值的反对数 -1）$\times 100$，即

$$(e^{\hat{\beta}_2} - 1) \times 100 \tag{2}$$

证明如下：由于对数和指数互为反函数，所以我们可以把（1）写成：

$$\ln Y_i = \beta_1 + \ln(e^{\beta_2 D_i}) \tag{3}$$

现在，当 $D=0$ 时，$e^{\beta_2 D_i}=1$，当 $D=1$ 时，$e^{\beta_2 D_i}=e^{\beta_2}$。因此，从状态 0 到状态 1，$\ln Y_i$ 变化了（$e^{\beta_2}-1$）。但一个变量对数的变化只是相对变化，乘以 100 后就得到百分数变化。因此百分数变化就如所要证明的那样为$(e^{\beta_2}-1)\times 100$。（注：$\ln_e e=1$ 即以 e 为底 e 的对数等于 1，就像以 10 为底 10 的对数等于 1 一样。记住，以 e 为底的对数被称为自然对数，而已 10 为底的对数被称为常用对数。）

【注释】

[1] 我们将在第 15 章讨论序数尺度变量。

[2] 对这方面证据的一个综述，可参见 Bruce E. Kaufman and Julie L. Hotchkiss, *The Economics of Labor Market*, 5 th ed., Dryden Press, New York, 2000。

[3] 虚拟变量取值 0 和 1 绝非必需。通过一个诸如 $Z=a+bD$（$b\neq 0$）之类的线性函数，就可以把数对（0, 1）转换成任意一个其他的数对，这里 a 和 b 都是常数，而且 $D=1$ 或 0。当 $D=1$ 时，$Z=a+b$，而当 $D=0$ 时，$Z=a$，于是数对（0, 1）就变成了数对（a, $a+b$）。比如 $a=1$ 和 $b=2$，虚拟变量就变成了（1, 3）。这个表达式表明，定性或虚拟变量不具有一个自然度量尺度。这就是把它们描述成名义尺度变量的原因。

[4] ANOVA 模型被用于评价定量回归子和定性或虚拟回归元之关系的统计显著性。它们通常可以用于比较两或多组（或类别）均值的差别，因此相对于只用于比较两组均值的 t 检验更一般。

[5] 至于应用处理，可参见 John Fox, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage Publications, 1997, Chap. 8。

[6] 实际上你将得到数据矩阵退化的提示信息。

[7] Peter Kennedy, A *Guide to Econometrics*, 4th ed., MIT Press, Cambridge, Mass., 1998, p. 223。

[8] 数据得自下书的数据盘：Arthur S. Goldberger, *Introductory Econometrics*, Harvard University Press, Cambridge, Mass., 1998。我们已经在第 2 章考虑过这些数据。

[9] 本节内容取自作者的如下论文："Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note," 和 "Use of Dummy Variables…A Generalization," 均发表在 *American Statistician*, vol. 24, nos. 1 and 5, 1970, pp.50－52 and 18－21。

[10] 如在邹至庄检验中一样，混合法假定同方差性，即 $\sigma_1^2=\sigma_2^2=\sigma^2$。

[11] 如果我们用高中以下学历、高中文化和高中以上学历来表示受教育程度，那我们就只需要两个虚拟变量来表示这三个类别。

[12] 一个时间序列可以包含四个成分：季节性、周期性、趋势和严格的随机性。

[13] 关于季节调整的各种模型，可参见 Francis X. Diebod, *Elements of Forecasting*, 2d ed., South-Western Publishers, 2001, Chap. 5。

[14] 注意一个技术上的问题。为每个季度指定一个虚拟变量的方法，假定了季节因素（若存在的话）是确定的而非随机的。当我们在本书第 4 篇讨论时间序列计量经济学时还会重新回到这个问题。

[15] 当然，这假定了虚拟变量方法是除去时间序列中季节性的适当方法，而且假定时间序列（TS）又能表示成：TS $=s+c+t+u$，其中 s 表示季节性，c 表示趋势，t 表示周期性，u 表示随机成分。但如果时间序列的形式是TS $=(s)(c)(t)(u)$，即四个成分是相乘的关系，则上述除季节性的方法不合适，因为上述方法假定时间序列的四个成分是相加的关系。我们在时间序列计量经济学的章节中还会对此谈得更多。

[16] 其证明可参见 Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, Lyme, U.K., 1995, pp. 150－152。

[17] 可临界值并不总是那么明显。作为权宜之计，可以将因变量相对解释变量描点，并观察二者之间的关系是否在 X（即 X^*）的某给定值前后有明显的变化。发现转折点的分析方法在所谓转换回归模型中可以见到。但这是一个高深的专题，对此进行讨论的教材可参见 Thomas Fomby, R. Carter Hill, and Stanley Johnson, *Advanced Econometric Methods*, Springer-Verlag, New York, 1984, Chap.14。

[18] 对间断回归的一个容易切入的讨论，可参见 Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, 3d ed., New York, 2001, pp.228－230。

[19] Robert Halvorsen and Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, vol. 70, no. 3, pp.474－475.

[20] 即便在出现异方差性的情况下，也能使用邹至庄检验程序，但那就必须使用瓦尔德检验。此检验背后涉及的数学多少有些复杂。在有关异方差性的章节，我们会再次讨论这个问题。

[21] P.A.V.B. Swamy, *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, Berlin, 1971.

[22] S. Goldfeld and R. Quandt, *Nonlinear Methods in Econometrics*, North Holland, Amsterdam, 1972.

[23] Richard E. Quandt, *The Econometrics of Disequilibrium*, Basil Blackwell, New York, 1988.

第2篇

放松经典模型的假定

335 在第1篇里，我们详细考虑了经典正态线性回归模型，并且阐述了怎样利用它来处理统计推断的孪生问题：估计与假设检验以及预测问题。但请回忆，此模型是建立在一些简单化了的假定基础之上的。这些假定是：

假定1. 回归模型对参数而言是线性的。

假定2. 诸回归元 X 的值在重复抽样中是固定的。

假定3. 对给定的 X，干扰项 u_i 的均值为零。

假定4. 对给定的 X，u_i 的方差不变或有同方差性。

假定5. 对给定的 X，干扰无自相关。

假定6. 如果 X 是随机的，则干扰项与（随机的）诸 X 是独立的或至少是不相关的。

假定7. 观测次数必定大于回归元的个数。

假定8. 回归元的取值必须有足够的变异性。

假定9. 回归模型是正确地设定的。

假定10. 回归元之间无准确的线性关系（即无多重共线性）。

假定11. 随机（干扰）项 u_i 是正态分布的。

336 在继续往下讲之前，让我们指出，在大多数教科书中，没有列举11个假定之多。例如假定7和8只是默认而不明示列出。我们把它们分别地加以叙述，是为了要区分哪些假定能使OLS具有良好的性质（如BLUE），而哪些条件能使OLS成为有用。例如，即使假定8得不到满足，OLS估计量仍是BLUE，但这时OLS估计量的标准误对其系数来说，将必然相对地大（即 t 比率必然是小的），以致难以评价一个或多个回归元对解释平方和的贡献。

如韦瑟里尔（Wetherill）所指出的，实际上在应用经典线性回归模型时有两类主要问题：（1）关于对模型设定和对干扰项 u_i 的假定问题，（2）关于对数据的假定问题。[1] 属于第一类的为假定1、2、3、4、5、9和11。属于第二类的有假定6、7、8和10。此外，诸如越轨值（outliers）（异常或非典型的观测值）和数据中的测量误差等数据问题，也属于第二类。

关于来自对干扰和模型设定的假定问题主要有三：（1）要偏离一个具体的假定多远才会产生不可忽视的差别？例如，如果 u_i 恰好不是正态分布的，那么，在认为OLS估计量的BLUE性质受到破坏之前，我们能容忍多大程度的正态性偏离呢？（2）在一个具体的问题里，我们怎样能发现某一假定是否被破坏？比方说，我们怎样在某项应用中判知干扰项是不是正态分布的？我们讨论过关于正态性的 χ^2 检验和雅克-贝拉检验。（3）如果一或多个假定是错误的，我们能采取什么补救性措施？例如，如果我们在某项应用中发现关于同方差性的假定不真实，怎么办？

关于数据的假定，我们也遇到类似的问题。（1）某一具体问题的严重性如何？例如，多重共线性是否严重到几乎无法进行估计和推断的地步？（2）怎样确定数据问题的严重性？例如，保留或除掉一些可能代表越轨的观测值会不会造成分析上的巨大差别？怎样做出决定？（3）一些数据缺陷是否容易

得到补救？例如，有没有办法找到原始数据以明确数据中测量误差的来源？

遗憾的是，无法对所有这些问题都给出令人满意的答案。在本篇的其余部分里，我们的做法将是，对某些假定给予更多的注意，而不是对所有的假
337 定进行全面的探讨，特别是，不去深究假定2、3、6和11中的问题，其理由如下：

假定2和6：固定的回归元与随机的回归元。 记得我们的回归分析是以回归元为非随机的，并在重复抽样中取固定值的这个假定为基础的。之所以采取这一策略，有充分的理由。如第1章所指出的，经济学家不同于物理科学方面的科学家，常依赖于第二手材料，也就是他人如政府或私人机构所收集的材料。因此，实际可行的策略是，假定解释变量的值对于头上的问题来说是给定的，即使变量本身实质上也许是随机的。因此，回归分析的结果，是以这些给定值为条件的。

但假若我们不能把这些 X 看成非随机的或固定的，就引出**随机回归元**的情形。这种情形会是相当复杂的。根据假定 u_i 是随机的，如果诸 X 也是随机的，则我们必须明确诸 X 和 u_i 是怎样分布的。如果我们愿意做假定6（即诸 X 虽是随机的，但它们的分布独立于 u_i，或至少与 u_i 不相关），则为了一切实际目的，我们仍可把诸 X 当作非随机的来看待。如克曼塔（Kmenta）所说：

> 这样，放弃 X 的非随机性假定，即代之以 X 随机但独立于 $[u]$ 的假定，不致改变最小二乘估计的优良性质与可行性。[2]

因此，直至我们在第4篇讨论联立方程之前，我们将保留假定2或假定6。[3]

假定3：u_i 的零均值。 回顾 k 变量线性回归模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{1}$$

现假定：

$$E(u_i \mid X_{2i}, X_{3i}, \cdots, X_{ki}) = w \tag{2}$$

其中 w 是一常数，注意在标准模型中 $w=0$，而现在 w 是任意常数。

338 取（1）的条件期望值得：

$$\begin{aligned} E(Y_i \mid X_{2i}, X_{3i}, \cdots, X_{ki}) &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + w \\ &= (\beta_1 + w) + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \\ &= \alpha + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \end{aligned} \tag{3}$$

其中 $\alpha = (\beta_1 + w)$，并且在取期望值时视诸 X 为常数。（为什么？）

因此，我们看到，如果假定3不满足，我们将无法估计原来的截距 β_1，我们所能得到的是含有 β_1 和 $E(u_i) = w$ 的 α。简言之，我们得到了 β_1 的一

个有偏误的估计。

但是，如同我们曾多次指出的，在许多实际场合中，截距项并不重要；更有意义的量是那些斜率系数，即使假定 3 不成立，它们也不受影响。[14]另外，在许多应用中，截距项并无实质性意义。

假定 11：u 的正态性。如果我们的目的仅在于估计，则此假定并不是非有不可。如在第 3 章所说的，不论 u_i 是不是正态分布的，OLS 估计量都是 BLUE。然而，有了正态性假定，我们就能肯定回归系数的 OLS 估计量遵循正态分布，$(n-k)\hat{\sigma}^2/\sigma^2$ 遵循 χ^2 分布，以及我们能用 t 和 F 统计量去检验各种统计假设而不论样本大小如何。

但若 u_i 不是正态分布的，又会出现什么情况？这时我们要依靠如下的中心极限定理的延伸；记得我们曾用中心极限定理说明正态性假定的合理性：

> 如果干扰项 u_i 是独立且同一分布的，并有零均值和不变方差，而解释变量在重复抽样中保持不变，则 OLS 系数估计量是渐近正态分布的，且其均值等于相应的诸 β。[15]

因此，通常的检验程序——t 和 F 检验——仍然渐近有效，也就是说，它在大样本中而不是在有限或小样本中有效。

如果干扰不是正态分布的，则 OLS 估计量（在同方差和固定 X 的假定下）仍然渐近于正态分布。——这一事实对常常得不到昂贵的大样本数据的
339 实践经济学家来说，并不是什么福音。因此，为了假设检验和预测的目的，正态性假定就变得极为重要。考虑到估计与假设检验的孪生问题，并鉴于在大多数经济分析中小样本情形是常规而不是例外，我们将继续利用正态性假定。[16]

当然，这意味着，当我们在用小样本时，我们必须明确地检验正态性假定。我们已经考虑过 **χ^2 平方拟合优度和雅克-贝拉正态性检验**。我们强烈敦促读者把这些或别的正态性检验应用到回归残差中。要记住，在有限样本中，没有正态性假定，通常的 t 和 F 变量未必遵循 t 和 F 分布。

我们还剩下假定 1、4、5、7、8、9 和 10。假定 7、8 和 10 是紧密相关的，将在多重共线性一章（第 10 章）中讨论。假定 4 将在异方差性一章（第 11 章）中讨论。假定 5 则在自相关一章（第 12 章）中讨论。最后，假定 9 放在模型设定和诊断检验一章（第 13 章）中讨论。由于其特殊的性质和数学上的需要，假定 1 作为第 3 篇的一个专题（第 14 章）进行讨论。

由于教学方法上的原因，在以下的每一章中，我们都遵循一个共同的格式，即（1）明确问题的性质；（2）分析它的影响；（3）提出侦察它的方法；（4）考虑补救措施，从而有可能导出具有第 1 篇所讨论的那些良好统计性质的估计量。

提一点告诫是合适的。如前面所指出的，对于因违反 CLRM 的假定而

引起的一切问题，并不总是存在有令人满意的解答。此外，对一具体问题的答案可能不止一个，而且常常弄不清楚哪个答案最好。再者，在一项具体应用中，可能出现违反多于 1 个 CLRM 假定的情形，例如，设定偏误，多重共线性以及异方差性可能同时存在。没有任何单一的、万能检验能同时解决所有的问题。[7]再其次，曾在一个时期广为使用的某一检验方法，由于后来有人发现了它原先的弊病，就不一定在以后仍继续风行。但这正说明科学在进步。计量经济学也不例外。

【注释】

[1] G. Barrie Wetherill, *Regression Analysis with Applications*, Chapman and Hall, New York, 1986, pp.14 – 15.

[2] Jan Kmenta, *Elements of Econometrics*, 2d ed., Macmillan, New York, 1986, p.338.（原文中加重语句。）

[3] 这里拟注释一个技术性问题，取代诸 X 与 u 相独立这个强假定，可用较弱的 X 变量与 u 无瞬时（同一点时间）相关的假定。但这时 OLS 估计量可能有偏误，然而它们是一致性的，即随着样本含量无限增大，估计量收敛于其真值。但若诸 X 与 u 瞬时相关，则 OLS 估计量既有偏误，又有非一致性的。在第 17 章中，我们将说明，有时可用**工具变量**法去获得这种情形的一致性估计量。

[4] 此叙述仅当对每一 i 都有$E(u_i)=w$ 时才是正确的，注意到这点是非常重要的。如果$E(u_i)=w_i$，即对每一 i 有一不同的常数，则偏斜率系数可能有偏误且非一致性的。这时，假定 3 是否成立将是关键性的。证明和更多的细节，见 Peter Schmidt, *Econometrics*, Marcel Dekker, New York, 1976, pp.36 – 39。

[5] Henri Theil, *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1978, p.240. 须知，对这一结论，假定固定的 X 和不变的 σ^2 是至关重要的。

[6] 顺便指出，文献中常在**稳健性估计**（robust estimation）这个标题下讨论偏离正态性的后果及相关论题。稳健性估计是一个超出本书论述范围的论题。

[7] 这并非缺少尝试。参见 A.K.Bera! and C.M.Jarque, “Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence,” *Economic Letters*, vol. 7, 1981, pp.313 – 318。

第 10 章　多重共线性：回归元相关会怎么样？

341 　联体字“多重—共线性”问题在计量经济学教科书中和在应用文献中的误用情况之多，再没有其他联体字能比得上的了。我们的许多解释变量本是高度共线性的，而这就是生活。毫无疑问，世上存在有实验的设计 $\boldsymbol{X}'\boldsymbol{X}$（即数据矩阵），会远远优于自然实验已为我们提供的设计（即手中的样本）。但怨天尤人完全无济于事。对一个坏的设计采取就事论事的治疗方法，诸如逐步回归（stepwise regression）或脊回归（ridge regression），可能招致灾难性的后果。正确的做法是，宁可接受事实：我们的非实验数据［即不是从经过设计的实验中得到的数据］有时不能对我们感兴趣的参数提供多少信息。[1]

经典线性回归模型的假定 10 说，包含在回归模型中的诸回归元之间无多重共线性。本章中，为了寻找下述问题的答案，我们对多重共线性假定做一严格的分析：

1. 多重共线性的性质是什么？
2. 多重共线性真的成为一个问题吗？
3. 它会引起一些什么实际后果？
4. 怎样去发现它？
5. 为了缓解多重共线性的问题，能采取哪些补救措施？

我们还将表明，假定 7 和 8 是怎样同无多重共线性假定相配合的。

342 我们在本章中还讨论 CLRM 的假定 7 和假定 8，假定 7 要求样本次数必须大于回归元的个数，假定 8 要求回归元的取值必须有足够的变异，它们都是对存在多重共线性这一假定的补充。戈德伯格已经将假定 7 命名为**微数缺测性**（micronumerosity）[2]问题，也就是小样本容量的问题。

§ 10.1 多重共线性的性质

多重共线性（multicollinearity）一词由弗里希引入。[3]它原先的含义是指一个回归模型中的一些或全部解释变量之间存在有一种“完全”或准确的线性关系。[4]对涉及解释变量 X_1，X_2，…，X_k（其中，为了把截距项考虑进来，在一切观测中取 $X_1=1$）的 k 变量回归来说，存在有一准确线性关系，如果以下条件得到满足：

$$\lambda_1 X_1+\lambda_2 X_2+\cdots+\lambda_k X_k=0 \tag{10.1.1}$$

其中 λ_1，λ_2，$\cdots\lambda_k$ 为常数，但不同时为零。[5]

然而，现在用多重共线性一词，有更广的含义，既包括（10.1.1）所表明的完全多重共线性情形，还包括诸 X 变量之间有交互相关，但又非完全相关的如下情形[6]：

$$\lambda_1 X_1+\lambda_2 X_2+\cdots+\lambda_k X_k+v_i=0 \tag{10.1.2}$$

其中 v_i 是随机误差项。

为了区分完全（perfect）和欠完全（less than perfect）多重共线性，且假定$\lambda_2\neq 0$。于是，（10.1.1）可写为：

$$X_{2i}=-\frac{\lambda_1}{\lambda_2}X_{1i}-\frac{\lambda_3}{\lambda_2}X_{3i}-\cdots-\frac{\lambda_k}{\lambda_2}X_{ki} \tag{10.1.3}$$

343 这表明 X_2 与其他变量有准确的线性关系，或者它能从其他 X 变量的线性组合推出。这时，变量 X_2 与方程（10.1.3）右端的线性组合之间的相关系数必定是 1。

类似地，如果 $\lambda_2\neq 0$，则方程（10.1.2）可写为：

$$X_{2i}=-\frac{\lambda_1}{\lambda_2}X_{1i}-\frac{\lambda_3}{\lambda_2}X_{3i}-\cdots-\frac{\lambda_k}{\lambda_2}X_{ki}-\frac{1}{\lambda_2}v_i \tag{10.1.4}$$

这表明 X_2 不是其他 X 的一个准确的线性组合，因为它还决定于随机误差项 v_i。

作为一个数值例子，考虑如下的人为数据：

X_2	X_3	X_3^*
10	50	52
15	75	75
18	90	97
24	120	129
30	150	152

很明显，$X_{3i}=5X_{2i}$。因此 X_2 与 X_3 之间有完全的共线性并且相关系数是 1。变量 X_3^* 不外是把从随机数表的读数 2，0，7，9，2 加到 X_3 上而产生的。但现在 X_2 与 X_3^* 之间不再有完全的共线性。然而，它们之间的相关系数是 0.995 9，所以两者是高度相关的。

以上对多重共线性的代数处理方法，可通过巴伦坦（回顾图 3.9）做简明的描述。在该图中圆圈 Y，X_2 和 X_3 分别代表 Y（因变量）和 X_2 及 X_3（解释变量）的变异。共线性的程度可用 X_2 和 X_3 两圆圈的重叠程度来衡量。在图 10.1a 中 X_2 与 X_3 无重叠，因而无共线性。从图 10.1b 到图 10.1e 共线性程度由低渐高。——X_2 与 X_3 的重叠部分越大（即阴影面积越大），共线性程度也越高。达到极端时，X_2 与 X_3 完全重合（或者 X_2 完全落在 X_3 内，或者反过来说），共线性就是完全的。

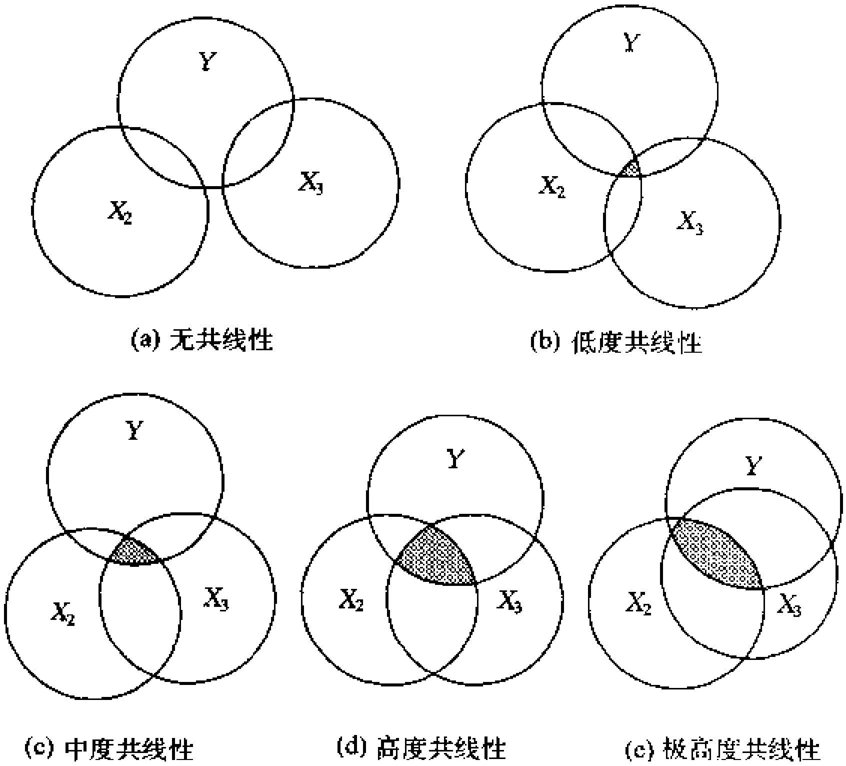


图 10.1　多重共线性的巴伦坦视图

顺便指出，我们定义的多重共线性仅对 X 变量之间的线性关系而言。此外，还可能有它们之间的非线性关系。例如，考虑以下回归模型：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad (10.1.5)$$

其中，比如说，Y = 生产总成本，而 X = 产出。变量 X_i^2（产出的平方）和
344 X_i^3（产出的立方）显然与 X_i 有函数关系，但这种关系是非线性的。因此，严格地说，像（10.1.5）这样的模型，并不违反无多重共线性假定。然而，在具体的应用中，通常测算的相关系数将表明 X_i、X_i^2 和 X_i^3 是高度相关的。如我们即将表明的，这种情形将使（10.1.5）的参数难以较准确地（即以较小标准误）估计。

为什么经典线性回归模型要假定诸 X 之间无多重共线性呢？可以这样去理解：**如果多重共线性是完全的，有如（10.1.1）那样，诸 X 变量的回归系数将是不确定的，并且它们的标准误为无穷大。如果多重共线性是欠完全的，像（10.1.2）那样，那么，虽然回归系数可以确定，却有较大的标准误（相对于系数本身来说），意思是，系数不能以很高的精确或准确度加以估计。**随后的几节将对此做出证明。

345 多重共线性有多种来源。按蒙哥马利（Montgomery）和佩克（Peck）的提法，多重共线性可能由以下因素导致。[7]

1. 数据采集所用的方法。例如，抽样限于总体中诸回归元所取值的一个有限的范围内。

2. 模型或从中取样的总体受到约束。例如，在做电力消费对收入（X_2）和住房面积（X_3）的回归时，总体中有这样的一种有形的约束，即一般地说收入较高的家庭比收入较低的家庭有较大的住房。

3. 模型设定。例如在回归中添加多项式项，尤其当 X 变量的变化范围（极差）较小时。

4. 一个过度决定的模型（overdetermined）。这种情况出现在模型的回归元个数大于观测次数时。例如，在医药研究中，可能只有少数病人，但却要在他们身上收集大量变元的信息。

多重共线性的另外一个原因（特别是在时间序列数据中）可能是模型中所包含的回归元具有**相同的时间趋势**，即它们同时随着时间而增减。于是，在消费支出对收入、财富和人口的回归中，回归元收入、财富和人口可能都以多少有些一致的速度递增，从而导致这些变量之间的共线性。

§10.2 出现完全多重共线性时的估计问题

前面说过，在完全多重共线性的情形中，回归系数是不确定的，并且其标准误是无穷大的。这一事实容易通过三变量回归模型加以说明。利用离差形式把三个变量都表示为偏离它们各自的样本均值的离差，我们就能把这三

变量回归模型写为：

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{10.2.1}$$

现在从第 7 章我们得到：

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \tag{7.4.7}$$

346 $$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \tag{7.4.8}$$

假定 $X_{3i} = \lambda X_{2i}$，这里 λ 是非零常数（比如 2、4、1.8，等等）。以此代入（7.4.7）可得：

$$\begin{aligned}\hat{\beta}_2 &= \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2) - (\lambda \sum y_i x_{2i})(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2) - \lambda^2 (\sum x_{2i}^2)^2} \\ &= \frac{0}{0} \end{aligned} \tag{10.2.2}$$

这是一个不定式。读者容易验证 $\hat{\beta}_3$ 也是不确定的。[8]

我们为什么会得到像（10.2.2）那样的结果呢？回想一下 $\hat{\beta}_2$ 的意义：它是在保持 X_3 不变的情况下，当 X_2 每改变一单位 Y 的平均值的变化率。但如果 X_3 和 X_2 是完全共线性的，就没有任何方法能保持 X_3 不变：随着 X_2 改变，X_3 也按一个倍数因子 λ 改变。这就意味着没有任何方法能从所给的样本中把 X_2 和 X_3 的各自影响分解开来。从实际方面考虑，X_2 和 X_3 是不可区分的。在应用计量经济学中，我们的宗旨就是要把每个 X 对应变量的偏影响分离开来，所以这个问题是最具破坏性的。

从另一角度看这个问题，把 $X_{3i} = \lambda X_{2i}$ 代入（10.2.1）得到以下方程[比较（7.1.10）]：

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3(\lambda x_{2i}) + \hat{u}_i \\ &= (\hat{\beta}_2 + \lambda\hat{\beta}_3) x_{2i} + \hat{u}_i \\ &= \hat{\alpha} x_{2i} + \hat{u}_i \end{aligned} \tag{10.2.3}$$

其中：

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) \tag{10.2.4}$$

对（10.2.3）应用平常的 OLS 公式可得：

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2} \tag{10.2.5}$$

因此，虽然 $\hat{\alpha}$ 可以惟一地估计出来，却无法惟一地估计 β_2 和 β_3；数学上，

$$\hat{\alpha} = \hat{\beta}_2 + \lambda\hat{\beta}_3 \tag{10.2.6}$$

347 是一个方程有两个未知数（注意 λ 是给定的），对给定的 $\hat{\alpha}$ 和 λ 值，

(10.2.6) 便有无穷多个解。为了把这个概念说得更具体，令 $\hat{\alpha}=0.8$ 和 $\lambda=2$。这样就得到：

$$0.8=\hat{\beta}_2+2\hat{\beta}_3 \tag{10.2.7}$$

或：

$$\hat{\beta}_2=0.8-2\hat{\beta}_3 \tag{10.2.8}$$

现在任意选一个 $\hat{\beta}_3$ 的值，将得到 $\hat{\beta}_2$ 的一个解。选另一个 $\hat{\beta}_3$ 的值又得到 $\hat{\beta}_2$ 的另一个解。不管你怎样尝试，都没有 $\hat{\beta}_2$ 的惟一值。

以上讨论的要点在于：对于完全多重共线性情形，我们无法得到个别回归系数的惟一解。但应注意到，我们能够得到这些系数的线性组合的惟一解。给定 λ 值，线性组合 $(\beta_2+\lambda\beta_3)$ 的惟一估计值是 α。[9]

顺便看到，对于完全多重共线性的情形，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的个别方差和标准误都是无穷大（见习题 10.21）。

§10.3 出现“高度”但“不完全”多重共线性时的估计问题

完全多重共线性情形只不过是一种极端的隐忧。通常，尤其是在涉及经济时间序列的数据中，X 变量之间并无准确的线性关系。拿 (10.2.1) 所给的离差形式的三变量模型来看，我们有的不是准确的多重共线性，而是：

$$x_{3i}=\lambda x_{2i}+v_i \tag{10.3.1}$$

其中 $\lambda\neq0$ 并且 v_i 是具有性质 $\sum x_{2i}v_i=0$ 的随机误差项。（为什么？）

顺便提一下，图 10.1b 至图 10.1e 的巴伦坦视图都代表不完全共线性的情形。

对于这种情形，回归系数 β_2 和 β_3 的估计是可能的。例如将 (10.3.1) 代入 (7.4.7) 得：

$$\hat{\beta}_2=\frac{\sum(y_ix_{2i})(\lambda^2\sum x_{2i}^2+\sum v_i^2)-(\lambda\sum y_ix_{2i}+\sum y_iv_i)(\lambda\sum x_{2i}^2)}{\sum x_{2i}^2(\lambda^2\sum x_{2i}^2+\sum v_i^2)-(\lambda\sum x_{2i}^2)^2} \tag{10.3.2}$$

其中利用了关系式 $\sum x_{2i}v_i=0$。对 $\hat{\beta}_3$ 也可推出类似的表达式。

348 现在，不同于 (10.2.2)，没有理由先验地认为 (10.3.2) 不可估计。当然，如果 v_i 充分地小，以至非常接近于零，则 (10.3.1) 表示几乎完全共线性。这时我们又将回到 (10.2.2) 的不定式情形。

§10.4 多重共线性：是庸人自扰吗？多重共线性的理论后果

回想一下，如果经典模型的假定得到满足，则回归系数的 OLS 估计量是 BLUE（或 BUE，如果加上正态性假定）。而现在可以证明，即使多重共线性是非常高的，如近似多重共线性，那么，OLS 估计量仍保持 BLUE 性质。[10]那么，大谈特谈多重共线性究竟为了什么？如 C. 阿肯（Christopher Achen）所说［并参考本章开头所引用利莫尔（Leamer）的话］：

> 初次接触方法论的学生们有时担心他们的自变量有相关性，即所谓多重共线性问题。但多重共线性并不违反回归假定。无偏的、一致性的估计值仍将出现，并且对它们的标准误仍将有正确的估计。多重共线性的惟一影响，是难于得到标准误小的系数估计值。然而，仅有少量的观测次数时也会出现这种影响，好比自变量的小方差性所造成的影响那样。（事实上，从理论的高度看，多重共线性，过少的观测次数以及过小的自变量方差，实质上是同一问题。）因此，"遇到多重共线性我该怎样办？"这个问题无异于"如果我没有很多的观测值该怎么办？"统计答案是没有的。[11]

为了彻底弄明白样本大小的重要性，戈德伯格构造**微数缺测性**一词以应对古怪的多音节名称：多重共线性。按照戈德伯格所说的，准确的**微数缺测性**（与准确多重共线性相对照），是指样本大小 n 等于零的情形。这时，任何种类的估计都是不可能的。近似微数缺测性，则好比近似完全多重共线性，指观测次数刚刚超过待估参数个数的情形。

利莫尔、阿肯和戈德伯格埋怨人们过少地注意样本大小的问题，而过多地注意多重共线性的问题。没有错，可惜的是，在应用研究工作中用到第二手资料（指由他人收集的数据资料，诸如由政府收集的 GNP 数据）的个人
349 研究者，对样本数据的多少似乎无能为力，而只好"把多重共线性视为对 CLR［经典线性回归］的破坏，以正视估计的问题"[12]。

第一，诚然，即使是近似多重共线性的情形，OLS 估计量仍然是无偏的。但无偏性是一种多维样本（multisample）或重复抽样的性质。意思是说，如果我们在 X 变量取固定值的情况下反复抽取样本，并对每一样本计算 OLS 估计量，那么，随着样本个数的增加，估计量的样本值的均值将收敛于它们的真实总体值。

第二，说共线性并不破坏最小方差性质，也没错。在所有线性无偏估计量一类中，OLS 估计量有最小方差。就是说，它们是有效的。但这并不意味着在任一给定的样本中，一个 OLS 估计量的方差一定是小的（相对于估计量的值而言），这点我们即将予以说明。

第三，**多重共线性本质上是一种样本（回归）现象**。意思是说，即使在总体中诸 X 变量没有线性关系，但在具体获得的样本中仍可能有线性关系：当我们设想一个理论或总体回归函数时，我们相信，所有包含在模型中的 X 变量对 Y 都有各自的独立影响。但有可能在任给的一个用以检验 PRF 的样本中，一些或全部 X 变量之间的共线性却是如此之高，以致我们无法分开它们对 Y 的各自影响。看来，是样本把我们难住了。尽管理论告诉我们，所有的 X 变量都重要，但我们的样本还不"富裕"到足以在分析中容纳全部 X 变量的境况。

作为一个说明，再考虑第 3 章的消费—收入例子。经济学家从理论上推知除收入外，消费者的财富也是消费支出的重要决定因素。于是我们可以写为：

$$\text{消费}_i = \beta_1 + \beta_2\,\text{收入}_i + \beta_3\,\text{财富}_i + u_i$$

但可能出现这样的情形：当我们获得收入和财富的数据时，这两个变量可能高度（如果不是完全）相关：较富有的人们一般倾向于有较高的收入。因此，虽然理论上收入和财富都是解释消费支出行为的合理备选变量，但实际上（即在样本中）要分开收入和财富对消费支出的影响，也许是困难的。

理想地，为了评论财富和收入对消费支出的个别影响，我们需要对财富
350 多而收入低以及财富少而收入高的人们进行足够多的样本观测（回顾假定 8）。虽然在横截面研究中（通过加大样本的方法）有可能做到这点，但在加总时间序列操作中却非常难以实现。

考虑到所有这些理由，多重共线性虽不影响 OLS 估计量的 BLUE 性质，但这点在实际上并没有什么值得宽慰的。我们必须意识到，在任给的一个样本中会出现什么情况。这就是下节要讨论的议题。

§10.5　多重共线性的实际后果

近似或高度多重共线性可能招致以下后果：

1. 虽然 OLS 估计量是 BLUE，但有大的方差和协方差，故难以做出精确的估计。

2. 由于后果 1，置信区间将要宽得多，以致接受"零虚拟假设"（即真实总体系数为零的假设）更为容易。

3. 仍由于后果 1，一或多个系数的 t 比率倾向于统计上不显著。

4. 虽然一或多个系数的 t 比率在统计意义上不显著，总的拟合优度 R^2 仍可能非常之高。

5. OLS 估计量及其标准误对数据的微小变化也会是敏感的。

上述后果可解释如下。

OLS 估计量的大方差与协方差

大方差和协方差可从模型（10.2.1）所给的 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差和协方差公式看到：

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)} \tag{7.4.12}$$

$$\operatorname{var}(\hat{\beta}_3)=\frac{\sigma^2}{\sum x_{3i}^2(1-r_{23}^2)} \tag{7.4.15}$$

$$\operatorname{cov}(\hat{\beta}_2,\hat{\beta}_3)=\frac{-r_{23}\sigma^2}{(1-r_{23}^2)\sqrt{\sum x_{2i}^2\sum x_{3i}^2}} \tag{7.4.17}$$

其中 r_{23}是 X_2 与 X_3 之间的相关系数。

从（7.4.12）和（7.4.15）显见，随着 r_{23}趋于 1，即随着共线性增加，两估计量的方差也增加。在达到极限 $r_{23}=1$ 时，方差变为无穷大。同样，由（7.4.17）显见，随着 r_{23}朝着 1 增大，两估计量的协方差在绝对值上也增大。[注：cov（$\hat{\beta}_2$，$\hat{\beta}_3$）≡cov（$\hat{\beta}_3$，$\hat{\beta}_2$）。]

351 方差和协方差增大的速度可由**方差膨胀因子**（variance-inflating factor，VIF）看出。后者定义为：

$$\text{VIF}=\frac{1}{(1-r_{23}^2)} \tag{10.5.1}$$

VIF 表明，估计量的方差由于多重共线性的出现而膨胀起来。随着 r_{23}^2趋于 1，VIF 趋于无穷大，即随着共线性程度的增加，估计量的方差也增加，并且在达到极限时，它可以变到无穷大。还容易看到，如 X_2 与 X_3 无共线性，VIF 将是 1。

利用这一定义，可将（7.4.12）和（7.4.15）表达为：

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_{2i}^2}\text{VIF} \tag{10.5.2}$$

$$\operatorname{var}(\hat{\beta}_3)=\frac{\sigma^2}{\sum x_{3i}^2}\text{VIF} \tag{10.5.3}$$

从而表明 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差同 VIF 成正比关系。

为了对方差和协方差随 r_{23}增加而增加的速度有所感知，表 10.1 对选定的 r_{23}值算出方差和协方差。如表所示，r_{23}的增加，对 OLS 估计量的方差和协方差估计值有剧烈的影响。当 $r_{23}=0.50$ 时，var（$\hat{\beta}_2$）1.33 倍于 r_{23}
351 为零时的方差，但到了 $r_{23}=0.95$ 时，它将近 10 倍于没有共线性的情形。再看，当 r_{23}从 0.95 增到 0.995 时，估计的方差值跃升到无共线性情形的 100 倍。类似的戏剧般的效应也见于所估计的协方差。图 10.2 展示了所有

这些。

表 10.1　　r_{23}增大对 $\mathrm{var}(\hat{\beta}_2)$和 $\mathrm{cov}(\hat{\beta}_2,\hat{\beta}_3)$的影响

r_{23}值 (1)	VIF (2)	$\mathrm{var}(\hat{\beta}_2)$ (3)*	$\frac{\mathrm{var}(\hat{\beta}_2)(r_{23}\neq 0)}{\mathrm{var}(\hat{\beta}_2)(r_{23}=0)}$ (4)	$\mathrm{cov}(\hat{\beta}_2,\hat{\beta}_3)$ (5)
0.00	1.00	$\frac{\sigma^2}{\sum x_{2i}^2}=A$	–	0
0.50	1.33	1.33 × A	1.33	0.67 × B
0.70	1.96	1.96 × A	1.96	1.37 × B
0.80	2.78	2.78 × A	2.78	2.22 × B
0.90	5.26	5.26 × A	5.26	4.73 × B
0.95	10.26	10.26 × A	10.26	9.74 × B
0.97	16.92	16.92 × A	16.92	16.41 × B
0.99	50.25	50.25 × A	50.25	49.75 × B
0.995	100.00	100.00 × A	100.00	99.50 × B
0.999	500.00	500.00 × A	500.00	499.50 × B

注：$A=\frac{\sigma^2}{\sum x_{2i}^2}$

$B=\frac{-\sigma^2}{\sqrt{\sum x_{2i}^2\sum x_{3i}^2}}$

× = 倍（乘号）

* 要看出 r_{23}增加对$\mathrm{var}(\hat{\beta}_3)$的影响，只须注意：当 $r_{23}=0$ 时，$A=\sigma^2/\sum x_{3i}^2$。方差和协方差的膨胀因子是一样的。

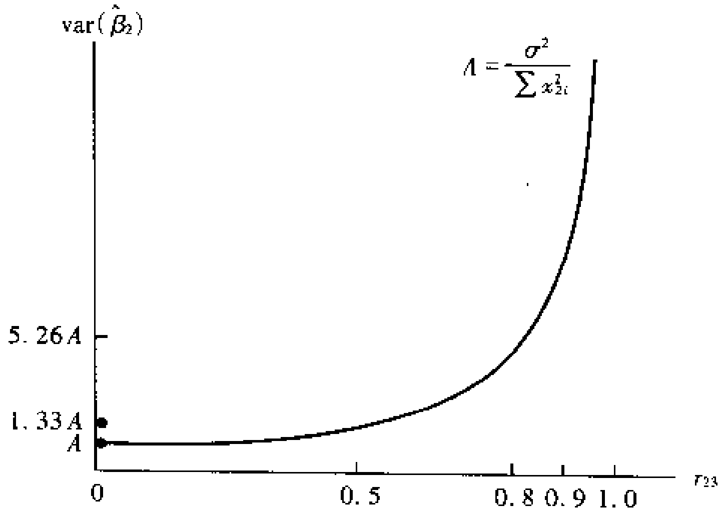


图 10.2　$\mathrm{var}(\hat{\beta}_2)$作为 r_{23}的一个函数的性态

刚才讨论的结果可容易地推广到 k 变量模型。在这样的一个模型中，第 k 个系数的方差可如（7.5.6）表示为

$$\operatorname{var}(\hat{\beta}_j)=\frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \tag{7.5.6}$$

其中 $\hat{\beta}_j$ = 回归元 X_j 的（估计）偏回归系数

$R_j^2 = X_j$ 对其余（$k-2$）个回归元进行回归的 R^2［注：在 k 变量回归模型中有（$k-1$）个回归元。］

$$\sum x_j^2=\sum (X_j-\bar{X}_j)^2$$

我们还可以把(7.5.6)写成

$$\operatorname{var}(\hat{\beta}_j)=\frac{\sigma^2}{\sum x_j^2}\mathrm{VIF}_j \tag{10.5.4}$$

如你从这个表达式所见，$\operatorname{var}(\hat{\beta}_j)$ 与 σ^2 和 VIF 成正比，但与 $\sum x_j^2$ 成反比。
353 因此，$\operatorname{var}(\hat{\beta}_j)$的大小取决于三个部分：(1) σ^2，(2) VIF 和 (3) $\sum x_j^2$。最后一个与经典模型的假定 8 相联系，它说明回归元的变异越大，在假定其他两个因素不变的情况下，该回归元系数的方差就越小，因此用它估计系数就越准确。

在进一步深入下去之前，注意 VIF 的倒数被称为**容许度**（TOL）。即

$$\mathrm{TOL}_j=\frac{1}{\mathrm{VIF}_j}=(1-R_j^2) \tag{10.5.5}$$

当 $R_j^2=1$（即完全共线性）时，$\mathrm{TOL}_j=0$，当 $R_j^2=0$（即不存在共线性）时，$\mathrm{TOL}_j=1$。由于 VIF 和 TOL 之间有密切关系，所以可以将它们互换使用。

更宽的置信区间

由于大的标准误，有关总体参数的置信区间随之变大。这可由表 10.2 看出，例如，当 $r_{23}=0.95$ 时，β_2 的置信区间要比 $r_{23}=0$ 时的大 $\sqrt{10.26}$ 或约 3 倍。

表 10.2　增加共线性对 β_2 的 95%置信区间"$\hat{\beta}_2 \pm 1.96\ \mathrm{se}(\hat{\beta}_2)$"的影响

r_{23} 值	β_2 的 95% 置信区间
0.00	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.50	$\hat{\beta}_2 \pm 1.96\ \sqrt{(1.33)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

0.95	$\hat{\beta}_2 + 1.96\ \sqrt{(10.26)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.99	$\hat{\beta}_2 + 1.96\ \sqrt{(100)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.999	$\hat{\beta}_2 \pm 1.96\ \sqrt{(500)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

注：为方便起见，假定 σ^2 已知，因此可用正态分布，从而说明用 1.96 作为正态分布下的 95% 置信因子。

与各 r_{23} 值相对应的标准误取自表 10.1。

因此，在高度多重共线性的情形中，样本数据可能与分歧很大的一些假设均无矛盾，这样就增加了接受错误假设（即第Ⅱ类错误）的概率。

"不显著"的 t 比率

354 记得在检验虚拟假设 $\beta_2=0$（比方说）中，我们使用 t 比率 $\hat{\beta}_2/\mathrm{se}(\hat{\beta}_2)$，并将估计的 t 值同从 t 表查出的临界 t 值相比。但我们已经看到，在高度共线性情形中，估计的标准误增加奇快，从而使 t 值迅速变小。因此，在这种情形中，我们会越来越多地接受有关真实总体值为零的虚拟假设。[13]

R^2 值高而显著的 t 比率少

考虑 k 变量线性回归模型：

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i$$

如同我们刚才说过的，在高度共线性情形中，有可能会发现一个或多个偏斜率系数，在 t 检验的基础上，个别地在统计意义上是不显著的，然而这时 $\bar{R}^2$ 却高达（比如说）0.9 以上，从而根据 F 检验，可令人信服地拒绝 $\beta_2=\beta_3=\cdots=\beta_k=0$ 的假设。其实，这就是重复共线性的一个信号——不显著的 t 值却带有一个高的总 R^2 值（并因而有一个显著的 F 值）！

下节里我们将展示这个信号。不过鉴于我们在第 8 章中关于个别检验者联合检验的讨论，这种信号的出现并没有什么奇怪。你也许会想到，这里的真正问题在于估计量之间的协方差。如公式（7.4.17）所表明的，这些协方差是与回归元之间的相关有关的。

OLS 估计量及其标准误对数据微小变化的敏感性

只要多重共线性还不是完全的，回归系数的估计就是可能的。然而，估计值及其标准误对数据中的哪怕是小小的变化，也会是非常敏感的。

为了看清楚这一点，考虑表 10.3。根据这些数据，我们得到如下的多元回归：

$$\begin{aligned}\hat{Y}_i &= 1.1939 + 0.4463X_{2i} + 0.0030X_{3i}\\ &\quad\ (0.7737)\qquad (0.1848)\qquad (0.0851)\\ t &= (1.5431)\qquad (2.4151)\qquad (0.0358)\\ R^2 &= 0.8101\quad r_{23}=0.5523\\ \operatorname{cov}(\hat{\beta}_2,\hat{\beta}_3) &= 0.00868\quad df=2\end{aligned}\tag{10.5.6}$$

表 10.3　Y，X_2 和 X_3 的人为数据

Y	X_2	X_3
1	2	4
2	0	2
3	4	12
4	6	0
5	8	16

355 回归（10.5.4）表明，个别地看，没有一个回归系数在通常的 1% 或 5% 显著水平上是显著的，虽然 $\hat{\beta}_2$ 在单尾 t 检验的 10% 显著性水平上是显著的。

再看表 10.4，它和表 10.3 的差别仅在于：X_3 的第 3 个值和第 4 个值互相对调了。现在用表 10.4 的数据我们得到：

$$\begin{aligned}\hat{Y}_i &= 1.2108 + 0.4014X_{2i} + 0.0270X_{3i}\\ &\quad\ (0.7480)\qquad (0.2721)\qquad (0.1252)\\ t &= (1.6187)\qquad (1.4752)\qquad (0.2158)\\ R^2 &= 0.8143\quad r_{23}=0.8285\\ \operatorname{cov}(\hat{\beta}_2,\hat{\beta}_3) &= -0.0282\quad df=2\end{aligned}\tag{10.5.7}$$

表 10.4　　Y，X_2 和 X_3 的人为数据

Y	X_2	X_3
1	2	4
2	0	2
3	4	0
4	6	12
5	8	16

数据的微小变化的结果是，原先在 10% 显著性水平上为统计显著的 $\hat{\beta}_2$，现在不再在该水平上显著了。还可注意到（10.5.4）中的 $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.008\,68$，而在（10.5.5）中它是 $-0.028\,2$，增加了 3 倍以上。所有这些都可归因于增大了的多重共线性：在（10.5.4）中，$r_{23} = 0.552\,3$，而在（10.5.5）中，它是 0.828 5。同理，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的标准误都在增大。这正是共线性的通常象征。

我们前面说过，在出现高度共线性时，我们无法精确估计个别的回归系数，但可以较精确地估计这些系数的某些线性组合。这一事实，可从回归（10.5.4）和（10.5.5）得以证实。在第一个回归中，两个偏斜率系数之和为 0.449 3，而在第二个回归中，此和为 0.428 4，基本一致。不仅如此，它们的标准差也实际上相差不大，分别是 0.155 0 和 0.182 3。[14]然而，要看到，X_3 的系数已戏剧般地从 0.003 变到 0.027。

微数缺测性的后果

356 仿照多重共线性的俚语，以鹦鹉学舌般的方式，戈德伯格根据他对过小样本分析，引出完全类似的微数缺测性的后果。[15]建议读者参阅戈德伯格本人做的分析，看看他为什么把微数缺测性看成同多重共线性一样重要（或不重要）的概念。

§10.6　一个说明性例子：消费支出与收入和财富的关系

为了说明至今所提的种种问题，让我们再来考虑第 3 章中的消费—收入一例。在表 10.5 中，我们复制了表 3.2 中的数据，并在上面加进了消费者

的财富数据。如果我们假定消费与收入和财富有线性关系，则根据表 10.5 我们得到如下回归：

$$
\begin{aligned}
\hat{Y}_i &= 24.7747 + 0.9415X_{2i} - 0.0424X_{3i} \\
&\quad\ (6.7525) \qquad (0.8229) \qquad (0.0807) \\
t &= \ (3.6690) \qquad (1.1442) \qquad (-0.5261) \qquad\qquad (10.6.1) \\
R^2 &= 0.9635 \quad R^2 = 0.9531 \quad \mathrm{df} = 7
\end{aligned}
$$

表 10.5　关于消费支出 Y，收入 X_2 和财富 X_3 的假想数据

Y，美元	X_2，美元	X_3，美元
70	80	810
65	100	1 009
90	120	1 273
95	140	1 425
110	160	1 633
115	180	1 876
120	200	2 052
140	220	2 201
155	240	2 435
150	260	2 686

回归（10.6.1）表明收入和财富一起解释了消费支出变异中的约 96%。然而没有一个斜率系数是个别地在统计意义上显著的。不但如此，财富变量
357 不但统计上不显著，而且带有错误的符号。人们会先验地预料到消费与财富之间有正的关系。虽然 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 个别地看都是统计上不显著的，但如果我们同时检验假设 $\beta_2=\beta_3=0$，如表 10.6 所示，我们就可以拒绝此假设。在通常的假定下，我们得到：

$$F=\frac{4\,282.777\,0}{46.349\,4}=92.401\,9 \qquad (10.6.2)$$

显然这个 F 值是高度显著的。

表 10.6　消费—收入—财富一例的 ANOVA 表

变异来源	SS	df	MSS
来自回归	8 565.554 1	2	4 282.777 0
来自残差	324.445 9	7	46.349 4

看看这一结果的几何意义是有趣的（见图 10.3）。根据回归（10.6.1），我们按照第 8 章讨论的通常程序建立 β_2 和 β_3 的个别的 95% 置信区间。这些区间表明，个别地看，每一区间都包含着零值。因此，个别而论，我们可以
358 接受两个偏斜率都是零的假设。但当我们建立联合置信区域以检验假设 $\beta_2=\beta_3=0$时，由于这实际上是一个椭圆形的联合置信（区）域，不含有原点，就不能接受此假设。[16]如同前面已指出的，在出现高度共线性时，对个别回归元的检验是不可靠的。这时要用总的 F 检验来察看 Y 是否与诸回归元有关。

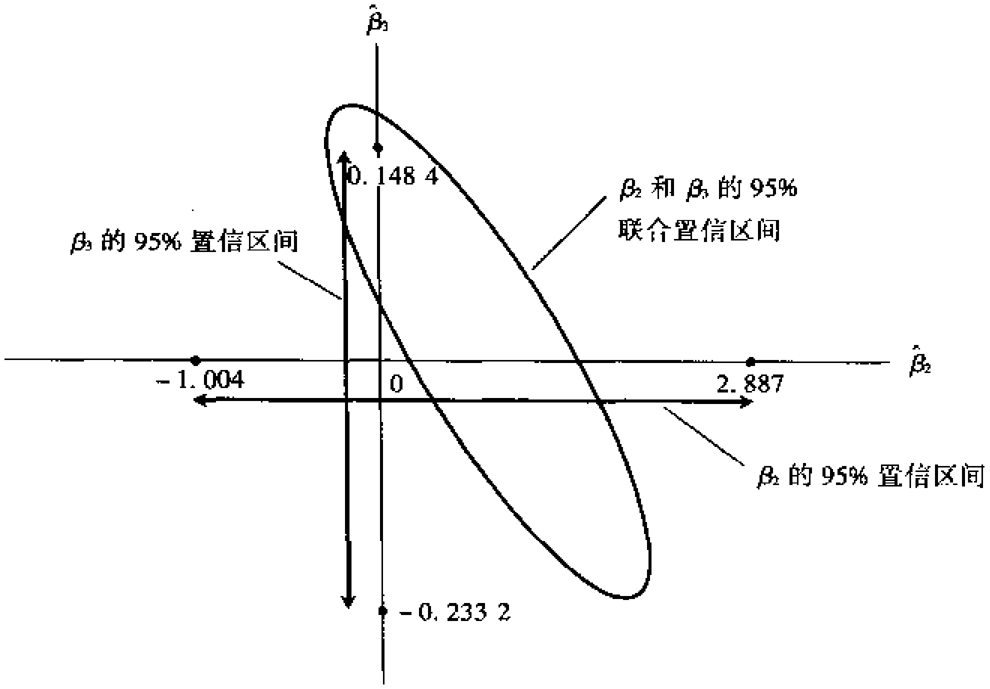


图 10.3　β_2 和 β_3 的个别置信区间与 β_2 和 β_3 的联合置信域（椭圆形）

我们的例子生动地表明多重共线性是怎么一回事。F 检验是显著的，而 X_2 和 X_3 的 t 值个别地看又是不显著的，这一事实本身就说明两变量的相关程度如此之高，以致无法辨别收入或财富对消息的个别的影响。事实上，如果我们做 X_3 对 X_2 的回归便得到：

$$
\begin{aligned}
\hat{X}_{3i} &= 7.545\ 4 \ + \ 10.190\ 9X_{2i} \\
&\quad\ (29.475\ 8) \qquad (0.164\ 3) \\
t &= \ (0.256\ 0)\ (62.040\ 5) \qquad R^2=0.997\ 9
\end{aligned}
\tag{10.6.3}
$$

这表明 X_3 和 X_2 之间有着几乎完全的共线性。

现在让我们做 Y 仅对 X_2 的回归，看看会出现什么情况：

$$
\begin{aligned}
\hat{Y}_i &= 24.454\ 5 \ + \ 0.509\ 1X_{2i} \\
&\quad\ (6.413\ 8) \qquad (0.035\ 7) \\
t &= \ (3.812\ 8)\ (14.243\ 2) \qquad R^2=0.962\ 1
\end{aligned}
\tag{10.6.4}
$$

在（10.6.1）中，收入变量是统计上不显著的，而现在则是高度显著的。如果不做 Y 对 X_2 而做 Y 对 X_3 的回归，则得到：

$$
\begin{aligned}
\hat{Y}_i &= 24.411 + 0.0498X_{3i} \\
&\quad\ (6.874) \qquad (0.0037) \qquad\qquad\qquad (10.6.5) \\
t &= \ (3.551) \quad (13.29) \qquad R^2 = 0.9567
\end{aligned}
$$

我们看到财富现在对消费支出也有显著的影响，而在（10.6.1）中却没有。

回归（10.6.4）和（10.6.5）非常明显地表示，在极端多重共线性的情况下，去掉一个高度共线的变量常常会使另一个 X 变量变得统计上显著。这个结果提示我们，解决极端共线性的一个方法，是扔掉共线的变量。关于这点在 10.8 节中我们还有话要说。

§ 10.7 多重共线性的侦察

359 在研究多重共线性的性质与后果之后，自然要问：在任一给定的情况下，特别是在涉及多于两个解释变量的模型中，我们怎样能知道有没有共线性？这里记住克曼塔的忠告是有益的。

> 1. 多重共线性是一个程度问题而不是有无的问题。有意义的区分不在于有与无之间，而在于它的不同的程度。
>
> 2. 由于多重共线性是对被假定为非随机的解释变量的情况而言，所以它是一种样本而非总体特征。
>
> 因此，我们不做"多重共线性的检验"，但如果我们愿意的话，可以测度它在任一具体样本中显现的程度。[17]

由于多重共线性本质上是一种样本现象，它来源于大多数社会科学中所收集的基本上是非实验性质的数据。我们没有侦破它或度量其强度的惟一方法。我们所有的是一些经验规则；正式的也好，非正式的也好，同样是经验规则。现在我们来考虑这些规则中的一些内容。

1. R^2 值高而显著的 t 比率小。 如同前面已注意到的，这是多重共线性的"经典"征兆。如果 R^2 值高，比方说，超过 0.8，F 检验在大多数情形中都会拒绝所有偏斜率系数同时为零的假设，但个别的 t 检验却表明，没有或很少有偏斜率系数是统计上异于零的。我们的消费—收入—财富一例已清楚地说明这一事实。

虽然这种诊断是可以理解的，但它过于强调"多重共线性的危害，仅仅在于（它使我们）无法把诸解释变量对 Y 的全部影响加以分解"[18]。

2. 回归元之间有高度的两两相关。 另一个可以提出的经验规则是，如果每两个回归元的零阶相关系数高，比方说，超过 0.8，则多重共线性的问

题是严重的。这一准则却带来一个疑问，虽然高的零阶相关表明了共线性，但并非所有情形都表明，只有高的零阶相关才带来共线性。更技术性地说，高的零阶相关是多重共线性存在的充分条件，而不是必要条件。即使零阶或简单相关系数比较低（比方说，低于0.5），多重共线性也可能存在。为了弄清楚这种关系，假使我们有一个四变量模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

360 并且假定有：

$$X_{4i} = \lambda_2 X_{2i} + \lambda_3 X_{3i}$$

其中 λ_2 和 λ_3 为常数，但不同时为零。显然，X_4 是 X_2 和 X_3 的一个准确的线性组合，从而给出 X_4 对 X_2 和 X_3 回归中的判定系数 $R^2_{4.2\,3}=1$。

回顾第7章中的公式（7.9.6），我们可以写：

$$R^2_{4.2\,3} = \frac{r^2_{4\,2} + r^2_{4\,3} - 2r_{4\,2}r_{4\,3}r_{2\,3}}{1 - r^2_{2\,3}} \tag{10.7.1}$$

但由于完全共线性，$R^2_{4.2\,3}=1$，于是得：

$$1 = \frac{r^2_{4\,2} + r^2_{4\,3} - 2r_{4\,2}r_{4\,3}r_{2\,3}}{1 - r^2_{2\,3}} \tag{10.7.2}$$

不难看出，取 $r_{4\,2}=0.5$，$r_{4\,3}=0.5$ 及 $r_{2\,3}=-0.5$ 这些不是很高的值，就能满足（10.7.2）。

因此，在涉及多于两个解释变量的模型中，简单或零阶相关并不提供判别多重共线性的一个准确无误的指南。当然，如果只有两个解释变量，看零阶相关也就够了。

3. 检查偏相关。由于刚才讲的仅仅依靠零阶相关所带来的问题，法勒（Farrar）和格劳伯（Glauber）建议我们去检查偏相关系数。[19]例如，在做 Y 对 X_2，X_3 和 X_4 的回归中，发现 $R^2_{1.2\,3\,4}$ 很高，而 $r^2_{1\,2.3\,4}$，$r^2_{1\,3.2\,4}$ 和 $r^2_{1\,4.2\,3}$ 都比较低时，可能表示变量 X_2，X_3 和 X_4 是高度交互相关的，并且至少其中一个变量是多余的。

虽然对偏相关的检查会有一定的用处，但不能保证偏相关能对多重共线性提供一个准确无误的指南。因为有可能 R^2 和全部偏相关都充分地高，多重共线性仍然出现。但更为重要的是，C.R. 威克斯（C.Robert Wichers）已证明[20]法勒-格劳伯的偏相关检验在下述意义上是无效的：给定的一种偏相关可以同不相同的多重共线性模式都无矛盾。法勒-格劳伯检验还受到T.K. 库马（T.Krishna Kumar）[21]、J. 奥黑根（John O'Hagan）和B. 麦凯布（Brendan McCabe）[22]的严厉批评。

361 **4. 辅助回归。**由于多重共线性来自回归元之一或多个是其余回归元的准确或近似线性组合，为了找出究竟哪一个 X 变量和其余 X 变量有这种关系，方法之一是，做每一 X_i 对其余 X 变量的回归，并算出相应的 R^2，记之

为 R_i^2；这样的回归叫做**辅助回归**（auxiliary regression）以辅助 Y 对诸 X 的主回归（main regression）。然后，按照（8.5.11）中建立的 F 与 R^2 之间的关系，变量遵循自由度为 $k-2$ 和 $n-k+1$ 的 F 分布。等式（10.7.3）中的 n 代表样本大小，k 代表包括截距项在内的解释变量的个数，以及 $R^2_{x_i \cdot x_2 x_3 \cdots x_k}$ 为变量 X_i 对其余 X 变量的回归中的判定系数。[23]

$$F_i = \frac{R^2_{x_i \cdot x_2 x_3 \cdots x_k}/(k-2)}{(1-R^2_{x_i \cdot x_2 x_3 \cdots x_k})/(n-k+1)} \tag{10.7.3}$$

如果计算值 F_i 超过选定显著水平的临界 F 值，我们就把它看作这个特定的 X_i 和其余的 X 有共线性；如果它不超过临界 F，就说 X_i 和其余 X 无共线性。这时可把该变量（X）保留在模型中。但如果 F_i 是统计上显著的，则 X_i 的去留问题仍待解决。在 10.8 节中我们将再回到此问题。

但是这种方法并非没有缺点，因为

> ……如果多重共线性仅涉及少数变量，辅助回归还不至于对广泛的多重共线关系有应接不暇之苦，那么所估计的系数也许能揭示回归元之间的线性关系的性质。不幸的是，如果遇上几个复杂的线性相关，做这种曲线拟合的练习就不一定有多少价值；要辨别各个不同的交互关系仍是困难的。[24]

除了对所有辅助 R^2 值做形式检验外，还可采取**克里安的经验法则**（Klien's rule of thumb）：仅当来自一个辅助回归的 R^2 大于得自 Y 对全部回归元的回归中的总 R^2 值时，多重共线性才算是一个麻烦的问题。[25] 当然，和别的经验法则一样，不可把这个经验法则当作法定的规则来运用。

5. **本征值**（eigenvalues）**与病态指数**（condition index）。如果你检查一
362 下附录 7A.5 中柯布-道格拉斯生产函数的 SAS 输出，你将看到 SAS 用本征值和病态指数（这里 condition 一词其实指 ill-conditioned。——译者注），去诊断多重共线性。这里我们不准备讨论本征值，否则将被引到超出本书范围的矩阵代数的专题讨论。但是通过本征值，我们可导出我们要讲的**病态数**（condition number）k，其定义为：

$$k = \frac{\text{最大本征值}}{\text{最小本征值}}$$

以及**病态指数**，其定义为：

$$\mathrm{CI} = \sqrt{\frac{\text{最大本征值}}{\text{最小本征值}}} = \sqrt{k}$$

于是，我们有这样的经验法则：如果 k 在 100 和 1 000 之间，就算有中强度多重共线性；如果 k 大于 1 000，就算有严重多重共线性。另一算法是：如果 CI（$=\sqrt{k}$）是在 10 与 30 之间，就算有中强度多重共线性，而如果 CI 在

30 之上，就算是严重多重共线性。

拿上面的说明例子为例，$k = 3.0/0.000\ 024\ 22$ 或约合 123 864，而 $CI = \sqrt{123\ 864}$ = 约 352；因此 k 和 CI 都表明多重共线性是严重的。当然，还可以计算最大本征值和任何其他本征值之间的 k 和 CI；参见打印结果。（注：打印结果虽然没有明算 k，但它无非是 CI 的平方。）顺便指出，有一个低的本征值（相对于最大本征值而言），一般都表明数据中有近似线性相关性。

一些作者认为病态指数是现有最好的多重共线性诊断法。但这一见解并未被广泛赞同。我们则认为，CI 也仅是一种经验法则，也许略为奥妙。要进一步了解细节，可阅读参考文献。[26]

6. 容许度与方差膨胀因子。我们已经介绍过 TOL 和 VIF。R_j^2 是 X_j 对其余（$k-2$）个回归元的（辅助）回归中的 R^2，而 VIF_j 是 10.5 节首次引进的方差膨胀因子。随着 R_j^2 朝 1 增大，也就是随着 X_j 与其他回归元的共线性增加，VIF 也增加，并且以无穷大为其极限。

因此，一些作者用 VIF 作为多重共线性的一个指标：VIF_j 值越大，变量 X_j 越“麻烦”或共线性越大。但究竟 VIF 要多高才会使回归陷入麻烦的地步呢？**作为一种经验规则**，如果一个变量的 VIF 超过 10（当 R_j^2 超过 0.90 时将发生这种情况），则说该变量是高度共线的。[27]

当然，鉴于 TOL_j 与 VIF_j 之间的密切联系，也能用 TOL_j 来度量多重共线性。TOL_j 越接近于 0，该变量与其他回归元之间共线性的程度就越大。
363 另一方面，TOL_j 越接近于 1，则 X_j 与其他回归元之间没有共线性的证据就越充分。

用 VIF（或容许度）去度量共线性也受到批评。如（10.5.4）所示，$\operatorname{var}(\hat{\beta}_j)$ 依赖于 3 个因子：σ^2，$\sum x_j^2$ 和 VIF_j。一个高的 VIF 可被一个低的 σ^2 或一个高的 $\sum x_j^2$ 抵消。换句话说，一个高的 VIF，既不是导致高的方差和高的标准误的必要条件，也不是其充分条件。因此，一个高的 VIF 度量出来的高多重共线性不一定就是高标准误的原因。在所有这些讨论中，高和低这两个用语是相对而言的。

在结束我们关于侦察多重共线性的讨论时，我们强调本节所讨论的各种方法，实质上都像是一次钓鱼式的出征（fishing expedition）。我们无法知道一种方法在任一特定的应用中是否灵验。真遗憾，想不出有什么好的办法，因为多重共线性就出现在研究者对于一个给定的样本施加不了多少控制的场合之中，尤其是数据本质上是非实验性质的——社会科学研究者所遇到的共同命运。

作为多重共线性的一种仿效，戈德伯格想出许多侦察微数缺测性的方法，比如说，制定样本大小的临界 n^* 值：仅当实际样本大小 n 小于 n^* 时，微数缺测性才会成为问题。戈德伯格仿效的意义，在于强调小样本含量以及缺少变异的解释变量会导致至少和多重共线性同等严重的问题。

§ 10.8 补救措施

如果多重共线性严重，怎么办？我们有两种选择：(1) 无为而治；(2) 采用某些经验规则。

无为而治

布兰查德（Blanchard）将“无为而治”流派表述如下[28]：

> 当学生在做他们的第一个普通最小二乘回归时，他们通常遇到的第一个问题就是多重共线性的问题。他们中许多人断定 OLS 有些问题；有些人求助于新的通常也有创造性的方法来避免这个问题。但我们告诉他们，那样做是错的，多重共线性是上帝的意志，而不是 OLS 或其他一般性统计方法的问题。

布兰查德所说的是，多重共线性实质上是一个数据不足的问题（微数缺测性），而我们有时候无法选择能用于经验分析的数据。

同样，并非回归模型中所有的系数都是统计上非显著的。此外，即使我们不能更准确地估计一或多个回归系数，但可以相对有效地估计它们的一个
364 线性组合（即可估计的函数）。如我们在（10.2.3）中所见，即使我们不能分别估计 α 的两个部分，但可以单独地估计 α。有时候，这是我们对给定的数据集最好的做法。[29]

经验程序

你也可以尝试用如下的经验法则来解决多重共线性问题，其成功与否取决于共线性问题的严重程度。

1. 先验（priori）信息。假使我们考虑模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

其中 Y = 消费，X_2 = 收入，X_3 = 财富。如前所见，收入与财富有高度共线的趋势。但若凭事先想像，认为 $\beta_3 = 0.10\beta_2$；就是说，消费对财富的变化率是对收入的相应变化率的 1/10。这样一来，我们就可做下面的回归：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + 0.10\beta_2 X_{3i} + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned}$$

其中 $X_i = X_{2i} + 0.1X_{3i}$。一旦估算出 $\hat{\beta}_2$，便可从想像中的 β_2 与 β_3 的关系式估计 $\hat{\beta}_3$。

怎样能获得先验信息呢？它可以来自先前遇到的同样严重共线问题的经验研究工作，或者来自该研究领域的有关基础理论。例如，在柯布-道格拉斯生产函数（7.10.1）中，如果人们预期不变的规模报酬生效，则有 $(\beta_2 + \beta_3) = 1$。这样就能做回归（8.7.13），即做产出劳力比对资本劳力比的回归。如果劳力和资本之间有共线性，好比大多数样本数据一般都会遇到的情形那样，这一变换就减轻或消除了共线性的问题。但这里提出一个关于施加此类先验约束的忠告是适宜的，“……因为，一般地说，我们宁愿检验经济理论上的先验预期而不是单纯地把这些未必合适的预期施加于数据之上”[30]。不管怎样，我们从 8.7 节知道了怎样去明确地检验这些约束的真实性。

2. **横截面与时间序列数据并用**。外部信息或先验信息法的一个变种，是横截面数据与时间序列数据的组合，称数据并用（pooling the data）。假如
365 我们要研究美国的汽车需求，并假定我们拥有车辆出售数、车辆平均价格和消费者收入的时间序列数据，还设定：

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$$

其中 Y = 车辆出售数，P = 平均价格，I = 收入，t = 时间。我们的目的是要估计价格弹性 β_2 和收入弹性 β_3。

在时间序列数据中，价格和收入变量一般都有高度共线的趋势。因此，如果我们做上述回归，我们将遇到通常的多重共线性问题。解决此问题的一个方法曾由托宾（Tobin）提出。[31]他说，如果我们拥有横截面数据［例如，由消费者定点追踪（consumer panels）产生的数据，或各种私人和政府机构举办的预算研究］，我们就能相当可靠地估计收入弹性 β_3。因为这些数据都产生于一点时间上，价格还不至于有多大变化。令收入弹性的横截面估计为 $\hat{\beta}_3$。利用这一估计值，就可将前述的时间序列回归写为：

$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t$$

其中 $Y^* = \ln Y - \hat{\beta}_3 \ln I$，即 Y^* 代表除去收入效应之后的 Y 值。现在就可从上面的回归中得到价格弹性的估计值。

虽然时间序列和横截面数据并用看来是一个很不错的方法，但刚才的做法可能引起解释方面的问题。因为这样做，无形地假定了收入弹性的横截面估计和从纯粹的时间序列分析中得到的估计是一样的。[32]不管怎样，数据并用技术已经在多种应用中使用，而且当横截面估计在不同截面之间无大变化时，这是一个值得考虑的方法。这种技术的一个例子见习题 10.26。

3. **剔除变量与设定偏误**。面对严重多重共线性，最简单的做法之一是剔除共线性诸变量之一。例如，在我们的消费—收入—财富一例中，当我们剔除财富变量时，得到回归（10.6.4），表明收入变量在原先的模型中是统计上不显著的，而现在则是“高度”显著的。

但从模型中删除一个变量，可能导致**设定偏误**或**设定误差**。设定偏误指

在分析中使用了不正确设定的模型。比如说，假如经济理论告诉我们，在解
366 释消费支出的模型中，应包括收入和财富两者，那么删去财富变量，就会构成设定偏误。

虽然我们将在第 13 章中专题讨论设定偏误，但在 7.7 节中也曾对它有过思考。我们曾看到如果真实模型是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

而我们错误地拟合了模型：

$$Y_i = b_1 + b_{12} X_{2i} + \hat{u}_i \tag{10.8.1}$$

那么，可以证明（见附录 13A.1）

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \tag{10.8.2}$$

其中 b_{32} = X_3 对 X_2 回归中的斜率系数。由（10.8.2）明显可见，只要 b_{32} 不为零（我们假定 β_3 异于零，否则在原始模型中包括 X_3 是没有意义的），b_{12}就必定是 β_2 的一个有偏误的估计。[33] 当然，如果 b_{32}是零，我们本来就没有多重共线性问题。从（10.8.2）还明显看到，如果 b_{32}和 β_3 都是正的，$E(b_{12})$将大于 β_2；从而平均而言，b_{32}高估了 β_2，即导致一个正的偏误；同理，如果乘积 $b_{32}\beta_3$ 是负的，则平均而言，b_{12}将低估了 β_2，即导致一个负的偏误。

由上述讨论可见，从模型中除掉一个变量以缓解多重共线性的问题会导致设定上的偏误。因此，在某些情形中，医治也许比疾病更糟糕，多重共线性虽有碍于对模型参数的准确估计，但剔除变量，则对参数的真值有严重的误导。应记得，在近似共线性情形下，OLS 估计量仍是 BLUE。

4. 变量代换。假使我们拥有消费支出、收入和财富的时间序列数据，数据中收入与财富有高度多重共线性的一个理由是，随着时间的演变这两个变量都朝同一方向变动。减少这种相依性的一个方法是按以下方法去做。

如果关系式：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \tag{10.8.3}$$

367 当时间 t 成立，那么它在时间 $t-1$ 也成立，因为，反正时间原点是任意的。因此又有：

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \tag{10.8.4}$$

如果从（10.8.3）减去（10.8.4），就得到：

$$Y_t - Y_{t-1} = \beta_2(X_{2t} - X_{2,t-1}) + \beta_3(X_{3t} - X_{3,t-1}) + v_t \tag{10.8.5}$$

其中 $v_t = u_t - u_{t-1}$。方程（10.8.5）被称为**一次差分形式**（first difference form），因为我们不是对原始变量做回归，而是对这些变量的相继差异做回归。

如我们在**时间序列计量经济学**（time series econometrics）的章节中所见，一阶差分变换的一个附带优点在于，它可以使非平稳时间序列变得平稳。我们在那些章节中将会看到平稳时间序列的重要性。如第 1 章曾指出的，粗略地讲，如果一个时间序列 Y_t 的均值和方差不随时间而系统地变化，那它就

是平稳的。

实践中另外一个常用的变换是**比率变换**（ratio transformation）。考虑模型：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \tag{10.8.6}$$

其中 Y 为以真实价格表示的消费支出，X_2 为 GDP，X_3 为总人口。由于 GDP 和人口都随时间而增长，所以它们可能会相关。对此问题的一种“解决办法”是，通过将（10.8.6）除以 X_3 得到以人均量为基础的模型：

$$\frac{Y_t}{X_{3t}} = \beta_1\left(\frac{1}{X_{3t}}\right) + \beta_2\left(\frac{X_{2t}}{X_{3t}}\right) + \beta_3 + \left(\frac{u_t}{X_{3t}}\right) \tag{10.8.7}$$

这样的变换可能会减少原有变量的共线性。

但一阶差分或比率变换都不是没有问题的。例如，（10.8.5）中的误差项 v_t 可能不满足经典线性回归模型的一个假定，即干扰的序列不相关性。我们在第 12 章将会看到，如果原来的干扰项 u_t 是序列无关的，那么上面得到的误差项 v_t 在多数情况下将会序列相关。因此，治疗比疾病更糟糕。而
368 且，还会因为差分过程而减少一个观测，并因此减少一个自由度。在小样本中，这可能是你起码要考虑的一个因素。另外，一阶差分程序在横截面数据中可能不太适合，因为横截面数据的观测不存在逻辑上的顺序。

类似地，在比率模型（10.8.7）中，如果误差项 u_t 是同方差的，那么误差项：

$$\left(\frac{u_t}{X_{3t}}\right)$$

将是异方差的，我们在第 11 章将会看到。同样，补救的办法比原来的问题更糟糕。

总之，在应用一阶差分或比率变换来解决多重共线性问题时应该尤其小心。

5. 补充新数据。由于多重共线性是一个样本特性，故有可能在关于同样变量的另一样本中共线性没有第一个样本那么严重。有时只须增大样本容量（如果可能）就能减轻共线性的问题。例如，在三变量模型中，我们看到：

$$\operatorname{var}(\hat\beta_2) = \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{23}^2)}$$

现在，随着样本增大，$\sum x_{2i}^2$ 一般地说都会增大。（为什么？）因此，对任给的 r_{23}^2，$\hat\beta_2$ 的方差将减小，从而降低标准误，以使我们能更准确地估计 β_2。

作为一个说明，考虑以下根据 10 次观测的消费支出 Y 对收入 X_2 和财富 X_3 的回归[34]：

$$\begin{aligned}\hat Y_i &= 24.377 + 0.8716 X_{2i} - 0.0349 X_{3i} \\ t &= (3.875) \quad (2.7726) \quad (-1.1595) \qquad R^2 = 0.9682\end{aligned} \tag{10.8.8}$$

回归中的财富系数不仅在 5% 的水平上不是统计上显著的，而且有错误的符号。但当样本大小增加到 40 次观测（微数缺测性？）时，我们得到以下结果：

$$\hat{Y}_i = 2.0907 + 0.7299X_{2i} + 0.0605X_{3i} \quad (10.8.9)$$
$$t=(0.8713)\quad(6.0014)\quad(2.0014)\quad R^2=0.9672$$

现在，财富系数不仅具有正确的符号，而且在5%的水平上是统计上显著的。

369 要获得补充数据或“更好”的数据，并不总是那么容易的。贾奇(Judge)等人曾说：

> 不幸的是，经济学家很少能取得补充数据而不花大本钱的。而要选取他们所希望的解释变量的值就更难了。此外，在非控制的情况下添加新变量，我们必须警惕，新增观测值的产生过程不同于原来数据集的产生过程；就是说，我们必须有把握看到，与新观测值相对应的经济结构和原来的结构是一样的。[35]

6. 在多项式回归中降低共线性。在7.10节中，我们曾讨论多项式回归模型。这种模型的一个特点是解释变量以不同的幂出现。例如，在总成本对产出、(产出)^{2}和(产出)^{3}的回归即所谓立方总成本函数(7.10.4)中，各产出项将是相关的，以致难以准确估计各个斜率系数。[36]然而，在实践中，我们发现，如果将解释变量表达为离差形式（即对均值的离差），多重共线性就可大为降低。但即使如此，问题仍然存在。[37]这时也许还可考虑诸如**正交多项式**(orthogonal polynomials)一类技术。[38]

7. 拯救多重共线性的其他方法。多元统计技术诸如**因子分析**(factor analysis)，**主元法**(principal components)或**脊回归**常被用来“解决”多重共线性问题。可惜这些技术都要利用矩阵代数才便于讨论。但这样做，就超出了本书的范围。[39]

§10.9 多重共线性一定是坏事吗？如果预测是惟一目的，就未必如此

前面说过，如果回归分析的惟一目的是预测或预报，则多重共线性不是一个严重的问题。因为，R^2值越高，预测越准。[40]但是，这也许是对的

370 “……只要预测值所对应的解释变量值和原始的设计［数据］矩阵X都遵从同样近似于准确的相依关系”[41]。比方说，如果在回归的估计中发现$X_2=2X_3$近似地成立，那么在一个用以预测Y的未来样本中，X_2也应近似地等于$2X_3$。但这是一个实际上难以满足的条件（参看章末注35）。由此可见，预测将变得越来越不确定。[42]此外，如果分析的目的不仅在于预测，而且还在于参数的可靠估计，那么，严重的多重共线性将成为一个问题，因为我们已经看到它将导致估计量的大标准误。

然而有一种情形多重共线性还不成为一个严重问题的，这就是R^2高同时回归系数由于较高的t值也都表现为个别地显著的情形。毕竟，多重共线

性诊断（如病态指数）表明了数据中有严重的共线性。那么，什么时候会出现这种“不成为严重问题的”情形呢？如约翰斯顿（Johnston）所说的：

> 如果每个个别系数都正好在数值上大大超过真值，那么尽管标准误膨胀了，效应依然显示出来，或者真值本来就是如此之大，即使估计值过低，仍然表现为显著的。[43]

§ 10.10 一个引申的例子：朗利数据

我们以分析朗利（Longley）所搜集的数据结束本章。[44] 尽管最初搜集的朗利数据只是为了在几个计算机程序中评价普通最小二乘估计值的计算精度，但现在又担负着说明包括多重共线在内的几个计量经济学问题的重任。表 10.7 复制了这些数据。这些数据是 1947—1962 年内的时间序列，Y = 被雇佣人数（以千人计）；X_1 = GNP 的暗含价格缩减指数；X_2 = GNP（以百万美元计）；X_3 = 以千人计的失业人数；X_4 = 军队中的人数；X_5 = 14 岁以上的非编制人口数；X_6 = 年份：1947 年取值 1，……，1962 年取值 16。

371 **表 10.7** **朗利数据**

观测	Y	X_1	X_2	X_3	X_4	X_5	时间
1947	60 323	830	234 289	2 356	1 590	107 608	1
1948	61 122	885	259 426	2 325	1 456	108 632	2
1949	60 171	882	258 054	3 682	1 616	109 773	3
1950	61 187	895	284 599	3 351	1 650	110 929	4
1951	63 221	962	328 975	2 099	3 099	112 075	5
1952	63 639	981	346 999	1 932	3 594	113 270	6
1953	64 989	990	365 385	1 870	3 547	115 094	7
1954	63 761	1 000	363 112	3 578	3 350	116 219	8
1955	66 019	1 012	397 469	2 904	3 048	117 388	9
1956	67 857	1 046	419 180	2 822	2 857	118 734	10
1957	68 169	1 084	442 769	2 936	2 798	120 445	11
1958	66 513	1 108	444 546	4 681	2 637	121 950	12
1959	68 655	1 126	482 704	3 813	2 552	123 366	13
1960	69 564	1 142	502 601	3 931	2 514	125 368	14
1961	69 331	1 157	518 173	4 806	2 572	127 852	15
1962	70 551	1 169	554 894	4 007	2 827	130 081	16

资料来源：见章末注 44。

假定我们的目标是基于这6个 X 变量来预测 Y。利用 Eviews3 软件，我们得到如下回归结果：

因变量：Y
样本：1947—1962

变量	系数	标准误	t 统计量	概率
C	−3 482 259.0	890 420.4	−3.910 803	0.003 6
X_1	15.061 87	84.914 93	0.177 376	0.863 1
X_2	−0.035 819	0.033 491	−1.069 516	0.312 7
X_3	−2.020 230	0.488 400	−4.136 427	0.002 5
X_4	−1.033 227	0.214 274	−4.821 985	0.000 9
X_5	−0.051 104	0.226 073	−0.226 051	0.826 2
X_6	1 829.151	455.478 5	4.015 890	0.003 0

R^2	0.995 479	因变量均值	65 317.00
校正 R^2	0.992 465	因变量标准差	3 511.968
回归标准误	304.854 1	赤池信息准则	14.577 18
残差平方和	836 424.1	施瓦茨准则	14.915 19
对数似然值	−109.617 4	F 统计量	330.285 3
DW 统计量	2.559 488	概率（F 统计量）	0.000 000

从结果一眼就看出存在多重共线性问题，因为 R^2 的值很高，但有几个变量是统计上非显著的（X_1，X_2 和 X_5），这是多重共线性的典型特征。为了更清楚地说明这一点，我们在表10.8中给出这6个回归元之间的相关关系。

372

表10.8 **相关关系**

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.000 000	0.991 589	0.620 633	0.464 744	0.979 163	0.991 149
X_2	0.991 589	1.000 000	0.604 261	0.446 437	0.991 090	0.995 273
X_3	0.620 633	0.604 261	1.000 000	0.177 421	0.686 552	0.668 257
X_4	0.464 744	0.446 437	0.177 421	1.000 000	0.364 416	0.417 245
X_5	0.979 163	0.991 090	0.686 552	0.364 416	1.000 000	0.993 953
X_6	0.991 149	0.995 273	0.668 257	0.417 245	0.993 953	1.000 000

此表给出了所谓的**相关矩阵**（correlation matrix）。此表中主对角线上的

格子（从左上角到右下角）给出了一个变量与其自身的相关系数，根据定义，都应该是1，而主对角线之外的格子给出了 X 变量两两之间的相关系数。此表的第一行给出了 X_1 与其他变量之间的相关系数。比如0.991 589就是 X_1 与 X_2 之间的相关系数，0.620 633是 X_1 与 X_3 之间的相关系数，等等。

你看到，这些两两之间的相关系数有几个很高，表明可能存在着严重的共线性问题。当然，记住前面给过的警告，这种两两相关可能是存在多重共线性的充分但非必要条件。

为了进一步了解多重共线性的性质，我们做辅助回归，即将每个 X 变量都对其余的 X 变量进行回归，为节省篇幅，我们只给出从这些回归中所得到的 R^2 值，由表10.9给出。由于辅助回归的 R^2 值很高（X_4 的回归可能例外），看来确实有严重的共线性问题。从容许度因子能得到相同的信息。前面曾指出，容许度因子越接近零，共线性的证据就越大。

表 10.9　　辅助回归的 R^2 值

因变量	R^2 值	容许度（TOL）$=1-R^2$
X_1	0.992 6	0.007 4
X_2	0.999 4	0.000 6
X_3	0.970 2	0.029 8
X_4	0.721 3	0.278 7
X_5	0.997 0	0.003 0
X_6	0.998 6	0.001 4

应用克莱因（Klein）的经验法则，我们看到，6个辅助回归中有3个回归得到的 R^2 值超过了总体 R^2 值（即从 Y 对所有 X 变量回归得到的 R^2 值）
378 0.9954，再次表明朗利数据确实被多重共线性问题所困扰。顺便提一句，应用（10.7.3）中给出的 F 检验，读者应该能够验证上表中给出的 R^2 值都是统计上显著地异于零的。

我们在前面曾指出，OLS估计量及其标准误对数据的微小变化都很敏感。习题10.32要求读者在去掉最后一个观测的情况下重做 Y 对所有6个 X 变量的回归，即对1947—1961年期间做回归。你将看到，仅去掉一年的观测，回归结果会如何变化。

既然我们已经证实存在多重共线性问题，那我们能采取什么“补救”措施呢？让我们考虑原来的模型。首先，我们可以不用名义GNP而用真实GNP，将名义GNP除以暗含的价格缩减指数即可。其次，由于14岁以上非机构人口因人口自然增长而随时间不断增长，所以它与我们模型中的时间变量 X_6 高度相关。因此，不再同时采用这两个变量，我们将留下 X_5 并去掉 X_6。最后，没有充分有力的理由把失业人数 X_3 包括进来；可能失业率是劳

动市场状况的一个更好的度量指标，但我们没有这方面的数据，故去掉变量 X_3。经过这些变化，我们得到如下回归结果（RGNP=真实 GNP）：[45]

因变量：Y

样本：1947—1962

变量	系数	标准误	t 统计量	概率
C	65 720.37	10 624.81	6.185 558	0.000 0
RGNP	9.736 496	1.791 552	5.434 671	0.000 2
X_4	−0.687 966	0.322 238	−2.134 965	0.054 1
X_5	−0.299 537	0.141 761	−2.112 965	0.056 2

R^2	0.981 404	因变量均值	65 317.00
校正 R^2	0.976 755	因变量标准差	3 511.968
回归标准误	535.449 2	赤池信息准则	15.616 41
残差平方和	3 440 470.	施瓦茨准则	15.809 55
对数似然值	−120.931 3	F 统计量	211.097 2
DW 统计量	1.654 069	概率（F 统计量）	0.000 000

尽管 R^2 值与原来的 R^2 相比略有下降，但仍然很高。现在，所有的估计系数都是统计上显著的，系数的符号也都有经济意义。

374 我们让读者自己构想另外一个模型并看其结果的变化。仍须记住以前我们听到对数据进行比率变换来解决共线性问题的警告。我们在第 11 章将再次讨论这个问题。

§ 10.11 要点与结论

1. 经典线性回归模型的假定之一，是诸解释变量 X 之间无多重共线性。大致地说，多重共线性指的是诸 X 变量之间有准确的或近似准确的线性关系。

2. 多重共线性有如下后果：如果诸 X 之间有完全共线性，则它们的回归系数是不确定的，并且它们的标准误没有定义。如果共线性是高度的而不是完全的，则回归系数的估计是可能的，但趋向于有大的标准误。其结果是，系数的总体值不能准确地加以估计。然而，如果目的在于估计这些系数的线性组合所谓可估函数，则虽有完全多重共线性也无妨。

3. 虽然没有识破共线性的十拿九稳的方法，却有几种指标可以利用，见下：

(a) 多重共线性的最明显信号是 R^2 异常高而回归系数在通常 t 检验的

基础上却没有一个是统计上显著的。当然，这是一个极端情形。

(b) 在仅有两个解释变量的模型中，检查两个变量之间的零阶或简单相关系数，会得到对共线性的一个相当好的认识。如果此相关值高，则通常可归咎于多重共线性。

(c) 然而，当模型涉及多于两个 X 解释变量时，低的零阶相关却可能给出高的多重共线性。对于这种情形，也许有必要检查偏相关系数。

(d) 如果 R^2 高而偏相关低，则多重共线性是可能的。这时一个或多个变量可能是多余的。但若 R^2 高且偏相关也高，则多重共线性也许不易识破。而且，如 C. 罗伯特，K. 库马，J. 奥黑根和 B. 麦凯布等人所指出的，法勒和格劳伯建议的偏相关检验有一些统计上的毛病。

(e) 因此，不妨拿模型中的每一 X_i 变量对所有其余 X 变量做一个回
375 归，并求出相应的判定系数 R_i^2。一个高的 R_i^2 将表明 X_i 和其余的 X 高度相关，从而可考虑把 X_i 从模型中清除出去，如果这样做，不致引起严重的设定偏误的话。

4. 多重共线性的侦察仅是整个战役的一半，另一半则是怎样解决多重共线性的问题。没有什么十拿九稳的办法。只有那么几条经验规则，其中的一些是：(1) 利用外部或先验信息；(2) 横截面与时间序列数据并用；(3) 剔除一个高度共线的变量；(4) 数据转换；(5) 获取补充或新的数据。当然，哪条规则在实践中灵验，要看数据的性质和共线性问题的严重程度。

5. 我们曾看到多重共线性在预测中的作用，并指出除非共线性的结构继续存在于未来的样本之中。否则，利用受到多重共线性困扰的回归估计去做预测，将是冒险的。

6. 虽然多重共线性在文献中备受关注（有人甚至说是过多的关注），但在经验研究中遇到的一个同等重要的问题是微数缺测性，即样本（大小）的微小性。按照戈德伯格的意见，当一篇研究论文在埋怨多重共线性时，读者应把“多重共线性”换为“微数缺测性”，看看这种埋怨是否有道理。[46] 他建议，读者须决定观测次数 n 要多小，才能说遇上了样本微小性问题，正好比人们须决定辅助回归中的 R_i^2 值要多高，才能宣称一个共线性问题是非常严重的。

习　题

问答题

10.1 在 k 变量模型中有 k 个正规方程用以估计 k 个未知数。这些正规方程见于附录 C。假定 X_k 是其余 X 变量的一个完全的线性组合，你怎样说明在这种情形中不可能估计这 k 个回归系数？

10.2 考虑表 10.10 中的一组假想数据。假如你要用模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

拟合数据，

a. 你能估计这 3 个未知数吗？为什么或为什么不？

b. 如果不能，那么你能估计这些参数是什么样的线性组合，即可估函数呢？说明必要的计算。

376 **表 10.10**

Y	**X_2**	**X_3**
−10	1	1
−8	2	3
−6	3	5
−4	4	7
−2	5	9
0	6	11
2	7	13
4	8	15
6	9	17
8	10	19
10	11	21

10.3 参照第8 章中讨论的儿童死亡率的例子。此例涉及儿童死亡率对人均 GNP 和妇女识字率的回归。现在假设我们增加变量总人口出生率。这就得到如下回归结果。

因变量:CM

变量	系数	标准误	统计量	概率
C	168.306 7	32.891 65	5.117 003	0.000 0
PGNP	−0.005 511	0.001 878	−2.934 275	0.004 7
FLR	−1.768 029	0.248 017	−7.128 663	0.000 0
TFR	12.868 64	4.190 533	3.070 883	0.003 2
R^2	0.747 372	因变量均值	141.500 0	
校正 R^2	0.734 740	因变量标准差	75.978 07	
回归标准误	39.131 27	赤池信息准则	10.232 18	
残差平方和	91 875.38	施瓦茨准则	10.367 11	
对数似然值	−323.429 8	F 统计量	59.167 67	
DW 统计量	2.170 318	概率(F 统计量)	0.000 000	

a. 将这些回归结果与方程(8.2.1)中给出的结果相比较。你看到了什么变化？你又如何解释这些变化？

b. 值得在模型中增加变量 TFR 吗？为什么？

c. 既然所有的个别 t 系数都是统计显著的，我们能否说此时不存在共线性问题？

10.4 如果关系式 $\lambda_1X_{1i}+\lambda_2X_{2i}+\lambda_3X_{3i}=0$ 对所有 λ_1,λ_2 和 λ_3 值都成立，试估计 $r_{12.3}, r_{13.2}$ 和 $r_{23.1}$。再求 $R^2_{1.23}, R^2_{2.13}$ 和 $R^2_{3.12}$。在此情形中多重共线性的程度为何？注：$R^2_{1.23}$ 是 Y 对 X_2 和 X_3 回归中的判定系数。类似地解释其他 R^2 值。

10.5 考虑以下模型：

$$Y_t=\beta_1+\beta_2X_t+\beta_3X_{t-1}+\beta_4X_{t-2}+\beta_5X_{t-3}+\beta_6X_{t-4}+u_t$$

其中 Y=消费，X=收入，t=时间。上述模型假定了时间 t 的消费支出不仅是时间 t 的收入，而且是以前多期的收入的函数。例如，1976 年第 1 季度的消费支出是同季度收入和 1975 年的四个季度收入的函数。这类模型叫做**分布滞后模型**(distributed lag models)。我们将在以后的一章中加以讨论。

a. 你预期在这类模型中有多重共线性吗？为什么？

b. 如果预期有多重共线性，你会怎样解决这个问题？

10.6 考虑10.6 节的说明性例子。你会怎样解释(10.6.1)和(10.6.4)所得的边际消费倾向中的差异？

10.7 在涉及诸如 GNP、货币供给、价格、收入、失业等时间序列的数据中，一般都疑虑有多重共线性，为什么？

10.8 设想在模型

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i$$

其中，X_2 和 X_3 之间的相关系数 r_{23} 为零。因此，某人建议你做如下的回归：

$$Y_i=\alpha_1+\alpha_2X_{2i}+u_{1i}$$

$$Y_i=\gamma_1+\gamma_3X_{3i}+u_{2i}$$

a. 会不会有 $\hat{\alpha}_2=\hat{\beta}_2$ 且 $\hat{\gamma}_3=\hat{\beta}_3$ 呢？为什么？

b. $\hat{\beta}_1$ 会等于 $\hat{\alpha}_1$ 或 $\hat{\gamma}_1$ 或两者的某个线性组合？

c. 会不会有 $\text{var}(\hat{\beta}_2)=\text{var}(\hat{\alpha}_2)$ 且 $\text{var}(\hat{\beta}_3)=\text{var}(\hat{\gamma}_3)$？

10.9 参照第7章的说明性例子。在该例中我们曾对台湾的农业部门拟合柯布-道格拉斯生产函数。由（7.9.4）给出的回归结果表明劳动力和资本系数两者都是个别地在统计上显著的。

a. 判明劳力和资本两变量是否高度相关。

b. 如果你对（a）的回答是肯定的，你会不会从模型中剔除劳动力变量（比方说），而仅对资本投入做产出变量的回归呢？

c. 如果你这样做，你将冒犯哪一种设定偏误？找出这种偏误的性质。

10.10 参照例 7.4。这个问题的相关矩阵如下：

	X_i	X_i^2	X_i^3
X_i	1	0.974 2	0.928 4
X_i^2		1.0	0.987 2
X_i^3			1.0

a. 对“由于零阶相关非常之高，必定有严重的多重共线性”。试加评论。

b. 你会从模型中剔除 X_i^2 和 X_i^3 吗？

c. 如果你把它们剔除，X_i 的系数值将会出现什么情况？

378 **10.11** **逐步回归**。为决定一个回归模型的“最优”解释变量集，研究者常用逐步回归的方法。在此方法中，既可采取每次引进一个 X 变量**逐步向前回归**（stepwise forward regression）的程序，也可先把所有可能的 X 变量都放在一个复回归中，然后逐一地把它们剔除**逐步向后回归**（stepwise backward regression）。加进或除掉一个变量，通常是根据 F 检验看它对 ESS 的贡献而做出决定的。根据你现在对多重共线性的认识，你赞成任何一种逐步程序吗？为什么或为什么不？[1]

10.12 以下的陈述是正确，是谬误，还是不肯定？说出理由。

a. 尽管有完全的多重共线性，OLS 估计量仍然是 BLUE。

b. 在高度多重共线性的情形中，要评价一或多个偏回归系数的个别显著性是不可能的。

c. 如果有某一辅助回归显示出高的 R_i^2 值，则高度共线性的存在是肯定无疑的了。

d. 变量的两两高度相关并不表示高度多重共线性。

e. 如果分析的目的仅仅是预测，则多重共线性是无害的。

f. 其他条件不变，VIF 越高，OLS 估计量的方差越大。

g. 和 VIF 相比，容许度是多重共线性的更好度量。

h. 如果在复回归中，根据通常的 t 检验，全部偏斜率系数个别而论都是统计上不显著的，你就不会得到一个高的 R^2 值。

i. 在 Y 对 X_2 和 X_3 的回归中，假如 X_3 的值很少变化，这就会使 $\text{var}(\hat{\beta}_3)$ 增大，在极端的情形下，如果全部 X_3 值都相同，$\text{var}(\hat{\beta}_3)$ 将是无穷大。

10.13 a. 证明：如果 $r_{1i}=0$ 对 $i=2、3、\cdots、k$，则：

$$R_{1.2\,3\cdots k}=0$$

b. 对于变量 $X_1(=Y)$ 对 X_2、X_3、…，X_k 的回归来说，这一发现有什么重要意义？

10.14 假如 $X_1(=Y)$、X_2、…X_k 的全部零阶相关系数都等于 r。

a. $R^2_{1.23\cdots k}$ 值为何？

b. 一阶相关系数的值为何？

***10.15** 在第 9 章中我们看到，用矩阵符号表示有：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

a. 当诸 X 变量之间有完全共线性时，$\hat{\beta}$ 会发生什么情况？

b. 你怎样知道有没有完全的共线性？

379 ***10.16** 用矩阵符号表示，我们在（9.3.13）中得到：

$$\text{var}-\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

a. 在有完全的多重共线性时，和 b. 共线性是高度的，但并非完全时，上述方差协方差矩阵会分别出现什么情况？

***10.17** 考虑如下的**相关矩阵**：

$$\mathbf{R} = \begin{array}{c} \\ X_2 \\ X_3 \\ \\ X_k \end{array}\begin{array}{c} \begin{array}{cccc} X_2 & X_3 & \cdots & X_k \end{array} \\ \begin{bmatrix} 1 & r_{23} & \cdots & r_{2k} \\ r_{32} & 1 & \cdots & r_{3k} \\ \cdots & \cdots & \cdots & \\ r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix} \end{array}$$

你怎样从相关矩阵看出是否（a）有完全共线性，（b）有少于完全的共线性，以及（c）诸 X 不相关。

提示：你可利用 $|\mathbf{R}|$ 来回答这些问题，这里 $|\mathbf{R}|$ 表示 $\mathbf{R}$ 的行列式。

***10.18** **正交解释变量**。假使在模型中，

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

从 X_2 到 X_k 各不相关。这样的变量，叫做正交变量。在这种情形中，

a. $(\mathbf{X}'\mathbf{X})$ 矩阵的结构将是怎样的？

b. 你将怎样求 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$？

c. $\hat{\boldsymbol{\beta}}$ 的方差协方差矩阵有怎样的性质？

d. 假如你在做完回归之后，想再引进另一正交变量 X_{k+1} 到模型中来，你需要重新计算先前的系数 $\hat{\beta}_1$ 至 $\hat{\beta}_k$ 吗？为什么或为什么不？

10.19 考虑如下模型：

---

* 选做题。

$$\mathrm{GNP}_t = \beta_1 + \beta_2 \mathrm{M}_t + \beta_3 \mathrm{M}_{t-1} + \beta_4(\mathrm{M}_t - \mathrm{M}_{t-1}) + u_t$$

其中，GNP_t = 时间 t 的 GNP，M_t = 时间 t 的货币供给，M_{t-1} = 时间$(t-1)$的货币供给以及$(\mathrm{M}_t - \mathrm{M}_{t-1})$ = 从时间$(t-1)$到时间 t 货币供给的变化。也就是，此模型设想时间 t 的 GNP 是时间 t 和时间$(t-1)$的货币供给以及这段时间货币供给变化的函数。

a. 假定你拥有估计上述模型的数据，你能成功地估计出模型的全部系数吗？为什么或为什么不？

b. 如果不能，那么什么系数可以估计？

380 c. 假使 $\beta_3 \mathrm{M}_{t-1}$ 一项不在模型中出现，你对(a)的回答仍然一样吗？

d. 重做(c)，但现在假定 $\beta_2 \mathrm{M}_t$ 不出现。

10.20 说明(7.4.7)和(7.4.8)还可表示为：

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2)(1 - r_{23}^2)}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2)(1 - r_{23}^2)}$$

其中 r_{23} 是 X_2 和 X_3 的相关系数。

10.21 利用(7.4.12)和(7.4.15)，说明当完全共线性存在时，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差是无穷大。

10.22 证实由(10.5.6)和(10.5.7)估计的斜率系数总和的标准误分别是 0.154 9 和 0.182 5。(见 10.5 节。)

10.23 对于 k 变量回归模型，可以证明(7.5.6)中给出的第 k 个($k=2,3,\cdots,k$)偏回归系数的方差可表示为[2]：

$$\operatorname{var}(\hat{\beta}_k) = \frac{1}{n-k} \frac{\sigma_y^2}{\sigma_k^2}\left(\frac{1-R^2}{1-R_k^2}\right)$$

其中 σ_y^2 = Y 的方差，σ_k^2 = 第 k 个解释变量的方差，$R_k^2 = X_k$ 对其余 X 变量的回归中的 R^2，以及 R^2 = 复回归即 Y 对全部 X 变量的回归中的判定系数。

a. 其他情况不变，如果 σ_k^2 增大，$\operatorname{var}(\hat{\beta}_k)$ 会出现什么情况？这时多重共线性问题有什么含义？

b. 如果共线性是完全的，上述公式会出现什么情况？

c. “$\hat{\beta}_k$ 的方差随 R^2 上升而下降，因此由高的 R_k^2 产生的影响可由高的 R^2 来抵消。”正确或谬误？

10.24 根据 1899—1922 年美国制造业部门的年度数据，多尔蒂(Dougherty)获得如下回归结果[3]：

$$\widehat{\log Y} = 2.81 - 0.53\log K + 0.91\log L + 0.047t \tag{1}$$
$$\text{se} = (1.38) \quad (0.34) \quad (0.14) \quad (0.021)$$
$$R^2 = 0.97 \qquad F = 189.8$$

其中 Y = 实际产出指数，K = 实际资本投入指数，L = 实际劳动力投入指数，t = 时间或趋势。

381 利用同样数据，他又获得以下回归：

$$\widehat{\log(Y/L)} = -0.11 + 0.11\log(K/L) + 0.006t \tag{2}$$
$$\text{se} = (0.03) \quad (0.15) \quad (0.006)$$
$$R^2 = 0.65 \qquad F = 19.5$$

a. 回归（1）中有没有多重共线性？你怎样知道的？

b. 在回归（1）中，$\log K$ 的先验符号是什么？结果是否与预期的一致？为什么或为什么不？

c. 你怎样替回归的函数形式（1）做辩护？（提示：柯布-道格拉斯生产函数。）

d. 解释回归（1）。在此回归中趋势变量的作用为何？

e. 估计回归（2）的道理何在？

f. 如果原先的回归（1）有多重共线性，是否已被回归（2）减弱？你怎样知道？

g. 如果回归（2）被看作回归（1）的一个受约束形式，作者施加的约束是什么呢？（提示：规模报酬。）你怎样知道这个约束是否正确？你用哪一种检验？说明你的计算。

h. 两个回归的 R^2 值是可比的吗？为什么或为什么不？如果它们现在的形式不可比，你会怎样使得它们可比？

10.25 批判性地评价如下命题：

a. "多重共线性实际上不是一个建模的错误，而是数据不充分的一种状况。"[4]

b. "如果不能得到更多的数据，那就必须接受数据包含有限信息量的事实并相应地设定模型。试图估计过分复杂的模型是经验丰富的应用计量经济学家最常见的错误之一。"[5]

c. "研究者通常认为，只要在回归结果中没有看到他们预先假设的符号，他们先验地推定重要的变量具有不显著的 t 值，或者去掉一个解释变量会导致各种回归结果都明显变化，那就是多重共线性在作怪。不幸的是，这些条件中没有一个是存在共线性的充分或必要条件，而且对于解决他们提出的估计问题需要什么样的额外信息没有提供任何有用的建议。"[6]

d. "……任何包含多于四个自变量的时间序列回归都会带来垃圾。"[7]

382 **解答题**

10.26 克莱因和戈德伯格试图对美国经济拟合如下回归模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

其中 Y = 消费，X_2 = 工资收入，X_3 = 非工资、非农场收入，及 X_4 = 农场收入。但他们预料 X_2，X_3 和 X_4 高度共线，故通过横截面分析把 β_3 和 β_4 估计为 $\beta_3 = 0.75\beta_2$ 和 $\beta_4 = 0.625\beta_2$。利用这些估计，他们重新建立他们的消费函数如下：

$$Y_i = \beta_1 + \beta_2(X_{2i} + 0.75X_{3i} + 0.625X_{4i}) + u_i = \beta_1 + \beta_2 Z_i + u_i$$

其中 $Z_i = X_{2i} + 0.75X_{3i} + 0.625X_{4i}$。

a. 用这个修改的模型去拟合表 10.11 所附数据，并估计 β_1 至 β_4。

b. 你会怎样解释变量 Z?

表 10.11

年份	Y	X_2	X_3	X_4	年份	Y	X_2	X_3	X_4
1936	62.8	43.41	17.10	3.96	1946	95.7	76.73	28.26	9.76
1937	65.0	46.44	18.65	5.48	1947	98.3	75.91	27.91	9.31
1938	63.9	44.35	17.09	4.37	1948	100.3	77.62	32.30	9.85
1939	67.5	47.82	19.28	4.51	1949	103.2	78.01	31.39	7.21
1940	71.3	51.02	23.24	4.88	1950	108.9	83.57	35.61	7.39
1941	76.6	58.71	28.11	6.37	1951	108.5	90.59	37.58	7.98
1945*	86.3	87.69	30.29	8.96	1952	111.4	95.47	35.17	7.42

* 战争年代 1942—1944 年的数据缺落。其他年份的数据以 1939 年 10 亿美元计。

资料来源：L.R. Klein and A. S. Goldberger, *An Economic Model of the United States, 1929—1952*, North Holland Publishing Company, Amsterdam, 1964, p. 131.

10.27 表 10.12 给出 1970—1998 年期间美国进口、GDP 和消费者价格指数数据。

383 **表 10.12** **1970—1998 年美国商品进口、GDP 和 CPI**

年份	CPI，全部项目 (1967 年 = 100)	GDP (10 亿美元)	商品进口 (百万美元)	年份	CPI，全部项目 (1967 年 = 100)	GDP (10 亿美元)	商品进口 (百万美元)
1970	38.8	1 039.7	39 866	1985	107.6	4 213.0	338 088
1971	40.5	1 128.6	45 579	1986	109.6	4 452.9	368 425
1972	41.8	1 240.4	55 797	1987	113.6	4 742.5	409 765
1973	44.4	1 385.5	70 449	1988	118.3	5 108.3	447 189

1974	49.3	1 501.0	103 811	1989	124.0	5 489.1	477 365
1975	53.8	1 635.2	98 185	1990	130.7	5 803.2	498 337
1976	56.9	1 823.9	124 228	1991	136.2	5 986.2	490 981
1977	60.6	2 031.4	151 907	1992	140.3	6 318.9	536 458
1978	65.2	2 295.9	176 002	1993	144.5	6 642.3	589 441
1979	72.6	2 566.4	212 007	1994	148.2	7 054.3	668 590
1980	82.4	2 795.0	249 750	1995	152.4	7 400.5	749 574
1981	90.9	3 131.3	265 067	1996	156.9	7 813.2	803 327
1982	96.5	3 259.2	247 642	1997	160.5	8 300.8	876 366
1983	99.6	3 534.9	268 901	1998	163.0	8 759.9	917 178
1984	103.9	3 932.7	332 418				

请你考虑以下模型：

$\ln$ 进口$_t = \beta_1 + \beta_2 \ln \mathrm{GDP}_t + \beta_3 \ln \mathrm{CPI}_t + u_i$

a. 用表中数据估计此模型的参数。

b. 你猜想数据中有多重共线性吗？

c. 利用病态指数，以分析共线性的性质。

d. 做回归：(1) $\ln$ 进口$_t = A_1 + A_2 \ln \mathrm{GDP}_t$

(2) $\ln$ 进口$_t = B_1 + B_2 \ln \mathrm{CPI}_t$

(3) $\ln \mathrm{GDP}_t = C_1 + C_2 \ln \mathrm{CPI}_t$

根据这些回归你能对数据中多重共线性的性质得出什么结论？

e. 假使数据有多重共线性，但 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 在 5% 水平上个别地显著，并且总的 F 检验也是显著的。对这样的情形，我们用不用顾虑共线性的问题？

10.28 参考习题 7.19 关于美国对子鸡的需求函数。

a. 利用对数线性或双对数模型估计各种辅助回归，一共有多少个这样的回归？

b. 你怎样从这些辅助回归决定哪些回归元是高度共线的？你用哪一种检验？说明你的计算细节。

c. 如果数据中有显著的共线性，你会除掉哪个（些）变量以减少共线性问题的严重性？如果你这样做，你会遇到什么计量经济学问题？

d. 除了删去变量之外，你有什么建议可以缓解共线性问题？做出解释。

10.29 表10.13 给出作为若干变量的函数的美国新客车出售数据。

a. 制定适当的线性或对数线性模型，以估计美国对汽车的需求函数。

b. 如果你决定用表中全部回归元作为解释变量，你预料会遇到

多重共线性的问题吗？为什么？

c. 如果你这样预料，你准备怎样解决这个问题？明确你的假定并说明全部计算。

384 **表 10.13**

年份	Y	X_2	X_3	X_4	X_5	X_6
1971	10 227	112.0	121.3	776.8	4.89	79 367
1972	10 872	111.0	125.3	839.6	4.55	82 153
1973	11 350	111.1	133.1	949.8	7.38	85 064
1974	8 775	117.5	147.7	1 038.4	8.61	86 794
1975	8 539	127.6	161.2	1 142.8	6.16	85 846
1976	9 994	135.7	170.5	1 252.6	5.22	88 752
1977	11 046	142.9	181.5	1 379.3	5.50	92 017
1978	11 164	153.8	195.3	1 551.2	7.78	96 048
1979	10 559	166.0	217.7	1 729.3	10.25	98 824
1980	8 979	179.3	247.0	1 918.0	11.28	99 303
1981	8 535	190.2	272.3	2 127.6	13.73	100 397
1982	7 980	197.6	286.6	2 261.4	11.20	99 526
1983	9 179	202.6	297.4	2 428.1	8.69	100 834
1984	10 394	208.5	307.6	2 670.6	9.65	105 005
1985	11 039	215.2	318.5	2 841.1	7.75	107 150
1986	11 450	224.4	323.4	3 022.1	6.31	109 597

注：Y = 新客车出售量，未经季节调整数据；

X_2 = 新车，消费者价格指数，1967 年 = 100，未经季节调整；

X_3 = 消费者价格指数，全部项目，全部城市消费者，1967 年 = 100，未经季节调整；

X_4 = 个人可支配收入，10 亿美元，未经季节调整；

X_5 = 利率，百分数，金融公司票据直接使用；

X_6 = 民间就业劳动人数（千人），未经季节调整。

资料来源：*Business Statistics*, 1986, A Supplement to the *Current Survey of Business*, U. S. Department of Commerce.

10.30 为了评价保障最低年工资（负收入税）政策的可行性，兰德（Rand）公司进行了一项研究，以评价劳动供给（平均工作小时数）对小时工资提高的反应。[8] 此研究中的数据取自6 000户男户主年收入低于15 000美元的一个国民样本。这些数据被分成 39 个人口组，并放在表 10.14 中。由于 4 个人口组中的某些

变量缺失，所以此表中只给出了 35 个组的数据。用于分析的各个变量的定义在表末给出。

a. 将该年度平均工作小时数对表中变量进行回归，并解释你的回归。

b. 数据中存在多重共线性的证据吗？你如何知道？

c. 计算各个回归元的方差膨胀因子和 TOL 度量。

d. 若存在多重共线性问题，那你会采用什么补救措施（如果有的话）？

e. 此研究对负收入税的可行性有何结论？

385 **表 10.14** **35 个人口组的工作小时数及其他数据**

观测	Hours	Rate	ERSP	ERNO	NEIN	Assets	Age	DEP	School
1	2 157	2.905	1 121	291	380	7 250	38.5	2.340	10.5
2	2 174	2.970	1 128	301	398	7 744	39.3	2.335	10.5
3	2 062	2.350	1 214	326	185	3 068	40.1	2.851	8.9
4	2 111	2.511	1 203	49	117	1 632	22.4	1.159	11.5
5	2 134	2.791	1 013	594	730	12 710	57.7	1.229	8.8
6	2 185	3.040	1 135	287	382	7 706	38.6	2.602	10.7
7	2 210	3.222	1 100	295	474	9 338	39.0	2.187	11.2
8	2 105	2.493	1 180	310	255	4 730	39.9	2.616	9.3
9	2 267	2.838	1 298	252	431	8 317	38.9	2.024	11.1
10	2 205	2.356	885	264	373	6 789	38.8	2.662	9.5
11	2 121	2.922	1 251	328	312	5 907	39.8	2.287	10.3
12	2 109	2.499	1 207	347	271	5 069	39.7	3.193	8.9
13	2 108	2.796	1 036	300	259	4 614	38.2	2.040	9.2
14	2 047	2.453	1 213	297	139	1 987	40.3	2.545	9.1
15	2 174	3.582	1 141	414	498	10 239	40.0	2.064	11.7
16	2 067	2.909	1 805	290	239	4 439	39.1	2.301	10.5
17	2 159	2.511	1 075	289	308	5 621	39.3	2.486	9.5
18	2 257	2.516	1 093	176	392	7 293	37.9	2.042	10.1
19	1 985	1.423	553	381	146	1 866	40.6	3.833	6.6
20	2 184	3.636	1 091	291	560	11 240	39.1	2.328	11.6
21	2 084	2.983	1 327	331	296	5 653	39.8	2.208	10.2
22	2 051	2.573	1 194	279	172	2 806	40.0	2.362	9.1
23	2 127	3.262	1 226	314	408	8 042	39.5	2.259	10.8

24	2 102	3.234	1 188	414	352	7 557	39.8	2.019	10.7
25	2 098	2.280	973	364	272	4 400	40.6	2.661	8.4
26	2 042	2.304	1 085	328	140	1 739	41.8	2.444	8.2
27	2 181	2.912	1 072	304	383	7 340	39.0	2.337	10.2
28	2 186	3.015	1 122	30	352	7 292	37.2	2.046	10.9
29	2 188	3.010	990	366	374	7 325	38.4	2.847	10.6
30	2 077	1.901	350	209	95	1 370	37.4	4.158	8.2
31	2 196	3.009	947	294	342	6 888	37.5	3.047	10.6
32	2 093	1.899	342	311	120	1 425	37.5	4.512	8.1
33	2 173	2.959	1 116	296	387	7 625	39.2	2.342	10.5
34	2 179	2.971	1 128	312	397	7 779	39.4	2.341	10.5
35	2 200	2.980	1 126	204	393	7 885	39.2	2.341	10.6

注：Hours＝该年度平均工作小时数；Rate＝平均小时工资（美元）；
ERSP＝配偶年均收入（美元）；ERNO＝其他家庭成员的年均收入（美元）；
NEIN＝年均非劳动收入；Assets＝平均家庭资产拥有量（银行存款等）（美元）；
Age＝被调查者的平均年龄；Dep＝平均赡养人数；
School＝平均完成的最高年级。

资料来源：D. H. Greenberg and M. Kosters, *Income Guarantees and the Working Poor*, The Rand Corporation, R-579-OEO, December1970.

10.31 表10.15 给出了 1960 年美国 47 个州的犯罪率数据。试着用一个适当的模型来解释犯罪与表中 14 个社会经济变量的关系。在给出你的模型时，特别注意共线性问题。

886 **表 10.15** **1960 年美国 47 个州的犯罪数据**

观测	R	Age	S	ED	EX_0	EX_1	LF	M	N	NW	U_1	U_2	W	X
1	79.1	151	1	91	58	56	510	950	33	301	108	41	394	261
2	163.5	143	0	113	103	95	583	1 012	13	102	96	36	557	194
3	57.8	142	1	89	45	44	533	969	18	219	94	33	318	250
4	196.9	136	0	121	149	141	577	994	157	80	102	39	673	167
5	123.4	141	0	121	109	101	591	985	18	30	91	20	578	174
6	68.2	121	0	110	118	115	547	964	25	44	84	29	689	126
7	96.3	127	1	111	82	79	519	982	4	139	97	38	620	168
8	155.5	131	1	109	115	109	542	969	50	179	79	35	472	206
9	85.6	157	1	90	65	62	553	955	39	286	81	28	421	239
10	70.5	140	0	118	71	68	632	1 029	7	15	100	24	526	174

11	167.4	124	0	105	121	116	580	966	101	106	77	35	657	170
12	84.9	134	0	108	75	71	595	972	47	59	83	31	580	172
13	51.1	128	0	113	67	60	624	972	28	10	77	25	507	206
14	66.4	135	0	117	62	61	595	986	22	46	77	27	529	190
15	79.8	152	1	87	57	53	530	986	30	72	92	43	405	264
16	94.6	142	1	88	81	77	497	956	33	321	116	47	427	247
17	53.9	143	0	110	66	63	537	977	10	6	114	35	487	166
18	92.9	135	1	104	123	115	537	978	31	170	89	34	631	165
19	75.0	130	0	116	128	128	536	934	51	24	78	34	627	135
20	122.5	125	0	108	113	105	567	985	78	94	130	58	626	166
21	74.2	126	0	108	74	67	602	984	34	12	102	33	557	195
22	43.9	157	1	89	47	44	512	962	22	423	97	34	288	276
23	121.6	132	0	96	87	83	564	953	43	92	83	32	513	227
24	96.8	131	0	116	78	73	574	1 038	7	36	142	42	540	176
25	52.3	130	0	116	63	57	641	984	14	26	70	21	486	196
26	199.3	131	0	121	160	143	631	1 071	3	77	102	41	674	152
27	34.2	135	0	109	69	71	540	965	6	4	80	22	564	139
28	121.6	152	0	112	82	76	571	1 018	10	79	103	28	537	215
29	104.3	119	0	107	166	157	521	938	168	89	92	36	637	154
30	69.6	166	1	89	58	54	521	973	46	254	72	26	396	237
31	37.3	140	0	93	55	54	535	1 045	6	20	135	40	453	200
32	75.4	125	0	109	90	81	586	964	97	82	105	43	617	163
33	107.2	147	1	104	63	64	560	972	23	95	76	24	462	233
34	92.3	126	0	118	97	97	542	990	18	21	102	35	589	166
35	65.3	123	0	102	97	87	526	948	113	76	124	50	572	158
36	127.2	150	0	100	109	98	531	964	9	24	87	38	559	153
37	83.1	177	1	87	58	56	638	974	24	349	76	28	382	254
38	56.6	133	0	104	51	47	599	1 024	7	40	99	27	425	225
39	82.6	149	1	88	61	54	515	953	36	165	86	35	395	251
40	115.1	145	1	104	82	74	560	981	96	126	88	31	488	228
41	88.0	148	0	122	72	66	601	998	9	19	84	20	590	144
42	54.2	141	0	109	56	54	523	968	4	2	107	37	489	170
43	82.3	162	1	99	75	70	522	996	40	208	73	27	496	224
44	103.0	136	0	121	95	96	574	1 012	29	36	111	37	622	162

45	45.5	139	1	88	46	41	480	968	19	49	135	53	457	249
46	50.8	126	0	104	106	97	599	989	40	24	78	25	593	171
47	84.9	130	0	121	90	91	623	1 049	3	22	113	40	588	160

变量定义：R = 犯罪率，每百万人口中向警察报告的违法次数；

Age = 每千人中年龄在 14～24 岁的男性人数；

S = 位于南方与否的指标变量（0 = 否，1 = 是）；

ED = 25 岁及 25 岁以上人口读书年数的均值乘以 10；

EX_0 = 1960 年州和地方政府对警方的人均支出；

EX_1 = 1959 年州和地方政府对警方的人均支出；

LF = 每 1 000 名 14～24 岁城镇男性居民的劳动力参与率；

M = 每 1 000 名女性对应的男性人数；

N = 以 10 万计的州人口规模；

NW = 每 1 000 人中非白人的人口数；

U_1 = 每 1 000 名 14～24 岁城镇男性的失业率；

U_2 = 每 1 000 名 35～39 岁城镇男性的失业率；

W = 以 10 美元计可转换商品和资产或家庭收入的中位数；

X = 每 1 000 户中挣到中位数收入一半的家庭数；

观测 = 州（1960 年的 47 个州）。

资料来源：W. Vandaele, "Participation in Illegitimate Activities: Erlich Revisted," in A. Blumstein, J. Cohen, and Nagin, D., eds., *Deterrence and Incapacitation*, National Academy of Sciences, 1978, pp. 270－335.

10.32 参照第10.10 节中给出的朗利数据。去掉 1962 年的数据重做表中的回归；即做 1947—1961 年期间的回归。比较这两个回归。从此题中你能得到什么一般性的结论？

【习题注释】

[1] 参看 Arthur S. Goldberg and D. B. Jochems, "Note on Stepwise Least-Squares," *Journal of the American Statistical Association*, vol. 56, March 1961, pp. 105－110. 比较一下你的理解是否和这些作者一致。

[2] 此公式见于 R. Stone, "The Analysis of Market Demand," *Journal of the Royal Statistical Society*, vol. B7, 1945, p. 297, 还可回顾(7.5.6)。进一步的讨论，见于 Peter Kennedy, *A Guide to Econometrics*, 2d ed., The MIT Press, Cambridge, Mass., 1985, p. 156。

[3] Christopher Dougherty, *Introduction to Econometrics*, Oxford University Press, New York, 1992, pp. 159－160.

[4] Samprit Chatterjee, Ali S. Hadi, and Betram Price, *Regression Analysis by Example*, 3d ed., John Wiley & Sons, New York, 2000, p. 226.

[5] Russel Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, p. 186.

[6] Peter Kennedy, *A Guide to Econometrics*, 4th ed., MIT Press, Cambridge,

Mass., 1998, p.187.

[7] 此段引文是已故计量经济学家 Zvi Griliches 所说，引自 Ernst R. Berndt, *The Practice of Econometrics: Classic and Contemporary*, Addison Wesley, Reading, Mass., 1991, p. 224。

[8] D. H. Greenberg and M. Kosters, *Income Guarantees and the Working Poor*, Rand Corporation, R-579-OEO, December 1970.

【注释】

[1] Edward E. Leamer, "Model Choice and Specification Analysis," Zvi Griliches and Michael D. Intriligator, eds., *Handbook of Econometrics*, vol. 1, North Holland Publishing Company, Amsterdam, 1983, pp.300-301.

[2] 见他写的 *A Course in Econometrics*, Harvard University Press, Cambridge, Mass., 1991, p.249。

[3] Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Institute of Economics, Oslo University, publ. no.5, 1934.

[4] 严格地说，多重共线性是指存在有多于 1 个的准确线性关系。而共线性，则指存在有单一的线性关系。但在实践中，这种区分很少得到遵循，从而多重共线性兼指两种情形。

[5] 在实践中要得到一个样本，它的回归元的数值是按这种方式联系起来的，这样的机会确实是很小的，除非有意设计。比如说，当观测次数小于回归元个数时，或者研究者陷入了以后第 15 章讨论的"虚拟变量陷阱"时，就会出现有（10.1.1）这种关系。

[6] 如果只有两个解释变量，交互相关就可由零阶或简单相关系数来度量。但若有多于两个 X 变量，则交互相关由偏相关系数或一个 X 变量对其余所有 X 变量的复回归系数 R 来度量。

[7] Douglas Montgomery and Elizabeth Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, 1982, pp.289-290. Seealso R. L. Mason, R. F. Gunst, and J. T. Webster, "Regression Analysis and Problems of Multicollinearity," *Communications in Statistics A*, vol.4, no.3, 1975, pp.277-292; R. F. Gunst, and R. L. Mason, "Advantages of Examining Multicollinearities in Regression Analysis," *Biometrics*, vol. 33, 1977, pp.249-260.

[8] 说明此问题的另一方法是：按定义 X_2 和 X_3 的相关系数 r_{23} 是 $\sum x_{2i}x_{3i}/\sqrt{\sum x_{2i}^2\sum x_{3i}^2}$。如果 $r_{23}^2=1$，即 X_2 和 X_3 完全共线，则(7.4.7)的分母将为零，从而使 β_2(或 β_3)的估计不可能。

[9] 在计量经济学(和数理统计学——译者注)文献中，称类似于($\beta_2+\lambda\beta_3$)的函数为**可估计的函数**(estimable function)。

[10] 由于近似多重共线性本身并不违背第 7 章所列的其他假定，OLS 估计仍是那里说的 BLUE。

[11] Christopher H. Achen, *Interpreting and Using Regression*, Sage Publications, Beverly Hills, Calif., 1982, pp.82-83.

[12] Peter Kennedy, *A Guide to Econometrics*, 3d ed., The MIT Press, Cambridge, Mass., 1992, p.177.

[13] 用置信区间表示，随着共线性程度的增大，$\beta_2=0$ 这个值将越来越多地落入接受区间内。

[14] 这些标准误得自公式：

$$\mathrm{se}(\hat{\beta}_2+\hat{\beta}_3)=\sqrt{\mathrm{var}(\hat{\beta}_2)+\mathrm{var}(\hat{\beta}_3)+2\mathrm{cov}(\hat{\beta}_2,\hat{\beta}_3)}$$

注意，增加共线性，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差也随之增加。但若两者有高的负协方差，则这些方差可能被抵消，如同我们的结果所表明的。

[15] Goldberger，前引文献，第 248～250 页。

[16] 第 5.3 节已表明，联合置信域的问题比较复杂，有兴趣的读者可参考那里所引的文献。

[17] Jan Kmenta, *Elements of Econometrics*, 2d ed., Macmillan, New York, 1986, p.431.

[18] 同上，第 439 页。

[19] D.E. Farrar and R.R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *Review of Economics and Statistics*, vol.49, 1967, pp.92－107.

[20] "The Detection of Multicollinearity: A Comment", *Review of Economics and Statistics*, vol.57, 1975, pp.365－366.

[21] "Multicollinearity in Regression Analysis," *Review of Economics and Statistics*, vol.57, 1975, pp.366－368.

[22] "Tests for the Severity of Multicollinearity in Regression Analysis: A Comment", *Review of Economics and Statistics*, vol.57, 1975, pp.368－370.

[23] 例如，$R^2_{x_2}$ 可通过做 X_{2i} 的如下回归得到：$X_{2i}=a_1+a_3X_{3i}+a_4X_{4i}+\cdots+a_kX_{ki}+\hat{u}_i$。

[24] George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, 1982, p.621.

[25] Lawrence R. Klein, *An Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, N.J., 1962, p.101.

[26] 特别是参考 D.A. Belsley, E. Kuh, and R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 1980, Chap.3。但此书不是写给初学者阅读的。

[27] See David G. Kleinbaum, Lawrence L. Kupper, and Keith E. Muller, *Applied Regression Analysis and other Multivariate Methods*, 2d ed., PWS-Kent, Boston, Mass., 1988, p.210.

[28] Blanchard, O.J., Comment, *Journal of Business and Economic Statistics*, vol.5, 1967, pp.449－451. 所引内容复制于 Peter Kennedy, *A Guide to Econometrics*, 4th ed., MIT Press, Cambridge, Mass., 1998, p.190。

[29] 对此一个有意思的讨论可参见 Conlisk, J., "When Collinearity is Desirable," *Western Economic Journal*, vol.9, 1971, pp.393－407。

[30] Mark B. Stewart and Kenneth F. Wallis, *Introductory Econometrics*, 2d ed., John Wiley & Sons, A Halstead Press Book, New York, 1981, p.154.

[31] J. Tobin, "A Statistical Demand Function for Food in the U.S.A.," *Journal of the Royal Statistical Society*, Ser. A, 1950, pp.113－141.

[32] 关于数据并用技术的一个透彻的讨论与应用，参见 Edwin Kuh, *Capital Stock Growth: A Micro-Econometric Approach*, North-Holland Publishing Company, Amsterdam, 1963, Chaps. 5 and 6。

[33] 再者，如果 b_{32} 不随样本无限增大而趋于零，则 b_{12} 不仅是有偏误，而且没有一致性。

[34] 感谢艾伯特·朱克(Albert Zucker)提供下述回归结果。

[35] Judge 等：前引文献，第 625 页。参见第 10.9 节。

[36] 如已指出的，因 X，X^2 和 X^3 是非线性关系，故严格地说，多项式回归并不违反经典模型中无多重共线性的假定。

[37] See R. A. Bradley and S. S. Srivastava, "Correlation and Polynomial Regression," *American Statistician*, vol. 33, 1979, pp. 11–14.

[38] See Norman Draper and Harry Smith, *Applied Regression Analysis*, 2d ed., John Wiley & Sons, New York, 1981, pp. 266–274.

[39] 一种具有可读性的、从应用观点说明这些技术的读物是 Samprit Chatterjee and Bertram Price, *Regression Analysis by Example*, John Wiley & Sons, New York, 1977, Chaps ruary 7 and 8。还参阅 H. D. Vinod, "A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares," *Review of Economics and Statistics*, vol. 60, Feb ruary 1978, pp. 121–131。

[40] 参看 R. C. Geary, "Some Results about Relations between Stochastic Variables: A Discussion Document", *Review of International Statistical Institute*, vol. 31, 1963, pp. 163–181。

[41] 贾奇等：前引文献，第 619 页。你还将看到，在此页上证明了为什么尽管有共线性，但如果现有的共线性结构继续存在于未来的样本中，人们就能得到较好的均值预测。

[42] E. Malinvaud, *Statistical Methods of Econometrics*, 2d ed., North-Holland Publishing Company, Amsterdam, 1970, pp. 220–221，有出色的讨论。

[43] J. Johnston, *Econometric Methods*, 3d ed., McGraw-Hill, New York, 1984, p. 249.

[44] Longley, J., "An Appraisal of Least-Squares Programs from the Point of the User," *Journal of the American Statisitical Association*, vol. 62, 1967, pp. 819–841.

[45] X_5 和 X_6 之间的相关系数约为 0.993 9，实际上是很高的相关度。

[46] Goldberger，前引文献，第 250 页。

第 11 章　异方差性：误差方差不是常数会怎么样？

387　经典线性回归模型的一个重要假定（假定 4）是，出现在总体回归函数中的干扰项 u_i 是同方差性的（homoscedastic）；就是说，它们都有相同的方差。本章中，我们分析这一假定的真实性，并探明如果此假定不成立将会出现什么情况。类似于第 10 章，我们寻求下述问题的答案：

1. 异方差性的性质是什么？
2. 它的后果是什么？
3. 怎样去侦察它？
4. 有什么补救措施？

§11.1　异方差的性质

如第 3 章中所指出的，经典线性回归模型的重要假定之一是，以解释变量的选定值为条件的每一干扰项 u_i 的方差是一个等于 σ^2 的常数。这就是同方差性假定。**同方差性**（homoscedasticity）意谓等同的分散程度，亦即相等的方差。用符号表示，

$$E(u_i^2) = \sigma^2 \quad i = 1,2,\cdots,n \tag{11.1.1}$$

从图解上看，双变量回归模型中的同方差性可表示为图 3.4。为方便起见，将该图重制为图 11.1。如图 11.1 所示，以给定 X_i 为条件的 Y_i 的条件方差（等于 u_i 的条件方差），不管变量 X 取什么值，都保持不变。

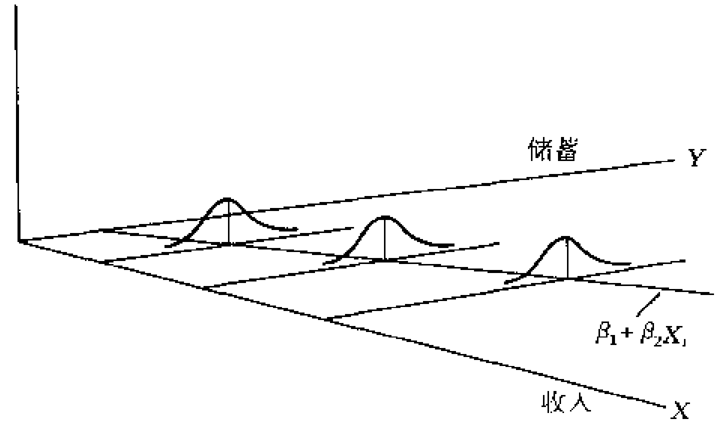


图 11.1　同方差性干扰

与此对照，考虑图 11.2。该图表明，Y_i 的条件方差随 X 增加而增加。这里，Y_i 的方差随 i 而变，从而有异方差性。符号上写为：

$$E(u_i^2) = \sigma_i^2 \tag{11.1.2}$$

注意，σ^2 的下标提醒我们，u_i 的条件方差（$= Y_i$ 的条件方差）不再是常数。

为了看清楚同方差性和异方差性的区别，假定在双变量模型 $Y_i = \beta_1 + \beta_2 X_i + u_i$ 中，Y 代表储蓄和 X 代表收入。图 11.1 和 11.2 都表明随着收入增加，储蓄平均来说也增加。但在图 11.1 中，储蓄的方程在所有的收入水
389 平上都保持不变。而在图 11.2 中，它却随收入增加而增加。看来在图 11.2 中，较高收入的家庭不仅比低收入的家庭平均而言有更多的储蓄，而且在他们的储蓄中有更大的变异。

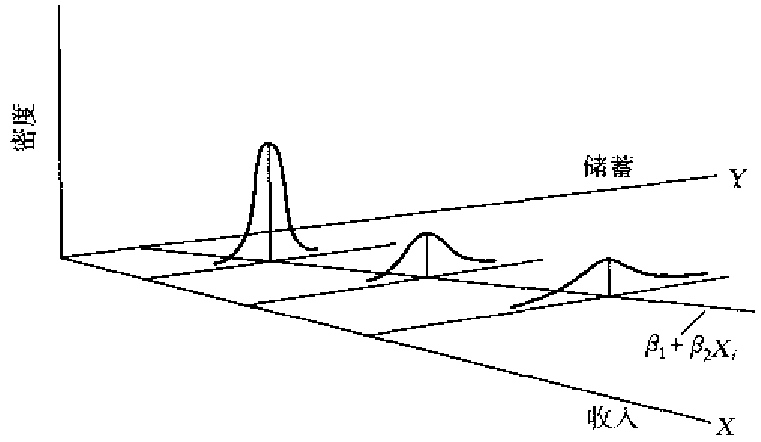


图 11.2　异方差性干扰

有几个理由说明为什么 u_i 的方差可能有变化。其中的一些有如下述[1]：

1. 按照边错边改学习模型，人们在学习的过程中，其行为误差随时间而减少。在这种情形中，预料 σ_i^2 会减小。作为一个例子，考虑图 11.3。该图描写一次测验；在给定的一段时间里，打字出错个数与用于打字练习的小时数的关系，如图所示。随着打字练习小时数增加，不仅平均打错个数而且打错个数的方差都有所下降。

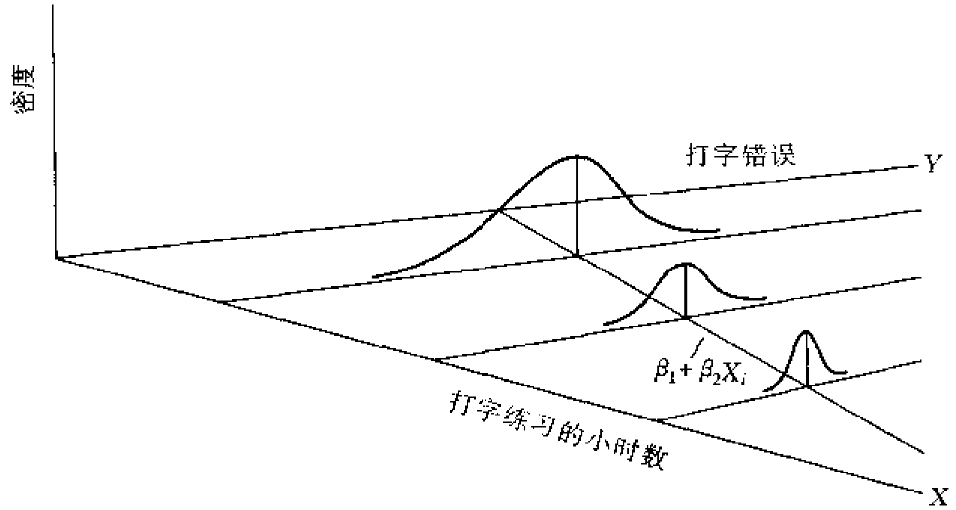


图 11.3　异方差性展示

2. 随着收入增长，人们有更多的备用收入（discretionary income）[2]，从而如何支配他们的收入有更大的选择范围。因此，在做储蓄对收入的回归时，很可能发现，由于人们对其储蓄行为有更多的选择，σ_i^2 与收入同时增加（如图 11.2 所示）。同理，利润较丰厚的公司在分红政策方面和利润微薄的公司相比，一般均可预料有较大的变化。而且以增长为导向的公司比之已发展定型的公司，在红利支付方面也可能表现更多的变异。

3. 随着数据采集技术的改进，σ_i^2 可能减小。例如，有精巧数据处理装
390 备的银行，在它们对账户的每月或每季收支说明书中，比之于没有这种装备的银行，会出现更少的差错。

4. 异方差性还会因为**异常值**（outliers）的出现而产生。一个超越正常范围的观测值或称异常值，是指和其他观测值相比相差很多（非常小或非常大）的观测值。包括或不包括这样的一个观测值，尤其是样本较小时，会在很大程度上改变回归分析的结果。作为一个例子，考虑图 11.4 中的散点图。此图根据习题 11.22 所给数据，对 20 个国家在第二次世界大战后直至 1969 年期间的股票价格（Y）和消费者价格（X）的百分率变化进行描点。图中，对智利的观测值 Y 和 X，远大于对其他国家的观测值，故可视为一个异常值。类似于这种情况，同方差性的假定就难以维持了。在习题 11.22 中，我们要求读者考虑，如果在分析中把对智利的观测值除掉，会出现怎样的回归结果。

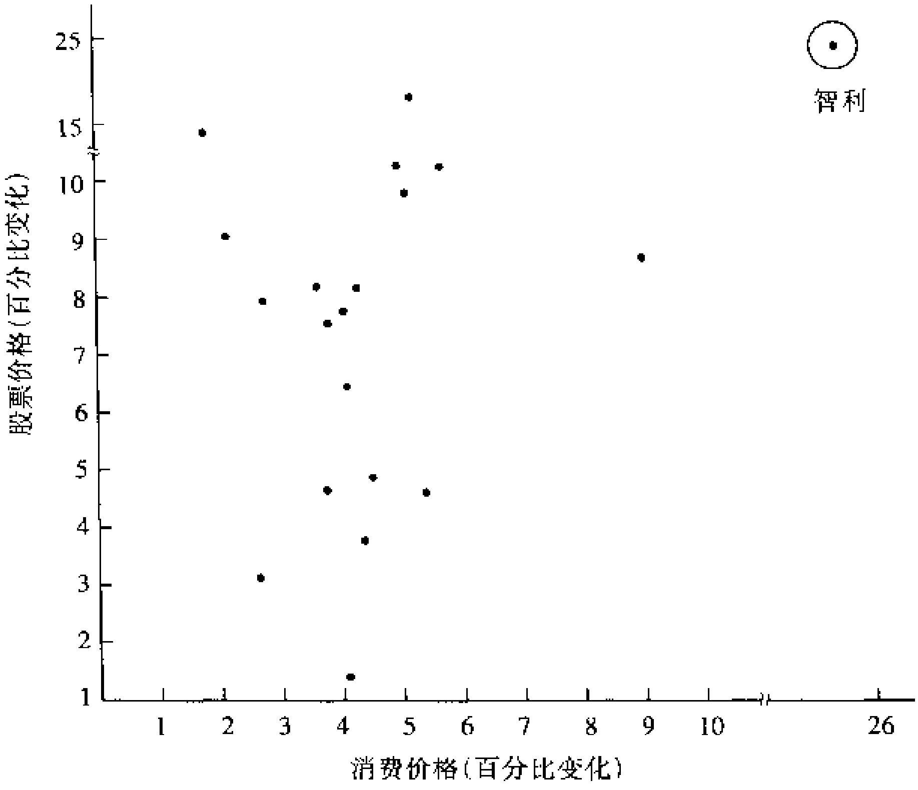
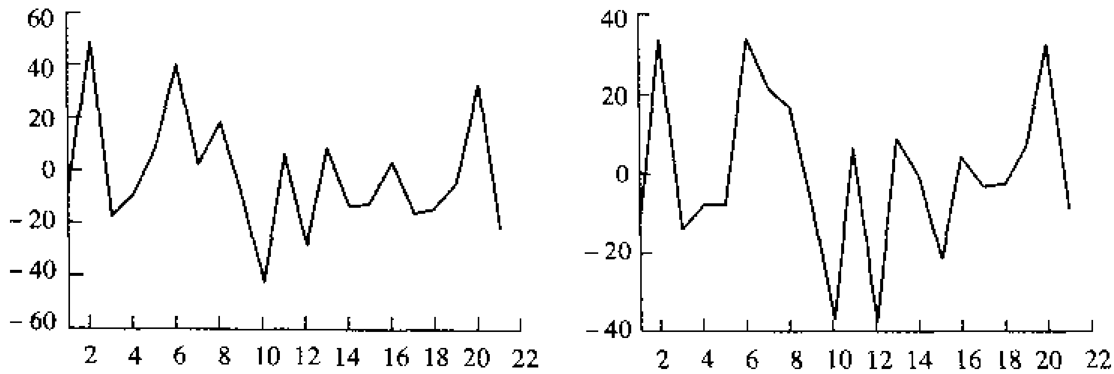


图 11.4　股票价格与消费价格的关系

391　5. 异方差性的另一来源来自 CLRM 的假定 9 的破坏，即回归模型的设定是不正确的。虽然我们将在第 13 章中对设定误差的问题做更全面的探讨，但常常看来像是异方差性的问题，其实是由于模型中的一些重要变量被忽略了。例如，在一个对商品的需求函数中，如果没有把有关的互补和（或）替代商品的价格包括进来（忽略变量偏误），则回归残差可能给人以异方差的表面印象；而当模型把所忽略的变量包括进来时，这种印象也许会消失。

作为一个简明的例子，回想我们对广告印象（Y）与广告支出（X）之关系的研究。(参见习题 8.32。) 若只将 Y 对 X 回归并观测此回归的残差，你会看到一种类型，但若将 Y 对 X 和 X^2 回归，你又会看到另一种类型，从图 11.5 中明显可以看出这一点。我们已经看到 X^2 属于此模型。(参见习题 8.32。)



(a) 广告印象对广告支出进行回归的残差　(b) 广告印象对 Adexp 和 $\mathbf{Adexp^2}$ 进行回归的残差

图 11.5

6. 异方差性的另一个来源是模型中一个或多个回归元的分布偏态(skewness)。诸如收入、财富和教育等经济变量都是很好的例子。众所周知，大多数社会中收入和财富的分配都是不匀称的，处在顶端的少数几个人拥有大部分的收入和财富。

7. 异方差性其他来源：如戴维·亨德里（David Hendry）所注意到的那样，由于（1）不正确的数据变形（如比率或一阶差分变换等）和（2）不正确的函数形式（如线性与对数线性模型的变换），同样能导致异方差性。[4]

注意，异方差性问题在横截面数据中比在时间序列数据中更为常见。在横截面数据中，人们通常在一个给定的时间点上对总体中的一些成员进行观

392 测，例如对个别的消费者或家庭、厂商、工业或地区（如州、农村或城市）等进行观测。而且，这些成员可能大小不一，例如厂家有大、中、小之分，收入有高、中、低之分。而另一方面，在时间序列数据中，人们经常收集同一实体在一个时期内的数据，例如美国在1950—2000年期间的GNP、消费支出、储蓄或就业数据。

作为横截面分析中常会遇到的异方差性的一个说明，且考虑表11.1。

表 11.1　按厂家职工人数划分的非耐用品制造行业的人均工薪，1958 年

行业	就业人数（平均职工人数）								
	1～4	5～9	10～19	20～49	50～99	100～249	250～499	500～999	1 000～2 499
食品干果	2 994	3 295	3 565	3 907	4 189	4 486	4 676	4 968	5 342
烟草产品	1 721	2 057	3 336	3 320	2 980	2 848	3 072	2 969	3 822
纺织品	3 600	3 657	3 674	3 437	3 340	3 334	3 225	3 163	3 168
器皿用具	3 494	3 787	3 533	3 215	3 030	2 834	2 750	2 967	3 453
纸张类	3 498	3 847	3 913	4 135	4 445	4 885	5 132	5 342	5 326
印刷与出版	3 611	4 206	4 695	5 083	5 301	5 269	5 182	5 395	5 552
化工产品	3 875	4 660	4 930	5 005	5 114	5 248	5 630	5 870	5 876
石油与煤炭	4 616	5 181	5 317	5 337	5 421	5 710	6 316	6 455	6 347
橡胶与塑料	3 538	3 984	4 014	4 287	4 221	4 539	4 721	4 905	5 481
皮革与皮革制品	3 016	3 196	3 149	3 317	3 414	3 254	3 177	3 346	4 067
平均工薪	3 396	3 787	4 013	4 014	4 146	4 241	4 388	4 538	4 843
标准差	743.7	851.4	727.8	805.06	929.9	1 080.6	1 243.2	1 307.7	1 112.5
平均生产力	9 355	8 584	7 962	8 275	8 389	9 418	9 795	10 281	11 750

资料来源：*The Census of Manufacturers*, U. S. Department of Commerce, 1958.（由本书作者计算。）

该表给出1958年按厂家或企业就业职工人数划分的10个非耐用品制造行业的每个职工工薪数据。表中还给出9个按职工人数分组的平均生产力数字。

393 虽然不同工业有不同的产出构成，但表 11.1 清楚地表明，平均而言，大的厂家比小的厂家平均支付更多的工薪。例如，职工人数在 1～4 人的厂家平均支付约3 396美元，而职工人数在1 000～2 499人的厂家平均支付约4 843美元。但应注意，在不同就业人数的类别之间，如估计的工资收入的标准差所表明的，工资收入上有相当大的变异性。这点还可以从图 11.6 看出，图 11.6 描出了每个职工人数组中薪金的标准差和平均薪金。我们清楚地看到，总体上，薪金的标准差随着薪金平均值的提高而提高。

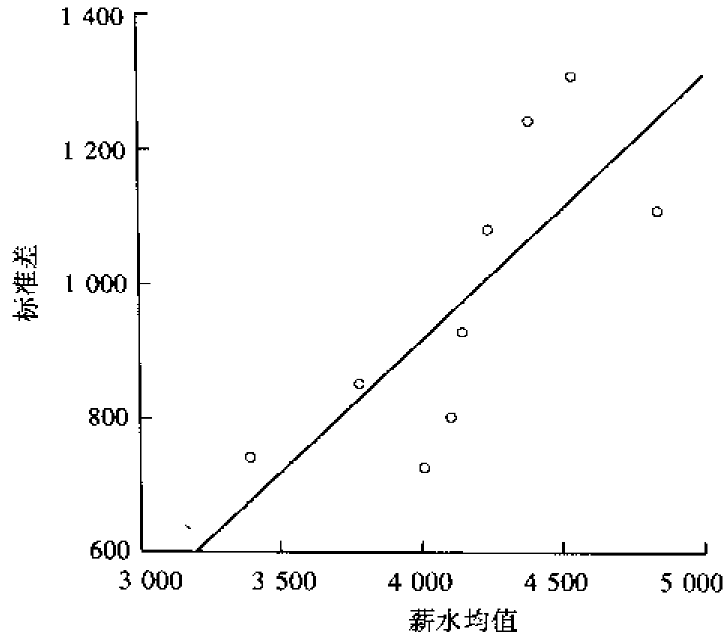


图 11.6 薪金的标准差与均值

§ 11.2 出现异方差性时的 OLS 估计

如果引进异方差性 $E(u_i^2)=\sigma_i^2$ 而保留经典模型的所有其他假定，OLS 估计量及其方差会出现什么变化？为了回答此问题，让我们回到双变量模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

按照惯常的公式，β_2 的 OLS 估计量是：

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}\end{aligned} \tag{11.2.1}$$

394 但现在它的方差是(参看附录 11A，第 11A.1 节)：

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sum x_i^2\sigma_i^2}{(\sum x_i^2)^2} \tag{11.2.2}$$

这显然不同于同方差性假定下的常用方差公式：

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_i^2} \tag{11.2.3}$$

当然，如果对每个 i 都有 $\sigma_i^2=\sigma^2$，那么这两个公式是相同的。(为什么?)

回顾一下，如果经典模型的诸假定，包括同方差性在内，全部成立，则 $\hat{\beta}_2$ 是最优线性无偏估计量。那么，当我们仅除掉同方差性假定而代之以异方差性的假定时，它还会不会是 BLUE 呢？容易证明，$\hat{\beta}_2$ 仍是线性和无偏的。事实上，如附录 3A，3A.2 节所表明的，为了确立 $\hat{\beta}_2$ 的无偏性，干扰项（u_i）的同方差性并不是必要的。的确，u_i 的方差，同方差或异方差，与无偏性的确定无关。记得在附录 3A 的 3A.7 节中，我们证明过，在经典线性回归模型的假定之下，$\hat{\beta}_2$ 是一个一致估计量。尽管我们不去证明它，但我们可以证明，$\hat{\beta}_2$ 在异方差情形下是一个一致估计量；即随着样本容量无限扩大，估计的 β_2 收敛于其真实值。而且，还可以证明，在一定的条件(被称为正则性条件）下，$\hat{\beta}_2$ 还是渐近正态分布的。当然，上述结论对多元回归模型中的其他参数仍成立。

认定 $\hat{\beta}_2$ 是线性无偏的，那它是不是“有效”或“最优”的，即是否在线性无偏估计量一类中有最小方差呢？并且这个最小方差由方程（11.2.2）给出？对两个问题的回答都是否定的：$\hat{\beta}_2$ 不再是最优的，而且最小方差也不由（11.2.2）给出。那么，在异方差性出现时，什么才是 BLUE？下节给出这个回答。

§11.3 广义最小二乘法

为什么（11.2.1）所给的 β_2 的常用 OLS 估计量虽然无偏但不最优呢？直观的理由可从图 11.5 看出。如图所示，各就业组之间的工薪收入有相当大的变异。假如要我们做每个职工的酬金对就业人数的回归，我们就应对酬金的这种组间变异知识加以利用。最理想的，是设计出这样一种估计方案：
395 对来自变异较大的总体的观测值做较小的加权，而对来自较小的总体的观测值做较大的加权。检查一下图 11.5，就知道，要对来自如同 10～19 人和20～49 人的就业组的观测值比对来自如同 5～9 人和 250～499 人的那些观测值做更大的加权。因为前一种观测值比较紧密地聚集在它们的均值周围，从而能使我们更准确地估计 PRF。

可惜的是，常用的 OLS 方法并不采取这种策略，因而对表 11.1 中的职工酬金这个因变量 Y 的不等变异所含的信息未加利用：它仍然对每一观测值同样重视或同等加权。而名为**广义最小二乘**（GLS）的一种估计方法则明确地利用了这一信息，因而能产生 BLUE 估计量。为了看清楚怎样做到这一点，让我们继续利用现在已经熟悉的双变量模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{11.3.1}$$

为便于代数上的处理，我们把上述模型写为：

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \tag{11.3.2}$$

其中对每个 i 均有 $X_{0i} = 1$。谅读者能看出这两种写法是完全相同的。

现在假定相异的方差 σ_i^2 已知。用 σ_i 通除（11.3.2）得：

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \tag{11.3.3}$$

为了易于阐述，将它写为：

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \tag{11.3.4}$$

其中带星号变量或转换变量为原始变量除以（已知的）σ_i。我们用符号 β_1^* 和 β_2^* 表示转换模型的参数，以区别于常用的 OLS 参数 β_1 和 β_2。

转换原始模型的用意何在？为说明这点，可注意变换误差项 u_i^* 的如下特点：

$$\begin{aligned}
\operatorname{var}(u_i^*) &= E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 \\
&= \frac{1}{\sigma_i^2}E(u_i^2) \quad \text{因为知道 } \sigma_i^2 \\
&= \frac{1}{\sigma_i^2}(\sigma_i^2) \quad \text{因为 } E(u_i^2) = \sigma_i^2 \\
&= 1
\end{aligned} \tag{11.3.5}$$

396 这是一常数。就是说，转换干扰项 u_i^* 的方差，现在有了同方差性。因为我们仍保留着经典模型的其他假定，所以 u_i^* 的这一同方差性的发现表明，如果我们把 OLS 应用到转换模型（11.3.3）上，将产生 BLUE 估计量。简言之，这时估计出来的 β_1^* 和 β_2^* 是 BLUE，而 OLS 估计量 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 则不是。

先将原始变量转换成满足经典模型假设的转换变量，然后对它们使用 OLS 程序，叫做广义最小二乘法。概括地说，**GLS 是对满足标准最小二乘假定的转换变量的 OLS**。如此得到的估计量叫做 **GLS 估计量**。这些估计量是 BLUE。

估计 β_1^* 和 β_2^* 的具体步骤如下：首先写下对应于（11.3.3）的 SRF：

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{\hat{u}_i}{\sigma_i}\right)$$

或，

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{11.3.6}$$

然后最小化：

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

即：

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i}\right)\right]^2 \tag{11.3.7}$$

以获得 GLS 估计量。附录 11A，11A.2 节给出最小化（11.3.7）的标准计算程序。如该节所表明的，β_2^* 的 GLS 估计量为：

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \tag{11.3.8}$$

它的方差为：

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \tag{11.3.9}$$

其中 $w_i = 1/\sigma_i^2$。

OLS 和 GLS 的差别

397 从第 3 章中看到，OLS 要求我们最小化：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{11.3.10}$$

而 GLS 要求我们最小化表达式（11.3.7），而该式又可写为：

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2 \tag{11.3.11}$$

其中 $w_i = 1/\sigma_i^2$［读者可证实（11.3.11）和（11.3.7）相同］。

可见，在 GLS 中，我们最小化以 $w_i = 1/\sigma_i^2$ 为权的一个加权残差平方和，而在 OLS 中我们最小化一个无权的或等权的残差平方和。如（11.3.7）所表明的，分配给每一观测值的权与它的 σ_i 成反比，即在最小化（11.3.11）的 RSS 的过程中，来自有较大 σ_i 的总体的观测值将得到较小的加权，而来自有较小 σ_i 的总体的观测值将得到较大的加权。为了看清楚 OLS 和 GLS 的差别，且考虑图 11.7 中这个假想的散点图。

在（无权的）OLS 中，点 A、B 和 C 处的误差在 RSS 的最小化过程中都得到相等的加权。显而易见，这时 C 的误差将支配 RSS。但在 GLS 中，这

个极端的观测值 C，和另外两个观测值相比，将得到相对小的加权。如前所
398 说，这个策略是正确的，因为为了更可靠地估计总体回归函数，我们应该给那些紧密围绕其（总体）均值的观测值比给那些远离均值的观测值以更大的加权。

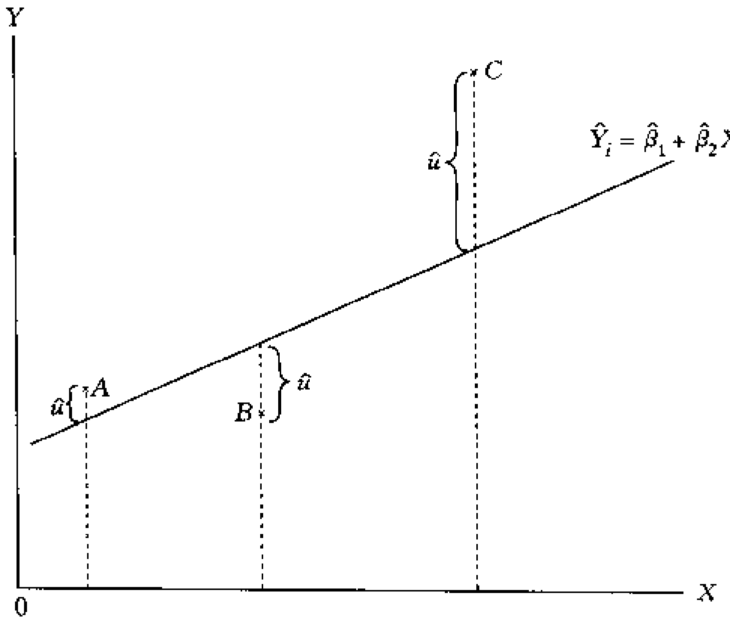


图 11.7　假想的散点图

因（11.3.11）最小化一个加权的 RSS，故适宜于称它为**加权最小二乘**（weighted least squares，WLS），并把由此得到的、并由（11.3.8）和（11.3.9）给出的估计量称为 **WLS 估计量**。但 WLS 只不过是更为一般的估计方法 GLS 的一种特殊情形。在异方差性的讨论中 WLS 和 GLS 两词可交换地使用。但在以后的章节中，我们会遇到 GLS 的其他特例。

顺便指出，如果对所有的 i，$w_i = w$，即为一常数，则 $\hat{\beta}_2^*$ 雷同于 $\hat{\beta}_2$，并且 $\mathrm{var}(\hat{\beta}_2^*)$ 也雷同于由（11.2.3）所给的常见的（即同方差性的）$\mathrm{var}(\hat{\beta}_2)$。这点是没有什么奇怪的。（为什么？）（参看习题 11.8。）

§11.4　出现异方差性时使用 OLS 的后果

正如我们已经看到的，$\hat{\beta}_2^*$ 和 $\hat{\beta}_2$ 两者都是（线性）无偏估计量：在重复抽样中，平均而言，$\hat{\beta}_2^*$ 和 $\hat{\beta}_2$ 都将等于真实 β_2，就是说，它们都是无偏估计量。但我们知道 $\hat{\beta}_2^*$ 才是有效的，即有最小方差的。那么，如果我们继续使用 OLS 估计量 $\hat{\beta}_2$，我们的置信区间、假设检验以及其他相关事宜会出现什么情况？我们分两种情形讨论。

考虑异方差性的 OLS 估计

假如我们使用 $\hat{\beta}_2$，但又明显考虑有异方差性而使用由（11.2.2）给出的方差公式。那么，按照该公式，假定 σ_i^2 为已知，是否就可利用平常的 t 和 F 检验建立置信区间并检验假设呢？一般地说，回答是否定的。因为，可以证明 $\mathrm{var}(\beta_2^*)\leqslant\mathrm{var}(\hat{\beta}_2)$。[5] 这就是说，根据 $\mathrm{var}(\hat{\beta}_2)$ 做出的置信区间将是无谓的过大。其结果是，t 和 F 检验很可能给我们提供了不准确的结果：因为明显过大的 $\mathrm{var}(\hat{\beta}_2)$ 会使本来（如果我们使用由 GLS 程序建立的正确的置信区间的话）显著的系数变成了统计上不显著的（因 t 值过小）。（之所以做这种比较，似是因为 $\hat{\beta}_2$ 无偏，但这种比较毕竟欠妥。——译者注）

忽视异方差性的 OLS 估计

如果在有或怀疑有异方差性的情形下，我们不但使用了 $\hat{\beta}_2$，而且继续使用（11.2.3）所给的常用（同方差性的）方差公式，情况就变得严重了：注意，这是我们所讨论的两种情形中尤为常见的一种。原因是：当我们使用
399 标准的 OLS 回归软件包时，忽略异方差性（或对异方差性无知）就会给出 $\hat{\beta}^2$ 的方差，一如（11.2.3）所给的。首先，（11.2.3）所给 $\mathrm{var}(\hat{\beta}_2)$ 是（11.2.2）所给 $\mathrm{var}(\hat{\beta}_2)$ 的一个有偏误的估计量。就是说，平均而言，前者不是高估就是低估了后者。而且，一般地说，我们无法告知这个偏误是正的（过高估计）还是负的（过低估计）。可从（11.2.2）清楚地看到，它依赖于 σ_i^2 的变化与解释变量 X 的取值之间的关系（参看习题 11.9），之所以有偏误，是因为当异方差性出现时，惯用的作为 σ^2 的估计量 $\hat{\sigma}^2=\sum\hat{u}_i^2/(n-2)$ 就不是 σ^2 的无偏估计量（注意，这时 $\sigma_i^2\neq\sigma^2$。——译者注）。忽视异方差性的结果是，我们不能再依赖通常计算的置信区间和通常使用的 t 和 F 检验。[6] **总之，如果我们忽视异方差性而一味使用惯常的检验程序，则无论我们得出什么结论或做出什么推断，都可能产生严重的误导。**

为使问题的讨论更加明朗，我们引用戴维森和麦金农所做的一个**蒙特卡罗实验**。[7] 他们考虑一个简单的模型，可用我们的符号表示如下：

$$Y_i=\beta_1+\beta_2X_i+u_i \tag{11.4.1}$$

他们假定 $\beta_1=1$，$\beta_2=1$ 和 $u_i\sim N(0,X_i^\alpha)$。最后的一个式子表明误差方差是异方差性的，并且它的值是回归元 X 值的 α 次方。例如，当 $\alpha=1$ 时，误差方差与 X 值成正比；当 $\alpha=2$ 时，误差方差与 X 值的平方成正比，等等。在 11.6 节中我们将讨论这种比例关系的逻辑性。根据 20 000 次重复实验，并令 α 取不同的值，他们得到使用 OLS［见方程（11.2.3）］和使用 OLS 但

考虑到有异方差性［见方程（11.2.2)］以及使用 GLS［见方程（11.3.9)］的两个回归系数的标准误。现对所选的一些 α 值，把它们的结果列表如下。

400 这些结果的最令人注目的特点是，不管是否考虑对异方差性的修正，OLS 一致地过高估计了由（正确的）GLS 程序得到的真实标准误，尤其以大的 α 值为甚，从而确立了 GLS 的优越性。这些结果还表明，如果我们不用 GLS 而只用 OLS，不管是否考虑到异方差性，情况都不是清晰的。常用的 OLS，相对于顾及异方差性的 OLS 来说，其标准误或者偏之于过大（对截距而言)，或者一般偏之于过小（对斜率系数而言)。一条明显的信息是：在出现有异方差性时要用 GLS。然而，在实际中，GLS 并不总是容易使用的，理由见后。同样，我们以后还要提到，除非异方差性很严重，否则，可能不应该放弃 OLS 而支持 GLS 或 WLS。

α 值	$\hat{\beta}_1$ 的标准误			$\hat{\beta}_2$ 的标准误		
	OLS	OLS_{het}	GLS	OLS	OLS_{het}	GLS
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.007 3	0.200	0.220	0.109
3.0	0.100	0.064	0.001 3	0.173	0.206	0.056
4.0	0.089	0.059	0.000 3	0.154	0.195	0.017

注：OLS_{het}表示考虑异方差性的 OLS。

由上讨论可见，异方差性是一个潜在的严重问题，研究者需要知道它在某一给定的情况中是否出现。如果发现有异方差性，就可采取纠正步骤，诸如使用加权最小二乘回归或某些其他技术。然而，在我们考虑各种纠正措施之前，我们必先判知在某一给定情况中，是否有或很可能有异方差性。这就是下节中要讨论的问题。

一个技术性注解

尽管我们曾经说过，在异方差情形中，GLS 是 BLUE，而 OLS 不是，但在有些例子中 OLS 在异方差情况下仍是 BLUE。[18] 只是这种例子在实践中并不多见。

§11.5 异方差性的侦察

和多重共线性一样，一个重要的实际问题是：我们怎样知道在一个具体的情况中是否有异方差性？而且也和多重共线性相类似，并不存在有侦破异方差性的严明的法则，只有少数的经验规则。但是这种结局是不可避免的，因为除非我们知道对应于选定 X 值的整个 Y 总体，如同表 2.1 或表 11.1 所
401 给的总体那样，否则 σ_i^2 是无从获知的。然而，在经济研究中这样的数据（总体）照例是得不到的，除非是例外。在这方面，计量经济学家不同于诸如农学或生物学等领域的科学家。农学或生物学的研究者们能很好地控制他们的研究主题。而在经济研究中，对应于一个具体的 X 值，多数情形都只有一个样本 Y 值。所以没有任何方法能从仅仅一个 Y 观测值去获知 σ_i^2。因此在大多数的计量经济调查研究中，异方差性不过是一种直觉，深思熟虑的猜测，先前经验或纯粹猜想。

有了上述告诫，现在就可以列举一些非正式或正式的侦察异方差性的方法。如下面的讨论将要表明的，大多数的方法都是基于对我们所能观测到的 OLS 残差 $\hat{u}_i$ 的分析，而不是对干扰项 u_i 的分析。我们寄希望于 $\hat{u}_i$ 是 u_i 的良好估计。当样本较大时，这一希望也许能兑现。

非正式方法

问题的性质。 往往根据所考虑问题的性质就能判别是否会遇到异方差性。例如，普雷斯（Prais）和霍撒克（Houthakker）在一项家庭预算研究中发现，围绕消费对收入的回归，残差的方差随收入增加而增加。仿效这一开拓性的工作，现在人们一般都假定在类似的调查中可以预期不同干扰之间有不相等的方差。[9] 事实上，在涉及不均匀单元的横截面数据中，异方差性可能是一种常规而不是例外。例如，在投资与销售量、利率等关系的横截面分析中，如果样本同时含有小、中和大型厂家，一般都预期有异方差性。

事实上，我们已经遇到过这种例子。我们在第 2 章讨论了美国小时工资均值与受教育年数的关系。而且，我们在那里还讨论了印度 55 个家庭的食物支出与总支出之间的关系（见习题 11.16）。

图解法。 如果对异方差性的性质没有任何先验或经验信息，实际上，可先在无异方差性的假定下做回归分析，然后对残差的平方 $\hat{u}_i^2$ 做一事后检查，看看这些 $\hat{u}_i^2$ 是否呈现任何系统性的样式。虽然 $\hat{u}_i^2$ 还不等于 u_i^2，但可作为
402 一种替代变量，特别是当样本含量足够大时。[10] 对 $\hat{u}_i^2$ 的检查可能出现诸如图 11.8 中的那些样式。

在图 11.8 中，$\hat{u}_i^2$ 是对应于 $\hat{Y}_i$ 而描绘的，$\hat{Y}_i$ 是从回归线读出的 Y_i 的估计值，其用意是要找出 Y 的估计均值是否与平方残差有任何系统联系。在图 11.8a 中，我们未发现这两个变量之间有任何系统性样式，表明了数据中也许没有异方差性。图 11.8b 到 e 则呈现一定的样式。例如，图 11.8c 表示 $\hat{u}_i^2$ 与 $\hat{Y}_i$ 之间的一个线性关系，而图 11.8d 和 e 则表示二次关系。利用这种虽然是非正式的知识，我们却有可能把数据转换成不表现有异方差性的（转换）数据。在 11.6 节中，我们将分析几种这样的转换。

除了将 $\hat{u}_i^2$ 对 $\hat{Y}_i$ 描点外，还可将它们对解释变量之一描点，特别是对 $\hat{Y}_i$ 描点的结果，像图 11.8a 那样看不出什么异方差性。如图 11.9 所示，对 X 描点的结果，会显示出类似于图 11.8 的图样。（在双变量模型的情形中，将 $\hat{u}_i^2$ 对 $\hat{Y}_i$ 描点等效于将它对 X_i 描点。因此，图 11.9 和图 11.8 必然是相似的。但当我们考虑两个或多个 X 变量的模型时，情况就不同了；这时可将 $\hat{u}_i^2$ 相对于模型中的任一个 X 变量描点。）

403

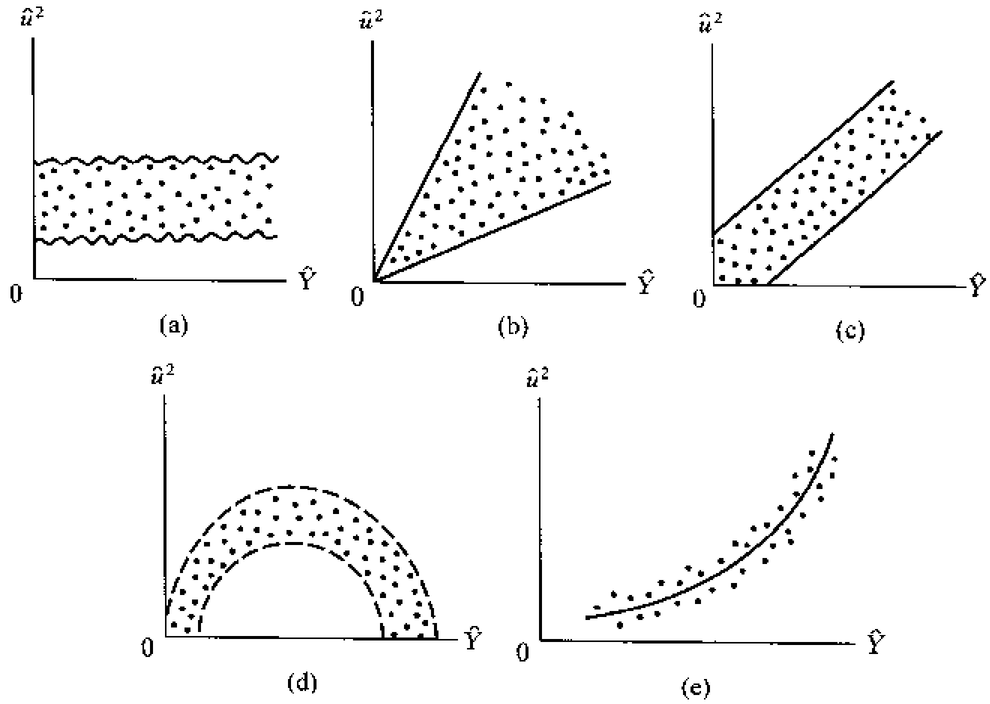


图 11.8　估计平方残差的假想图样

一个类似于图 11.9 的图形可能表明干扰项的方差与 X 变量有线性关系。因此，如果在储蓄对收入的回归中，发现有如同图 11.9c 那样的图样，就表明相异的方差可能正比于收入变量的取值。这一知识有助于我们将数据转换，使得对转换后的数据做回归时，干扰项的方差变成了同方差性的。下节我们还将回到这个问题上。

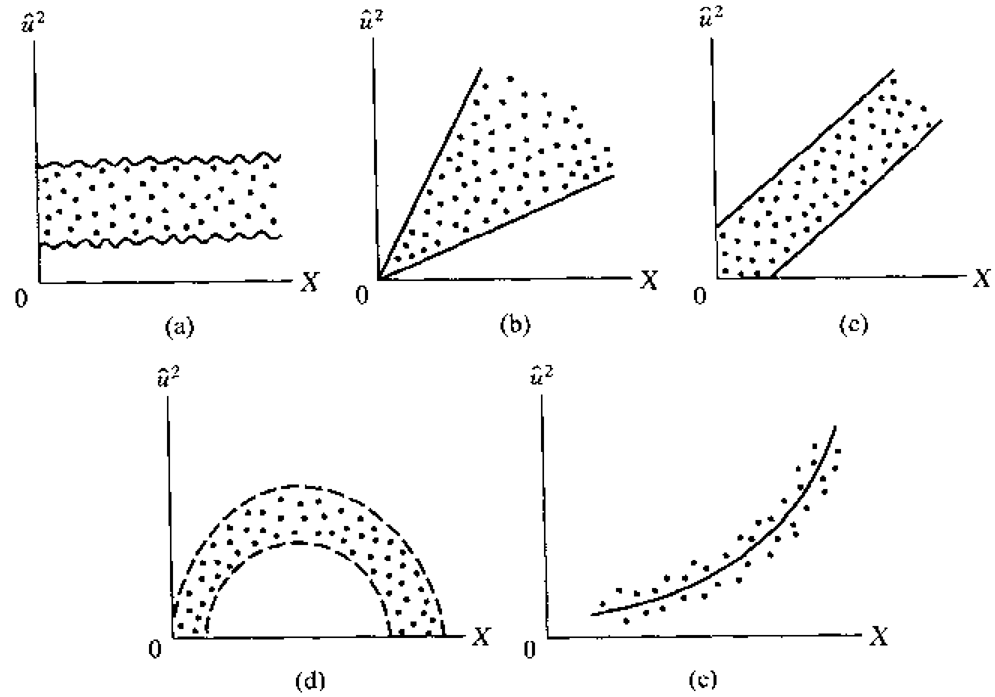


图 11.9　对应于 X 的估计平方残差散点图

正式方法

404　**帕克检验**。[11]帕克提出 σ_i^2 是解释变量 X_i 的某个函数，从而把图解法公式化。他所建议的函数形式是：

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{v_i}$$

或：

$$\ln\sigma_i^2 = \ln\sigma^2 + \beta\ln X_i + v_i \tag{11.5.1}$$

其中 v_i 是随机干扰项。

由于 σ_i^2 通常是未知的，帕克建议用 $\hat{u}_i^2$ 作为替代变量并做如下的回归：

$$\begin{aligned}\ln\hat{u}_i^2 &= \ln\sigma^2 + \beta\ln X_i + v_i \\ &= \alpha + \beta\ln X_i + v_i \end{aligned} \tag{11.5.2}$$

如果 β 表现为统计上显著的，就表明数据中有异方差性。如果它不显著，则可接受同方差性假设。可见，帕克检验是一个两阶段程序。在第一阶段中，我们做 OLS 回归，而不考虑异方差性问题。我们从这一回归获得 $\hat{u}_i$，然后在第二阶段中做回归（11.5.2）。

虽然帕克检验从经验上看颇有魅力，却遇到一些问题，戈德菲尔德和匡特曾称辩说，进入（11.5.2）的误差项 v_i 可能不满足 OLS 假设而且本身还可能是异方差的。[12]然而作为一个纯粹探索性的方法，帕克检验还是可以使

用的。

例 11.1 酬金与生产力的关系。 为说明帕克方法，我们利用表 11.1 的数据做如下回归：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

其中 Y = 以千美元计的平均酬金，X = 以千美元计的平均生产力，i = 第 i（类）企业就业人数。回归的结果如下：

$$\begin{aligned}\hat{Y}_i &= 1\,992.345\,2 \quad + \quad 0.232\,9X_i \\ \text{se} &= (936.479\,1) \quad (0.099\,8) \\ t &= \quad (2.127\,5) \quad (2.333) \qquad R^2 = 0.437\,5\end{aligned} \tag{11.5.3}$$

结果表明所估斜率系数在单尾 t 检验的基础上达到 5% 的显著水平。方程表示劳动生产力每增加 1 美元，劳动报酬平均约增加 23 美分。

将得自回归（11.5.3）的残差用于方程（11.5.2）中对 X_i 的回归，给出如下结果：

$$\begin{aligned}\ln\hat{u}_i^2 &= \ 35.817 \quad - \quad 2.809\,9\ln X_i \\ \text{se} &= (38.319) \quad (4.216) \\ t &= \ (0.934) \ (-0.667) \qquad R^2 = 0.059\,5\end{aligned} \tag{11.5.4}$$

显然，两个变量之间无统计上的显著关系。按照帕克检验，便可下结论说，在误差的方差中没有异方差性。[13]

405 **格莱泽（Glejser）检验。**[14]格莱泽检验在观念上类似于帕克检验。格莱泽建议，在从 OLS 回归取得残差 $\hat{u}_i$ 之后，用 $\hat{u}_i$ 的绝对值对被认为与 σ_i^2 密切相关的 X 变量做回归。在他的实验中，他使用以下的多种函数形式：

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

其中 v_i 是误差项。

格莱泽检验仍然可作为一种经验或实际处理方法加以使用。但戈德菲尔德和匡特指出误差项的若干问题，如非零的期望值，序列相关性以及有讽刺意味的异方差性。[15]格莱泽方法的另一困难是，像

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i \text{ 和 } \quad |\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

这样的模型对参数而言是非线性的，因而不能用平常的 OLS 程序加以估计。

格莱泽发现，在异方差性的侦察中，上列模型中的前四个，对大样本来说，一般都能给出令人满意的结果，因此，从实际方面考虑，格莱泽技术可用于大样本。而在小样本中，则仅可作为异方差性摸索的一种定性技巧。关于格莱泽方法的一个应用，见 11.7 节。

例 11.2　薪金与生产力的关系：格莱泽检验

继续例 11.1，将回归（11.5.3）所得到残差的绝对值对平均生产力（X）回归，得到如下结论：

$$
\begin{aligned}
|\hat{u}_i| &= 407.2783 - 0.0203X_i \\
\text{se} &= (633.1621) \quad (0.0675) \qquad r^2 = 0.0127 \\
t &= (0.6432) \quad (-0.3012)
\end{aligned}
\tag{11.5.5}
$$

诚如你从此回归中所见，残差的绝对值与回归元平均生产力之间没有关系，这就加强了基于帕克检验所得到的结论。

406 **斯皮尔曼的等级相关经验。**在习题 3.8 中，我们曾定义斯皮尔曼的等级相关系数为：

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)}\right] \tag{11.5.6}$$

其中 d_i = 第 i 单元或现象的两种不同特性所处的等级之差，而 n = 带有级别的单元或现象的个数。上述等级相关系数可按下述方法用于侦察异方差性：假定 $Y_i = \beta_0 + \beta_1 X_i + u_i$。

步骤 1. 对 Y 和 X 的数据做回归拟合并求出残差 $\hat{u}_i$。

步骤 2. 忽视 $\hat{u}_i$ 的符号，也就是取其绝对值 $|\hat{u}_i|$，同时将 $|\hat{u}_i|$ 和 X_i（或 $\hat{Y}_i$）按递升或递降次序划等级，然后计算上述斯皮尔曼的等级相关系数。

步骤 3. 假定总体等级相关系数 ρ_s 为零且 $n>8$，样本 r_s 的显著性可通过 t 检验按下述方法加以检验[16]：

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \tag{11.5.7}$$

其自由度 df $= n-2$。

407 如果计算的 t 值超过临界 t 值就可接受异方差性假设；否则拒绝。如果回归模型涉及多于一个 X 变量，则可在 $|\hat{u}_i|$ 与每一 X 变量之间分别地计算 r_s，再用方程（11.5.7）中的 t 检验做统计显著性检验。

例 11.3　等级相关检验的说明。为说明等级相关检验，考虑表 11.2 中的数据。这些数据包含了 10 个共同基金的平均年回报（E,%）及其标准差（σ_i,%）。

投资组合理论中的资本市场线假定期望收益（E_i）和风险（用标准差 σ 来度量）之间有如下线性关系：

$$E_i = \beta_i + \beta_2 \sigma_i$$

利用表 11.2 中的数据，估计上述模型，并从中计算出残差。由于数据涉及规模与投资目标都不相同的 10 个共同（互助）基金，人们会先验地预料着异方差性。为检验此假设，现利用等级相关技术。必要的计算也见于表 11.2。

应用公式（11.5.6）得：

$$\begin{aligned} r_s &= 1 - 6\frac{110}{10\ (100-1)} \\ &= 0.333\ 3 \end{aligned} \tag{11.5.8}$$

表 11.2　　**异方差性的等级相关检验**

共同基金名称	E_i，平均年回报 %	σ_i，年回报标准差，%	$\hat{E}_i^*$	$\lvert \hat{u}_i \rvert^{+}$ 残差 $\lvert (E_i - \hat{E}_i) \rvert$	$\lvert \hat{u}_i \rvert$ 的等级	σ_i 的等级	d，两等级之差	d^2
波士顿基金	12.4	12.1	11.37	1.03	9	4	5	25
特拉华基金	14.4	21.4	15.64	1.24	10	9	1	1
权益基金	14.6	18.7	14.40	0.20	4	7	−3	9
基本投资基金	16.0	21.7	15.78	0.22	5	10	−5	25
互有投资基金	11.3	12.5	11.56	0.26	6	5	1	1
卢米斯销售互有投资基金	10.0	10.4	10.59	0.59	7	2	5	25
麻省信赖投资基金	16.2	20.8	15.37	0.83	8	8	0	0
新英格兰基金	10.4	10.2	10.50	0.10	3	1	2	4
波士顿普塔姆基金	13.1	16.0	13.16	0.06	2	6	−4	16
惠灵顿基金	11.3	12.0	11.33	0.03	1	3	−2	4
总计							0	110

* 得自回归：$\hat{E}_i = 5.819\ 4 + 0.459\ 0\sigma_i$。

+ 残差的绝对值。

注：按递升次序编级。

再用（11.5.7）中的 t 检验得：

$$t=\frac{(0.3333)(\sqrt{8})}{\sqrt{1-0.1110}}$$
$$=0.9998 \tag{11.5.9}$$

对于8个自由度，即使在10%的显著水平上，这个 t 值也是不显著的；p 值是0.17。因此，没有迹象表明解释变量与残差绝对值之间有任何系统的联系，故可认为没有异方差性。

408 **戈德菲尔德-匡特检验**。[17]这一广为流传的方法适用于异方差性方差 σ_i^2 同回归模型中解释变量之一有正向关系的情形。为简单起见，考虑通常的双变量模型：

$$Y_i=\beta_1+\beta_2X_i+u_i$$

假使 σ_i^2 的正向关系为：

$$\sigma_i^2=\sigma^2X_i^2 \tag{11.5.10}$$

其中 σ^2 是一常数。[18]

假定（11.5.10）设想 σ_i^2 与 X 变量的平方成正比。普雷斯和霍撒克在其家庭预算研究中曾发现这种假定甚为有用。（见11.6节。）

如果（11.5.10）得当，则意味着 X_i 值越大，σ_i^2 也越大。如果情况正是如此，则模型中有异方差性是最为可能的。为做出明确的检验，戈德菲尔德和匡特提出如下的步骤：

步骤1. 从最小 X 值开始，按 X 值的大小顺序将观测值排列。

步骤2. 略去居中的 c 个观测值，其中 c 是预定的，并将其余（$n-c$）个观测值分成两组，每组（$n-c$）/2个。

步骤3. 分别对头（$n-c$）/2个观测值和末（$n-c$）/2个观测值各拟合一个回归，并分别获得残差平方和 RSS_1 和 RSS_2，RSS_1 代表对较小 X_i 值所做回归的RSS（小方差组），而 RSS_2 代表对较大 X_i 值所做回归的RSS（大方差组）。这些RSS各有：

$$\frac{(n-c)}{2}-k\text{ 或}\left(\frac{n-c-2k}{2}\right)\text{个自由度（df）}$$

其中 k 是包括截距在内的待估参数个数。（为什么？）当然，对于双变量情形 $k=2$。

步骤4. 计算比率

$$\lambda=\frac{RSS_2/\text{df}}{RSS_1/\text{df}} \tag{11.5.11}$$

如果 u_i 是正态分布的（我们常做这种假定），并且如果同方差性假定真实，则可以证明（11.5.11）的 λ 遵循分子和分母自由度各为（$n-c-2k$）/2的 F 分布。

409 如果在一项应用中，计算的 λ（$=F$）值大于选定显著性水平的临界 F 值，就可拒绝同方差性假设，就是说异方差性看来很可能出现了。

在举例阐明此检验前，谈谈所省略的 c 个居中观测值是适宜的。这些观测值的省略是为了突出或激化小方差组（即 RSS_1）与大方差组（即 RSS_2）之间的差异，但戈德菲尔德-匡特检验之所以能成功地做到这一点，有赖于怎样选好 c。[19] 对于双变量模型，戈德菲尔德和匡特所做的蒙特卡罗实验表明，当样本大小为 60 时，c 约为 16。但贾奇等人却提出，在实践中发现，当 $n=30$ 时，取 $c=4$；当 $n=60$ 时，取 $c=10$ 为宜。[20]

在继续往下讲之前，还要提请注意，当模型中有多于 1 个 X 变量时，在检验的步骤 1 中，就可按任一个 X 的大小顺序将观测值排列。例如，在模型 $Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\beta_4X_{4i}+u_i$ 中，可按这些 X 中的任一个将数据排序。如果我们事先没有把握哪个 X 变量合适，则可对每一 X 变量进行检验，或者通过对每一个 X 轮流做帕克检验。

例 11.4 做戈德菲尔德-匡特检验。 为说明戈德菲尔德-匡特检验，我们在表 11.3 中展现一个 30 户家庭的横截面相对于收入的消费支出数据。假如我们公设：消费支出与收入有线性关系，而数据中有异方差性，并且进一步公设：异方差的性质有如（11.5.10）所设。为了进行检验，我们把重新排序的数据也展现在表 11.3 中。

表 11.3　为说明戈德菲尔德-匡特检验的假想消费支出 Y（美元）与收入 X（美元）数据

		按 X 值排序的数据	
Y	X	Y	X
55	80	55	80
65	100	70	85
70	85	75	90
80	110	65	100
79	120	74	105
84	115	80	110
98	130	84	115
95	140	79	120
90	125	90	125
75	90	98	130
74	105	95	140
110	160	108	145
113	150	113	150

125	165	110	160	居中的4个观测值
108	145	125	165	
115	180	115	180	
140	225	130	185	
120	200	135	190	
145	240	120	200	
130	185	140	205	
152	220	144	210	
144	210	152	220	
175	245	140	225	
180	260	137	230	
135	190	145	240	
140	205	175	245	
178	265	189	250	
191	270	180	260	
137	230	178	265	
189	250	191	270	

略去居中的4个观测值后，对开头的13个和末尾的13个观测值分别做OLS回归，并计算相应的残差平方和，具体结果见下（括号中为标准误）。

对头13个观测值做回归：

$$\hat{Y}_i = 3.409\,4 + 0.696\,8X_i$$
$$(8.704\,9)\quad(0.074\,4)\qquad r^2 = 0.888\,7$$
$$\mathrm{RSS}_1 = 377.17$$
$$\mathrm{df} = 11$$

对末13个观测值做回归：

$$\hat{Y}_i = -28.0272 + 0.794\,1X_i$$
$$(30.642\,1)\quad(0.131\,9)\qquad r^2 = 0.7681$$
$$\mathrm{RSS}_2 = 1\,536.8$$
$$\mathrm{df} = 11$$

从这些结果我们得到：

$$\lambda = \frac{\mathrm{RSS}_2/\mathrm{df}}{\mathrm{RSS}_1/\mathrm{df}} = \frac{1\,536.8/11}{377.17/11}$$
$$\lambda = 4.07$$

对于11个分子自由度和11个分母自由度，5%显著水平的临界F值是2.82。由于估计的F（$=\lambda$）值超过此临界值，故可做结论：误差的方差中有异方差性。然而，如果我们把显著性水平定在1%上，则我们未必拒绝同方差性假定。（为什么？）注意，观测到λ的p值是0.014。

411 **布劳殊-培干-戈弗雷（Breusch-Pagan-Godfrey）检验。**[21] 戈德菲尔德-匡特检验的成功不仅依赖于 c 值（被省略的居中观测值个数），还依赖于用以排序的 X 变量的正确识别。如果我们考虑**布劳殊-培干-戈弗雷（BPG）检验，则可避免这种检验的局限性。**

为说明这种检验，考虑 k 变量线性回归模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{11.5.12}$$

假定误差方差 σ_i^2 有如下函数关系：

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi}) \tag{11.5.13}$$

即 σ_i^2 是诸非随机变量 Z 的某个函数；部分或全部 X 可用作 Z。具体地说，假定：

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} \tag{11.5.14}$$

即 σ_i^2 是诸 Z 的一个线性函数。如果 $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$，则 $\sigma_i^2 = \alpha_1$，此为一常数。因此，为了检验 σ_i^2 是否同方差性，就可检验假设 $\alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$。这就是布劳殊-培干检验的基本思想。具体检验步骤如下。

步骤 1. 用 OLS 估计(11.5.12)并得到残差 $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_n$。

步骤 2. 计算 $\tilde{\sigma}^2 = \sum \hat{u}_i^2 / n$。回顾第 4 章知这是 σ^2 的最大似然估计量。[注：OLS 估计量是 $\sum \hat{u}_i^2 / (n - k)$。]

步骤 3. 按以下定义构造 p_i：

$$p_i = \hat{u}_i^2 / \tilde{\sigma}^2$$

这不外是将每个平方残差除以 $\tilde{\sigma}^2$。

步骤 4. 将如此构造的 p_i 对诸 Z 回归：

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i \tag{11.5.15}$$

其中 v_i 是回归的残差项。

步骤 5. 从（11.5.15）求出 ESS（解释平方和）并定义：

$$\Theta = \frac{1}{2}(\text{ESS}) \tag{11.5.16}$$

假定 u_i 是正态分布的。可以证明，如果有同方差性，当样本大小 n 无限增大时，则：

$$\Theta \underset{\text{asy}}{\sim} \chi^2_{m-1} \tag{11.5.17}$$

412 就是说，Θ 遵循自由度为（$m-1$）的 χ^2 分布。（注：asy 意为渐近地。）

因此，在一项应用中，如果所计算的 Θ（$=\chi^2$）超过选定显著性水平的临界 χ^2 值，就可拒绝同方差性假设；否则不拒绝。

读者可能想知道，为什么 BPG 选取 $\frac{1}{2}$ESS 作为检验统计量。其原因略为复杂，留待查阅参考文献。[22]

例 11.5 做布劳殊–培干–戈弗雷检验。作为一个例子，我们再回到曾用来说明戈德菲尔德–匡特异方差性检验的数据（表 11.3）。将 Y 对 X 回归得到：

步骤 1.

$$\hat{Y}_i = 9.290\ 3 + 0.637\ 8X_i$$
$$\text{se} = (5.231\ 4)(0.028\ 6) \qquad \text{RSS} = 2\ 361.153$$
$$R^2 = 0.946\ 6 \tag{11.5.18}$$

步骤 2.

$$\tilde{\sigma}^2 = \sum \hat{u}_i^2/30 = 2\ 361.153/30 = 78.705\ 1$$

步骤 3. 用 78.705 1 除得自回归（11.5.18）的残差，以构造变量 p_i。

步骤 4. 假定 p_i 按（11.5.14）所设与 X_i（$=Z_i$）有线性关系，我们算得回归：

$$\hat{p}_i = -0.742\ 6 + 0.010\ 1X_i$$
$$\text{se} = (0.752\ 9)(0.004\ 1) \qquad \text{ESS} = 10.428\ 0 \tag{11.5.19}$$
$$R^2 = 0.18$$

步骤 5.

$$\Theta = \frac{1}{2}(\text{ESS}) = 5.214\ 0 \tag{11.5.20}$$

在 BPG 检验的假定下，（11.5.20）中的 Θ 渐近地遵循 1 个自由度的 χ^2 分布。[注：（11.5.19）中只有一个回归元。] 现在从 χ^2 表中我们找到，对于 1 个自由度，5% 临界 χ^2 值是 3.841 4 和 1% 临界 χ^2 值是 6.634 9。由此知所测算的 χ^2 值 5.214 0 在 5% 显著性水平上显著，但不在 1% 显著性水平上显著。因此，我们得到如同戈德菲尔德–匡特检验一样的结论。但应记住，严格地说，BPG 检验是一种渐近性或大样本检验，而在本例中 30 个观测值还未必构成一大样本。还应指出，在小样本中，该检验对干扰项 u_i 的正态性假定是灵敏的。当然，我们可通过先前讨论的 χ^2 检验或雅克–贝拉检验去检验正态性假定。[23]

418 **怀特的一般异方差性检验。** 戈德菲尔德–匡特检验要求按照被认为是引起异方差性的 X 变量把观测值重新排序，而 BGP 检验则易受偏离正态性假定的影响。怀特所提出的检验，不同于这两个检验，并不要求排序也不依赖于正态性假定，而且易于付诸实施。[24] 为说明其基本思想，考虑如下的三变量回归模型（对 k 变量模型的推广是显然的）：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{11.5.21}$$

怀特检验进行如下：

步骤 1. 对给定的数据，估计（11.5.21）并获得残差 $\hat{u}_i$。

步骤 2. 再做如下（辅助）回归：

$$\hat{u}_i^2=\alpha_1+\alpha_2X_{2i}+\alpha_3X_{3i}+\alpha_4X_{2i}^2+\alpha_5X_{3i}^2+\alpha_6X_{2i}X_{3i}+v_i \tag{11.5.22}$$[25]

就是将得自原始回归的平方残差对原始诸 X 变量或回归元、它们的平方和交叉乘积做回归，还可引进回归元的高次方。注意方程中有一常数项，即使原始回归不一定含有它。从这个（辅助）回归求 R^2。

步骤 3. 在无异方差性的虚拟假设下，可以证明，从辅助回归算得的 R^2 乘以样本大小（n），渐近地遵循自由度等于辅助回归中的回归元（不包括常数项）个数的 χ^2 分布，即：

$$n\cdot R^2 \underset{\text{asy}}{\sim} \chi^2_{\text{df}} \tag{11.5.23}$$

其中 df 的定义如前。在本例中，因辅助回归中有 5 个回归元，故有 5 个自由度。

步骤 4. 如果（11.5.23）中算得的 χ^2 值超过选定显著性水平的临界 χ^2 值，结论就是有异方差性。如果不超过，就算没有异方差性，也就是说，在辅助回归（11.5.21）中，$\alpha_2=\alpha_3=\alpha_4=\alpha_5=\alpha_6=0$（参看章末注 25）。

414 **例 11.6 做怀特异方差性检验。**根据 41 个国家的横截面数据，S. 刘易斯（Stephen Lewis）估计了如下回归模型[26]：

$$\ln Y_i=\beta_1+\beta_2\ln X_{2i}+\beta_3\ln X_{3i}+u_i \tag{11.5.24}$$

其中 Y＝贸易税收（进口与出口税收）与政府总收入之比，X_2＝进出口总和与 GNP 之比，X_3＝人均 GNP；ln 表示自然对数。他的假设是 Y 与 X_2 有正向关系（贸易额越高，贸易税收越高），并且 Y 与 X_3 有负向关系（随着收入增大，政府发现直接税——如所得税——比贸易税更易于征收）。

经验结果支持了这些假设。对我们来说，重要的问题是数据中有没有异方差性。由于数据是涉及多个相异国家的横截面数据，人们会先验地预期误差方差中的异方差性。将**怀特的异方差性检验**应用于从回归（11.5.24）得到的残差，得到如下结果[27]：

$$\begin{aligned}\hat{u}_i^2=&-5.8417+2.5629\ln\text{Trade}_i+0.6918\ln\text{GNP}_i\\&-0.4081(\ln\text{Trade}_i)^2-0.0491(\ln\text{GNP}_i)^2\\&+0.0015(\ln\text{Trade}_i)(\ln\text{GNP}_i)\end{aligned} \tag{11.5.25}$$

$$R^2=0.1148$$

注：因标准误不切合我们这里的目的，故未给出。

现在 $n\cdot R^2=41(0.1148)=4.7068$，它渐近地遵循自由度为 5 的 χ^2 分

布（为什么?）。对于5个自由度，5%临界 χ^2 值是11.070 5；10%临界值是9.236 3；25%临界值是6.625 68。为了一切实际的目的，我们都可下结论说，根据怀特检验，这里不存在有异方差性。

对怀特检验作一评语是适宜的。如果模型有多个回归元，那么引进所有的回归元，它们的平方（或更高次方）项以及它们的交叉乘积就会迅速消耗掉许多的自由度。因此，在使用怀特检验时要保持警觉。[28]

在（11.5.2）中给出的怀特统计量统计显著的情形下，异方差性并非必然的原因，也可能是设定误差，在第13章中我们将更详细地讨论设定误差（回顾11.1节中的第5点理由）。换句话说，**怀特检验可能是（纯粹）异方差性的一个检验，或者是设定错误的一个检验，或者两者兼有**。已经被证明，若怀特检验程序中没有出现交叉项，则是对纯粹异方差性的检验；若出现交叉项，则既是对异方差性又是对设定偏误的检验。[29]

415 **其他异方差性检验**。还有若干个其他异方差性检验，每个都依赖于一定的假定。有兴趣的读者可参阅有关文献。[30]我们只提出其中的一种，因为它特别简单。这就是**寇因克-巴塞特（Koenker-Bassett）检验**（KB test）。与异方差性的帕克、布劳殊-培干-戈弗雷和怀特检验相似，KB检验也是基于残差的平方 $\hat{u}_i^2$，但不是对一或多个回归元做回归，而是对回归子估计值的平方进行回归。具体而言，若原模型是

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{11.5.26}$$

估计此模型并从中得到 $\hat{u}_i$，然后估计

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 (\hat{Y}_i)^2 + v_i \tag{11.5.27}$$

其中 $\hat{Y}_i$ 是从模型（11.5.26）中得到的估计值。虚拟假设是 $\alpha_2 = 0$。若未被拒绝，则可以断定不存在异方差性。利用通常的 t 检验或 F 检验就能检验虚拟假设。（注：$F_{1,k} = t_k^2$。）若模型（11.5.26）是双对数模型，则将残差平方对 $(\log \hat{Y}_i)^2$ 进行回归。KB检验的另一个优点在于，即便原模型（11.5.26）中的误差项不是正态分布的，它仍能适用。你若将KB检验用于例11.1，你将发现，将（11.5.3）中得到的残差的平方对（11.5.3）中估计的 $\hat{Y}_i^2$ 进行回归，斜率系数统计上异于零，从而加强了帕克检验。由于此例中只有一个回归元，所以这个结果无足为奇，KB检验在一或多个回归元的情况下都能适用。

§11.6 补救措施

正如已经看到的，异方差性虽然不损坏OLS估计量的无偏性和一致性，但却使它们不再是有效的，甚至不是渐近（即在大样本中）有效的。效率的缺乏使得通常的假设检验程序变成可疑。因此，补救措施显然是需要的。补

救方法可分两种：σ_i^2 为已知和 σ_i^2 为未知。

当 σ_i^2 为已知：加权最小二乘法

正如在 11.3 节所看到的，如果已知 σ_i^2，纠正异方差性的最明显方法，就是采取加权最小二乘，因为这样一来，得到的估计量是 BLUE。

416 **例 11.7 加权最小二乘法说明。** 为说明此法，假定我们要针对表 11.1 中的数据，研究酬金与就业人数之间的关系。为简单起见，我们用 1 表示就业人数（1～4 个职工），2 表示（5～9 个职工），……9 表示（1 000～2 499 个职工）。我们还可用表中各组就业人数的组中值表示就业人数。（见习题 11.21。）

现令 Y 代表平均每职工酬金（美元），而 X 代表就业人数，我们做以下回归［参看方程（11.3.6）]：

$$Y_i/\sigma_i=\hat{\beta}_1^*(1/\sigma_i)+\hat{\beta}_2^*(X_i/\sigma_i)+(\hat{u}_i/\sigma_i) \tag{11.6.1}$$

其中 σ_i 为表 11.1 中报道的薪金的标准差。计算此回归所必需的原始数据由表 11.4 给出。

表 11.4　　加权最小二乘回归的说明

酬金 Y	就业人数 X	σ_i	Y_i/σ_i	X_i/σ_i
3 396	1	743.7	4.566 4	0.001 3
3 787	2	851.4	4.448 0	0.002 3
4 013	3	727.8	5.513 9	0.004 1
4 104	4	805.06	5.097 8	0.005 0
4 146	5	929.9	4.458 5	0.005 4
4 241	6	1 080.6	3.924 7	0.005 5
4 387	7	1 243.2	3.528 8	0.005 6
4 538	8	1 307.7	3.470 2	0.006 1
4 843	9	1 112.5	4.353 2	0.008 1

注：在回归（11.6.2）中，因变量是（Y_i/σ_i）而自变量是（$1/\sigma_i$）和（X_i/σ_i）。

资料来源：Y 和 σ_i（酬金的标准差）数据来自表 11.1。就业人数：1＝1～4 个职工，2＝5～9 个职工，等等。后边的数据来自表 11.1。

在讨论回归的结果之前，要注意（11.6.1）没有截距项。（为什么？）因此，有必要利用过原点回归模型去估计 β_1^* 和 β_2^*，这是第 6 章

中已讨论过的一个问题。但当今大多数计算机软件包都有缩减截距项的选择（例如参看 Minitab, Eviews）。还应注意（11.6.1）的另一有趣的特点：它有两个解释变量（$1/\sigma_i$）和（X_i/σ_i），但如果我们在做酬金对就业人数的回归时用 OLS，则回归中只有一个解释变量 X_i。（为什么？）

WLS 的回归结果如下：

$$
\begin{aligned}
(\widehat{Y_i/\sigma_i}) &= 3\,406.639(1/\sigma_i) + 154.153(X_i/\sigma_i) \\
&\quad\;(80.983) \qquad\qquad (16.959) \qquad\qquad (11.6.2)\\
t &= \;(42.066) \qquad\qquad (9.090)\\
R^2 &= 0.999\,3^{[31]}
\end{aligned}
$$

为了比较，我们给出平常的或不加权的 OLS 回归结果如下：

$$
\begin{aligned}
\hat{Y}_i &= 3\,417.833 + 148.767X_i\\
&\quad\;(81.136)\quad(14.418)\\
t &= \;(42.125)\quad(10.318) \qquad R^2 = 0.938\,3 \qquad (11.6.3)
\end{aligned}
$$

在习题 11.7 中我们要求读者去比较两种回归。

当 σ_i^2 为未知

417 如前所说，若已知真实的 σ_i^2。我们可用 WLS 法得到 BLUE 估计量。由于真实的 σ_i^2 鲜为人知，是否有任何方法，即使在有异方差性的情形下，也能获得 OLS 估计量的方差和协方差的（统计上）一致性估计呢？答案是肯定的。

怀特的"异方差性相一致"的方差与标准误。怀特曾证明，可以做出这样一种估计，它可以对真实的参数值做出渐近（即大样本）有效的统计推断。[32]我们将不讨论数学上的细节，以免超出本书的讨论范围。但附录 11A.4 勾勒了怀特程序。目前有若干计算机软件包（如 TSP, ET, SHAZAM）在给出平常的 OLS 方差和协方差的同时，也给出怀特的经异方差性校正的方差和标准误。[33]顺便提一句，怀特的经异方差性校正的标准差又被称为**稳健标准误**（robust standard errors）。

例 11.8　怀特程序的说明。作为一个例子，我们引用格林（Greene）的一些结果如下[34]：

$$
\begin{aligned}
\hat{Y}_i &= 832.91 \quad - \;1\,834.2\,(\text{收入}) \;+ 1\,587.04\,(\text{收入})^2\\
\text{OLS se} &= (327.3) \qquad (829.0) \qquad\qquad (519.1)\\
t &= \;(2.54) \qquad\quad (2.21) \qquad\qquad\; (3.06) \qquad (11.6.4)\\
\text{怀特 se} &= (460.9) \qquad (1\,243.0) \qquad\quad (830.0)\\
t &= \;(1.81) \qquad (-1.48) \qquad\qquad (1.91)
\end{aligned}
$$

其中 $Y=1979$ 年各州公共学校人均支出，收入 $=1979$ 年各州人均收入。样本由 50 个州及华盛顿特区构成。

以上数字结果表明，经（怀特）异方差性校正的标准误比 OLS 标准误大得多，因而所估计的 t 值比得自 OLS 的要小得多。根据后者，两个回归元都在 5% 水平上统计上显著，而根据怀特估计量则不然。但应指出，怀特
418 的经异方差校正的标准误可能大于或小于未校正标准误。

由于当今现成的回归软件包都备有方差的怀特异方差性相一致估计量，建议读者做回归时予以报道。如华莱士（Wallace）和西尔弗（Silver）所说的：

> 一般地说，经常地使用［回归程序中备有的］怀特选择（White option）大概是个好主意，也许通过怀特输出同 OLS 输出相比，可以看出在一组特定的数据中异方差性是否构成一个严重的问题。[35]

关于异方差性模式的可能假定。 怀特程序除了本身是一个大样本程序外，还有一个缺点，就是这样得到的估计量，不如先按异方差性的类型做数据变换，再做估计来得有效。为说明这点，让我们再回到双变量回归模型：

$$Y_i=\beta_1+\beta_2X_i+u_i$$

我们现在考虑关于异方差性模式的几种假定。

> 假定 1：误差方差正比于 X_i^2：
>
> $$E(u_i^2)=\sigma^2X_i^2 \tag{11.6.5}$$ [36]

如果作为一种"猜测"或通过描图或帕克和格莱泽方法，认为 u_i 的方差正比于解释变量 X 的平方（见图 11.10），则可对原模型做如下变换：用 X_i 通除原模型：

$$\begin{aligned}\frac{Y_i}{X_i}&=\frac{\beta_1}{X_i}+\beta_2+\frac{u_i}{X_i}\\&=\beta_1\frac{1}{X_i}+\beta_2+v_i\end{aligned} \tag{11.6.6}$$

其中 v_i 是变换后的干扰项，等于 u_i/X_i。现在容易证实，

$$\begin{aligned}E(v_i^2)&=E\left(\frac{u_i}{X_i}\right)^2=\frac{1}{X_i^2}E(u_i^2)\\&=\sigma^2\qquad \text{利用(11.6.5)}\end{aligned}$$

419 从而 v_i 的方差是同方差性的，并可对变换方程（11.6.6）施行 OLS，做 Y_i/X_i 对 $1/X_i$ 的回归。

注意，在变换的回归中，截距项 β_2 是原方程中的斜率系数，而斜率系

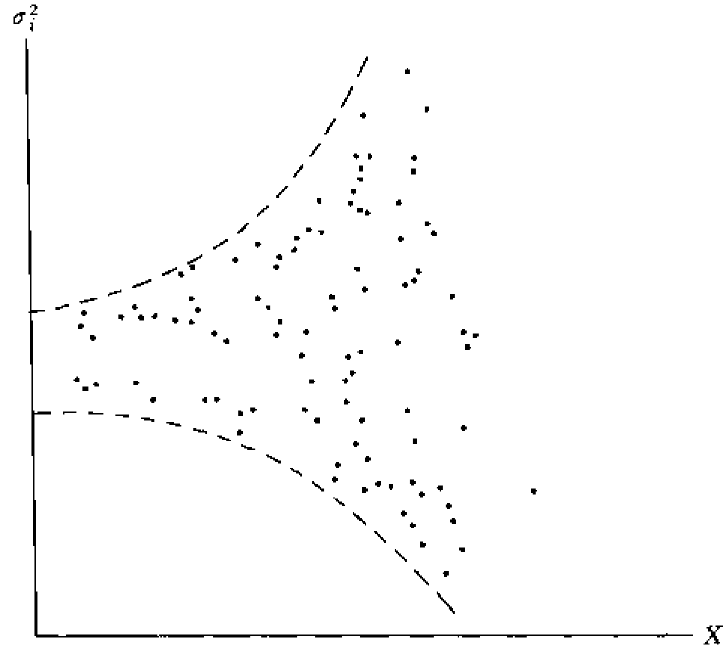


图 11.10 误差方差正比于 X^2

数 $\boldsymbol{\beta}_1$ 则是原始模型中的截距项。因此，要回到原来模型，须用 X_i 乘以估得的（11.6.6）。习题 11.20 给出这种变换的一个应用。

> 假定 2：误差方差正比于 X_i。**平方根变换：**
>
> $$E(u_i^2)=\sigma^2 X_i \tag{11.6.7}$$

如果认为 u_i 的方差不是正比于 X_i 的平方而是正比于 X_i 本身，就可将原始模型变换如下（参看图 11.11）：

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}}&=\frac{\beta_1}{\sqrt{X_i}}+\beta_2\sqrt{X_i}+\frac{u_i}{\sqrt{X_i}}\\&=\beta_1\frac{1}{\sqrt{X_i}}+\beta_2\sqrt{X_i}+v_i\end{aligned} \tag{11.6.8}$$

其中 $v_i=u_i/\sqrt{X_i}$ 且 $X_i>0$。

420 在假定 2 下容易证实，$E(v_i^2)=\sigma^2$ 为同方差性情形，因此，可按 OLS 对（11.6.8）进行 $Y_i/\sqrt{X_i}$ 对 $1/\sqrt{X_i}$ 和 $\sqrt{X_i}$ 的回归。

注意变换模型的一个重要特点：它没有截距项。因此要用过原点回归模型去估计 β_1 和 β_2。在做完回归（11.6.8）之后，只须乘以 $\sqrt{X_i}$，即回到原始模型。

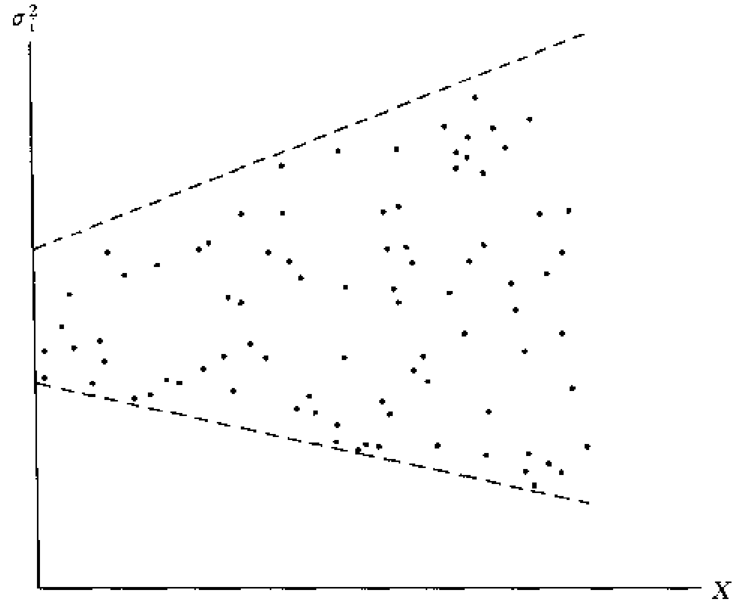


图 11.11　误差方差正比于 X

> 假定 3：误差方差正比于 Y 均值的平方。
>
> $$E(u_i^2)=\sigma^2[E(Y_i)]^2 \tag{11.6.9}$$

方程（11.6.9）假定了 u_i 的方差正比于 Y 的期望值的平方（参看图 11.8e）。现在：

$$E(Y_i)=\beta_1+\beta_2X_i$$

若将原模型变换如下：

$$\begin{aligned}\frac{Y_i}{E(Y_i)}&=\frac{\beta_1}{E(Y_i)}+\beta_2\frac{X_i}{E(Y_i)}+\frac{u_i}{E(Y_i)}\\&=\beta_1\left(\frac{1}{E(Y_i)}\right)+\beta_2\frac{X_i}{E(Y_i)}+v_i\end{aligned} \tag{11.6.10}$$

421 其中 $v_i=u_i/E(Y_i)$，则可表明 $E(v_i^2)=\sigma^2$；即干扰项 v_i 是同方差性的。从而回归（11.6.10）满足经典线性回归模型的同方差性假定。

然而，由于 $E(Y_i)$ 依赖于未知的 β_1 和 β_2，变换（11.6.10）是不可操作的。当然，我们知道 $\hat{Y}_i=\hat{\beta}_1+\hat{\beta}_2X_i$，这是 $E(Y_i)$ 的一个估计量。因此，可按两步进行：首先，暂且忽略异方差性的问题，做平常的 OLS 回归并获得 $\hat{Y}_i$。然后利用估计得的 $\hat{Y}_i$ 做如下的模型变换：

$$\frac{Y_i}{\hat{Y}_i}=\beta_1\left(\frac{1}{\hat{Y}_i}\right)+\beta_2\left(\frac{X_i}{\hat{Y}_i}\right)+v_i \tag{11.6.11}$$

其中 $v_i=(u_i/\hat{Y}_i)$。在第 2 步中，我们做回归（11.6.11）。虽然 $\hat{Y}_i$ 并不正好

等于$E(\hat{Y}_i)$，但 $\hat{Y}_i$ 是一致性估计量；当样本无限地增大时，它们将趋于真 $E(Y_i)$。因此，如果样本含量合理地大，变换（11.6.11）实际上会有良好的表现。

> 假定 4：和回归 $Y_i=\beta_1+\beta_2X_i+u_i$ 相比，诸如：
>
> $$\ln Y_i=\beta_1+\beta_2\ln X_i+u_i \tag{11.6.12}$$

这样一个对数变换常常能减低异方差性。

这样的结果之所以出现，是因为对数变换压缩了测量变量的尺度，把两个值的 10 倍之差降低到约 2 倍之差。例如，数值 80 十倍于数值 8。但 ln80（$=4.328\ 0$）仅约两倍于 ln8（$=2.079\ 4$）。

对数变换的另一优点是，斜率系数 β_2 测出 Y 对 X 的弹性，即对应于 X 的 1% 的变化，Y 的百分比变化。例如，Y 是消费而 X 是收入，（11.6.12）中的 β_2 将测出收入弹性，而在原始模型中，β_2 仅测出对应于收入的一单位变化，平均消费的变化率。这就是为什么对数模型在经验计量经济学中广为应用的一个理由。（至于对数变换所带来的一些问题，参看习题 11.4。）

在结束我们对补救措施的讨论之时，我们再次强调，以上所有讨论的变换都是一种权宜之计（ad hoc）。我们基本上是在猜测 σ_i^2。在所讨论的变换中哪一种能行之有效，要看问题的性质和异方差性的严重程度。应记住，我

422 们所考虑的变换还有其他的一些问题：

1. 当我们越出双变量模型的范围时，我们也许先验上不知道应选择哪一个 X 变量去变换数据。[37]

2. 假定 4 中讨论的对数变换，当某些 Y 和 X 值为零或负数时便不适用。[38]

3. 然后，还有一个**谬误相关**（spurious correlation）的问题。该词来自 K. 皮尔逊，指的是即使原始变量是不相关或随机的，但变量的比率却被发现有相关关系的情形。[39] 例如在模型 $Y_i=\beta_1+\beta_2X_i+u_i$ 中，Y 和 X 也许相关。但在变换模型 $Y_i/X_i=\beta_1(1/X_i)+\beta_2$ 中，Y_i/X_i 和 $1/X_i$ 却常常被发现有相关关系。

4. 当 σ_i^2 无法直接得知而要从前面讨论的一个或多个变换中做出估计时，所有用到 t 检验、F 检验等检验程序，严格地说，都只在大样本中有效。因此，在小的或有限的样本中，如何根据各种变换去解释所得到的结果，必须多加小心。[40]

§ 11.7 总结性的例子

在结束我们对异方差性的讨论时，我们用两个例子说明侦察它的各种方法以及对它的一些补救措施。

例 11.9 再看儿童死亡率一例

让我们再次回到曾考虑过几次的儿童死亡率一例。我们从 64 个国家的数据得到方程 (8.2.1) 中所示的回归结果。由于数据是横截面数据，涉及的国家在儿童死亡率上有不同的表现，所以很可能会出现异方差性。为探明究竟，首先看从方程 (8.2.1) 中得到的残差。这些残差画在图 11.12 中。从图上看，残差没有显示出任何存在异方差性的明显形式。尽管如此，表象仍可能有欺骗性。所以，我们用帕克、格莱泽和怀特检验，看是否有异方差性的证据。

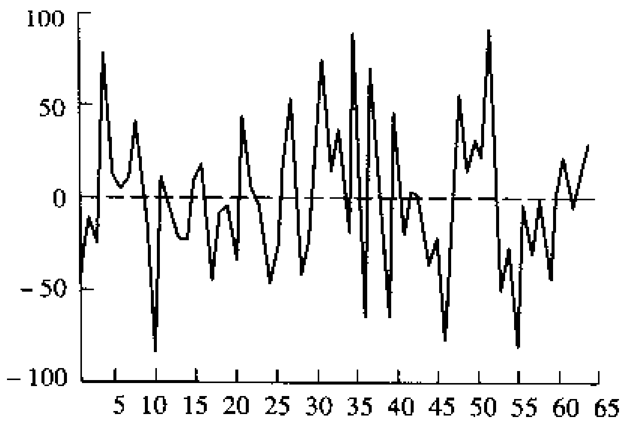


图 11.12 回归 (8.2.1) 中的残差

帕克检验

由于有两个回归元 GNP 和 FLR，所以我们可以将回归 (8.2.1) 中残差的平方对其中任意一个回归，或者将它们对回归 (8.2.1) 中估计出来的 CM 值 ($=\widehat{CM}$) 做回归。利用后者，我们得到如下结论。

$$\hat{u}_i^2 = 854.4006 + 5.7016\,\widehat{CM}_i$$
$$t = \ (1.2010)(1.2428) \qquad r^2 = 0.024 \qquad (11.7.1)$$

注：$\hat{u}_i$ 是从回归 (8.2.1) 中得到的残差，$\widehat{CM}$是从回归 (8.2.1) 中估计出来的 CM 值。

423 如此回归所示，残差的平方与估计的 CM 之间没有系统的关系（为什

么?)，这就表明同方差性可能站得住脚。顺便一提，将残差平方值的对数对$\hat{CM}$的对数做回归，也不会改变这个结论。

格莱泽检验

将 (8.2.1) 中所得到残差的绝对值对同一回归所估计的 CM 值做回归，得到如下结论：

$$|\hat{u}_i| = 22.3127 + 0.0646\,\hat{CM}_i$$
$$t = (2.8086)(1.2622) \qquad r^2 = 0.0250 \tag{11.7.2}$$

同样，由于斜率系数的 t 值并非统计显著，所以残差的绝对值与估计的 CM 值之间没有系统的关系。

怀特检验

应用含有和不含交叉项的怀特异方差性检验，我们没有发现任何异方差性证据。我们也重新估计 (8.2.1) 以得到怀特异方差一致的标准误和 t 值，结论与方程 (8.2.1) 中给出的那些结论十分相似，从我们前面所做的各种异方差性检验来看，无足为奇。

总之，儿童死亡率回归 (8.2.1) 看来不存在异方差性的问题。

例 11.10　1988 年美国 18 个产业群体的 R&D 支出、销售额和利润

表 11.5 给出了美国 18 个产业群体研发支出、销售额和利润方面的数据，所有数据都以百万美元计。由于表中横截面数据差别很大，所以在 R&D 对销售额（或利润）的回归中就可能出现异方差性。回归结果如下：

$$\widehat{R\&D}_i = 192.9931 + 0.0319\,\text{Sales}_i$$
$$\text{se} = (533.9317)(0.0083)$$
$$t = (0.3614)(3.8433) \qquad r^2 = 0.4783 \tag{11.7.3}$$

无足为奇，R&D 与销售额之间有明显的正相关关系。

表 11.5　美国的创新：1988 年美国研究与开发（R&D）费用支出

（所有数字均以百万美元计）

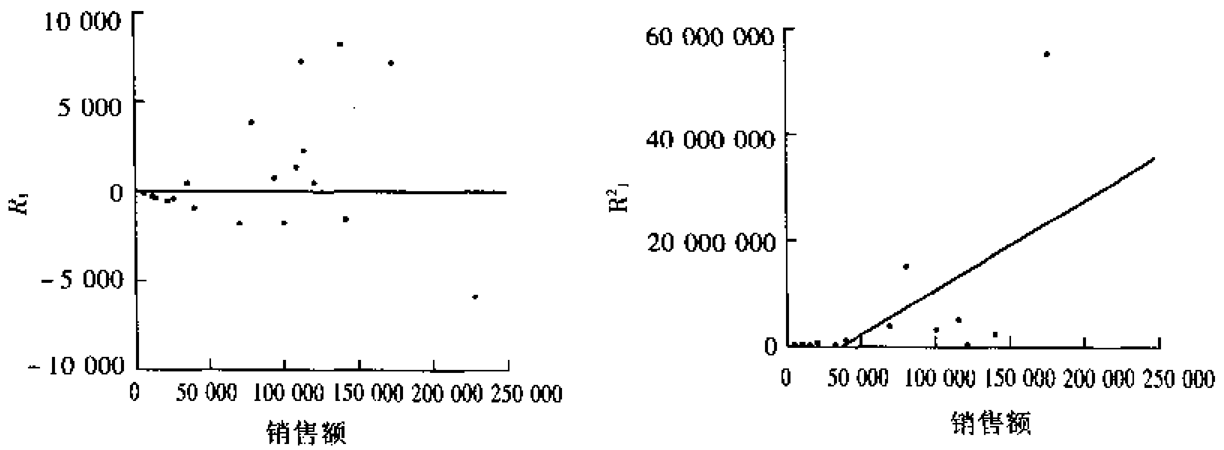
工业群体	销售量	R&D 费用	利润
1. 容器与包装	6 375.3	62.5	185.1
2. 非银行业金融	11 626.4	92.9	1 569.5
3. 服务行业	14 655.1	178.3	276.8
4. 金属与采矿	21 869.2	258.4	2 828.1

5. 住房与建筑	26 408.3	494.7	225.9
6. 一般制造业	32 405.6	1 083.0	3 751.9
7. 休闲娱乐	35 107.7	1 620.6	2 884.1
8. 纸张与林木产品	40 295.4	421.7	4 645.7
9. 食品	70 761.6	509.2	5 036.4
10. 卫生保健	80 552.8	6 620.1	13 869.9
11. 宇航	95 294.0	3 918.6	4 487.8
12. 消费者用品	101 314.1	1 595.3	10 278.9
13. 电器与电子产品	116 141.3	6 107.5	8 787.3
14. 化工产品	122 315.7	4 454.1	16 438.8
15. 五金	141 649.9	3 163.8	9 761.4
16. 办公设备与计算机	175 025.8	13 210.7	19 774.5
17. 燃料	230 614.5	1 703.8	22 626.6
18. 汽车	293 543.0	9 528.2	18 415.4

注：工业群体按销售量递增顺序排列。

资料来源：*Business Week*, Special 1989 Bonus Issue, R&D Scorecard, pp.180－224.

为了看出回归（11.7.3）是否遇到异方差性的问题，我们从上述回归中得到残差 $\hat{u}_i$ 及其平方 $\hat{u}_i^2$，并相对销售额进行描点，示于图 11.13。从此图
424 来看，残差及其平方与销售额之间有系统关系，可能表明存在异方差性。为规范地进行检验，我们使用帕克、格莱泽和怀特检验，并给出如下结果：



残差得自 R&D 对销售额的回归　　残差平方得自 R&D 对销售额的回归

图 11.13　残差 R_1 及其平方（R_1^2）对销售额的散点图

帕克检验

$$
\begin{aligned}
\hat{u}_i^2 &= -974\,469.1 \quad + \quad 86.232\,1\text{Sales}_i \\
\text{se} &= (4\,802\,343) \qquad (40.362\,5) \qquad r^2=0.221\,9 \qquad (11.7.4) \\
t &= \quad (-0.202\,9) \ (2.136\,4)
\end{aligned}
$$

帕克检验表明，残差平方与销售额之间存在着统计上显著的正相关。

格莱泽检验

$$
\begin{aligned}
|\hat{u}_i| &= \quad 578.571\,0 + \quad 0.011\,9\text{Sales}_i \\
\text{se} &= (678.695\,0) \quad (0.005\,7) \qquad r^2=0.214 \qquad (11.7.5) \\
t &= \quad (0.852\,4) \quad (2.087\,7)
\end{aligned}
$$

格莱泽检验也表明，残差的绝对值与销售额之间也有系统的关系，从而增加了回归（11.7.3）存在异方差性的可能性。

425 **怀特检验**

$$
\begin{aligned}
\hat{u}_i^2 &= -6\,219\,665 + 229.350\,8\text{Sales}_i - 0.000\,537\text{Sales}_i^2 \\
\text{se} &= \quad (6\,459\,809)(126.219\,7) \qquad (0.000\,4) \qquad (11.7.6) \\
t &= \quad (0.962\,8) \ (1.817\,0) \quad (-1.342\,5) \qquad R^2=0.289\,5
\end{aligned}
$$

利用 R^2 值和 $n=18$，我们得到 $nR^2=5.212\,4$，在不存在异方差性的虚拟假设之下，服从自由度为 2 的 χ^2 分布［由于（11.7.6）中有两个回归元］。得到大于等于 5.212 4 的一个 χ^2 值的 p 值约为 0.074。若认为这个 p 值足够
426 低，则怀特检验也表明存在异方差性。

总之，基于残差图及帕克、格莱泽和怀特检验来看，我们在（11.7.3）中所做的 R&D 回归遇到了异方差性的问题。由于真实的误差方差未知，所以欲得到对异方差进行修正的标准误和 t 值，我们还不能使用加权最小二乘法。于是，对于误差方差的性质，我们必须做出某种有经验的猜测。

从图 11.13 中的残差图来看，误差方差似乎如（11.6.7）一般与销售额成比例，即平方根变换。做这种变换，我们得到如下结果：

$$
\begin{aligned}
\frac{\widehat{\text{R\&D}}}{\sqrt{\text{Sales}}} &= -246.676\,9\frac{1}{\sqrt{\text{Sales}_i}} + 0.036\,7\sqrt{\text{Sales}_i} \\
\text{se} &= (381.128\,5) \qquad (0.007\,1) \qquad R^2=0.364\,8 \ (11.7.7) \\
t &= (-0.647\,2) \qquad (5.169\,0)
\end{aligned}
$$

你若愿意，还可以将上述方程两边同时乘以 $\sqrt{\text{Sales}_i}$ 以得到原模型。将（11.7.7）与（11.7.3）对比，你会发现，这两个方程的斜率系数大致相同，但它们的标准误不同：在（11.7.3）中为0.008 3，而在（11.7.7）中只有 0.007 1，约下降 14%。

作为本例的结束，我们如在第 11.6 节中所讨论的那样给出如下怀特异方差一致的标准误：

$$\widehat{R\&D}_i = 192.993\,1 + 0.031\,9\text{Sales}_i$$
$$\text{se} = (533.993\,1)(0.010\,1) \qquad r^2 = 0.478\,3 \tag{11.7.8}$$
$$t = (0.361\,4)(3.158\,4)$$

与原回归（11.7.3）（即不对异方差性进行修正）相比，我们看到，尽管参数的估计值没有变化（与我们的预料相一致），但截距系数的标准误下降了，而斜率系数的标准误则略有上升。但须声明，怀特程序是一个严格的大样本程序，而我们这里只有 18 个观测。

§11.8 谨防对异方差性反应过度

回到上一节中所讨论的 R&D 例子，我们看到，当我们对原模型（11.7.3）进行平方根变换来修正其异方差性时，斜率系数的标准误下降，其 t 值上升。这一变化显著到值得担心的程度了吗？换言之，我们什么时候应该真正担心异方差性的问题呢？如一位作者所言："一个好的模型，绝不会因异方差性的原因而被抛弃。"[41]

这里，牢记约翰·福克斯（John Fox）的警言会有所帮助。

427 ……只有在问题严重的时候，误差方差不相等的问题才值得去修正。

误差方差不是常数，对普通最小二乘估计量的有效性和最小二乘推断的可靠性所产生的影响，取决于如下几个因素：样本容量、σ_i^2 中变异的程度、X（即回归元）值的结构及误差方差与 X 之间的关系。因此，就异方差所导致的危害而言，不可能得到一个纯粹一般性的结论。[42]

回顾模型（11.3.1），我们已看到斜率估计量的方差 $\text{var}(\hat{\beta}_2)$ 由（11.2.3）所示的常用表达式给出。在广义最小二乘法下，斜率估计量的方差 $\text{var}(\hat{\beta}_2^*)$ 由（11.3.9）给出。我们知道，后者比前者更有效。但二者之间的差别到底要多大，我们才应该真正担心呢？作为一个经验法则，福克斯建议，"当最大方差比最小方差的 10 倍还大时"[43]，我们就要担心这个问题。于是，回到前面提到戴维森和麦金农的蒙特卡罗模拟结果，考虑 $\alpha = 2$ 的情况。所估计 β_2 的方差在 OLS 下为 0.04，而在 GLS 下为 0.012，前者约为后者的 3.33 倍。[44] 根据福克斯法则，这种情况下，异方差性的严重程度不足以引起担心。

还要记住，尽管有异方差性的问题，但 OLS 估计量仍是线性无偏和渐近（即在大样本中）正态分布的（在一般条件下）。

在我们讨论其他违背经典线性回归模型之假定情况时，我们将会看到，在本节敲响的警钟作为一个一般规则也是适当的。否则，就有可能反应过度。

§11.9 要点与结论

1. 经典线性回归模型的一个关键性的假定是干扰项 u_i 都有相同的方差 σ^2。如果此假定不成立，则说有异方差性。

2. 异方差性并不破坏 OLS 估计量的无偏性和一致性性质。

3. 但这些估计量不再是最小方差或有效的，就是说，它们不是 BLUE。

4. 如果相异的误差方差 σ_i^2 已知，加权最小二乘法可给出 BLUE 估计量。

5. 当异方差性出现时，OLS 估计量的方差并不由常用的 OLS 公式给
428 出。如果我们一味地使用 OLS 公式，则以这些公式为依据的 t 和 F 检验可能严重误导，以致引出错误的结论。

6. 列举异方差性的后果容易，侦察异方差的工作困难。现有若干诊断性检验，但无法告知在特定情况中哪一检验能行之有效。

7. 即使异方差性受到怀疑并且被侦察出来，如何纠正并非易事。如果样本够大，则可获取 OLS 估计量的怀特异方差校正标准误并以之作为统计推断的依据。

8. 另外，可根据 OLS 残差，合理地猜测异方差性的可能模式，以便将原始数据变换成没有异方差性的变换数据。

习　题

问答题

11.1 用简明的理由说明以下的陈述是正确的、谬误的或者不确定的。

a. 当异方差性出现时，OLS 估计量是有偏误的和非有效的。

b. 如果异方差性出现，则惯用的 t 和 F 检验无效。

c. 在异方差性的情况下，常用的 OLS 法必定高估了估计量的标准误。

d. 如果 OLS 回归的残差表现出系统性样式，这就说明数据中有异方差性。

e. 没有任何一般性的异方差性检验能独立于误差项与某一变量相关的假定。

f. 如果一个回归模型误设（比如说，漏掉一个重要变量），则 OLS 残差必定表现出明显的样式。

g. 如果模型漏掉一个有非恒定方差的回归元，则 OLS 残差将是异方差性的。

11.2 在对一个含有 30 个厂商的随机样本做的平均薪金（W）对职工

人数（N）的回归中，得到如下的回归结果[1]：

$$\hat{W} = 7.5 + 0.009N$$
$$t = \text{n.a.}(16.10) \qquad R^2 = 0.90 \tag{1}$$
$$\hat{W}/N = 0.008 + 7.8(1/N)$$
$$t = (14.43) \quad (76.58) \qquad R^2 = 0.99 \tag{2}$$

a. 你怎样解释这两个回归？

b. 从方程（1）到（2）作者做了什么假定？他曾否担心过异方差性？你怎样知道？

429 c. 怎样能把这两个模型的截距和斜率（各一双）联系起来？

d. 你能比较两个模型的 R^2 值吗？为什么或为什么不？

11.3 a. 你能用 OLS 法估计模型：

$$|u_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$
$$|u_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

中的参数吗？为什么或为什么不？

b. 如果不能，你能提出一个估计这些模型的参数的方法吗？正式的或非正式的。

11.4 虽然对数模型，如方程（11.6.12）所示，常能降低异方差性，但须对这种模型的误差项的性质给予细心的关注。例如，模型

$$Y_i = \beta_1 X_i^{\beta_2} u_i \tag{1}$$

可以写为：

$$\ln Y_i = \ln\beta_1 + \beta_2 \ln X_i + \ln u_i \tag{2}$$

a. 如果 $\ln u_i$ 要有零期望值，u_i 的分布必须是什么？

b. 如果 $E(u_i) = 1$，会不会有 $E(\ln u_i) = 0$？为什么或为什么不？

c. 如果 $E(\ln u_i)$ 不为零，怎样能使它等于零？

11.5 说明（11.3.8）中的 β_2^* 还可表达为：

$$\beta_2^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{2*}}$$

以及（11.3.9）中的 $\text{var}(\beta_2^*)$ 又可表达为：

$$\text{var}(\beta_2^*) = \frac{1}{\sum w_i x_i^{2*}}$$

其中 $y_i^* = Y_i - \bar{Y}^*$ 和 $x_i^* = X_i - \bar{X}^*$ 代表对加权均值 $\bar{Y}^*$ 和 $\bar{X}^*$：

$$\bar{Y}^* = \sum w_i Y_i / \sum w_i$$
$$\bar{X}^* = \sum w_i X_i / \sum w_i$$

的离差。

11.6 为了教学的目的哈努谢克（Hanushek）和杰克逊（Jackson）估计如下模型：

$$C_t = \beta_1 + \beta_2 \mathrm{GNP}_t + \beta_3 D_t + u_i \tag{1}$$

其中 C_t = 年度 t 的私人总消费支出，GNP_t = 年度 t 的国民生产总值，以及 D = 年度 t 的国防支出。分析的目的在于研究国防开支对经济中其他开支的影响。

430 他们假拟 $\sigma_t^2 = \sigma^2(\mathrm{GNP}_t)^2$，从而将(1)变换如下，并加以估计：

$$C_t/\mathrm{GNP}_t = \beta_1(1/\mathrm{GNP}_t) + \beta_2 + \beta_3(D_t/\mathrm{GNP}_t) + u_t/\mathrm{GNP}_t \tag{2}$$

根据 1946—1975 年的数据得到的经验结果如下（括号中为标准误）[2]：

$$\hat{C}_t = 26.19 + 0.624\,8\ \mathrm{GNP}_t - 0.439\,8 D_t$$

$$(2.73)(0.006\,0) \qquad (0.073\,6) \quad R^2 = 0.999$$

$$\widehat{C_t/\mathrm{GNP}_t} = 25.92\ (1/\mathrm{GNP}_t) + 0.624\,6 - 0.431\,5\ (D_t/\mathrm{GNP}_t)$$

$$(2.22) \qquad (0.006\,8)(0.059\,7)$$

$$R^2 = 0.875$$

a. 作者对异方差性的性质做了什么假定？你能说明其中理由吗？

b. 比较两个回归的结果。你对原始模型的变换是否使结果有所改进，就是说，降低了估计的标准误？为什么或为什么不？

c. 你能比较这两个 R^2 值吗？为什么或为什么不？（提示：检查应变量。）

11.7 参照估计的回归（11.6.2）和（11.6.3）两回归的结果十分相近。这是什么原因？

11.8 证明若对每一 i，$w_i = w$ 为一常数，则 β_2^* 和 $\hat{\beta}_2$，以及它们的方差都是相同的。

11.9 参照公式（11.2.2）和（11.2.3）。假定：

$$\sigma_i^2 = \sigma^2 k_i$$

其中 σ^2 是常数，而 k_i 是已知的、但不一定相等的权数。

利用此假定，说明（11.2.2）中的方差可表达为：

$$\mathrm{var}(\hat{\beta}_2) = \frac{\sigma_2}{\sum x_i^2} \cdot \frac{\sum x_i^2 k_i}{\sum x_i^2}$$

右端第一项即是（11.2.3）中的方差公式，也就是在同方差性下的$\mathrm{var}(\hat{\beta}_2)$。你能说出在异方差性下的$\mathrm{var}(\hat{\beta}_2)$和在同方差性下的$\mathrm{var}(\hat{\beta}_2)$之间的关系有些什么性质吗？（提示：分析以上公式的右端第二项。）你能对（11.2.2）和（11.2.3）之间的关系做出任何一般性的结论吗？

481

11.10 在模型 $Y_i = \beta_2 X_i + u_i$（注意：没有截距）中你被告知 $\mathrm{var}(u_i) = \sigma^2 X_i^2$。

证明 $\mathrm{var}(\hat{\beta}_2) = (\sigma^2 \sum X_i^4)/(\sum X_i^2)^2$。

解答题

11.11 用表 11.1 的数据做平均酬金 Y 对平均生产力 X 的回归，把就业人数当作观测单元。解释你的结果，并查看你的结果是否和 (11.5.3) 给出的结果一致。

a. 从上面的回归算出残差 $\hat{u}_i$。

b. 按照帕克检验，将 $\ln \hat{u}_i^2$ 对 $\ln X_i$ 回归，并验证回归 (11.5.4)。

c. 仿照格莱泽方法，将 $|\hat{u}_i|$ 对 X_i 回归，再将 $|\hat{u}_i|$ 对 $\sqrt{X_i}$ 回归。然后评述你的结果。

d. 求 $|\hat{u}_i|$ 对 X_i 的等级相关，然后对数据中是否有异方差性及其性质做出评论。

11.12 表 11.6 给出 1971 年第 1 季度至 1973 年第 4 季度期间美国制造行业按公司的资产规模分类的销售/现金比率数据。（数据是按季度报告的。）销售/现金比率可看作公司部门收入流速的一个度量，也就是一个美元的周转次数。

表 11.6 资产规模（百万美元计）

年度与季度	1～10	10～25	25～50	50～100	100～250	250～1 000	1 000 +
1971 - Ⅰ	6.696	6.929	6.858	6.966	7.819	7.557	7.860
- Ⅱ	6.826	7.311	7.299	7.081	7.907	7.685	7.351
- Ⅲ	6.338	7.035	7.082	7.145	7.691	7.309	7.088
- Ⅳ	6.272	6.265	6.874	6.485	6.778	7.120	6.765
1972 - Ⅰ	6.692	6.236	7.101	7.060	7.104	7.584	6.717
- Ⅱ	6.818	7.010	7.719	7.009	8.064	7.457	7.280
- Ⅲ	6.783	6.934	7.182	6.923	7.784	7.142	6.619
- Ⅳ	6.779	6.988	6.531	7.146	7.279	6.928	6.919
1973 - Ⅰ	7.291	7.428	7.272	7.571	7.583	7.053	6.630
- Ⅱ	7.766	9.071	7.818	8.692	8.608	7.571	6.805
- Ⅲ	7.733	8.357	8.090	8.357	7.680	7.654	6.772
- Ⅳ	8.316	7.621	7.766	7.867	7.666	7.380	7.072

资料来源：*Quarterly Financial Report for Manufacturing Corporations*, Federal Trade Commission and the Securities and Exchange Commission, U.S.government、各期（计算结果）。

a. 对每一资产规模计算销售/现金比的均值与标准差。

b. 将 (a) 中计算的均值对标准差描点，把资产规模当作观测

单元。

c. 通过适当的回归模型，判明比率的标准差是否随均值增加而增加。如果说没有这种关系，怎样自圆其说？

432 d. 如果两者有统计上显著的关系，你会怎样变换数据以使异方差性不复存在？

11.13 **巴特利特（Bartlett）的方差匀检验。**[3] 假使有自由度为 f_1、f_2、$\cdots f_k$ 的 k 个独立样本方差 s_1^2、s_2^2、$\cdots s_k^2$，各来自以 u 为均值和 σ_i^2 为方差的正态分布的总体。再假使我们要检验虚拟假设 H_0：$\sigma_1^2=\sigma_2^2=\cdots=\sigma_k^2=\sigma^2$；即每一样本方差都是同一总体方差 σ^2 的一个估计值。

如果虚拟假设真实，则：

$$s^2=\frac{\sum_{i=1}^{k} f_i s_i^2}{\sum f_i}=\frac{\sum f_i s_i^2}{f}$$

给出总体方差 σ^2 的共同（联合）估计的一个估计值，其中 $f_i=(n_i-1)$，而 u_i 为第 i 组的观测个数，并且 $f=\sum_{i=1}^{k} f_i$。

巴特利特证明，虚拟假设可通过近似于 $k-1$ 个自由度的 χ^2 分布的比率 A/B 加以检验，其中：

$$A=f\ln s^2-\sum(f_i\ln s_i^2)$$

及

$$B=1+\frac{1}{3(k-1)}\left[\sum\left(\frac{1}{f_i}\right)-\frac{1}{f}\right]$$

对表 11.1 的数据做巴特利特检验并验证在 5% 显著水平上不能拒绝假设：每一厂商的职工人数组都有相同的总体酬金方差。

注：因每一样本（即就业组）的 n_i 都是 10，故每一样本方差的自由度 f_i 都是 9。

11.14 考虑如下过原点的回归模型：

$$Y_i=\beta X_i+u_i \quad i=1,\ 2$$

告诉你 $u_1\sim N(0,\sigma^2)$ 和 $u_2\sim N(0,2\sigma^2)$，而且它们相互独立。若 $X_1=+1$，$X_2=-1$，计算 β 的加权最小二乘估计值及其方差。若你此时不正确地假定了误差方差相同（比方说都等于 σ^2），β 的 OLS 估计量是什么？其方差又是多少？与用 WLS 方法得到的估计值相比，你能得到什么一般性结论？[4]

11.15 表 11.7 给出了 81 辆汽车在 MPG（每加仑里程数）、HP（发动机马力）、VOL（驾驶空间的立方英尺数）、SP（最高时速）和 WT（以 100 磅为单位的车身重量）方面的数据。

433 表 11.7 客车单位耗油量行驶里程数据

观测	MPG	SP	HP	VOL	WT	观测	MPG	SP	HP	VOL	WT
1	65.4	96	49	89	17.5	42	32.2	106	95	106	30.0
2	56.0	97	55	92	20.0	43	32.2	109	102	92	30.0
3	55.9	97	55	92	20.0	44	32.2	106	95	88	30.0
4	49.0	105	70	92	20.0	45	31.5	105	93	102	30.0
5	46.5	96	53	92	20.0	46	31.5	108	100	99	30.0
6	46.2	105	70	89	20.0	47	31.4	108	100	111	30.0
7	45.4	97	55	92	20.0	48	31.4	107	98	103	30.0
8	59.2	98	62	50	22.5	49	31.2	120	130	86	30.0
9	53.3	98	62	50	22.5	50	33.7	109	115	101	35.0
10	43.4	107	80	94	22.5	51	32.6	109	115	101	35.0
11	41.1	103	73	89	22.5	52	31.3	109	115	101	35.0
12	40.9	113	92	50	22.5	53	31.3	109	115	124	35.0
13	40.9	113	92	99	22.5	54	30.4	133	180	113	35.0
14	40.4	103	73	89	22.5	55	28.9	125	160	113	35.0
15	39.6	100	66	89	22.5	56	28.0	115	130	124	35.0
16	39.3	103	73	89	22.5	57	28.0	102	96	92	35.0
17	38.9	106	78	91	22.5	58	28.0	109	115	101	35.0
18	38.8	113	92	50	22.5	59	28.0	104	100	94	35.0
19	38.2	106	78	91	22.5	60	28.0	105	100	115	35.0
20	42.2	109	90	103	25.0	61	27.7	120	145	111	35.0
21	40.9	110	92	99	25.0	62	25.6	107	120	116	40.0
22	40.7	101	74	107	25.0	63	25.3	114	140	131	40.0
23	40.0	111	95	101	25.0	64	23.9	114	140	123	40.0
24	39.3	105	81	96	25.0	65	23.6	117	150	121	40.0
25	38.8	111	95	89	25.0	66	23.6	122	165	50	40.0
26	38.4	110	92	50	25.0	67	23.6	122	165	114	40.0
27	38.4	110	92	117	25.0	68	23.6	122	165	127	40.0
28	38.4	110	92	99	25.0	69	23.6	122	165	123	40.0
29	46.9	90	52	104	27.5	70	23.5	148	245	112	40.0
30	36.3	112	103	107	27.5	71	23.4	160	280	50	40.0
31	36.1	103	84	114	27.5	72	23.4	121	162	135	40.0
32	36.1	103	84	101	27.5	73	23.1	121	162	132	40.0
33	35.4	111	102	97	27.5	74	22.9	110	140	160	45.0
34	35.3	111	102	113	27.5	75	22.9	110	140	129	45.0
35	35.1	102	81	101	27.5	76	19.5	121	175	129	45.0
36	35.1	106	90	98	27.5	77	18.1	165	322	50	45.0

37	35.0	106	90	88	27.5	78	17.2	140	238	115	45.0
38	33.2	109	102	86	30.0	79	17.0	147	263	50	45.0
39	32.9	109	102	86	30.0	80	16.7	157	295	119	45.0
40	32.3	120	130	92	30.0	81	13.2	130	236	107	55.0
41	32.2	106	95	113	30.0						

注：VOL＝驾驶空间的立方英尺数

HP＝发动机马力

MPG　每加仑耗油量行驶的里程数

SP＝最高时速，里/小时

WT＝车身重量，百磅

观测＝汽车观测次数（车名未记）

资料来源：美国环境保护署，1991年，报告序列号为EPA/AA/CTAB/91-02。

434 a. 考虑如下模型：

$$MPG_i = \beta_1 + \beta_2 SP + \beta_3 HP + \beta_4 WT + u_i$$

估计模型参数并对结论做出解释。这些结论有经济意义吗？

b. 你认为上述模型中的误差方差存在异方差性吗？为什么？

c. 用怀特检验来检查误差方差是否存在异方差性。

d. 求出怀特异方差一致的标准误和 t 值，并与从OLS得到的结论进行比较。

e. 若证明存在异方差性，你如何对数据变形，以使变形后的数据是同方差的？给出必要的计算。

11.16 **印度的食物支出。**在表2.8中，我们已经给出印度55个家庭的食物支出和总支出数据。

a. 将食物支出对总支出回归，检查回归所得到的残差。

b. 将得到的残差对总支出描点，看是否存在系统关系。

c. 若描点图显示存在异方差性，用帕克、格莱泽和怀特检验来看这些检验是否支持从图中观察所得到的异方差性印象。

d. 求出怀特异方差一致标准误，并与OLS标准误进行比较。判断此例中是否值得修正异方差性。

11.17 重做习题11.16，只是把食物支出的对数对总支出的对数做回归。如果你在习题11.16的线性模型中得到异方差性，但在对数线性模型中没有发现，你会得出什么结论？给出所有必要的计算。

11.18 **怀特检验的简易途径。**正文中提到，如果有 n 个回归元，而且我们引入所有回归元及其平方项和交叉乘积项，怀特检验会消耗太多的自由度。因此，为什么要估计像（11.5.22）那样的回归，而不简单地做如下回归呢？

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 \hat{Y}_i + \alpha_3 \hat{Y}_i^2 + v_i$$

其中 $\hat{Y}_i$ 为你从模型中估计出来的 Y 值（回归子的估计值）。毕竟，$\hat{Y}_i$ 无非就是回归元的加权平均，只是以估计的回归系数为权重。

从上述回归中求出 R^2 值，并用（11.5.22）检验不存在异方差性的假设。

在习题 11.16 中，对食物支出的例子应用上述检验。

11.19 回到第11.7 节中所讨论的 R&D 一例。以利润为回归元重做一遍。
435 据经验，你预期你的结果会与以销售额为回归元的结果不同吗？为什么？

11.20 表 11.8 给出了美国研究型大学中统计学正教授在 2000—2001 年薪水的中位数数据。

表 11.8　2000—2001 年统计学教授薪水的中位数

任职年限	人数	薪水中位数（美元）
0～1	11	69 000
2～3	20	70 500
4～5	26	74 050
6～7	33	82 600
8～9	18	91 439
10～11	26	83 127
12～13	31	84 700
14～15	15	82 601
16～17	22	93 286
18～19	23	90 400
20～21	13	98 200
22～24	29	100 000
25～27	22	99 662
28～32	22	116 012
33 以上	11	85 200

资料来源：American Statistical Association, "2000－2001 Salary Report of Academic Statisticians," *Amstat News*, Issue 282, December 2000, p.4.

a. 将薪水中位数对担任教授年限（作为工作经验的一种度量）描点。

为便于描点，假定薪水中位数对应于任职年限的中点。于是任职 4～5 年的薪水 74 050 美元对应于任职 4.5 年，依此类推。对最后一组，假定其范围是 33～35。

b. 考虑如下回归模型：

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \tag{1}$$

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + v_i \tag{2}$$

其中 Y = 薪水中位数，X = 任职年数（任职年数的中点数据度量），u 和 v 为误差项。你能证明模型（2）比（1）更合适吗？利用所给数据估计这两个模型。

c. 若在模型（1）中观测到异方差性，而在模型（2）中没有，你会得出什么结论？给出必要的计算。

d. 若在模型（2）中观测到异方差性，你将如何对数据做变换，以消除异方差性？

436 **11.21** 给出以下数据：

从头 30 个观测值算出的 $RSS_1 = 55$，自由度 df = 25；

从末 30 个观测值算出的 $RSS_1 = 140$，自由度 df = 25；

完成显著性水平为 5% 的戈德菲尔德-匡特异方差性检验。

11.22 表 11.9 给出 20 个国家的股票价格和消费者价格年百分率变化的一个横截面数据。

第二次世界大战后（直至 1969 年）期间股票与消费者价格

国　家	变化率	%每年
	股票价格 Y	消费者价格 X
1. 澳大利亚	5.0	4.3
2. 奥地利	11.1	4.6
3. 比利时	3.2	2.4
4. 加拿大	7.9	2.4
5. 智利	25.5	26.4
6. 丹麦	3.8	4.2
7. 芬兰	11.1	5.5
8. 法国	9.9	4.7
9. 德国	13.3	2.2
10. 印度	1.5	4.0
11. 爱尔兰	6.4	4.0
12. 以色列	8.9	8.4
13. 意大利	8.1	3.3
14. 日本	13.5	4.7
15. 墨西哥	4.7	5.2
16. 荷兰	7.5	3.6
17. 新西兰	4.7	3.6
18. 瑞典	8.0	4.0
19. 英国	7.5	3.9
20. 美国	9.0	2.1

资料来源：Phillip Cagan, *Common Stock Values and Inflation: The Historical Record of Many Countries*, National Bureau of Economic Research, Suppl., March1974, Table.1, p.4.

a. 将数据描在散点图上。

b. 将 Y 对 X 回归并分析回归中的残差。你观察到什么？

c. 因智利的数据看来有些异常（异常值?），去掉智利数据后，重做（b）中的回归。分析从此回归得到的残差，你会看到什么？

d. 如果根据（b）的结果你将得到有异方差性的结论，而根据（c）的结果你又得到相反的结论。那么，你能做出什么一般性结论呢？

【习题注释】

[1] See Dominick Salvatore, *Managerial Economics*, McGraw-Hill, New York, 1989, p.157.

[2] Eric A. Hanushek and John E. Jackson, *Statistical Methods for Social Scientists*, Academic, New York, 1977, p.160.

[3] See "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London A*, vol.160, 1937, p.268.

[4] 节选自 F.A.F.Seber, *Linear Regression Analysis*, John Wiley & Sons, New York, 1977, p. 64。

附录 11A

11A.1 方程（11.2.2）的证明

437 由附录 3A，第 3A.3 节 我们有：

$$\begin{aligned}\operatorname{var}(\hat{\beta}_2) &= E(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2 + 2\text{ 交叉乘积项})\\ &= E(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2)\end{aligned}$$

这是因为，由于无序列相关假定，交叉乘积项的期望值为零。再由于 k_i 是已知的（为什么?），且 $E(u_i^2) = \sigma_i^2$，

$$\operatorname{var}(\hat{\beta}_2) = k_1^2E(u_1^2) + k_2^2E(u_2^2) + \cdots + k_n^2E(u_n^2)$$

即：

$$\operatorname{var}(\hat{\beta}_2) = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2 \qquad (\text{由于 } E(u_i^2) = \sigma_i^2)$$

或：

$$\begin{aligned}\operatorname{var}(\hat{\beta}_2) &= \sum k_i^2\sigma_i^2\\ &= \sum\left[\left(\frac{x_i}{\sum x_i^2}\right)^2\sigma_i^2\right] \quad \text{因 } k_i = \frac{x_i}{\sum x_i^2}\end{aligned}$$

$$= \frac{\sum x_i^2\sigma_i^2}{(\sum x_i^2)^2} \tag{11.2.2}$$

11A.2 加权最小二乘法

为说明方法，我们利用双变量模型 $Y_i = \beta_1 + \beta_2 X_i + u_i$。为求得估计值，不加权的最小二乘法要求最小化：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{1}$$

而加权最小二乘法要求最小化加权残差平方和：

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \tag{2}$$

其中 β_1^* 和 β_2^* 为加权最小二乘估计量，并且权数 w_i 与给定 X_i 的 u_i 或 Y_i 的条件方差成反比：

$$w_i = \frac{1}{\sigma_i^2} \tag{3}$$

438 这里，不言而喻，$\mathrm{var}(u_i \mid X_i) = \mathrm{var}(Y_i \mid X_i) = \sigma_i^2$。

将(2)对 $\hat{\beta}_1^*$ 和 $\hat{\beta}_2^*$ 微分得：

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_1^*} = 2\sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_2^*} = 2\sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-X_i)$$

令上式为零，即得以下两个正规方程：

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \tag{4}$$

$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2 \tag{5}$$

注意这两个正规方程和不加权的最小二乘正规方程的相似性。

解联立方程得：

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \tag{6}$$

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2}$$

$$(7) = (11.3.8)$$

(11.3.9) 给出的 $\hat{\beta}_2^*$ 的方差，可仿照附录 3A，3A.3 节求 $\hat{\beta}_2$ 方差的方法得到。

注：$\bar{Y}^* = \sum w_i Y_i / \sum w_i$ 和 $\bar{X}^* = \sum w_i X_i / \sum w_i$。容易验证，当对所有 i，$w_i = w$ 为一常数时，这些加权均值便化为平常的或不加权的均值 $\bar{Y}$ 和 $\bar{X}$。

11A.3 出现异方差时 $E(\hat{\sigma}^2) \neq \sigma^2$ 的证明

考虑双变量模型

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{1}$$

其中 $\text{var}(u_i) = \sigma_i^2$

现在，

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum [\beta_1 + \beta_2 X_i + u_i - \hat{\beta}_1 - \hat{\beta}_2 X_i]^2}{n-2}$$

$$= \frac{\sum [-(\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2) X_i + u_i]^2}{n-2}$$

439 注意 $(\hat{\beta}_1 - \beta_1) = -(\hat{\beta}_1 - \beta_2)\overline{X} + \bar{u}$，以之代入(2)并对两边求期望便得到

$$E(\hat{\sigma}^2) = \frac{1}{n-2}\left\{-\sum x_i^2 \text{var}(\hat{\beta}_2) + E\left[\sum (u_i - \bar{u})^2\right]\right\}$$

$$= \frac{1}{n-2}\left[-\frac{\sum x_i^2 \sigma_i^2}{\sum x_i^2} + \frac{(n-1)\sum \sigma_i^2}{n}\right] \tag{3}$$

其中用到（11.2.2）。

如你从（3）中所见，若存在同方差性，即对每个 i 都有 $\sigma_i^2 = \sigma^2$，则 $E(\hat{\sigma}^2) = \sigma^2$。因此，惯常计算的期望值 $\hat{\sigma}^2 = \sum \hat{u}^2/(n-2)$ 在异方差情况下不再等于真实的 σ^2。[1]

11A.4 怀特稳健标准误

为了解怀特对异方差修正后的标准误，考虑双变量回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad \text{var}(u_i) = \sigma_i^2 \tag{1}$$

如（11.2.2）所示，

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{2}$$

由于 σ_i^2 不能直接观测，所以怀特建议用每个 i 的残差平方 $\hat{u}_i^2$ 取代 σ_i^2，并估计 $\text{var}(\hat{\beta}_2)$ 如下：

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \hat{u}_i^2}{\left(\sum x_i^2\right)^2} \tag{3}$$

怀特证明，（3）是（2）的一致估计量，即随着样本容量无限增加，（3）

收敛于（2）。[2]

顺便提醒注意，你的软件包中并不包含怀特稳健标准误程序，你可以首先做通常的 OLS 回归，从中得到残差，然后利用（3）式即可。

怀特程序可推广至 k 变量回归模型

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{4}$$

440 任何一个偏回归系数（比方说 $\hat{\beta}_j$）的方差都可如下求得：

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sum \hat{w}_{ji}^2 \hat{u}_i^2}{(\sum \hat{w}_{ji}^2)^2} \tag{5}$$

其中 $\hat{u}_i$ 为从原回归（4）中得到的残差，$\hat{w}_j$ 为将回归元 X_j 对（4）中其余回归元做（辅助）回归时得到的残差。

显而易见，这是一个费时的程序，因为你必须对每个 X 变量都做回归（5）。当然，你如果有一个统计软件包进行例行计算，所有这些劳动则都可以避免：诸如 PcGive，Eviews，Microfit，Shazam，Stata 和 Limdep 等现在都能十分轻易地求出怀特异方差的稳健标准误。

【附录注释】

[1] 详细情况可参阅 Jan Kmenta，*Elements of Econometrics*，2d.ed.，Macmilln，New York，1986，pp.276－278。

[2] 更准确地讲，n 乘以（3）概率收敛于$E[(X_i-\mu_X)^2u_i^2]/(\sigma_X^2)^2$，即 n 乘以（2）的概率极限，其中 n 为样本容量，μ_X 为 X 的期望值，σ_X^2 为 X 的（总体）方差。更详细情形，参见 Jeffrey M. Wooldridge，*Introductory Econometrics：A Modern Approach*，South-Western Publishing，2000，p.250。

【注释】

[1] See Stefan Valavanis，*Econometrics*，McGraw-Hill，New York，1959，p. 48.

[2] 像 Valavanis 说的，“收入增加了，人们现在只在意一元钱怎样用，而过去则在意一角钱怎样用”，同上，第 48 页。

[3] 感谢迈克尔·麦卡利尔（Michael McAleer）向我指出这一点。

[4] David F. Hendry，*Dynamic Econometrics*，Oxford University Press，1995，p. 45.

[5] 一个正式的证明见于 Phoebus J.Dhrymes，*Introductory Econometrics*，Springer-Verlag，New York，1978，pp.110－111。顺便指出，$\hat{\beta}_2$ 的效率损失[指 $\operatorname{var}(\hat{\beta}_2)$超过 $\operatorname{var}(\hat{\beta}_2^*)$多少]有赖于诸 X 变量的样本值以及 σ_i^2 值。

[6] 由（5.3.6）我们知道 β_2 的$100(1-\alpha)\%$置信区间是$[\hat{\beta}_2 \pm t_{\alpha/2}\mathrm{se}(\hat{\beta}_2)]$。但若 $\mathrm{se}(\hat{\beta}_2)$不能无偏地加以估计，我们还能对通常计算的置信区间给予多少信赖呢？

[7] Russell Davidson and James G. MacKinnon，*Estimation and Inference in Econometrics* Oxford University Press，New York，1993，pp.549－550.

[8] 其原因在于，高斯-马尔可夫定理为 OLS 的有效性提供了充分（而非

必要）条件。OLS为BLUE的充分必要条件由**克鲁斯卡尔定理**（Kruskal's theorem）给出。但这个问题超出了本书的范围。感谢 Michael McAleer 使我注意到这一点。详尽分析，可参见 Denzil G. Fiebig, Michael McAleer, and Robert Bartels, "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances," *Journal of Econometrics*, vol.54, no.1－3, Oct.－Dec., 1992, pp.321－334。对喜欢数学的同学而言，我会在附录C中用矩阵代数进一步讨论这个问题。

[9] S.J.Prais and H.S.Houthakker, *The Analysis of Family Budgets*, Cambridge University Press, New York, 1955.

[10] 关于 $\hat{u}_i$ 与 u_i 之间的关系，参看 E.Malinvaud. *Statistical Methods of Econometrics*, North Holland Publishing Company, Amsterdam, 1970, pp.88－89。

[11] R.E.Park, "Estimation with Heteroscedastic Error Terms," *Econometrica*, vol.34, no.4, October 1966, p.888. 帕克检验是 A.C.Harvey 提出的一般检验的一种特殊情形。见 A.C.Harvey, "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, vol.44, no.3, 1976, pp.461－465。

[12] Stephen M.Goldfeld and Richard E.Quandt, *Nonlinear Methods in Econometrics*, North Holland Publishing Company, Amsterdam, 1972, pp.93－94.

[13] 帕克所选的特殊函数形式仅是他的一种建议。另一种函数形式可能表明有显著关系。例如，不妨试用 $\hat{u}_i^2$ 代替 $\ln\hat{u}_i^2$ 作为应变量。

[14] H.Glejser, "A New Test for Heteroscedasticity," *Journal of the American Statistical Association* vol.64, 1969, pp.316－323.

[15] 详见戈德菲尔德与匡特，前引文献，第3章。

[16] See G.Udny Yule and M.G.Kendall, *An Introduction to the Theory of Statistics*, Charles Griffin & Company, London, 1953, p.455.

[17] Goldfeld 与 Quandt，前引文献，第3章。

[18] 这仅是一个可取的假定。实际上，所要求的是 σ_i^2 与 X_i 有单调关系。

[19] 用专业语言说，检验的**功效**依赖于 C 的选择。在统计学中，一个检验的功效由拒绝非真的虚拟假设的概率［即 1－Prob（第Ⅱ类错误）］来衡量。这里，虚拟假设是：两组方差相同，即有同方差性。进一步的讨论，见 M.M.Ali and C.Giaccotto, "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, vol.26, 1984, pp.355－373。

[20] George G.Judge 等，前引文献，第422页。

[21] T.Breusch and A.Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, vol.47, 1979, pp.1287－1294. See also L.Godfrey, "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics*, vol.8, 1978, pp.227－236. 由于两者的相似性，故将其命名为异方差性的布劳殊-培干-戈弗雷检验。

[22] See Adrian C. Darnell, *A Dictionary of Econometrics*, Edward Elgar, Cheltenham, U.K., 1994, pp.178－179.

[23] 关于这个问题，参看 R.Koenker, "A Note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics*, vol.17, 1981, pp.1180－1200。

[24] H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity," *Econometrica*, vol.48, 1980, pp.817 - 818.

[25] 隐含于这一步骤的假定是误差 u_i 的方差 σ_i^2 与诸回归元，它们的平方以及它们的交叉乘积有函数关系。如果这个回归的全部偏斜率系数同时等于零，则误差方差是一个等于 α_1 的同方差性常数。

[26] Stephen R. Lewis, "Government Revenue from Foreign Trade," *Manchester School of Economics and Social Studies*, vol.31, 1963, pp.39 - 47.

[27] 这些结果复制自 William F. Lott and Subhash C. Ray, *Applied Econometrics: Problems with Data Sets*, Instructor's Manual, Chap. 22, pp.137 140，但用了不同的符号。

[28] 为节省自由度，有时候也可以把这个检验加以修改，见习题 11.18。

[29] See Richard Harris, *Using Cointegration Analysis in Econometrics Modelling*, Prentice Hall & Harvester Wheatsheat, U.K., 1995, p.68.

[30] See M.J. Harrison and B.P. McCabe, "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals," *Journal of the American Statistical Association*, vol.74, 1979, pp.494 - 499, J. Szroeter, "A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models," *Econometrica*, vol.46, 1978, pp.1311 - 1327; M.A. Evans and M.L. King, "A Further Class of Tests for Heteroscedasticity," *Jourant of Econometrics*, vol.37, 1988, pp.265 - 276, R. Koenker and G. Bassett, "Robust Tests for Heteroscedasticity Based on Regression Quantiles," Econometnica, vol.50, 1982, pp.43 - 61.

[31] 如第 6 章注释 [3] 所表明的，过原点回归的 R^2 和带截距模型的 R^2 不可直接相比。报道的 $R^2 = 0.999\ 3$ 已把有无截距的差异加以考虑。(关于如何修正因无截距而造成 R^2 的差异，详见 SAS 软件包，还可参阅附录 6A，第 6A.1 节。)

[32] 参看 H. White，前引文献。

[33] 更专门化的名称是**异方差性相一致协方差矩阵估计量**（heteroscedasticity-consistent covariance matrix estimators, HCCME)。

[34] William H. Greene, *Econometric Analysis*, 2d ed., Macmillan, New York, 1993, pp.385.

[35] T. Dudley Wallace and J. Lew Silver, *Econometrics: An Introduction*, Addison-Wesley, Reading, Mass., 1988, p.265.

[36] 我们回想在戈德菲尔德-匡特检验的讨论中曾遇见这种假定。

[37] 然而，一种实际的作法也许是将 $\hat{u}_i^2$ 对每一变量描图，然后决定用哪一个 X 变量去变换数据。(见图 11.9。)

[38] 有时可用 $\ln(Y_i + k)$ 或 $\ln(X_i + k)$，其中 k 是一个有待选择的正数，目的是要所有的 Y 和 X 值都变换成正数。

[39] 例如，如果 X_1，X_2 与 X_3 相互无关：$r_{12} = r_{13} = r_{23} = 0$；而我们发现比值 X_1/X_3 与 X_2/X_3 相关，则称此为谬误相关，"说得更一般，如果相关不出现在原始数据中，而是由于数据的处理方法而引发出来，则这种相关可以说是谬误的。" 见 M.G. Kendall and W.R. Buckland, *A Dictionary of Statistical Terms*, Hafner Publishing, New York, 1972, p.143。

[40] 更多的细节，见 George G. Judge 等，前引文献，14.4 节，第 415－420 页。

[41] N. Gregory Mankiw, "A Quick Refresher Course in Macroeconomics," *Journal of Economic Literature*, vol. XXVIII, December, 1990, p. 1648.

[42] John Fox, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage Publications, California, 1997, p. 306.

[43] 前引文献第 307 页。

[44] 注意我们为了得到方差已经把标准误平方了。

第 12 章　自相关：误差项相关会怎么样？

441 读者或许记得，经验分析通常有三种数据可用：（1）横截面数据；（2）时间序列数据；（3）横截面数据与时间序列数据的综合，即混和数据。我们在第 1 篇提出经典线性回归模型时做了几个假定，第 7.1 节有所讨论。不过，我们注意到，不是所有这些假定对每种数据类型都成立。事实上，我们在上一章看到，同方差或误差方差相等的假定在横截面数据中并非总是合理的假定。换言之，横截面数据时常受到异方差问题的纠缠。

不过，由于在横截面研究中，数据的搜集通常都基于横截单位的一个随机样本，比如（消费函数分析中的）家庭或（投资研究分析中的）企业，所以没有先验理由认为，一个家庭或企业的误差项与另一个家庭或企业的误差项会相关。如果碰巧在横截单位中观察到了这种相关，则称之为**空间自相关**（spatial autocorrelation），即空间而非时间上的相关。然而，重要的是要记得，在横截面分析中，为了使判断（空间）自相关是否存在的道理讲得过去，数据的顺序必须有某种逻辑或经济上的意义。

然而，当我们处理时间序列数据时，情况就极为不同了。因为这种数据中的观测服从时间上的一种自然顺序，所以连续的观测很可能表现出内部的
442 相关，特别是在连续观测的时间区间很短时，如一天、一周、一个月而非一年。如果你每天观测诸如道琼斯或标准普尔 500 等股票价格指数，发现这些指数接连几天上涨或下降再平常不过了。显然，在这种情形下，假定在

CLRM 背后的误差项不存在自相关或序列相关就不合常理。

本章中我们从严分析这一假定，以期得到对下列问题的回答：

1． 自相关的性质是什么？
2． 自相关会导致什么理论上和实际上的后果？
3． 由于非自相关假定涉及不可观测的干扰项 u_i，怎样能知道在一个任给的情况中是否有自相关？
4． 怎样补救自相关的问题？

读者将发现本章在许多方面类似于前章对异方差性的讨论。因为，**在自相关情况下，和在异方差情况下一样，平常的 OLS 估计量虽然仍是线性、无偏和渐近（即在大样本中）正态分布的[1]，但不再是所有线性无偏估计量中的最小方差者。简言之，相对其他线性无偏估计量而言，它不再是有效的，换言之，OLS 估计量不再是 BLUE。结果，通常的 t、F 和 χ^2 都不再成立。**

§ 12.1 问题的性质

自相关（autocorrelation）一词可定义为按时间（如在时间序列数据中）或空间（如在横截面数据中）排序的观测值序列的成员之间的相关。[2]在回归的论述中，经典线性回归模型假定在干扰项 u_i 之间不存在有自相关，用符号表示：

$$E(u_iu_j) = 0 \qquad i \neq j \tag{3.2.5}$$

简单地说，经典模型假定，任一次观测的干扰项都不受任何其他观测的干扰项影响。例如，在做产出对劳动力和资本投入的回归中，我们用了季度时间序列数据。如果某一季度的产出受到罢工的影响，却没有理由认为这一生产
448 中断会持续到下一季度，就是说，即使本季度产出下降，却没有理由预期下一季度的产出也下降。同理，如果我们在家庭消费支出对家庭收入的回归中，用了横截面数据，那么，一个家庭的收入增加对其消费支出的影响，应该不会波及到另一个家庭的消费支出。

然而，如果存在有这种相依性，我们就有了自相关。用符号表示：

$$E(u_iu_j) \neq 0 \qquad i \neq j \tag{12.1.1}$$

这时，由本季罢工引起的生产中断很可能影响下季的产出；或者，一个家庭的消费支出增加，很可能促使另一个家庭为了同邻居攀比也随之增大其消费支出。

在我们寻求自相关存在的原因前，澄清一些名词术语方面的问题是必要

的。虽然现在把名词**自相关**（autocorrelation）和**序列相关**（serial correlation）看作同义语已成为习惯，但一些作者比较喜欢把两者区分开来。例如，廷特纳（Tintner）定义自相关为"一给定序列同它自身滞后若干期的序列的滞后相关"。而与此同时，他把序列相关一词保留作"两个不同序列的滞后相关"[3]。例如，时间序列 u_1，u_2，…，u_{10}和它自身滞后一期的序列 u_2，u_3，…，u_{11}之间的相关叫做自相关，而两个不同时间序列 u_1，u_2，…，u_{10}和 v_2，v_3，…，v_{11}之间的相关叫做序列相关。尽管两名词的区分有一定用处，本书中将把它们看作同义语。

让我们想像自相关和无自相关的一些可能的模式，如图 12.1 所示。图 12.1a 到 d 各显示 u_i 中的一个可分辨的模式。图 12.1（a）显示一个周期模样；图 12.1（b）和（c）分别表明干扰中的一个上升或下降的线性趋势；而图 12.1（d）表明干扰中兼有线性和二次趋势项。惟有图 12.1（e）表示无系统性模样，符合于经典线性回归模型的无相关假定。

一个自然的问题是：为什么会出现序列相关？有种种理由，其中的一些如下：

惯性。大多数经济时间序列都有一个明显的特点，就是它的惯性或粘
445 滞。众所周知，GNP、价格指数、生产、就业和失业等时间序列都呈现（商业）循环。从衰退的谷底开始，当经济开始复苏时，大多数经济序列开始上升。在此上升期间，序列在每一时刻的值都高于前一时刻的值。看来有一种内在的动力驱使这一势头继续下去，直至某些情况（如利率或课税或两者同时升高）出现才把它拖慢下来。因此，在涉及时间序列的回归中，相继的观测值很可能是相依赖的。

设定偏误：应含而未含变量的情形。在经验分析中，研究者常从一个较好但不一定"最好"的回归模型开始。经过回归分析，研究者做事后检查，看结果是否与事先预期的相一致。如不一致，便要开始动手术。例如，研究者可能将拟合回归的残差描图，并可能观察到类似于图 12.1（a）到（d）那样的模式。这些残差（为 u_i 的替代变量）也许表明，某些原来待选的、却未被包含的变量，由于种种理由应被包含到模型中来。这就是**应含而未含变量**（excluded variable）的设定偏误。但这些变量包含进来常常能消除干扰中所观察到的相关模式。例如，假如我们有以下的需求模型：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \tag{12.1.2}$$

其中 Y_t = 牛肉需求量，X_2 = 牛肉价格，X_3 = 消费者收入，X_4 = 猪肉价格，t = 时间。[4]然而，为了某种理由，我们做了下述回归：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \tag{12.1.3}$$

现在，如果（12.1.2）是"正确"的模型或"真理"或真实的关系式，则回归（12.1.3）就无异于令 $v_t = \beta_4 X_{4t} + u_t$。于是在猪肉价格影响牛肉消费的限度内，误差或干扰项 v 将反映出一种系统性模式，从而造成（错误的）自相关。真的有自相关吗？一个简单的检验是，同时做回归（12.1.2）和

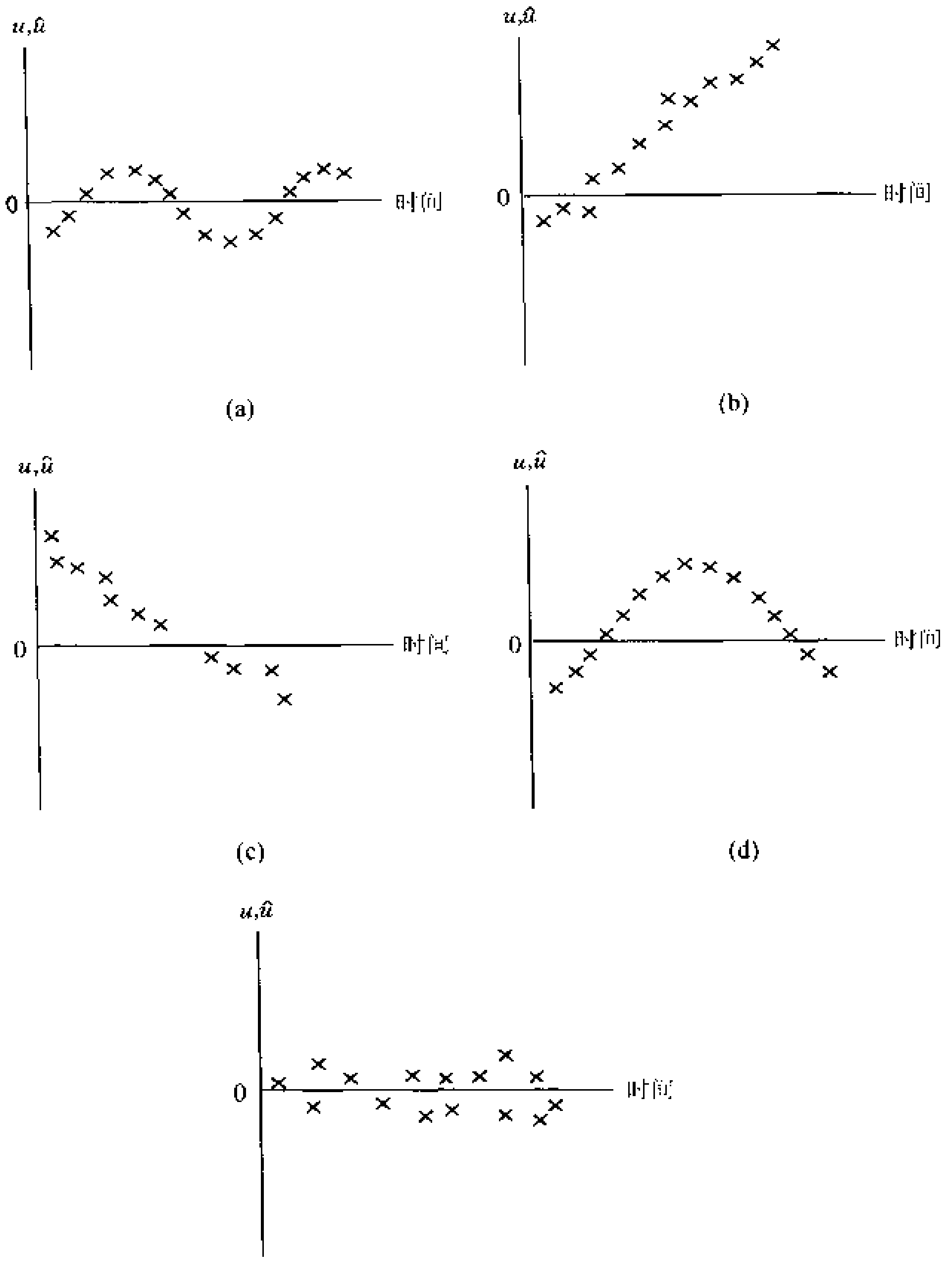


图 12.1　自相关模式和非自相关类型

(12.1.3)，看看在做（12.1.2）时，原先在模型（12.1.3）中观测到的自相关（如果观测有的话）是否消失。[5]

侦察自相关的实际步骤将在 12.5 节中讨论。届时我们将看到，用回归（12.1.2）的残差来描图，常常会在很大程度上有助于看清楚自相关的问题。

446 **设定偏误：不正确的函数形式。**假使在成本—产出研究中“真实”或“正确”的模型如下：

$$边际成本_i = \beta_1 + \beta_2 产出_i + \beta_3 产出_i^2 + u_i \qquad (12.1.4)$$

但我们拟合了以下的模型：

$$边际成本_i = \alpha_1 + \alpha_2 产出_i + v_i \qquad (12.1.5)$$

如图 12.2，连同“不正确”的线性成本曲线一起，展示了“真实”模型的边际成本曲线。

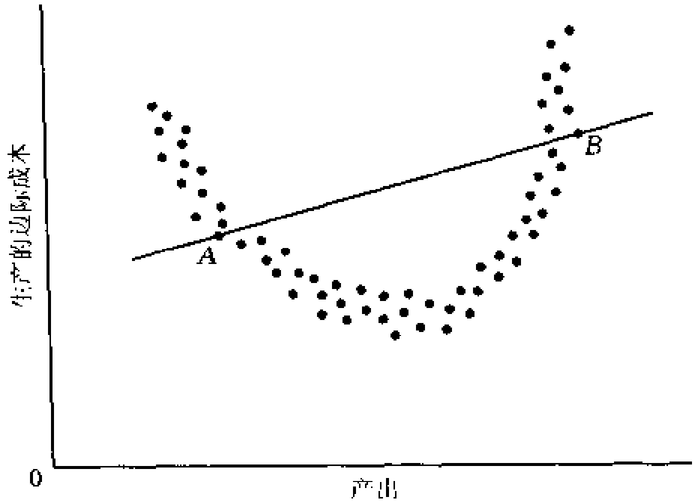


图 12.2 设定偏误：不正确的函数形式

如图 12.2 所示，在点 A 与 B 之间的线性边际成本曲线一直在高估真实的边际成本，而在这两点之外则一致地低估真实的边际成本。这种结果是可以预料到的，因为干扰项 v_i 事实上等于产出$^2 + u_i$，从而包含了（产出）^{2} 对边际成本的系统影响。在这种情形中，由于错误函数形式的使用，v_i 将要反映出自相关。在第 13 章中，我们将考虑侦察设定偏误的若干方法。

蛛网现象。许多农产品的供给反映一种所谓的蛛网现象（Cobweb phenomenon）。供给对价格的反应要滞后一个时期，是因为供给需要经过一定的时间才能实现（有一孕育期）。例如，今年年初的作物种植是受去年流行的价格影响的。因此，相关的供给函数是：

$$供给_t = \beta_1 + \beta_2 P_{t-1} + u_t \qquad (12.1.6)$$

假使时期 t 的期末价格 P_t 低于 P_{t-1}，农民就很可能决定在时期 $t+1$ 生产比时期 t 更少的东西。显然，在这种情形中，农民由于在年度 t 的过量生产很
447 可能在年度 $t+1$ 削减他们的产量。诸如此类的现象，就不能期望干扰项 u_t 是随机的，从而导致一种蛛网模式。

滞后效应（lags）。在一个消费支出对收入的时间序列回归中，人们常常发现当前时期的消费支出除了依赖于其他变量外，还依赖于前期的消费支出，就是：

$$消费_t = \beta_1 + \beta_2 收入_t + \beta_3 消费_{t-1} + u_t \qquad (12.1.7)$$

像（12.1.7）这样的回归，由于解释变量之一是因变量的滞后值而被称为**自回归**（autoregression）。（我们将在第 17 章中研究这种类型。）关于（12.1.7）这样一类模型的合理性是简单的。由于心理上、技术上以及制度上的原因，消费者是不会轻易改变他们的消费习惯的。如果现在我们忽略了（12.1.7）中的滞后项，所造成的误差项将由于滞后消费对当前消费的影响而反映出一种系统性模式。

数据的"编造"。在经验分析中，原始数据往往是经过"编造的"。例如，在用到季度数据的时间序列回归中，这些数据通常来自每月数据，不过是把 3 个月的观测值加在一起除以 3 罢了。这种平均的计算减弱了每月数据的波动而引进了数据中的匀滑性。因此，用季度数据描绘的图形要比用月度数据看来匀滑得多。这种匀滑性本身就能使干扰项中出现有系统性样式，从而导入自相关。数据编造的另一来源是数据的**内插**（interpolation）或**外推**（extrapolation）。例如在美国每 10 年举行一次人口普查。最近的一次在 1990 年，而此前的一次在 1980 年。假如现在需要两个普查年间 1980—1990 年的某年数据，通常的做法就是，根据某些特殊的假定进行内插。所有这些数据"糅合"技术，都会给数据带来原始数据所没有的系统性样式。[6]

数据变换。作为一个例子，考虑如下模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.1.8}$$

其中 Y = 消费支出，X = 收入。由于（12.1.8）在每个时期都成立，所以它在上一时期（$t-1$）也成立。于是，可以把（12.1.8）写成

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \tag{12.1.9}$$

448 Y_{t-1}、X_{t-1}和 u_{t-1}分别被称为 Y、X 和 u 的**滞后值**（lagged values），这里滞后一期。我们在本章稍后部分及本书其他几个地方将会看到滞后值的重要性。

现在，如果我们从（12.1.8）中减去（12.1.9），则得到

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \tag{12.1.10}$$

其中 Δ 表示**一阶差分算子**（first difference operator），告诉我们对所讨论变量连续取差分。因而，$\Delta Y_t = (Y_t - Y_{t-1})$，$\Delta X_t = (X_t - X_{t-1})$，$\Delta u_t = (u_t - u_{t-1})$。出于经验目的，我们把（12.1.10）写成

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \tag{12.1.11}$$

其中 $v_t = \Delta u_t = u_t - u_{t-1}$。

方程（12.1.9）被称为**水平型**（level form），方程（12.1.10）被称为**一阶差分型**（first difference form）。经验分析中，这两种形式都经常使用。比如在（12.1.9）中，若 Y 和 X 表示消费支出和收入的对数，则（12.1.10）中的 ΔY 和 ΔX 将表示消费支出和收入的对数的变化。但如我们所知，一个变量的对数的变化是其相对变化或百分比变化（如果将前者乘以 100）。所以，除了研究变量水平值之间的关系外，我们或许还对其增长型关系感兴趣。

现在，如果（12.1.8）中的误差项满足标准OLS假定，特别是无自相关的假定，那么，可以证明，（12.1.11）中的误差项 v_t 将会自相关。（附录12A的第12A.1节给出了证明。）这里或许注意到，像（12.1.11）这样的模型被称为**动态回归模型**（dynamic regression models），即含有滞后回归子的模型。我们将在第17章深入研究这种模型。

上述例子旨在说明，自相关有时候是作为对原模型进行变换的结果而产生。

非平稳性。我们在第1章曾提到，在处理时间序列数据时，我们必须明确一个给定的时间序列是否平稳。尽管我们在本书第4篇全面讨论时间序列计量经济学的章节中将会讨论非平稳时间序列的专题，但这里粗略地讲，如果一个时间序列的特征（如均值、方差和协方差）**不随时间而变化**（time invariant），那它就是一个平稳的时间序列。否则，就是一个非平稳时间序列。

如我们在第4篇所讨论的那样，在一个诸如（12.1.8）之类的回归模型中，Y 和 X 很可能都是非平稳的，因此误差 u 也是非平稳的。[7] 此时，误差项将表现出自相关。

449 总之，有很多原因导致一个回归模型中的误差项自相关。在本章的余下部分，我们将较详细地讨论自相关所带来的问题及其解决办法。

还应指出，虽然多数经济时间序列都因在一个较长时期内或者上升或者下降而表现出正的自相关，而不像图12.3（b）那样表现为一上一下的恒常运动，但自相关可以是正的或负的。

§12.2 自相关出现时的OLS估计量

如果我们在干扰中通过假定 $E(u_t u_{t+S}) \neq 0 (S \neq 0)$ 引进自相关，但保留
450 经典模型的全部其他假定，对OLS估计量及其方差来说，会出现什么情况呢？[8] 再次注意到，我们在干扰右下角用角标 t 来强调我们在处理时间序列数据。

我们再一次回到双变量模型，以解释所涉及的基本概念。就是，回到 $Y_t = \beta_1 + \beta_2 X_t + u_t$。要取得任何进展都必须对产生 u_t 的机制做出假定，因为 $E(u_t u_{t+s}) \neq 0 (s \neq 0)$ 这个假定过于一般，不会有任何实际用处。作为一个起点或一阶近似，不妨假定干扰是这样产生的：

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \tag{12.2.1}$$

其中 ρ 被称为**自协方差系数**（coefficient of autocovariance），并且 ε_t 是满足以下标准OLS假定的随机干扰：

$$\begin{aligned} &E(\varepsilon_t) = 0 \\ &\operatorname{var}(\varepsilon_t) = \sigma_\varepsilon^2 \\ &\operatorname{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0 \qquad s \neq 0 \end{aligned} \tag{12.2.2}$$

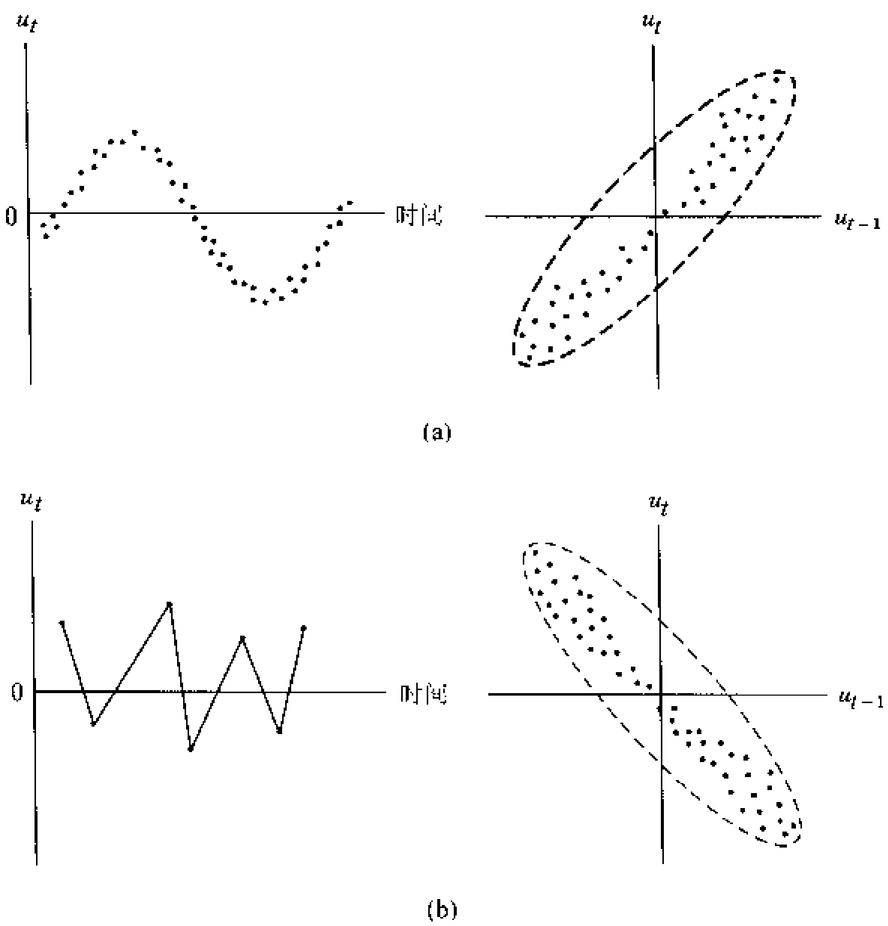


图 12.3　(a) 正的自相关和 (b) 负的自相关

在工程文献中，具有上述性质的误差项通常被称为**白噪音误差项**（white noise error term）。(12.2.1) 阐明，t 时期干扰项的值等于 ρ 乘以其上一期干扰值与一个纯粹随机误差项之和。

方案 (12.2.1) 被称为**马尔可夫一阶自回归模式**（Markov first-order autoregressive scheme）或简称**一阶自回归模式**，常记为**AR**(1)。由于 (12.2.1) 可解释为 u_t 对其自身滞后一期的回归，取名为自回归是适宜的。因仅涉及 u_t 及其最近的过去值，即最大滞后是 1，故称一阶。如果模型设为 $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$，它将是AR(2)或二阶自回归模式，如此类推。我们在第 4 篇有关时间序列计量经济学的章节中将再次考查这种高阶模式。

顺便指出，自协方差系数 ρ 又可解释为一阶自相关系数（first-order coefficient at autocorrelation），或说得更准确，滞后 1 期的自相关系数（coefficient of autocorrelation at lag 1）。[9]

451　给定AR(1)模式，可以证明（参见附录 12A、第 12A.2 节）：

$$\operatorname{var}(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.3}$$

$$\operatorname{cov}(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{12.2.4}$$

$$\operatorname{cor}(u_t, u_{t+s}) = \rho^s \tag{12.2.5}$$

其中$\operatorname{cov}(u_t, u_{t+s})$指相差 s 期的误差项之间的协方差，$\operatorname{cor}(u_t, u_{t+s})$指相差 s 期的误差项之间的相关系数。注意，由于协方差和相关系数的对称性，所以 $\operatorname{cov}(u_t, u_{t+s}) = \operatorname{cov}(u_t, u_{t-s})$，$\operatorname{cor}(u_t, u_{t+s}) = \operatorname{cor}(u_t, u_{t-s})$。

由于 ρ 是一个介于 -1 与 1 之间的常数，所以（12.2.3）表明，在AR(1)模式下，u_t 的方差仍是同方差的，但 u_t 不仅与其过去一期的值相关，与过去几期的值也相关。关键是注意到，$|\rho|<1$，即 ρ 的绝对值小于 1。比如，若 $\rho=1$，上述的方差和协方差都没有定义。若 $|\rho|<1$，我们说（12.2.1）中给出的AR(1)过程是平稳的；即 u_t 的均值、方差和协方差不随时间而变化。若$|\rho|<1$，则从（12.2.4）可见，协方差的值将随着我们越往过去看而越小。我们以后会看到上述结论的作用。

我们使用AR(1)过程的原因之一，不仅因为其相对高阶模式的简单性，还因为在许多应用中都证明了它相当有用。此外，对AR(1)模式所做的理论和经验研究数量也相当可观。

现在回到我们的双变量回归模型：$Y_t = \beta_1 + \beta_2 X_t + u_t$。我们从第 3 章知道，斜率系数的 OLS 估计量为：

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{12.2.6}$$

其方差为：

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2} \tag{12.2.7}$$

其中小写字母如通常一样表示对均值的离差。

现在，在AR(1)模式下，可以证明此估计量的方差为：

$$\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right] \tag{12.2.8}$$

其中 $\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}$表示 $\hat{\beta}_2$ 在一阶自回归模式下的方差。

452 （12.2.8）与（12.2.7）的比较表明，前者等于后者乘上带方括号的一项，这一项既取决于 ρ，又取决于回归元 X 在各种滞后值之间的样本自相关系数。[10]一般而言，我们不能预先知道$\operatorname{var}(\hat{\beta}_2)$是小于还是大于$\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}$［但参见下面的方程（12.4.1）］。当然，若 $\rho=0$，则这两个公式完全一样。（为什么？）而且，若回归元的连续值之间的相关系数很小，则斜率估计量通常的 OLS 方差将不会严重偏误。但作为一个一般原则，这两个方差不应该相同。

为了解（12.2.7）和（12.2.8）中给出的方差之间的差别，假定回归元

X 也服从含有自相关系数 r 的一阶自回归模式。那么，可以证明，(12.2.8) 简化成：

$$\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}=\frac{\sigma^2}{\sum x_t^2}\left(\frac{1+r\rho}{1-r\rho}\right)=\operatorname{var}(\hat{\beta}_2)_{\mathrm{OLS}}\left(\frac{1+r\rho}{1-r\rho}\right) \tag{12.2.9}$$

比如，若 $r=0.6$ 和 $\rho=0.8$，我们利用（12.2.9）可以验证 $\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}=2.846\ 1\operatorname{var}(\hat{\beta}_2)_{\mathrm{OLS}}$。换言之，$\operatorname{var}(\hat{\beta}_2)_{\mathrm{OLS}}=\frac{1}{2.846\ 1}\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}=0.351\ 3\operatorname{var}(\hat{\beta}_2)_{\mathrm{AR1}}$。即通常的 OLS 公式［即（12.2.7）］将 $(\hat{\beta}_2)_{\mathrm{AR1}}$ 的方差低估了约 65%。你或许意识到，这个答案是给定 r 和 ρ 的值的情况下的特定结果。但这个练习的目的是给你一个警告，盲目地使用通常的 OLS 公式来计算 OLS 估计量的方差和标准误，可能会给出严重误导性的结论。

假使我们继续使用 OLS 估计量 $\hat{\beta}_2$ 并在平常的方差公式中把 AR（1）模式考虑进来而将方差加以适当调整。就是说，我们使用（12.2.6）中的 $\hat{\beta}_2$，但使用（12.2.8）所给的方差公式。这时 $\hat{\beta}_2$ 的性质如何？容易证明，$\hat{\beta}_2$ 仍然是线性和无偏的。事实上，如附录 3A 第 3A.2 节所表明，无序列相关性和无异方差性一样，都不是证明 $\hat{\beta}_2$ 为无偏所必需的。$\hat{\beta}_2$ 是否仍然是 BLUE 呢？可惜不是；它在线性无偏估计量一类中不是最小方差的。总之，$\hat{\beta}_2$ 虽然线性无偏，但欠缺有效性（当然，相对而言）。读者必定注意到这一发现非常类似于发现在异方差性出现时，$\hat{\beta}_2$ 是较为低效的情形。我们已经看到，作为广义最小二乘估计量的特殊情形，由（11.3.8）给出的加权最小二乘估计量 $\hat{\beta}_2^*$ 是有效的。在自相关的情形中，我们能找到一个是 BLUE 的估计量吗？回答是肯定的。这可从下节的讨论中看出。

§12.3 自相关出现时的 BLUE 估计量

453 继续利用双变量模型并假定AR(1)过程，可以证明 β_2 的 BLUE 估计量由下式给出[11]：

$$\hat{\beta}_2^{\mathrm{GLS}}=\frac{\sum_{t=2}^{n}(x_t-\rho x_{t-1})(y_t-\rho y_{t-1})}{\sum_{t=2}^{n}(x_t-\rho x_{t-1})^2}+C \tag{12.3.1}$$

其中 C 是一校正因子，在实际中可以忽略。注意下标从 $t=2$ 变到 $t=n$。从而它的方差是：

$$\operatorname{var}\hat{\beta}_2^{\mathrm{GLS}}=\frac{\sigma^2}{\sum_{t=2}^{n}(x_t-\rho x_{t-1})^2}+D \tag{12.3.2}$$

其中 D 也是实际中可以忽略的一个校正因子。(参看习题 12.18。)

估计量 $\hat{\beta}_2^{GLS}$ 正如其上标所表明的，是由 GLS 法得到的。在第 11 章中我们看到，在 GLS 中我们通过变量转换把额外的信息包括到估计程序中去，而在 OLS 中我们并不直接考虑这种附加的信息。读者可以看到，(12.3.1) 中的 β_2 的 GLS 估计量把自相关参数 ρ 包含在估计公式中，而 (12.2.6) 的 OLS 公式则对 ρ 干脆不理睬。直观上，这就是为什么 GLS 估计量是 BLUE，而 OLS 估计量不是。——GLS 估计量最大地利用了现有的信息。[12] 不言而喻，如果 $\rho=0$ 就没有额外的信息需要考虑，从而 GLS 和 OLS 两估计量是相同的。

总之，在自相关情形中，由 (12.3.1) 给出的 GLS 估计量才是 BLUE，并且这时的最小方差由 (12.3.2) 给出。(12.2.8)，尤其是 (12.2.7)，都没有给出最小方差。

454 **一个技术性注解。**我们在上一节曾指出，马尔可夫定理只给出 OLS 为 BLUE 的充分条件。上一章还提到，OLS 为 BLUE 的充要条件由**克拉斯卡尔 (Krushkal) 定理**给出。因此，在有些情况下，尽管存在自相关，OLS 仍可能是 BLUE 的，只是这种情况在实践中不经常发生而已。

如果我们忽视自相关，一厢情愿地使用 OLS 程序，会出现什么情况？请看下节分解。

§12.4 自相关出现时使用 OLS 的后果

如同异方差性情形，在自相关出现时，OLS 估计量仍是线性无偏和一致性的，但不再是有效（亦即最小方差）的了。那么，如果我们继续使用 OLS 估计量，我们平常的假设检验程序会遇到什么问题呢？再次，如同异方差性的情形，我们区分两种情况。为了教学的目的，我们仍利用双变量模型。虽然如此，把以下的讨论推广到复回归并无多少困难。[13]

考虑到自相关的 OLS 估计

正如已指出的，$\hat{\beta}_2$ 不是 BLUE，即使我们使用 $\operatorname{var}(\hat{\beta}_2)_{AR1}$，由此得来的置信区间很可能比根据 GLS 程序得到的要宽些。如克曼塔所表明的，即使样本无限地增大，结果也很可能如此。[14] 就是说，$\hat{\beta}_2$ 并非渐近有效的。这一发现对假设检验的含义是明显的：我们很可能宣称一个系数是统计上不显著的（即无异于零），尽管事实上（即根据正确的 GLS 程序）它也许是显著的。这一差别可从图 12.4 明显看出。图中我们做出在真 $\beta_2=0$ 假设下的
455 95% OLS [AR (1)] 和 GLS 两置信区间。考虑 β_2 的一个具体的估计值，比方说，b_2。由于 b_2 落在 OLS 置信区间内，我们可以 95% 的置信度接受真 β_2

为零的假设。但如果我们使用（正确的）GLS 置信区间，由于 b_2 落入拒绝域，我们就会拒绝真 β_2 为零的虚拟假设。

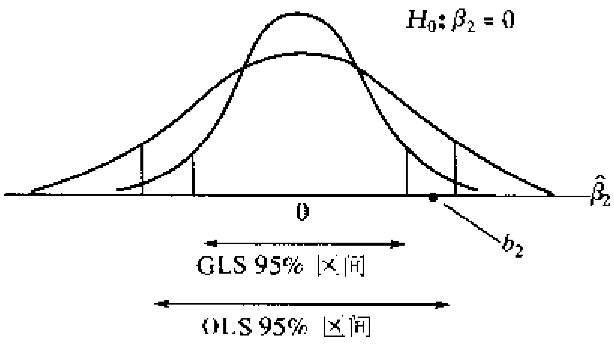


图 12.4　GLS 和 OLS 95%置信区间

一条信息：尽管 OLS 估计量是无偏和一致性的，但为了建立置信区间并检验假设，要用 GLS 而不用 OLS。（然而，参见后面的第 12.11 节。）

忽视自相关的 OLS 估计

如果我们不但使用 $\hat{\beta}_2$ 而且继续使用 $\mathrm{var}(\hat{\beta}_2) = \sigma^2 / \sum x_i^2$，完全不考虑自相关的问题，也就是我们错误地认为通常关于经典模型假定成立，情况就会变得非常严重。错误将出自多种原因：

1. 残差方差 $\hat{\sigma}^2 = \sum \hat{u}_t^2/(n-2)$ 很可能低估了真实的 σ^2。

2. 结果，我们很可能高估了 R^2。

3. 即使没有低估 σ^2，$\mathrm{var}(\hat{\beta}_2)$ 也可能低估了 $\mathrm{var}(\hat{\beta}_2)_{\mathrm{AR1}}$〔方程 12.2.8〕这个（一阶）自相关情形下的方差，虽然 $\mathrm{var}(\hat{\beta}_2)_{\mathrm{AR1}}$ 和 $\mathrm{var}(\hat{\beta}_2)^{\mathrm{GLS}}$ 比较起来是低效的。

4. 因此，通常的 t 和 F 显著性检验都变成无效的了。如果仍然使用这些检验，就很可能对所估计的回归系数做出有严重错误的统计显著性结论。

为了确认这些命题中的一些问题，让我们回到双变量模型。由第 3 章知，在经典假定下：

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)}$$

给出 σ^2 的一个无偏估计量，即 $E(\hat{\sigma}^2) = \sigma^2$。但若自相关出现，为 AR(1) 的情形，则可以证明：

$$E(\hat{\sigma}^2) = \frac{\sigma^2\{n - [2/(1-\rho)] - 2\rho r\}}{n-2} \tag{12.4.1}$$

其中 $r = \sum_{t=1}^{n-1} x_t x_{t-1} / \sum_{t=1}^{n} x_t^2$，可解释为 X 的相继观测值之间的（样本）相关系数。[15] 如果 ρ 和 r 都是正的（这对大多数经济时间序列来说都是一个

456 适当的假定)，则从（12.4.1）显见$E(\hat{\sigma}^2)<\sigma^2$；就是说，通常的残差方差公式平均而言低估了真$\sigma^2$。换言之，$\hat{\sigma}^2$偏之于过低。不言而喻，$\hat{\sigma}^2$中的这一偏误将被传递到$\text{var}(\hat{\beta}_2)$中来，因为实际上后者是由公式$\hat{\sigma}^2/\sum x_t^2$估计的。

但即使σ^2未被低估，$\text{var}(\hat{\beta}_2)$也是$\text{var}(\hat{\beta}_2)_{\text{AR1}}$的一个偏误估计量。这很容易通过（12.2.7）和（12.2.8）两个不同公式的比较看出。[16]事实上，如果ρ是正的（对大多数经济时间序列来说都是对的）并且X值是正相关的(大多数经济时间序列也都如此)，那么显然有：

$$\text{var}(\hat{\beta}_2)<\text{var}(\hat{\beta}_2)_{\text{AR1}} \tag{12.4.2}$$

就是说，通常的$\hat{\beta}_2$的OLS方差低估了在AR(1)下的$\hat{\beta}_2$的方差，见方程(12.2.9)。因此，如果我们使用$\text{var}(\hat{\beta}_2)$，我们就夸大了估计量$\hat{\beta}_2$的精密度(即低估了它的标准误)。结果在$t$比率的计算$t=\hat{\beta}_2/\text{se}(\hat{\beta}_2)$中（在$\beta_2=0$的假设下)，我们过高地估计了$t$值，从而夸大了$\hat{\beta}_2$的显著性。如果再加上$\sigma^2$的低估，情况就会变得更糟。

为了看到σ^2和$\hat{\beta}_2$的方差是怎样常被OLS低估的，让我们做如下的**蒙特卡罗实验**。假使在一双变量模型中我们“知道”真$\beta_1=1$和$\beta_2=0.8$，那么这个随机的PRF是：

$$Y_t=1.0+0.8X_t+u_t \tag{12.4.3}$$

从而，

$$E(Y_t\mid X_t)=1.0+0.8X_t \tag{12.4.4}$$

给出了真的总体回归线。现假定u_t由如下的一阶自回归模式产生：

$$u_t=0.7u_{t-1}+\varepsilon_t \tag{12.4.5}$$

其中ε_t满足所有的OLS假定。为方便起见，我们进一步假定ε_t是有零均值和单位方差的正态变量。方程（12.4.5）拟定，相继干扰是正相关的，其自相关系数是一个相当高的值+0.7。

现在用一张方差为1均值为零的正态随机数表，按照方案（12.4.5）产生10个随机数，如表12.1所示。为了启动此实验，还需要给定u的一个初始值，比方说，$u_0=5$。

457

表 12.1　正自相关误差项的一个人为例子

	ε_t^*	$u_t=0.7u_{t-1}+\varepsilon_t$
0	0	$u_0=5$(假设)
1	0.464	$u_1=0.7(5)+0.464=3.964$
2	2.026	$u_2=0.7(3.964)+2.026\ 2=4.800\ 8$
3	2.455	$u_3=0.7(4.801\ 0)+2.455=5.815\ 7$
4	-0.323	$u_4=0.7(5.815\ 7)-0.323=3.748\ 0$
5	-0.068	$u_5=0.7(3.748\ 0)-0.068=2.555\ 6$

6	0.296	$u_6=0.7(2.555\ 6)+0.296=2.084\ 9$
7	-0.288	$u_7=0.7(2.084\ 9)-0.288=1.171\ 4$
8	1.298	$u_8=0.7(1.171\ 4)+1.298=2.118\ 0$
9	0.241	$u_9=0.7(2.118\ 0)+0.241=1.723\ 6$
10	-0.957	$u_{10}=0.7(1.723\ 6)-0.957=0.249\ 5$

* 资料取自 *A Million Random Digits and One Hundred Thousand Deviates*, Rand Corporation, Santa Monica, Calif., 1950。

将表 12.1 中产生的 u_t 描点，我们得到图 12.5。该图表明，开始时每一相继的 u_t 大于它前面的值，后来，一般地说，便小于它前面的值。这种情
458 形通常表明一种正的自相关。

现在假如把 X 值固定在 1，2，3，…，10。那么，给定这些 X 值就能按（12.4.3）产生 10 个 Y 值，连同表 12.1 中所给的 u 值，一并放到表 12.2 中，如果我们利用表 12.2 的数据做 Y 对 X 的回归，就会得到以下的（样本）回归：

$$\begin{aligned}\hat{Y}_t &= 6.545\ 2+0.305\ 1X_t \\ &\quad(0.615\ 3)(0.099\ 2) \\ t &= (10.636\ 6)(3.076\ 3) \\ &\qquad r^2=0.541\ 9 \quad \hat{\sigma}^2=0.811\ 4\end{aligned} \tag{12.4.6}$$

而真回归线则由（12.4.4）给出。这两条回归线由图 12.6 一同示出。该图清楚地表明，所拟合的回归线在多大程度上歪曲了真回归线；它严重地低估了真斜率系数而高估了真截距。（但应注意，OLS 估计量仍然是无偏的。）

表 12.2　　　　Y 样本值的发生

X_t	u_t^*	$Y_t=1.0+0.8X_t+u_t$
1	3.964 0	$Y_1=1.0+0.8(1)\ +3.964\ 0=5.764\ 0$
2	4.801 0	$Y_2=1.0+0.8(2)\ +4.800\ 8=7.400\ 8$
3	5.815 7	$Y_3=1.0+0.8(3)\ +5.815\ 7=9.215\ 7$
4	3.748 0	$Y_4=1.0+0.8(4)\ +3.748\ 0=7.948\ 0$
5	2.555 6	$Y_5=1.0+0.8(5)\ +2.555\ 6=7.555\ 6$
6	2.084 9	$Y_6=1.0+0.8(6)\ +2.084\ 9=7.884\ 9$
7	1.171 4	$Y_7=1.0+0.8(7)\ +1.171\ 4=7.771\ 4$
8	2.118 0	$Y_8=1.0+0.8(8)\ +2.118\ 0=9.518\ 0$
9	1.723 6	$Y_9=1.0+0.8(9)\ +1.723\ 6=9.923\ 6$
10	0.249 5	$Y_{10}=1.0+0.8(10)+0.249\ 5=9.249\ 5$

* 得自表 12.1。

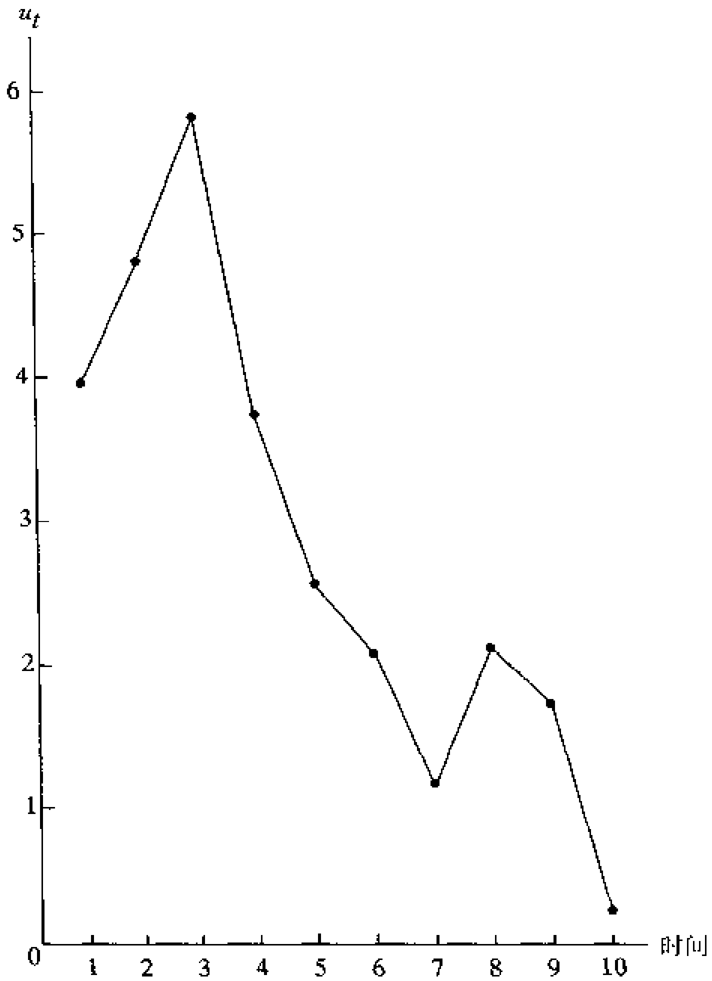


图 12.5　由模式 $u_t = 0.7u_{t-1} + \varepsilon_t$ 产生的相关（表 12.1）

图 12.6 还表明为什么 u_i 的真方差常被算自 $\hat{u}$ 的估计量 $\hat{\sigma}^2$ 所低估。$\hat{u}_t$ 一般都靠近拟合线（由于 OLS 程序），然而却远离了真 PRF。因此，$\hat{u}_i^2$ 没有给出 $\hat{u}_i$ 的正确图像。为了洞察真 σ^2 被低估的程度，假令我们做另一抽样实验，仍保留表 12.1 和 12.2 中的 X_t 值和 ε_t 值，但假定 $\rho = 0$，即无自相关。由此产生的 Y 的新样本值见于表 12.3。

表 12.3　零序列相关的 Y 样本值

X_t	$\varepsilon_t = u_t^*$	$Y_t = 1.0 + 0.8X_t + \varepsilon_t$
1	0.464	2.264
2	2.026	4.626
3	2.455	5.855
4	−0.323	3.877
5	−0.068	4.932
6	0.296	6.096
7	0.288	6.312

8	1.298	8.698
9	0.241	8.441
10	−0.957	8.043

* 因为没有自相关，所以 u_t 和 ϵ_t 相同。这些 ϵ_t 来自表 12.1。

459

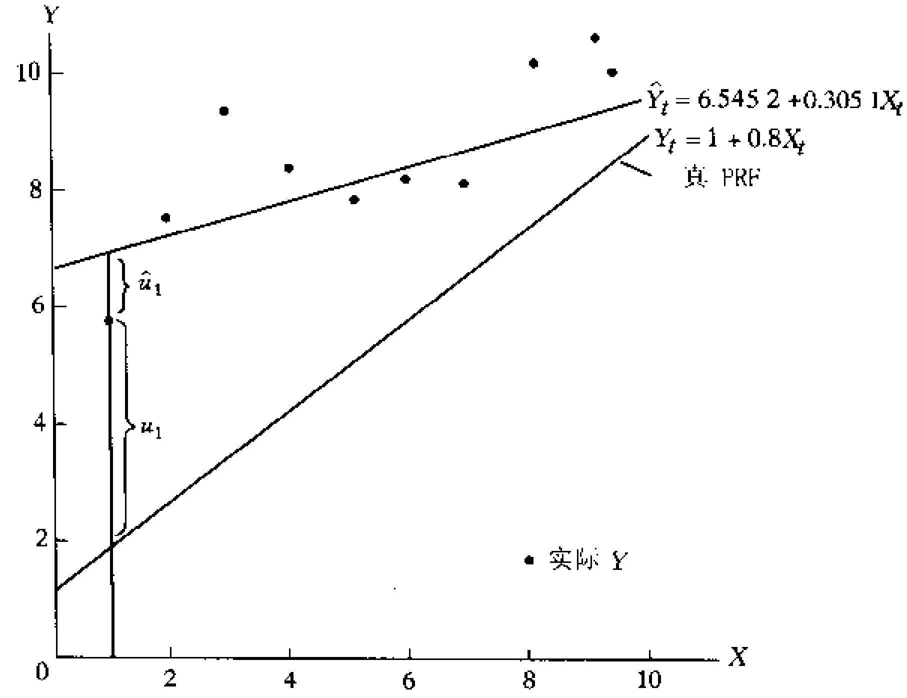


图 12.6　表 12.2 所给数据的真 PRF 与估计的回归线

根据表 12.3 得到的回归如下：

$$
\begin{aligned}
\hat{Y}_t &= 2.534\,5 + 0.614\,5X_t \\
&\quad (0.679\,6)\ (0.108\,7) \\
t &= (3.791\,0)\ (5.654\,1) \\
&\qquad r^2 = 0.799\,7 \qquad \hat{\sigma}^2 = 0.975\,2
\end{aligned}
\tag{12.4.7}
$$

460　因为现在的 Y 基本上是随机的，故此回归跟"真实情况"靠近得多。注意，$\hat{\sigma}^2$ 已从 0.811 4（$\rho = 0.7$）增至 0.975 2（$\rho = 0$），并且 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的标准误也都已增大。这结果是和先前考虑的理论结果相一致的。

§ 12.5　美国商业部门 1959—1998 年间工资与生产率之间的关系

至此我们已讨论了自相关的后果，一个显然的问题是，我们如何发现这个问题并加以纠正？在讨论这些专题之前，最好先考虑一个简明的例子。表

12.4 给出了美国商业部门在 1959—1998 年间人均真实工资（Y）和人均产出（X）指数的数据，以 1992 年的指数为基年，取值 100。

首先把 Y 和 X 的数据描点得到图 12.7。既然预期真实工资与劳动生产率之间的关系为正，那这两个变量相关也就无足为奇。令人吃惊之处在于，二者的关系几乎是线性的，尽管有迹象表明，在生产率的值较高时二者之间
461 的关系略显非线性。因此，我们决定估计一个线性模型，一个对数线性模型，结论如下：

$$\begin{aligned}\hat{Y}_t &= 29.5192 + 0.7136X_t \\ \text{se} &= (1.9423)\quad (0.0241) \\ t &= (15.1977)\quad (29.6066) \\ & r^2 = 0.9584 \quad d = 0.1229 \quad \hat{\sigma} = 2.6755\end{aligned} \tag{12.5.1}$$

其中 d 为德宾-沃森统计量，稍后将讨论这个统计量。

$$\begin{aligned}\widehat{\ln Y_t} &= 1.5239 + 0.6716\ln X_t \\ \text{se} &= (0.0762)\quad (0.0175) \\ t &= (19.9945)\quad (38.2892) \\ & r^2 = 0.9747 \quad d = 0.1542 \quad \hat{\sigma} = 0.0260\end{aligned} \tag{12.5.2}$$

为便于讨论，我们称（12.5.1）和（12.5.2）为工资—生产率回归。

462 定性而言，这两个模型的结论类似。很高的 t 值表明，在这两种情况下，所估计的系数都“高度”显著。在线性模型中，若生产率指数上升一个单位，则工资指数平均上升约 0.71 个单位。在对数线性模型中，斜率系数表示弹性（为什么?），我们发现，若生产率指数上升 1%，则真实工资指数平均上升 0.67%。

若存在自相关，（12.5.1）和（12.5.2）给出的结论可靠性如何呢？如前所述，若存在自相关，则估计的标准误就有偏误，结果估计的 t 比率就不可靠。显然，我们需要弄清楚，我们的数据是否受到自相关问题的困扰。在下一节，我们讨论侦查自相关的几种方法。我们只用线性模型（12.5.1）来解释这些方法，而把对数线性模型（12.5.2）留作练习。

表 12.4　　1959—1998 年美国人均真实工资与人均产出指数

年份	Y	X	年份	Y	X
1959	58.5	47.2	1979	90.0	79.7
1960	59.9	48.0	1980	89.7	79.8
1961	61.7	49.8	1981	89.8	81.4
1962	63.9	52.1	1982	91.1	81.2
1963	65.3	54.1	1983	91.2	84.0
1964	67.8	54.6	1984	91.5	86.4
1965	69.3	58.6	1985	92.8	88.1
1966	71.8	61.0	1986	95.9	90.7

1967	73.7	62.3	1987	96.3	91.3
1968	76.5	64.5	1988	97.3	92.4
1969	77.6	64.8	1989	95.8	93.3
1970	79.0	66.2	1990	96.4	94.5
1971	80.5	68.8	1991	97.4	95.9
1972	82.9	71.0	1992	100.0	100.0
1973	84.7	73.1	1993	99.9	100.1
1974	83.7	72.2	1994	99.7	101.4
1975	84.5	74.8	1995	99.1	102.2
1976	87.0	77.2	1996	99.6	105.2
1977	88.1	78.4	1997	101.1	107.5
1978	89.7	79.5	1998	105.1	110.5

注：X = 每小时产出指数，商业部分。

Y = 每小时真实工资，商业部分

资料来源：*Economic Report of the President*, 2000, Table B-47, p.362.

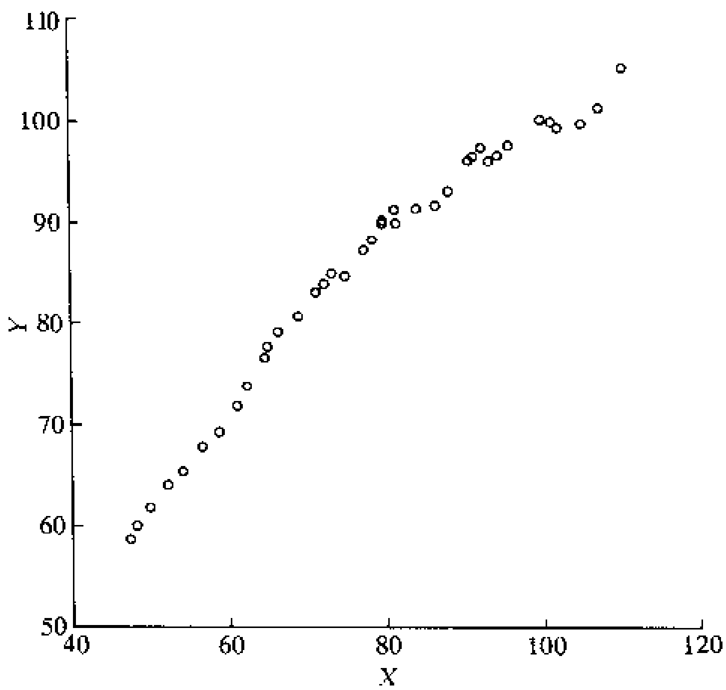


图 12.7　美国工资指数（Y）与生产力指数（X）：1959—1998

§12.6　侦察自相关

图解法

我们说过，经典模型的非自相关（nonautocorrelation）假定是对不可自

接观测的总体干扰项 u_t 而言，而我们所能获知的，则是得自通常 OLS 程序的、用以替代 u_t 的残差 $\hat{u}_t$。虽然 $\hat{u}_t$ 不同于 u_t[17]，但对 $\hat{u}_t$ 做一图像检查往往能对 u_t 中可能存有的自相关提供一些线索。其实，对 $\hat{u}_t$ 或 ($\hat{u}_t^2$) 的图像检查不仅为自相关而且为异方差性（如我们在上一章中看到的）甚至为模型适宜性或设定偏误（我们将在下章中看到）都能提供有用的信息。如一位作者所说：

> [残差] 图像的产生和分析，作为统计分析的一个标准部分，其重要性无论怎样强调，都不为过分。除了有时对复杂的问题提供了简易的要领外，还能在清楚地展现个别事例的性态的同时，使我们能把数据看作一个集合体，同时进行多方面的检查。[18]

有多种检查残差的方法。像图 12.8 那样，可以将残差对时间描点而得到一幅**时间顺序图**（time sequence plot）。图 12.8 展现了根据附录 12A 的数据做的美国工资—生产力回归（12.5.1）的残差。这些残差值与其他数据一起列在表 12.5 中。

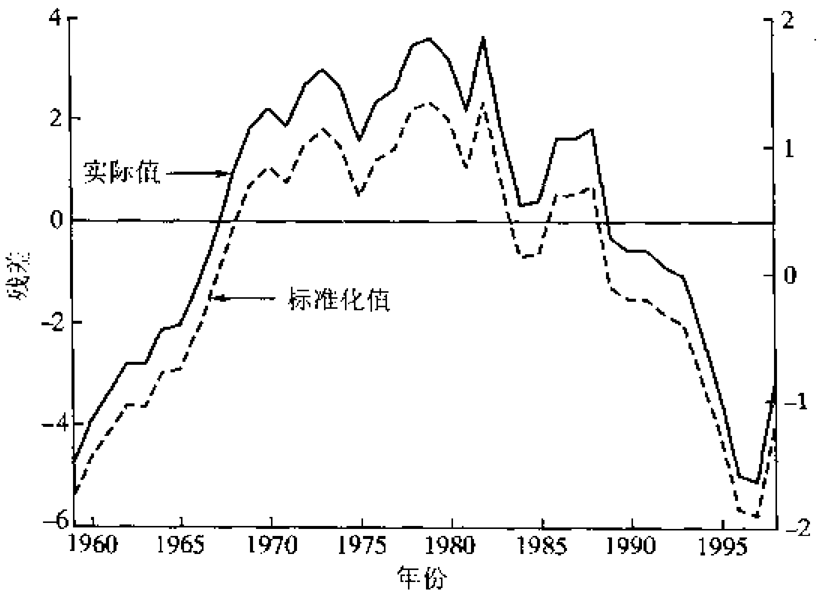


图 12.8　工资—生产力回归（12.5.1）中的残差与标准化列差值对比

464　另一方法是，将**标准化残差**对时间描图，也都见于图 12.7 和表 12.4。标准化残差无非是 $\hat{u}_t$ 除以估计值的标准误 $\hat{\sigma}$ $(=\sqrt{\hat{\sigma}^2})$。注意到 $\hat{u}_t$ 和 $\hat{\sigma}$ 的测量单位均与回归子 Y 的测量单位相同，这样，$\hat{u}_t/\hat{\sigma}$ 的值就是纯数（无测量单位的数），并因此可用来同其他回归的标准化残差相比较。此外，标准化残差和 $\hat{u}_t$ 一样，有零均值（为什么？）并且有近似于 1 的方差。[19] 在大样本中，($\hat{u}_t/\hat{\sigma}$) 近似地遵循零均值和单位方差的正态分布。就我们的例子而言，$\hat{\sigma}=2.675\ 5$。

通过对图 12.8 中的时间顺序图的分析，我们观察到 $\hat{u}_t$ 和标准化 $\hat{u}_t$ 都呈现一种类似于图 12.1（d）的模样，表明 u_t 也许不是随机的。

可从另一角度看待这一问题。将 $\hat{u}_t$ 对 $\hat{u}_{t-1}$ 描点，即时刻 t 的残差对它在时刻 $_{t-1}$ 的值描点，这是对AR(1)方案的一种经验检验。如果残差是非随机

465 的，我们将会得到类似于图 12.3 那样的图形。当我们用上述工资对生产力回归的 $\hat{u}_t$ 对 $\hat{u}_{t-1}$ 描点时，我们得到的图形，如图 12.9 所示，其所依据的数据，由表 12.5 给出。该图表明大多数残差都聚集在第一（东北）和第三（西南）象限内，有力地说明残差中有正的相关。

表 12.5　　残差的实际值、标准化值及滞后值

年份	RES1	SERS1	RES1(−1)	年份	RES1	SERS1	RES1(−1)
1959	−4.703 979	−1.758 168		1979	3.602 089	1.346 324	3.444 821
1960	−3.874 907	−1.448 293	−4.703 979	1980	3.230 723	1.207 521	3.602 089
1961	3.359 494	−1.255 651	−3.874 907	1981	2.188 868	0.818 116	3.230 723
1962	−2.800 911	−1.046 874	−3.359 494	1982	3.631 600	1.357 354	2.188 868
1963	−2.828 229	−1.057 084	−2.800 911	1983	1.733 354	0.647 862	3.631 600
1964	−2.112 378	−0.789 526	−2.828 229	1984	0.320 571	0.119 817	1.733 354
1965	−2.039 697	−0.762 361	−2.112 378	1985	0.407 350	0.152 252	0.320 571
1966	−1.252 480	−0.468 129	−2.039 697	1986	1.651 836	0.617 393	0.407 350
1967	−0.280 237	−0.104 742	−1.252 480	1987	1.623 640	0.606 855	1.651 836
1968	0.949 713	0.354 966	−0.280 237	1988	1.838 615	0.687 204	1.623 640
1969	1.835 615	0.686 083	0.949 713	1989	−0.303 679	−0.113 504	1.838 615
1970	2.236 492	0.835 815	1.835 615	1990	−0.560 070	−0.209 333	−0.303 679
1971	1.880 977	0.703 038	2.236 492	1991	−0.559 193	−0.209 005	−0.560 070
1972	2.710 926	1.013 241	1.880 977	1992	−0.885 197	−0.330 853	−0.559 193
1973	3.012 241	1.125 861	2.710 926	1993	−1.056 563	−0.394 903	−0.885 197
1974	2.654 535	0.992 164	3.012 241	1994	−2.184 320	−0.816 416	−1.056 563
1975	1.599 020	0.597 653	2.654 535	1995	−3.355 248	−1.254 064	−2.184 320
1976	2.386 238	0.891 885	1.599 020	1996	−4.996 226	−1.867 399	−3.355 248
1977	2.629 847	0.982 936	2.386 238	1997	−5.137 643	−1.920 255	−4.996 226
1978	3.444 821	1.287 543	2.629 847	1998	3.278 621	−1.225 424	−5.137 643

注：RES1 = 回归(12.5.1)得到的残差。

SRES1 = 标准化残差 = RES1/2.675 5。

RES(−1) = 滞后一期的残差值。

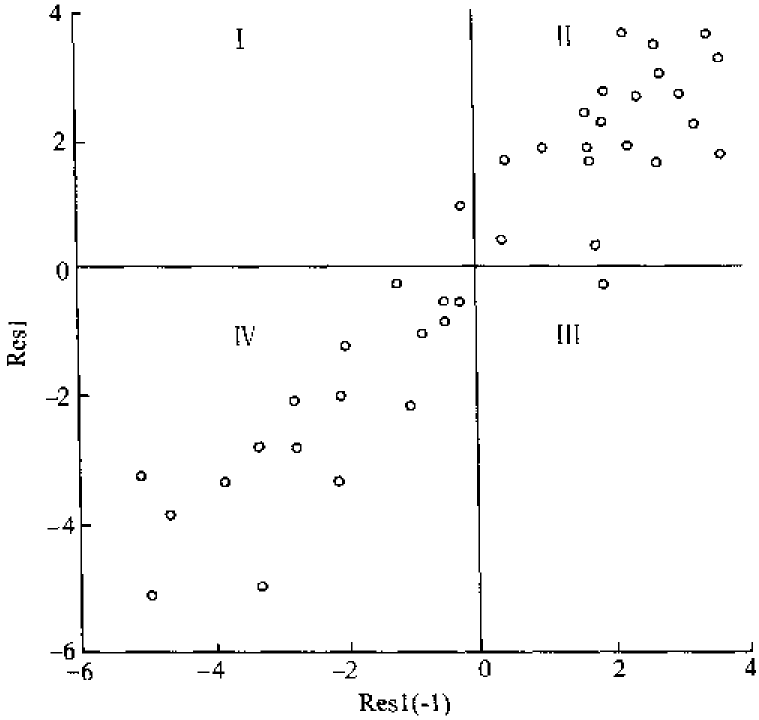


图 12.9 当期残差与滞后残差对比

刚才讨论的图解法，从性质上看，基本上是主观的或定性的。但有若干定量检验可用来补充这种纯粹定性的方法。现在让我们来介绍其中的一些。

游程检验

如果我们再次检查图 12.8，我们会注意到一种奇异的特点：开始时，我们有好几个残差都是负的，然后是一连串的正的残差，最后又是几个都是负的残差。如果这些残差是纯粹随机的，我们会观察到这样的模样吗？直觉地想，似乎不大可能。这种直觉可通过所谓**游程检验**（runs test）加以核实。这种检验有时又称**吉尔里（Geary）检验**，是一种非参数检验。[20]

为解释这种检验，让我们仅把表 12.5 第一列的工资—生产力回归残差的符号记录下来。

$$(---------)(+++++++++++++++++++++)(----------) \tag{12.6.1}$$

即先有 9 个（次）负残差，随之有 21 个正残差，再随之有 10 个负残差，共 40 次观测。现在我们定义一个**游程**为同一符号或属性（诸如 + 或 − ）的一个不中断历程（uninterrupted sequence）。我们再定义**游程的长度**（length of a run）为游程中的元素个数。在（12.6.1）所示的顺序中共有 3 个游程：一个 9 次负号游程（长度为 9 的游程），1 个 21 次正号游程（即长度为 21），

和 1 个 10 次负号游程（即长度为 10）。为了更好的视觉效果，我们用括号包住各个游程。

在一个严格随机的观测顺序中会出现怎样的游程？通过这一问题的分析，人们能推出游程的一个随机性检验。现在问：在我们的 40 次观测值的说明性例子中，我们看到有 3 个游程，如果拿它同一个严格随机的 40 次观测顺序中所预期的游程个数相比，是太多了还是太少了？如果太多，就是说在我们的例子中 $\hat{u}_t$ 变号太频繁，因而象征着一种负的序列相关（参看图 12.3b）。同理，如果游程太少，则表示可能有正的相关（如图 12.3a 所示）。从而，先验地，图 12.8 表示残差中有正的相关。

466 令：

N = 总观测个（次）数 = $N_1 + N_2$

N_1 = + 号个数（即 + 残差）

N_2 = − 号个数（即 − 残差）

R = 游程个数

于是，在相继结果（这里指残差）互相独立的虚拟假设下，并且假定 $N_1>10$ 和 $N_2>10$，游程个数将遵循（渐近地）正态分布，其

$$\text{均值}: E(R)=\frac{2N_1N_2}{N}+1$$

$$\text{方差}: \sigma_R^2=\frac{2N_1N_2(2N_1N_2-N)}{N^2(N-1)} \tag{12.6.2}$$

如果随机性假设是可维持的，则可预期在一个问题中所得到的游程个数将以 95% 的置信度落入如下区间内

$$\text{Prob}[E(R)-1.96\sigma_R\leqslant R\leqslant E(R)+1.96\sigma_R]=0.95 \tag{12.6.3}$$

因此我们得到这样的规则：

决策规则。在 95% 置信度下，若游程个数 R 落在上述置信区间内，就不要拒绝随机性虚拟假设；如果估计的 R 落在此范围外就拒绝虚拟假设。

回到我们的例子中，负号个数 $N_1=19$ 和正号个数 $N_2=21$，$R=3$。利用（12.6.2）中的公式我们可以得到：

$$\begin{aligned}E(R)&=10.975\\ \sigma_R^2&=9.6936\\ \sigma_R&=3.1134\end{aligned} \tag{12.6.4}$$

从而 95% 置信区间是：

$$[10.975\pm1.96(3.1134)]=(4.8728,17.0722)$$

由于游程个数3明显落在此区间之外，按 95% 置信水平，便可拒绝工资—生产力回归中残差的随机性假设。由于观测个数对上述正态检验来说可能小了

467 一些，诚望读者利用附录D表D.6所给的临界游程值，证实我们会得到同样的结论，即所测顺序不是随机的。换句话说，残差表现出自相关。作为一般原则，若存在正自相关，则游程数将很小，而若存在负相关，则游程数会很多。当然，我们从（12.6.2）可以发现，游程数是太多还是太少。

如果 N_1 或 N_2 小于20，斯威德（Swed）和艾森哈特（Eisenhart）曾研制出一个专用表，给出在 N 次观测的一个随机顺序中预期游程个数的临界值。此表见于附录D中的表D.6。利用这些表，读者可以验证，我们工资—生产力回归中的残差确实是非随机的，它们实际上正相关。

德宾-沃森 d 检验[21]

用以侦察序列相关的、最著名的检验是由统计学家德宾（Durbin）和沃森（Watson）研制出来的。人们普遍地称它为**德宾-沃森 d 统计量**（常简化为D-W d 统计量——译者注），其定义如下：

$$d=\frac{\sum_{t=2}^{t=n}(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_{t=2}^{t=n}\hat{u}_t^2} \tag{12.6.5}$$

它不外是相继残差的差异平方和与RSS之比。注意，由于取相继差异时损失一个观测值，在 d 统计量的分子中只有 $n-1$ 次观测值。

d 统计量的一大优点是，它仅依赖于估计的残差值，而后者在回归分析中照例都已被算出。正因为这一优点，现在把D－W d 统计量连同 R^2，校正 R^2，t 比率等等摘要统计量一起报告，已成为一种惯例。虽然现在 **d 统计量用得很经常，但记住它的一些基本假定是重要的：**

1. 回归含有截距项。如果没有截距项，像过原点回归那样，就要重新做带有截距项的回归，以求得RSS。[22]

2. 诸解释变量 X 是非随机的，或者在重复抽样中被固定的。

3. 干扰项 u_t 是按一阶自回归模式 $u_t=\rho u_{t-1}+\varepsilon_t$ 产生的。因此，它不能用于更高阶自回归模式的侦察。

4. 误差项被假定为正态分布。

468 5. 回归模型不把滞后因变量当作解释变量之一。因此，该检验在如下类型的模型中就不适用：

$$Y_t=\beta_1+\beta_2X_{2t}+\beta_3X_{3t}+\cdots+\beta_kX_{kt}+\gamma Y_{t-1}+u_t \tag{12.6.6}$$

其中 Y_{t-1} 是 Y 的一期滞后值，d 统计量便不适用。这样的模型叫做**自回归模型**。在第17章中我们将全面地分析它。

6. 没有缺落数据。例如，在我们对时期1959—1998年的工资—生产力回归中，如果出于某种原因，1978年和1982年（比方说）的观测值缺落，

则 d 统计量对这种失落数据没有补偿办法。[23]

如德宾和沃森曾表明的，由于 d 统计量的分布与出现在给定样本中的 X 值有复杂的关系。要导出它的准确的抽样或概率分布是困难的。[24]这种困难是可以理解的。因为，d 要从 $\hat{u}_t$ 算出，而 $\hat{u}_t$ 必然依赖于给定的诸 X。因此，它不同于 t，F 或 χ^2 检验，没有惟一的临界值可以导致拒绝或接受虚拟假设：干扰项 u_t 中无一阶序列相关。然而，他们成功地导出了临界值的一个下限 d_L 和一个上限 d_U，如果从（12.6.5）算出的 d 值落在这些临界值的范围之外，就可做出是否有正或负序列相关的决定。此外，这些界限值仅依赖于观测值的个数 n 以及解释变量的个数，却不依赖于这些解释变量取什么值。德宾和沃森已对从 6 到 200 的 n 值和多至 20 个解释变量的临界值编制成表，见附录 D 表 D.5（解释变量可多至 20 个）。

实际检验步骤，可借助于图 12.10 做出更好的解释。该图表明 d 的两个极限值是 0 和 4，这可由以下的展开式证实。将（12.6.5）展开得：

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.7}$$

因 $\sum \hat{u}_t^2$ 和 $\sum \hat{u}_{t-1}^2$ 只有一次观测之差，故可看作约为相等。因此，令：

$$\sum \hat{u}_{t-1}^2 \approx \sum \hat{u}_t^2$$

(12.6.7) 便可写为：

$$d \approx 2\left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \tag{12.6.8}$$

其中≈表示近似等于。

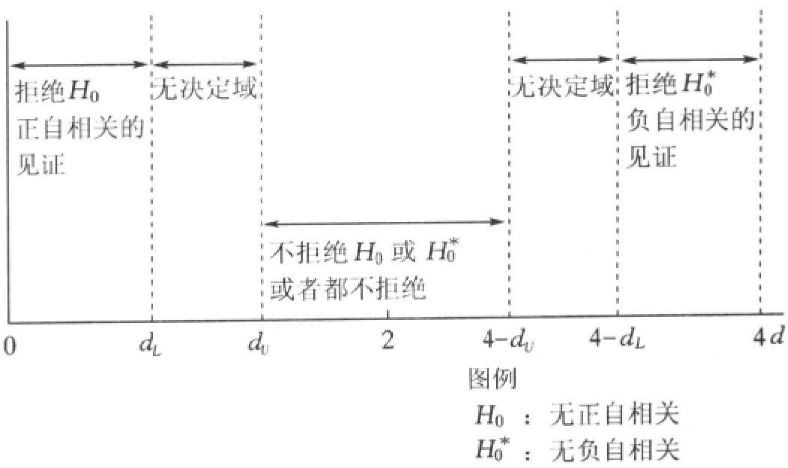


图 12.10　德宾-沃森 d 统计量

469　现定义：

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \tag{12.6.9}$$

为样本一阶自相关系数，作为 ρ 的一个估计量（参看章末注 9）。利用（12.6.9），可将（12.6.8）表达成：

$$d \approx 2(1-\hat{\rho}) \tag{12.6.10}$$

但因 $-1 \leqslant \rho \leqslant 1$，故（12.6.10）意味着

$$0 \leqslant d \leqslant 4 \tag{12.6.11}$$

这就是 d 的界限；d 的任何估计值必须落入这些界限内。

由方程（12.6.10）显见，若 $\hat{\rho}=0$，则 $d=2$；就是说，如果没有（一阶）序列相关，则预期 d 约为 2。因此，作为一种经验法则，如果在一项应用中求出 d 等于 2，便可认为没有一阶自相关，不管是正的或负的。如果 $\hat{\rho}=+1$，表示残差中有完全的正相关，则 $d \approx 0$。因此，d 越接近 0，正序列相关的迹象越明显。这种关系还能从（12.6.5）看出。因为，如果有正自相关，那些 $\hat{u}_t$ 就会被束缚在一起，以使它们的差异趋于微小，其结果将是分子平方和相对地小，而分母在任一给定的回归中保持着一个惟一的值。

470 如果 $\hat{\rho}=-1$ 即相继残差中有完全的负相关，则 $d \approx 4$。因此，d 越接近 4，负序列相关的迹象越明显。再从（12.6.5）看，这也是可以理解的。因为，如果有负自相关，就会有这样一种趋向：一个正的 $\hat{u}_t$ 将随后有一个负的 $\hat{u}_t$；反之亦然，以致 $|\hat{u}_t-\hat{u}_{t-1}|$ 常常比 $|\hat{u}_t|$ 大。因此，d 的分子比较而言将要大于其分母。

德宾-沃森检验的操作步骤如下：假定该检验的基本假定成立，

1. 做 OLS 回归并获取残差。
2. 按（12.6.5）计算 d。（现今的计算机程序大多数都给出 d 值。）
3. 对给定样本大小和给定解释变量个数找出临界 d_L 和 d_U 值。
4. 按照表 12.6 的决策规则行事。为了易于参照，再将这些规则描述如图 12.10。

为了说明步骤，再回到我们的工资—生产力回归。由表 12.5 的数据，估计的 d 值为 0.122 9，表明残差中有正的序列相关。（为什么？）由德宾-沃森表我们找出，对于 40 个观测和 1 个解释变量（不包括截距），在 5% 水平上，$d_L=1.44$ 和 $d_U=1.54$。由于估计值 0.122 9 低于 d_L，我们不能拒绝残差中有正序列相关的假设。

表 12.6　　德宾-沃森 d 检验：决策规则

虚拟假设	决策	如果
无正自相关	拒　绝	$0<d<d_L$
无正自相关	无决定	$d_L \leqslant d \leqslant d_U$

无负自相关	拒　绝	$4-d_L<d<4$
无负自相关	无决定	$4-d_U\leqslant d\leqslant 4-d_L$
无自相关，正或负	不拒绝	$d_U<d<4-d_U$

d 检验虽然极为流行，却有一大缺陷：如果它落入无决定区域（indecisive zone）或称无知域（region of ignorance），我们就无法论定自相关是否存在。为解决此问题，一些作者曾提出对德宾-沃森 d 检验的修改建议，但涉及的方法相当复杂，超过了本书的论述范围。[25]然而，在许多情况中，人们发现上限 d_U 差不多就是真实的显著性界限，因而，如果 d 的估计值落入无决定域，就不妨使用以下修订的 **d 检验程序**。给定显著性水平 α，

1. H_0：$\rho=0$ 对 H_1：$\rho>0$。如果估计的 $d<d_U$，则在水平 α 上拒绝 H_0；就是存在着统计上显著的正相关。

471 2. H_0：$\rho=0$ 对 H_1：$\rho<0$。如果估计的 $(4-d)<d_U$，则在水平 α 上拒绝 H_0；存在着统计上显著的负自相关。

3. H_0：$\rho=0$ 对 H_1：$\rho\neq 0$。如果 d 估计的 $d<d_U$ 或 $(4-d)<d_U$，则在水平 2α 上拒绝 H_0；存在着统计上显著的自相关。

或许可以指出，随着样本容量的扩大，无决定域逐渐变小，从德宾-沃森表可以清楚地看出这一点。比如，对 4 个回归元和 20 次观测而言，显著性水平为 5%时，d 值的下界和上界分别是 0.894 和 1.828，但当样本容量扩大到 75 时，这些值分别是 1.515 和 1.739。

计算机程序 SHAZAM 能做出（一种）**精确的 d 检验**（exact d test），即给出 d 计算值的准确概率或 p 值。利用现代计算工具，不难求出计算 d 统计量的 p 值。在我们的工资—生产力回归中，利用 SHAZAM 第 9 版，我们求出 d 计算值 0.122 9 的 p 值实际上为零，因而加强了我们前面基于德宾-沃森表得到的结论。

德宾-沃森 d 检验变得如此神圣，以致实践者常常忘记其背后的假定。特别是假定（1）解释变量或回归元是非随机的，（2）误差项服从正态分布，（3）回归模型不包括回归子的滞后值，这些对使用 d 检验都至关重要。

如果一个回归模型包含了回归子的滞后值，这种情形下的 d 值常为 2 左右，从而表明这种模型中不存在（一阶）自相关。因此，在这种模型中有一个有碍于发现序列相关的内在偏误。这并不意味着自回归模型就没有自相关的问题。事实上，德宾曾提出用所谓的 **h 检验**来检验这种模型中的序列相关。但从统计学意义上看，这一检验不如稍后讨论的**布劳殊-戈弗雷检验**（Breusch-Godfrey test）那样有效力，所以就没有使用 h 检验的必要。但鉴于其历史上的重要性，在习题 12.36 将讨论它。

此外，如果误差项 u_t 不是 NIID，那么，惯常使用的 d 检验可能就不可靠。[26]由此看来，前面讨论过的游程检验因不需要对误差项做任何分布上的

假定而占据优势。然而，若样本容量很大（技术性地讲，就是无穷大），那我们就可以使用德宾-沃森 d 检验，因为可以证明[27]：

$$\sqrt{n}\left(1-\frac{1}{2}d\right)\approx N(0,1) \tag{12.6.12}$$

472 也就是说，大样本容量下，按（12.6.12）变换后的 d 统计量服从标准正态分布。顺便一提，根据 d 和估计的一阶自相关系数 $\hat{\rho}$［见（12.6.10）］之间的关系可得到

$$\sqrt{n}\hat{\rho}\approx N(0,1) \tag{12.6.13}$$

即在大样本情形下，样本容量的方根与估计的一阶自相关系数之积也服从标准正态分布。

用我们的工资—生产力例子来说明这一检验，我们发现 $n=40$ 时 $d=0.122\ 9$。因此，我们从（12.6.12）求出

$$\sqrt{40}\left(1-\frac{0.122\ 9}{2}\right)\approx 5.94$$

渐近地讲，若零（一阶）自相关的虚拟假设为真，则得到一个大于或等于5.94的 Z 值（即标准化的正态变量）的概率极小。记住，对标准正态分布而言，（双侧）5%的临界 Z 值只有1.96，而1%的临界 Z 值也约为2.58。尽管我们的样本容量仅为40，但实际上足以使用正态近似。结论仍相同，即工资—生产力回归中的残差存在自相关的问题。

但 d 检验最严重的问题是回归元非随机性的假定，即回归元的值在重复抽样中保持不变的假定。若非如此，则无论是有限或少样本，还是大样本，d 检验都不成立。[28]而由于这一假定在涉及时间序列数据的经济模型中通常难以维持，因此有位作者认为，在涉及时间序列数据的计量经济学中，德宾-沃森统计量或许没有用武之地。[29]按照他的观点，仍有更有用的自相关检验可用，但它们都以大样本为前提。我们下面就讨论一个这种检验：**布劳殊-戈弗雷检验**。

自相关的一般性检验：布劳殊-戈弗雷检验[30]

为了避免自相关的德宾-沃森 d 检验所存在的隐患，统计学家布劳殊和戈弗雷（BG）提出了一种自相关检验。这种检验容许（1）非随机回归元，如回归子的滞后值；（2）高阶自回归模式，如AR(1)，AR(2)等；(3)白噪音

478 误差项[如(12.2.1)中的 ε_t]的简单或更高阶**移动平均**(moving averages)[31]，从这种意义上看，它比前面的检验更具有一般性。

尽管从参考文献中可以查到详尽的数学推导，但撇开这些，我们仍可以做**BG 检验**(也被称为 **LM 检验**)如下[32]：尽管可以在模型中添加许多回归元，但我们仍以双变量回归模型来说明这个检验。此外，回归子的滞后值也可以放到模型中。令

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.6.14}$$

假定误差项 u_t 服从如下 p 阶自回归AR(p)模式：

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t \tag{12.6.15}$$

其中 ε_t 为前面讨论过的白噪音项。你将会意识到，这只是对AR(1)模式的推广。

欲检验的虚拟假设 H_0 是

$$H_0:\ \rho_1 = \rho_2 = \cdots = \rho_p = 0 \tag{12.6.16}$$

即不存在任何阶数的序列相关。BG 检验包含如下步骤：

1. 用 OLS 估计（12.6.14）并得到残差 $\hat{u}_t$。

2. 将 $\hat{u}_t$ 对原 X_t（若原模型中不止一个 X 变量，则都包括进来）和第一步所估计的残差滞后值 $\hat{u}_{t-1}$，$\hat{u}_{t-2}$，…，$\hat{u}_{t-p}$做回归。因此，若 $p=4$，则我们就在模型中引入残差的 4 个滞后值作为额外的回归元。注意，在做这个回归时，我们将只有（$n-p$）次观测（为什么?）。简言之，做如下回归：

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t \tag{12.6.17}$$

并从这个（辅助）回归中得到 R^2。[33]

3. 若样本容量很大（技术上讲是无限样本），则布劳殊和戈弗雷证明了：

$$(n-p)R^2 \sim \chi^2_p \tag{12.6.18}$$

474 即（$n-p$）乘以从辅助回归（12.6.17）中得到的 R^2 值服从自由度为 p 的 χ^2 分布。在应用中，若在选定的显著性水平下$(n-p)R^2$ 超过临界 χ^2 值，我们就拒绝虚拟假设，此时（12.6.15）中至少有一个 ρ 在统计上显著异于零。

对于 BG 检验，有如下实际要点须注意：

1. 回归模型所包含的回归元中或许有回归子 Y 的滞后值，即 Y_{t-1}，Y_{t-2}等可能作为解释变量而出现。相比之下，德宾-沃森检验则要求回归元中不含回归子滞后值的约束。

2. 前面曾指出，即使干扰项服从 p 阶**移动平均**（MA）过程，即 u_t 如下生成：

$$u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \cdots + \lambda_p \varepsilon_{t-p} \tag{12.6.19}$$

其中 ε_t 为白噪音误差项，即这些误差项满足全部经典假定，BG 检验也可适用。

在有关时间序列计量经济学的章节中，我们将详尽研究 p 阶自回归和移动平均过程。

3. 若（12.6.15）中 $p=1$（即一阶自回归），则 BG 检验可称为**德宾的 M 检验**（Durbin's M test）。

4. BG 检验的一个缺陷在于，滞后长度 p 值不能先验地设定。这就不可

避免对 p 值的多次试验。人们有时候也能用所谓的**赤池**和**施瓦茨**信息准则来筛选滞后长度。我们将在第 13 章及后面有关时间序列计量经济学的章节中讨论这些准则。

对 BG 检验的说明：工资—生产力关系

为说明此检验，我们在前面说明性的例子中应用一下这个检验。利用一个AR(6)模式，我们得到习题 12.25 中所示的结论。从那里给出的回归结果可以看出，$(n-p)=34$，$R^2=0.8920$。因此，二者相乘则得到一个 χ^2 值 30.328。对 6 个自由度（为什么?）而言，得到一个 χ^2 值大于或等于 30.328 的概率极小；附录 D.4 中的 χ^2 表表明，得到一个约等于 30 的 χ^2 值的概率一定极小。事实上，实际的 p 值几乎为零。

于是，对我们的例子来说，结论是这 6 个自相关系数中至少有一个非零。

试遍从 1 到 6 的滞后长度，我们发现，只有AR(1)系数显著，从而表明没有必要考虑多于一个的滞后。实质上，此时的 BG 检验就是**德宾的 *M* 检验**。

为什么有这么多的自相关检验?

对这个问题的回答是："……没有某个特定的检验被证明绝对最好（即在统计意义上更有功效），因而分析家在考虑一系列侦察自相关的存在或结
475 构的检验程序时，仍处在十分尴尬的境地。"[34]当然，上一章讨论异方差性的各种检验时同样也有这个问题。

§12.7 发现自相关该怎么办：补救措施

如果我们利用上一节中讨论的一或多种自相关诊断检验，并发现存在自相关问题，那该怎么办呢？我们有 4 种选择：

1. 尽力查明自相关是否纯粹自相关，而不是模型误设的结果。我们在第 12.1 节讨论过，我们有时候观察到的残差形式是模型误设（即排除了某些重要变量）或函数形式不正确所导致的。

2. 若是纯粹自相关，则可对原模型做适当的变换，使变换后的模型不存在（纯粹）自相关问题。如同出现异方差时一样，我们必须使用某种**广义最小二乘法**。

3. 在大样本下，我们可以用**尼威-韦斯特（Newey-West）**方法，以得到 OLS 估计量在对自相关加以修正之后的标准误。这一方法实际上是对我们在上一章中讨论过的怀特异方差一致标准误方法的推广。

4. 在某些情形下，我们可以继续使用 OLS 方法。

鉴于这里每个专题都很重要，我们将分节加以讨论。

§12.8 模型误设与纯粹自相关

让我们回到（12.5.1）中给出的工资—生产力回归中来。我们在那里看到，d 值为0.122 9，而且我们基于德宾-沃森 d 检验断定，误差项中存在着正的自相关。这种自相关会因我们的模型被不正确地设定而引起吗？由于回归（12.5.1）背后的数据是时间序列数据，所以工资与生产力很可能都表现
476 出时间趋势。若然，我们就必须在模型中包括时间或趋势变量 t，查看工资和生产力在除去趋势因素后的关系。

为对此进行检验，我们在（12.5.1）中包含时间趋势变量并得到如下结果：

$$
\begin{aligned}
\hat{Y}_t = &\quad 1.475\,2 \quad + \quad 1.305\,7X_t \quad - \quad 0.903\,2t \\
\text{se} = &\ (13.18) \qquad (0.276\,5) \qquad (0.420\,3) \\
t = &\ (0.111\,9) \qquad (4.723\,0) \qquad (-2.149\,0) \\
& \qquad\qquad R^2 = 0.963\,1; d = 0.204\,6
\end{aligned}
\tag{12.8.1}
$$

对此模型的解释直截了当：真实工资指数逐年递减约 0.90 单位。在容许包含趋势变量后，生产力指数每提高一个单位，真实工资指数则平均增加约 1.30 个单位，当然这个数字并非统计上异于 1。（为什么？）有趣的是，即便容许出现趋势变量，d 值依然很低，这表明（12.8.1）存在纯粹自相关的问题，而不一定是设定误差的问题。

我们如何知道（12.8.1）被正确地设定了呢？为检验这一点，我们将 Y 对 X 和 X^2 做回归，来检验真实工资指数与生产力指数有非线性相关的可能性。回归结果如下：

$$
\begin{aligned}
\hat{Y}_t = &\quad -16.218\,1 \quad + \quad 1.948\,8X_t \quad - \quad 0.007\,9X_t^2 \\
t = &\quad (-5.489\,1) \quad (24.986\,8) \quad (-15.936\,3) \\
& \qquad\qquad R^2 = 0.994\,7 \quad d = 1.02
\end{aligned}
\tag{12.8.2}
$$

这些结论很有意思。所有这些系数在统计上都高度显著，p 值也极小。从负的二次项系数来看，尽管真实工资指数随着生产率指数的增加而增加，但增加的速度递减。但从 d 值来看，它仍表明残差中存在正的自相关，因为 $d_L = 1.391$，$d_U = 1.60$，而所估计的 d 值低于 d_L。

从上述分析可以安全地断定，我们的工资—生产力回归可能遇到纯粹自相关问题，而不一定会是设定偏误的问题。知道自相关的后果，我们就想采取某种行动来修正。我们稍后便这么做。

顺便一提，对我们上面给出的所有工资—生产力回归而言，我们用**雅克-贝拉正态值检验**发现残差是正态分布的，由于 d 检验假定了误差项的正态性，所以这一结果令人欣慰。

§ 12.9 （纯粹）自相关的修正：广义最小二乘法

477 知道自相关的后果之后，特别是知道 OLS 估计量缺乏效率之后，我们或许要补救这个问题。补救措施取决于对干扰之间相互依赖的性质的了解，即对自相关结构的了解。

首先，考虑双变量回归模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{12.9.1}$$

并假定误差项服从AR(1)模式，即

$$u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \tag{12.9.2}$$

现在我们考虑两种情况：(1) ρ 已知，(2) ρ 未知并有待估计。

ρ 已知

若一阶自相关系数 ρ 已知，则序列相关问题就可轻易解决。如果 (12.9.1) 在时刻 t 成立，则在时刻 $t-1$ 也成立。从而，

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \tag{12.9.3}$$

用 ρ 乘 (12.9.3) 两边，得：

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \tag{12.9.4}$$

从 (12.9.1) 减去 (12.9.4) 便给出：

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \tag{12.9.5}$$

其中 $\varepsilon_t = u_t - \rho u_{t-1}$。

可将 (12.9.5) 表达成：

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \tag{12.9.6}$$

其中 $\beta_1^* = \beta_1(1-\rho)$，$Y_t^* = (Y_t - \rho Y_{t-1})$，$X_t^* = (X_t - \rho X_{t-1})$，$\beta_2^* = \beta_2$。

由于 (12.9.6) 中的 ε_t 满足全部 OLS 假定，故可直接对转换变量 Y^* 和 X^* 应用 OLS 并获得具有全部最优性质的估计量，即 BLUE。其实，做回归 (12.9.6) 就等于应用上一章中讨论的广义最小二乘法。记住，GLS 无非就是把 OLS 用于变换后满足经典假定的变换模型。

478 回归 (12.9.5) 叫做**广义** (generalized) 或**拟** (quasi) **差分方程** (difference equation)。它不是原来的形式，而是以拟差分形式将 Y 对 X 回归。这一差分形式，是从一个变量的现期值减去它的前期值的一个比例 ($-\rho$ 的)

部分得到的。在这个取差分的过程中，由于第一次观测值没有先前值，所以失去了一次观测。为弥补这一损失，将对 Y 和 X 的第一次观测转换为 $Y_1\sqrt{1-\rho^2}$ 和 $X_1\sqrt{1-\rho^2}$。[35] 这一转换被称为**普雷斯-温斯坦变换**（Prais-Winsten transformation）。

ρ 未知

广义差分回归的应用虽然直接明了，但因 ρ 实际上鲜为人知，故一般而言难于实现。从而需要另想办法。其中的一些方法有如下述。

一阶差分法。因 ρ 落在 0 与 ± 1 之间，故可从两个极端开始尝试。在一个极端上可假定 $\rho=0$ 即无序列相关，而在另一个极端上则置 $\rho=\pm 1$ 即完全正的或负的相关。其实，当我们做一个回归时，我们通常假定没有自相关，然后通过德宾-沃森或其他检验以表明这种假定是否合理。而当 $\rho=+1$ 时，广义差分方程（12.9.5）便化为**一阶差分方程**如下：

$$\begin{aligned}Y_t - Y_{t-1} &= \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})\\ &= \beta_2(X_t - X_{t-1}) + \varepsilon_t\end{aligned}$$

或
$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \tag{12.9.7}$$

其中 Δ 是（12.1.10）中曾引入的是一阶差分运算子。

由于（12.9.7）中的误差项没有（一阶）序列相关的问题（为什么?），所以为了做回归（12.9.7），惟一要做的就是形成回归子和回归元的一阶差分，并对这些一阶差分做回归。

如果自相关系数很高（比方说大于 0.8）或德宾-沃森 d 统计量很低，那么进行一阶差分变换可能合适。曼达拉（Maddala）曾提出一个粗略的经
479 验法则：只要 $d < R^2$ 就能用一阶差分形式。[36] 在我们的工资—生产力回归（12.5.1）中，$d=0.1229$，$r^2=0.9584$，就属于这种情况。对其做的一阶差分回归稍后给出。

一阶差分模型（12.9.7）的一个有趣特征是，**它不含有截距项**。因此，为了估计（12.9.7），你必须使用**过原点的回归**（即去掉截距项），现在大多数软件包都能做到这些。不过，如果你忘记从模型中去掉截距项，并估计如下包含截距项的模型：

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t \tag{12.9.8}$$

那么，原模型必定有趋势变量，而 β_1 表示它的系数。[31] 因此，在一阶差分模型中引入截距项的一个“意外”好处是，检验原模型中是否该出现趋势变量。

回到我们的工资—生产力回归（12.5.1），给定AR(1)模式和相对 r^2 来说较低的 d 值，用一阶差分的形式并不含截距。重做（12.5.1）；记住，

(12.5.1) 是水平值形式的。结果如下[38]：

$$\widehat{\Delta Y_t} = 0.719\,9\Delta X_t$$
$$t = (9.207\,3) \quad r^2 = 0.361\,0 \quad d = 1.509\,6 \tag{12.9.9}$$

与水平值形式的回归 (12.5.1) 相比，我们看到，斜率系数没多大变化，但 r^2 值下降相当多。由于取一阶差分时，我们实质上研究的是变量在其（线性）趋势值附近的行为，所以通常都会如此。当然，因为 (12.9.9) 和 (12.5.1) 这两个模型中的因变量都不同，所以我们不能将它们的 r^2 值直接进行比较。[39]同时还须注意到，与原来的回归相比，d 值明显提高，这可能表明，在一阶差分回归中没有什么自相关。[40]

一阶差分变换的另一个有趣方面与时间序列背后的平稳性质有关。回到描述AR(1)模式的方程 (12.2.1)。现在，若 ρ 实际上等于 1，则从方程
480 (12.2.3) 和 (12.2.4) 明显可见，序列 u_t 是非平稳的，因为方差和协方差都变成无穷大。这正是我们为什么在讨论此专题时总施加 $|\rho|<1$ 的约束。但从 (12.2.1) 明显可见，若自相关系数实际上为 1，则 (12.2.1) 变成：

$$u_t = u_{t-1} + \varepsilon_t$$

或者

$$(u_t - u_{t-1}) = \Delta u_t = \varepsilon_t \tag{12.9.10}$$

即 u_t 的一阶差分变成平稳序列了，因为它等于白噪音误差项 ε_t。

以上讨论的要点是，若原时间序列是非平稳的，那么其一阶差分很可能变成平稳序列。因此，一阶差分变换起到一箭双雕的作用，既可能会消除（一阶）自相关，又使时间序列变得平稳。我们在更深入地讨论时间序列计量经济学的第 4 篇将会重新探讨这一专题。

我们曾提到，一阶差分变换在 ρ 很高或 d 很低时都适用。严格地讲，一阶差分变换只有在 $\rho=1$ 时才能成立。事实上，有一个被称为**贝伦布鲁特–韦布检验** (Berenblutt-Webb test)[41]的方法，可用来检验 $\rho=1$ 的假设。其所用的检验统计量被称为 **g 统计量** (g statistic)，定义如下：

$$g = \frac{\sum_2^n \hat{e}_t^2}{\sum_1^n \hat{u}_t^2} \tag{12.9.11}$$

其中 $\hat{u}_t$ 为原回归（即水平值形式的回归）中得到的残差，而 e_t 则为从一阶差分回归中所得到的 OLS 残差。记住，采用一阶差分形式时没有截距项。

为检验 g 统计量的显著性，假定水平形式回归中包含截距项，那我们就可以使用德宾–沃森表，只是现在的虚拟假设是 $\rho=1$，而不是德宾–沃森的假设 $\rho=0$。

重新回到我们的工资—生产力回归，我们从原回归 (12.5.1) 得到 $\sum \hat{u}_t^2 = 272.022\,0$，从一阶差分回归 (12.7.11) 中得到 $\sum \hat{e}_t^2 = 0.334\,270$。代入 (12.9.11) 所示的 g 统计量得到：

$$g=\frac{0.334\ 270}{272.022\ 0}=0.001\ 2 \qquad (12.9.12)$$

查阅德宾-沃森表中 39 次观测和 1 个解释变量，我们发现 $d_L=1.435$，$d_U=1.540$（显著性水平为5%）。由于所得到的 g 值低于 d 值的下限，所以我们
481 不能拒绝真实的 ρ 为 1 的假设。**牢记，尽管我们使用同样的德宾-沃森统计表，但现在的虚拟假设是** $\rho=1$ **而不是** $\rho=0$。鉴于这一发现，(12.9.9) 所给出的结论或许可以接受。

基于德宾-沃森 d 统计量的 ρ。若因 ρ 与 1 不够接近而不能使用一阶差分变换，那我们从前面在（12.6.10）中构建的 d 和 ρ 之间的关系中能找到一个简单的估计方法，我们可以如下估计出 ρ 来：

$$\hat{\rho}\approx 1-\frac{d}{2} \qquad (12.9.13)$$

因此，在样本充分大时，便可从（12.9.13）中估计出 ρ 来，并如广义差分方程（12.9.5）那样用它来对数据进行变换。记住，(12.9.13) 中给出的 ρ 和 d 的关系对小样本情形可能不成立，泰尔和纳加（Nagar）对此提出了修正意见，可参见习题 12.6。

我们在工资—生产力回归（12.5.1）中得到 $d=0.122\ 9$，代入（12.9.13）则得到 $\hat{\rho}\approx 0.938\ 6$。我们可以利用这个估计的 ρ 值来估计（12.9.5）。所须做的只是将 Y 和 X 的当期值都减去其前一期值的 0.938 6 倍，然后按（12.9.6）对如此变换的数据做 OLS 回归，其中 $Y_t^*=(Y_t-0.938\ 6Y_{t-1})$，$X_t^*=(X_t-0.938\ 6X_{t-1})$。

从残差中估计出来的 ρ。若AR(1)模式 $u_t=\rho u_{t-1}+\varepsilon_t$ 成立，一个估计 ρ 的简单方法就是将残差 $\hat{u}_t$ 对 $\hat{u}_{t-1}$ 做回归，因为前面曾指出，$\hat{u}_t$ 是 u_t 的一致估计量。即做如下回归：

$$\hat{u}_t=\hat{\rho}\cdot\hat{u}_{t-1}+v_t \qquad (12.9.14)$$

其中 $\hat{u}_t$ 为从原（水平型）回归中所得到的残差，v_t 为此回归的误差项。注意，在（12.9.14）中没有引入截距项的必要，因为我们知道 OLS 残差的总和为零。

（12.5.1）中工资—生产力回归的残差已在表 12.5 中给出。利用这些残差，我们可以得到如下回归结果：

$$\begin{aligned}\hat{u}_t&=0.914\ 2u_{t-1}\\ t&=(16.228\ 1) \qquad r^2=0.873\ 6\end{aligned} \qquad (12.9.15)$$

如此回归所示，$\hat{\rho}=0.914\ 2$。利用这个估计值，可如同（12.9.6）那样对原模型进行变换。由于此程序所估计的 ρ 与从德宾-沃森 d 统计量所得到
481 的 ρ 大致相同，所以利用（12.9.15）中的 ρ 进行回归的结果，与利用德宾-沃森 d 统计量估计的 ρ 所得到的回归结果，不应该有很大不同。我们留给读者来验证这一点。

估计 ρ 的迭代方法。前面讨论的所有估计 ρ 的方法都只为我们提供了 ρ 的一个估计值。但有些所谓**迭代法**（iterative methods）则可多次估计出 ρ 来，即从 ρ 的某个初始值开始，通过逐次逼近，反复估计 ρ 值。这些方法中值得一提的有：科克伦-奥克特（Cochrane-Orcutt）迭代法、科克伦-奥克特两步法、德宾两步法和**希尔德雷思-卢**（Hildreth-Lu）扫描或搜寻程序等。其中，最流行的是科克伦-奥克特迭代法。为节省篇幅，迭代法通过习题加以讨论。记住，这些方法的最终目标是给出 ρ 的一个估计值，使之能用于得到参数的 GLS 估计值。科克伦-奥克特迭代法的优越性之一，它不仅能用于 AR(1)模式，也能用于更高阶的自回归模式，比如 AR(2)：$\hat{u}_t=\hat{\rho}_1\hat{u}_{t-1}+\hat{\rho}_2\hat{u}_{t-2}+v_t$。得到两个 ρ 之后，很容易就能推广应用广义差分方程(12.9.6)。当然，计算机可以做到所有这些。

回到工资—生产力回归中来，并假定AR(1)模式，我们使用科克伦-奥克特迭代法得到 ρ 的如下估计值：0.914 2，0.905 2，0.899 2，0.895 6，0.893 5，0.892 4 和 0.891 9。最后一个值 0.891 9 现可用于如（12.9.6）般对原模型进行变换，并用 OLS 估计变换后的模型。当然，对变换模型做 OLS 无非就是做 GLS。其结果如下：

去掉第一个观测。由于第一个观测前面再没有其他观测，所以在估计(12.9.6) 时，我们去掉第一个观测。回归结果如下：

$$
\begin{aligned}
\hat{Y}_t^* &= 45.105+0.550\,3X_t^* \\
\text{se} &= (6.190)(0.065\,2) \qquad\qquad (12.9.16)\\
t &= (7.287)(8.433) \qquad r^2=0.995\,9
\end{aligned}
$$

将此结果与（12.5.1）中给出的原回归结果相比，我们看到，斜率系数明显下降。对（12.9.16）须注意两点。首先，(12.9.16) 中的截距系数为 $\beta_1(1-\rho)$，因为我们知道 $\rho=0.891\,9$，所以 β_1 很容易倒推出来。其次，变换模型（12.9.16）的 r^2 与原模型（12.5.1）的 r^2 不能直接比较，因为这两个模型中的因变量不同。

483 **按照普雷斯-温斯坦方式保留第一次观测**。我们在前面曾指出，尽管在大样本中去留第一次观测无所谓，但在小样本中则情况迥异。

按照普雷斯-温斯坦方式保留第一次观测，我们会得到如下回归结果[42]：

$$
\begin{aligned}
\hat{Y}_t^* &= 26.454 \quad + \quad 0.724\,5X_t^* \\
\text{se} &= (5.452\,0) \quad (0.061\,2) \qquad\qquad (12.9.17)\\
t &= (4.852\,1) \quad (11.838\,2) \qquad r^2=0.994\,9
\end{aligned}
$$

(12.9.16) 和 (12.9.17) 之间的差异告诉我们，去掉或留下第一次观测对回归结果有明显影响。同时还注意到，(12.9.17) 中的斜率系数与(12.5.1) 中的系数近乎一致。

一般评论。对用上述各种方法修正自相关问题要明确如下几点：

第一，由于在大样本情况下，即便存在自相关问题，OLS 估计量仍是一

致的，所以无论我们是从德宾-沃森 d、从当期残差对前期残差的回归、还是从科克伦-奥克特程序中估计出 ρ 来，并没有多大差别，因为这些方法也都是给出真实 ρ 的一致估计值。

第二，上述方法基本上都是两步法。我们在第一步得到未知 ρ 的一个估计值，第二步用这个估计值变换变量去估计广义差分方程（实质上就是 GLS)。但由于我们用的是 $\hat{\rho}$ 而非真正的 ρ，所以在文献中所有这些估计方法都被称为**可行 GLS**（FGLS）或**估计的 GLS**（EGLS）**法**。

第三，重要的是要指出，只要我们用 FGLS 或 EGLS 法估计变换模型的参数，估计系数都不一定具有通常经典模型所具有的优良性质（比如 BLUE)，特别是在小样本情况下。无须复杂的技术性知识，便可总结一个**一般原则：只要我们用的是估计量而非真实值，所估计的 OLS 系数在大样本下或渐近地具有通常的性质。同时，严格地讲，通常的假设检验程序也是渐近有效的。因此，在小样本下，必须小心地解释估计结果。**

第四，在使用 EGLS 时，若不包括第一次观测（如最初使用科克伦-奥
484 克特程序一样），不仅估计量的数值，就连其有效性都要受到不利的影响，特别是当样本容量很小或回归元并非严格随机的情况下。[43]因此，在小样本下，按照普雷斯-温斯坦的方式保留第一次观测很重要。当然，如果样本容量足够大，EGLS 在有或没有第一次观测的情况下都给出类似结果。顺便指出，以普雷斯-温斯坦方式进行变换的 EGLS 在文献中被称为**完全 EGLS**（full EGLS），或简称为 **FEGLS**。

§12.10 修正 OLS 标准误的尼威-韦斯特方法

除了使用上一节讨论的 FGLS 方法之外，我们仍可以使用 OLS，只是需要用尼威和韦斯特提出的方法对自相关问题修正标准误。[44]这是对上一章讨论的怀特异方差一致标准误的推广。修正的标准误被称为 **HAC（异方差一自相关一致）标准误**［HAC（heteroscedasticity and auto correlation-consistent）standard errors］，或简称**尼威-韦斯特标准误**（Newey-West standard errors）。我们不再给出尼威-韦斯特程序背后复杂的数学知识。[45]但大多数现代计算机软件现在都能计算尼威-韦斯特标准误。但重要的是须指出，严格地讲，尼威-韦斯特程序只对大样本有效，对小样本可能不合适。但在大样本情形下，我们现在有一种能对自相关修正其标准误的方法，所以就没有必要担心上一章讨论的 EGLS 变换。因此，如果样本足够大，在同时存在自相关和异方差的情况下，就应该使用尼威-韦斯特程序来修正 OLS 标准误，因为 HAC 法与专为异方差而设计的怀特方法不同，它能同时处理这两种问题。

我们再次回到工资—生产力回归（12.5.1)。我们知道此回归存在自相

关的问题。40 次的观测样本也足够地大，所以我们可用 HAC 程序。我们用 Eviews 4 得到如下回归结果：

$$
\begin{aligned}
\hat{Y}_t &= 29.5192 + 0.7136\hat{X}_t \\
se &= (4.1180)^* \quad (0.0512)^* \qquad (12.10.1)\\
& r^2 = 0.9584 \quad d = 0.1229
\end{aligned}
$$

其中 * 表示 HAC 标准误。

485 将此回归与（12.5.1）相比较，我们发现两个方程中的估计系数的 r^2 都相同。但必须注意，HAC 标准误比 OLS 标准误大得多，因此 HAC t 比率比 OLS t 比率小得多。这就表明，OLS 实际上低估了真实标准误。令人惊奇的是，(12.5.1) 和 (12.10.1) 中 d 统计量是一样的。但不必担心，HAC 程序在修正 OLS 标准误时已对此加以考虑。

§ 12.11 OLS 与 FGLS 和 HAC

研究者所面临的实际问题是：在出现自相关问题时，OLS 估计量尽管无偏、一致且渐近正态分布，但仍不是有效的。因此，通常基于 t，F 和 χ^2 的推断程序就不再适合。另一方面，虽然 FGLS 和 HAC 能给出有效的估计量，但这些估计量的有限或小样本性质并没有得到很好地证明。这就意味着，在小样本下，FGLS 和 HAC 实际上可能还不如 OLS。事实上，格里利谢斯和饶（Rao）在一项蒙特卡罗研究中发现[46]，若样本相对较小，且自相关系数低于 0.3，则 OLS 至少和 FGLS 一样好。于是，从实践的角度看，在估计的 ρ 低于 0.3 的小样本中，或许使用 OLS 较好。当然，样本大小是个相对问题，必须从实践中加以判断。如果只有 15～20 次观测，样本就很小，但如果有 50 次以上的观测，样本就足够大了。

§ 12.12 含有自相关误差时的预测

我们在第 5.10 节中在双变量回归模型的背景下讨论了预测的基本知识，当时的模型满足全部经典假定。若存在自相关，结果会怎么样呢？尽管这个问题一般在经济预测课上都要讨论，但我们在这里还是略看一下。为明确起见，我们继续采用双变量模型并假定AR(1)模式。因此，

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (12.12.1)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \qquad (12.12.2)$$

其中 ε_t 为白噪音误差项。

将（12.12.2）代入（12.12.1），我们得到：

$$Y_t = \beta_1 + \beta_2 X_t + \rho u_{t-1} + \varepsilon_t \tag{12.12.3}$$

486 如果我们想预测下一期（$t+1$）的 Y，我们有：

$$Y_{t+1} = \beta_1 + \beta_2 X_{t+1} + \rho u_t + \varepsilon_{t+1} \tag{12.12.4}$$

因此，对下一期的预测由如下三部分构成：（1）其期望值（$\beta_1 + \beta_2 X_{t+1}$），（2）$\rho$ 与上一期误差项的乘积，（3）期望值为零的纯粹白噪音项。给定 X_{t+1} 的值，我们用 $\hat{\beta}_1 + \hat{\beta}_2 X_{t+1}$ 来估计（1），其中的 OLS 估计量都是从一个给定样本中得到。我们用 $\hat{\rho}\hat{u}_t$ 估计（2），其中 $\hat{\rho}$ 由第 12.9 节中讨论的一种方法估计而来。在第（$t+1$）期，$\hat{u}_t$ 的值已知。因此，（12.12.4）中 Y_{t+1} 的估计值为：

$$\hat{Y}_{t+1} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+1} + \hat{\rho}\hat{u}_t \tag{12.12.5}$$

依此类推，第 2 期的预测值为：

$$\hat{Y}_{t+2} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+2} + \hat{\rho}^2 \hat{u}_t \tag{12.12.6}$$

我们在第 5.10 节中所做的预测被称为**静态预测**（statistic forecasting），而（12.12.5）和（12.12.6）所代表的预测则称为动态预测（dynamic forecasting），因为我们在做这些预测时考虑了过去的预测中所犯的错误。

与第 5.10 节一样，我们需要计算（12.12.5）和（12.12.6）的预测（标准）误。但这个公式十分复杂。由于大多数现代计量软件（如 Microfit，Eviews，Shazam 等）都能给出预测标准误，所以这里就不必给出其计算公式。

作为一个解说，回到我们的工资—生产力回归。记得我们的样本数据是 1959—1998 年，我们只用 1959—1996 年的数据重新估计这个模型，而留下两个观测用于预测。利用 Microfit 4.1，我们从 1959—1996 年间的估计回归中得到 Y 在 1997 年和 1998 年的如下预测值，既有静态预测，又有动态预测。

	1997 年	1998 年
实际 Y 值	101.1	105.1
Y 的静态预测	107.24（2.64）	109.45（2.67）
静态预测误差	−6.14	4.35
动态预测	100.75（1.08）	101.95（1.64）
动态预测误差	0.35	3.14

注：括号中的数字为预测值的估计标准误。

487 从这个练习中可见，动态预测比静态预测更接近其实际值，动态预测的标准误也比静态预测的标准误小。所以，从预测的角度来看，包含AR(1)模

式（或更高阶模式）或许有好处。不过须注意，对两种预测而言，1998 年预测的标准误比 1997 年的标准误更大，这不必吃惊，因为预测远期可能更有风险。

§ 12.13 自相关的其他方面

虚拟变量与自相关

我们在第 9 章考虑了虚拟变量回归模型。具体而言，回想我们在（9.5.1）中给出的美国 1970—1995 年间储蓄—收入回归模型，为方便起见，复制如下：

$$Y_t = \alpha_1 + \alpha_2 + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \tag{12.13.1}$$

其中　Y = 储蓄

X = 收入

$D = 1$，对 1982—1995 年间的观测

$D = 0$，对 1970—1981 年间的观测

基于此模型的回归结果在（9.5.4）中给出。当然，此模型是在通常的 OLS 假定下估计的。

但现在假设 u_t 服从一阶自回归AR(1)模式，即 $u_t = \rho u_{t-1} + \varepsilon_t$。通常若 ρ 已知或很容易用上述方法之一估计出来，我们就能用广义差分方法在没有（一阶）自相关的情况下估计模型参数。但虚拟变量 D 的出现提出了一个特殊问题：注意虚拟变量无非是一次观测被划到第一或第二期间。我们如何对此进行变换呢？我们可以按如下程序进行。[48]

1. 在（12.13.1）中，第一期间所有观测的 D 值都是零；第二期间 D 的**第一个**观测值是$1/(1-\rho)$而不是 1，其余的观测都是 1。

2. 变量 X_t 变换成（$X_t - \rho X_{t-1}$）。注意，我们在变换时丢失了一次观测，除非像前面曾指出的那样，用**普雷斯-温斯坦变换**来修复第一次观测。

3. 第一期间中所有观测的 $D_t X_t$ 值都是 0（注：第一期间中的 D_t 都为零）；在第二期间，第一个观测取值 $D_t X_t = X_t$，而其余观测都取值（$D_t X_t - D_t \rho X_{t-1}$）=（$X_t - \rho X_{t-1}$）。（注：第二期间 D_t 的值为 1。）

488 前面的讨论曾指出，**临界观测**是第二期间中的第一次观测。如果像刚才建议的那样考虑到这一点，那么在AR(1)自回归约束下估计（12.13.1）这样的回归就不应该有什么问题。在习题 12.37 中，要求读者对第 9 章给出的美国储蓄和收入数据进行这种变换。

ARCH 和 GARCH 模型

和在AR(1)模式中 t 期的误差项与（$t-1$）期的误差项相关或AR(p)模式中 t 期的误差项与各滞后误差项相关一样，t 期的方差 σ^2 是否也会与其一期或多期滞后值自相关呢？致力于预测股票价格、通货膨胀率和汇率等金融时间序列的研究者已经观察到了这种自相关。若误差方差与前一项误差的平方相关，则取一个吓人的名字——**自回归条件异方差**（autoregressive conditional heteroscedasticity，ARCH)，若误差方差与过去几期误差项的平方都相关，则命名为**广义自回归条件异方差**（generalized autoregressive conditional heteroscedasticity，GARCH)。由于该专题属于时间序列计量经济学的一般辖域，所以我们将在时间序列计量经济学的有关章节中深入探讨。我们这里的目标，只是指出自相关不仅限于当期与过去误差项之间的关系，还包括当期与过去误差方差之间的相关。

自相关与异方差的共存

如果一个回归模型同时遇到异方差和自相关的问题会怎么样呢？我们能否依次解决问题，即先考虑异方差再考虑自相关呢？事实上，有位作者认为“自回归只能在控制了异方差性之后才能侦察出来”[49]。但我们能否给出一种同时解决这两个及其他问题（如模型设定）的万能检验方法呢？回答是肯定的，存在这种检验，但对其讨论离题太远，最好把它们留在参考文献中。[50]

§12.14 要点与结论

1. 当经典线性回归模型的假定“进入总体回归模型的误差或干扰项 u_t 是随机的或不相关的”不成立时，就有序列相关或自相关的问题。

489 2. 自相关的出现有种种原因，诸如经济时间序列的惯性或粘滞性，模型遗漏了应包含的重要变量或使用了错误的函数形式所导致的设定偏误、蛛网现象、数据糅合、数据变换，等等。于是，区分纯粹自相关和因为刚才提到的一或多个因素所“引致”的自相关就很重要。

3. 虽然在自相关出现时 OLS 估计量仍是无偏和一致性的，但不再是有效的。结果，常用的显著性 t 和 F 检验都不能有效地应用。因此，需要有补救措施。

4. 如何补救与干扰项 u_t 中的相依性质有关。而由于这些 u_t 是不可观测

的，通常都要假定它们有某种生成机制。

5. 通常假定这种机制是马尔可夫一阶自回归模式，即假定现期的干扰与前期的干扰项有线性关系，自相关系数表示着它们之间相互依赖的程度。这种机制被称为AR(1)模式。

6. 如果AR(1)模式真实且自相关系数已知，则序列相关问题可通过数据变换，按照广义差分程序迎刃而解。AR(1)模式可容易地推广到一个AR(p)模式上。还可假定一个移动平均（MA）机制或AR与MA两种模式的混合，叫做ARMA。这个专题将在时间序列计量经济学的有关章节中讨论。

7. 即使在我们使用一个AR(1)模式时，自相关系数 ρ 并不是先验地预知的。我们考虑了估计 ρ 的几种方法如：德宾-沃森 d，泰尔-纳加修正 d，科克伦-奥克特两步程序，科克伦-奥克特迭代程序，以及德宾两步法。在大样本中，这些方法一般地说会产生类似的估计值；而在小样本中，它们的表现各不相同。在实践中，科克伦-奥克特迭代法用得很普遍。

8. 使用刚刚讨论过的任何一种方法，我们都能通过OLS用广义差分方法估计变换模型的参数，这种方法实质上就是GLS。但由于我们用了估计的 $\rho(=\hat{\rho})$，所以我们把这种估计方法称为可行或估计的GLS，简称为FGLS或EGLS。

9. 在使用EGLS时，去掉第一个观测时必须小心，因为在小样本情形下，保留还是去掉第一个观测对结果有显著影响。因此，在样本容量很小时，建议按照普雷斯-温斯坦程序对第一个观测进行变换。但对大样本而言，是否包括第一个观测没什么差别。

10. 必须强调指出，EGLS方法只有在大样本条件下才具有通常的优良统计性质。在小样本情况下，OLS实际上可能比EGLS更好，特别是在 $\rho<0.3$ 时。

490 11. 我们也可以不用EGLS，而使用OLS，只是要用尼威-韦斯特HAC程序对自相关问题修正其标准误。严格地讲，这一程序在大样本条件下有效。HAC程序的优势之一是，它不仅修正了自相关，还在出现异方差时修正了异方差问题。

12. 当然，自相关的侦察工作要先于补救措施。侦察的方法有正式和非正式两种。非正式方法中，可以对实际或标准化的残差描点，或者将当期残差对历史残差描点。正式的方法，可以使用游程检验、德宾-沃森 d 检验、渐近正态检验、贝伦布鲁特-韦布检验和布劳殊-戈弗雷检验。其中，最流行而又惯常使用的是德宾-沃森 d 检验，尽管它年高德劭，但仍有几方面的局限性。由于BG检验同时容许AR和MA误差结构，以及容许回归子的滞后值作为解释变量而出现，所以最好使用更一般性的BG检验。但必须牢记，它仍是一个大样本检验。

13. 我们在本章还十分简要地讨论了出现虚拟变量时自相关的侦察。自相关误差对预测的作用、以及ARCH和GARCH等专题。

习　题

问答题

12.1 判明以下陈述的真伪，简单地申述你的理由。

a. 当自相关出现时，OLS 估计量是偏误的和非有效的。

b. 德宾-沃森 d 检验假定误差项 u_i 的方差有同方差性。

c. 用一阶差分变换消除自相关是假定自相关系数 ρ 为 -1。

d. 如果一个是一阶差分形式的回归，而另一个是水平形式的回归，那么，这两个模型的 R^2 值是不可直接比较的。

e. 一个显著的德宾-沃森 d 不一定意味着一阶自相关。

f. 在自相关出现时，通常计算的预期值的方差和标准误就不是有效的。

g. 把一个（或多个）重要的变量从回归模型排除出去可能导致一个显著的 d 值。

h. 在AR(1)模式中，假设 $\rho = 1$ 既可通过贝伦布鲁特-韦布 g 统计量，也可通过德宾-沃森 d 统计量来检验。

i. 如果在 Y 的一阶差分对 X 的一阶差分的回归中有一常数项和一线性趋势项，就意味着在原始模型中有一个线性和一个二次趋势项。

491 **12.2** 给定一个含有 50 项观测的样本和 4 个解释变量，如果 (a) $d = 1.05$，(b) $d = 1.40$，(c) $d = 2.50$，(d) $d = 3.97$，你能对自相关的问题说些什么？

12.3 在研究生产中的劳动在增值（value added）中所占份额（即劳动份额）的变动时，古扎拉蒂考虑如下模型[1]：

$$\text{模型 A:}\quad Y_t = \beta_0 + \beta_1 t + u_t$$

$$\text{模型 B:}\quad Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$$

其中 Y = 劳动份额，t = 时间。根据 1949—1964 年数据，对初级金属工业得到如下结果：

$$\text{模型 A: } \hat{Y}_t = 0.452\,9 - 0.004\,1t \qquad R^2 = 0.528\,4 \quad d = 0.825\,2$$
$$(-3.960\,8)$$

$$\text{模型 B: } \hat{Y}_t = 0.478\,6 - 0.012\,7t + 0.000\,5t^2$$
$$(-3.272\,4)\ (2.777\,7)$$
$$R^2 = 0.662\,9 \qquad d = 1.82$$

其中括弧中的数字是 t 比率。

a. 模型 A 中有没有序列相关？模型 B 呢？

b. 怎样说明序列相关？

c. 你会怎样区分“纯粹”自相关和设定偏误？

12.4 **侦察自相关：冯诺伊曼检验**（Von Neumann ratio test）。[2]假定残差 $\hat{u}_t$ 是从正态分布随机抽取的，冯诺伊曼曾证明，对于大的 n，比率：

$$\frac{\delta^2}{s^2}=\frac{\sum(\hat{u}_i-\hat{u}_{i-1})^2/(n-1)}{\sum(\hat{u}_i-\bar{\hat{u}})^2/n} \qquad \text{注：OLS 的}\ \bar{\hat{u}}=0。$$

称冯诺伊曼比率，近似于正态分布，其均值为：

$$E\,\frac{\delta^2}{S^2}=\frac{2n}{n-1}$$

而方差为：
$$\operatorname{var}\frac{\delta^2}{s^2}=4n^2\frac{n-2}{(n+1)(n-1)^3}$$

a. 如果 n 足够大，你会怎样利用冯诺伊曼比率来检验自相关？

b. 德宾-沃森 d 和此比率有什么关系？

492 c. d 统计量落在 0 与 4 之间。冯诺伊曼的相应界限是什么？

d. 此比率依赖于假设“$\hat{u}$ 是从正态分布随机抽取的”，对 OLS 残差来说，这一假定的真实性如何？

e. 假使在观测次数为 100 的一项应用中发现此比率为 2.88。检验数据中无序列相关的假设。

注：B.I. 哈特（Hart）曾对多至 60 次观测的样本大小编制了冯诺伊曼比率的临界值表。[3]

12.5 在 17 个残差的一个排序中有 11 个正值和 6 个负值。游程个数是 3。这是否表明有自相关的迹象？如果游程个数是 14，你会改变答案吗？

12.6 **根据 d 统计量的泰尔-纳加 ρ 估计。**泰尔和纳加曾建议，在小样本中不把 ρ 估计为 $(1-d/2)$，而把它估计为：

$$\hat{\rho}=\frac{n^2(1-d/2)+k^2}{n^2-k^2}$$

其中 n = 观测总个数，d = 德宾-沃森 d，及 k = 待估系数个数（包括截距）。说明对于大的 n、ρ，这个估计值将化为较简单的公式 $(1-d/2)$。

12.7 **估计希尔德雷思-卢扫描或搜寻程序。**[4]由于在一阶自回归方式：

$$u_t=\rho u_{t-1}+\varepsilon_t$$

中，预期 ρ 落在 -1 与 $+1$ 之间，为确定它的位置，希尔德雷思-卢提出一种系统的“扫描”或搜寻程序。他们建议在 -1 与 $+1$ 之间按一定的间隔，比方说，每隔 0.1 单位试选 ρ 值，并通过广义差分方程（12.6.5）对数据做变换。就是，把 ρ 选为 -0.9、

−0.8、…，0.8、0.9。对每一选取的 ρ 值，做一个广义差分回归并得到相应的 RSS：$\sum \hat{u}_t^2$。希尔德雷思-卢建议最后选择使 RSS 最小（从而使 R^2 最大）的 ρ 值。如果需要更精细的结果，则还可采用更小的单位间隔，如每隔 0.01 单位而把 ρ 取为 −0.99、−0.98、…，0.90、0.91，等等。

a. 希尔德雷思-卢程序有何优越性？

b. 怎样知道最后选取的 ρ 值所做的数据转换事实上能保证 $\sum \hat{u}_t^2$ 最小？

493 **12.8** 估计 $\boldsymbol{\rho}$：科克伦-奥克特迭代程序。[5]

作为对此程序的一个说明，考虑双变量模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t \tag{1}$$

及AR(1)模式：

$$u_t = \rho u_{t-1} + \epsilon_t, \qquad -1 < \rho < 1 \tag{2}$$

于是科克伦和奥克特推荐如下步骤来估计 ρ。

1. 用通常的 OLS 方法估计（1）并得到残差 $\hat{u}_t$。顺便指出，你可以在模型中包含不止一个 X 变量。
2. 利用第 1 步得到的残差，做如下回归：

$$\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + v_t \tag{3}$$

这是（2）在实证上相对应的内容。

3. 利用（3）中得到的 $\hat{\rho}$，估计广义差分方程（12.9.6）。
4. 由于事先不知道（3）中得到的 $\hat{\rho}$ 是不是 ρ 的最佳估计值，所以把第 3 步中得到的 $\hat{\beta}_1^*$ 和 $\hat{\beta}_2^*$ 值代入原回归（1），并得到新的残差 $\hat{u}_t^*$ 为：

$$\hat{u}_t^* = Y_t - \hat{\beta}_1^* - \hat{\beta}_2^* X_t \tag{4}$$

由于 Y_t，X_t，$\hat{\beta}_1^*$ 和 $\hat{\beta}_2^*$ 皆已知，故很容易计算出来。

5. 现在估计如下回归：

$$\hat{u}_t^* = \hat{\rho}^* \hat{u}_{t-1}^* + w_t \tag{5}$$

它类似于（3），并给出 ρ 的第二轮估计值。

由于我们不知道 ρ 的第二轮估计值是不是真实 ρ 的最佳估计值，所以我们进入第三轮估计，如此等等。这正是称科克伦-奥克特程序为迭代程序的原因。我们该把这种（愉快的）轮回操作进行到什么程度呢？一般的建议是，当 ρ 的相邻两个估计值相差很小（比如不足 0.01 或 0.005）时，便可停止迭代。在工资—生产力一例中，在停止之前约需要 7 次迭代。

a. 利用你选择的软件，验证方程（12.9.16）估计的 ρ 值0.891 9 和方程（12.9.17）估计的 ρ 值 0.961 0 近似正确。

b. 科克伦-奥克特程序所得到的 ρ 值能保证全域最小，还是只能保证局部最小？

c. 选做：将科克伦-奥克特方法应用于（12.5.2）中给出的对数线性工资—生产力回归，保留第一次观测和去掉第一次观测各做一次。把结果与回归（12.5.1）中的结果做比较。

12.9 科克伦-奥克特两步法。这是科克伦-奥克特迭代程序的一个简化版。第一步，我们从第一次迭代［即从上题中的方程（3）］中估
494 计出 ρ。第二步就用所估计的 ρ 做上题方程（4）中的广义差分方程。实践中有时候，这种两步法给出的结果，与多次煞费苦心的迭代所得到的结果十分相似。

在正文给出的说明性工资—生产力回归中使用科克伦-奥克特两步法，并将结果与迭代法所得到的结果进行比较。特别注意在变换中对第一次观测的处理。

12.10 估计 ρ：**德宾的两步法**。[6]为了解释此方法，我们可以把广义差分方程（12.9.5）等价地写成：

$$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + \varepsilon_t \qquad (1)$$

德宾建议用如下两步法估计 ρ：第一步，把（1）作为一个多元回归模型将 Y_t 对 X_t，X_{t-1} 和 Y_{t-1} 做回归，并把 Y_{t-1} 回归系数的估计值（$=\hat{\rho}$）作为对 ρ 的一个估计值。第二步，得到 $\hat{\rho}$ 之后，用它估计广义差分方程（12.9.5）或与之等价的（12.9.6）中的参数。

a. 在正文中所讨论的工资—生产力例子中应用德宾两步法，并把结果与科克伦-奥克特迭代程序和科克伦-奥克特两步法所得到的结果相比较。评论你所得到结果的"质量"。

b. 检查上述方程（1），你会观察到 X_{t-1} 的系数等于 X_t 的系数（$=\beta_2$）和 Y_{t-1} 的系数（$=\rho$）之积的 -1 倍。你如何对这种系数约束进行检验？

12.11 在测量电力供给中的规模报酬时，纳洛夫（Nerlove）利用 1955 年美国 145 个私有公用事业的横截面数据，做了对数总成本对对数产出、对数工资、对数资本价格和对数燃料价格的回归。他发现由此回归估计得到的残差从德宾-沃森 d 看来，呈现"序列"相关。为了寻求补救方法，他将所估计的残差相对于对数产出描图而得到图 12.11。

a. 图 12.11 说明什么？

b. 在上述情况下，你怎样摆脱"序列"相关的问题？

12.12 将一个回归的残差对时间描图，得到图 12.12 中的散点图。打上圆圈的一个"极端"残差表示一个异常值。所谓异常值，是指远远超出样本中其他观测值的一个观测值，也许离开所有观测值的平均值 3～4 个标准差之多。

a. 出现异常值的原因是什么？

b. 如果有异常值（一或多个），是否应把它（们）剔除，然后再对其余的观测值做回归？

c. 在出现异常值的情况中，德宾-沃森 d 是否适用？

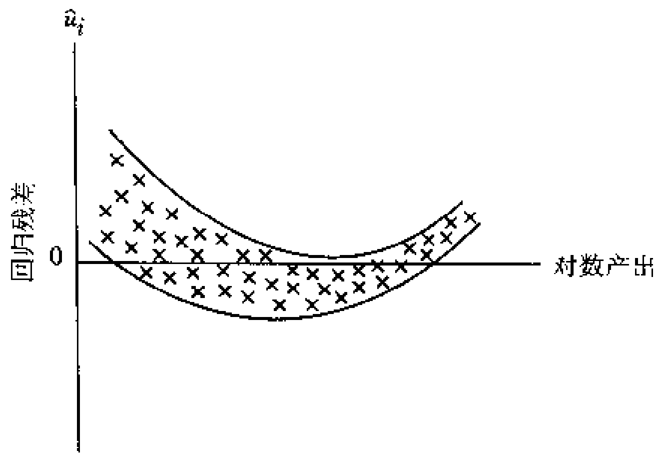


图 12.11　来自纳洛夫研究的残差

资料来源：Marc Nerlove, "Return to Scale in Electric Supply," in Carl F. Christ etal. *Measurement in Economics*, Stanford University Press, Stanford, Calif., 1963.

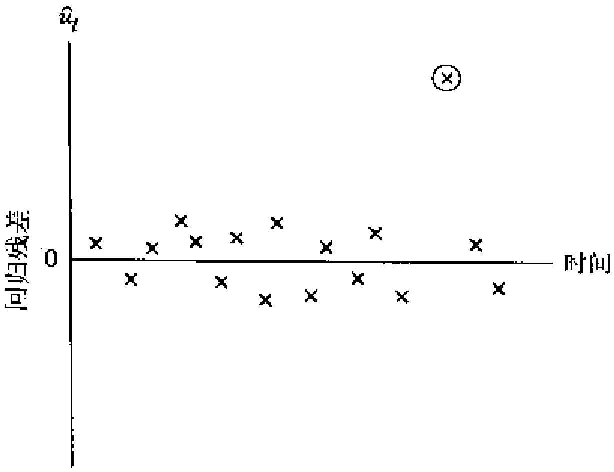


图 12.12　假想回归残差对时间描点

495 **12.13**　根据德宾-沃森 d 统计量，你怎样区分"纯粹"自相关和设定偏误？

12.14　假令在模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

中，u_t 确实序列无关。那么，如果我们假定了 $u_t = \rho u_{t-1} + \varepsilon_t$，并使用了广义差分回归：

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \varepsilon_t$$

将会出现什么情况？特别讨论干扰项 ε_t 的性质。

496

12.15 在对联合王国如何按要素成本对最终产品定价的一项研究中，根据1951—1969年期间的年度数据，得到如下的结果：

$$\widehat{PF}_t = 2.033 + 0.273W_t - 0.521X_t + 0.256M_t + 0.028M_{t-1} + 0.121PF_{t-1}$$
$$\text{se} = (0.992)(0.127) \quad (0.099) \quad (0.024) \quad (0.039) \quad (0.119)$$
$$R^2 = 0.984 \qquad d = 2.54$$

其中PF＝按要素成本的最终产品价格，W＝每个雇员的工薪，X＝每个就业人员的国内总产值，M＝进口价格，M_{t-1}＝滞后一年的进口价格，以及PF_{t-1}＝前一年按要素成本的最终产品价格。[5]

"因对18个观测值和5个解释变量，d值的5%下限和上限分别是0.71和2.06，而估计的d值是2.54，故表明无正的自相关。"试加评论。

12.16 在什么情况下以下各种估计一阶自相关系数ρ的方法是合适的：

a. 一阶差分回归。

b. 移动平均回归。

c. 泰尔-纳加变换。

d. 科克伦-奥克特迭代程序。

e. 希尔德雷思-卢搜寻程序。

f. 德宾两步法。

12.17 考虑模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

其中 $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

即误差项服从AR(2)模式，其中ε_t为白噪音误差项。在考虑二阶自回归情况下，勾勒出估计此模型的步骤。

12.18 把校正因子C包括进来，(12.3.1)所给的$\hat\beta_2^{\text{GLS}}$的计算公式是：

$$\hat\beta_2^{\text{GLS}} = \frac{(1-\rho^2)x_1y_1 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{(1-\rho^2)x_1^2 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})^2}$$

给定此公式和(12.3.1)，求校正因子C的表达式。

12.19 说明估计(12.9.5)等价于估计12.3节所讨论的、不包含对Y和X的第一次观测的GLS。

497

12.20 对于回归(12.7.2)，估计的残差值有如下的符号：

$$(++++)(-)(+++++++)(-)(++++)(--)$$
$$(+)(--)(+)(--)(++)(-)(+)(-----------)(+)$$

说明根据游程检验可以拒绝在这些残差中没有自相关的假设吗？

***12.21** **检验高阶序列相关。** 假使我们拥有季度时间序列数据，在涉及季度数据的回归中，更合适的假定将不是(12.2.1)所给的AR(1)模式，而是如下的AR(4)模式：

$$u_t = \rho_4 u_{t-4} + \varepsilon_t$$

也就是，假定当前的干扰项与上年同季度的干扰相关，而不是与上季度的干扰相关。

为检验 $\rho_4=0$ 的假设，沃利斯（Wallis）[9]建议做以下的修正德宾-沃森 d 检验：

$$d_4=\frac{\sum_{t=5}^{n}(\hat{u}_t-\hat{u}_{t-4})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

检验程序与文中讨论的通常 d 检验步骤相仿。沃利斯曾编制有 d_4 表，可在他的原始论文中找到。

现在假令我们拥有每月数据，能否将德宾-沃森检验加以推广，以适合于这种数据？如果能够，试写出适当的 d_{12} 公式。

12.22 假使你估计以下回归：

$$\Delta\ln \text{产出}_t=\beta_1+\beta_2\Delta\ln L_t+\beta_3\Delta\ln K_t+u_t$$

其中 Y=产出，L=劳动力投入，K=资本投入，而 Δ 为一阶差分运算子。你怎样解释此模型中的 β_1？可否把它解释为技术变化的一个估计值？说出你的理由。

12.23 文中曾指出，曼达拉曾建议在德宾-沃森 d 小于 R^2 时应该做一阶差分形式的回归。其背后的逻辑根据是什么？

498 **12.24** 参照方程（12.4.1）。假定 $r=0$ 但 $\rho\neq0$。若(a)$0<\rho<1$，(b)$-1<\rho<0$，其对 $E(\hat{\sigma}^2)$有何影响？$\hat{\sigma}^2$ 的偏误何时充分地小？

12.25 将（12.5.1）给出的工资—生产力回归所得到的残差对其过去6期滞后残差［即AR(6)］做回归，得到如下结果：

因变量：RESI

方法：最小二乘

样本（调整后）：1965—1998

包含观测：在调整端点后共34个。

变量	系数	标准误	t 统计量	概率
C	5.590 462	1.963 603	2.847 043	0.008 5
X	−0.066 605	0.023 469	−2.838 058	0.008 7
RES1 (−1)	0.814 971	0.216 231	3.768 978	0.000 9
RES1 (−2)	−0.268 651	0.273 887	−0.980 882	0.335 7
RES1 (−3)	−0.106 017	0.272 780	−0.388 652	0.700 7
RES1 (−4)	0.305 630	0.273 258	1.118 467	0.273 6
RES1 (−5)	−0.064 375	0.280 577	−0.229 438	0.820 3
RES1 (−6)	0.216 156	0.222 160	0.972 976	0.339 5

$R^2=0.892\ 0$　　德宾-沃森 d 统计量　1.758 9

$\bar{R}^2=0.862\ 9$

a. 你从上述结果中能得出工资—生产力数据中自相关的什么性质？

b. 你若认为AR(1)机制刻画了数据中的自相关，你会用一阶差分变换除掉自相关吗？说出你的理由。

解答题

12.26 参考表12.7中的铜工业数据：

表12.7 1951—1980年美国国内铜价格的决定因素

年份	C	G	I	L	H	A
1951	21.89	330.2	45.1	220.4	1 491.0	19.00
1952	22.29	347.2	50.9	259.5	1 504.0	19.41
1953	19.63	366.1	53.3	256.3	1 438.0	20.93
1954	22.85	366.3	53.6	249.3	1 551.0	21.78
1955	33.77	399.3	54.6	352.3	1 646.0	23.68
1956	39.18	420.7	61.1	329.1	1 349.0	26.01
1957	30.58	442.0	61.9	219.6	1 224.0	27.52
1958	26.30	447.0	57.9	234.8	1 382.0	26.89
1959	30.70	483.0	64.8	237.4	1 553.7	26.85
1960	32.10	506.0	66.2	245.8	1 296.1	27.23
1961	30.00	523.3	66.7	229.2	1 365.0	25.46
1962	30.80	563.8	72.2	233.9	1 492.5	23.88
1963	30.80	594.7	76.5	234.2	1 634.9	22.62
1964	32.60	635.7	81.7	347.0	1 561.0	23.72
1965	35.40	688.1	89.8	468.1	1 509.7	24.50
1966	36.60	753.0	97.8	555.0	1 195.8	24.50
1967	38.60	796.3	100.0	418.0	1 321.9	24.98
1968	42.20	868.5	106.3	525.2	1 545.4	25.58
1969	47.90	935.5	111.1	620.7	1 499.5	27.18
1970	58.20	982.4	107.8	588.6	1 469.0	28.72
1971	52.00	1 063.4	109.6	444.4	2 084.5	29.00
1972	51.20	1 171.1	119.7	427.8	2 378.5	26.67
1973	59.50	1 306.6	129.8	727.1	2 057.5	25.33
1974	77.30	1 412.9	129.3	877.6	1 352.5	34.06
1975	64.20	1 528.8	117.8	556.6	1 171.4	39.79
1976	69.60	1 700.1	129.8	780.6	1 547.6	44.49
1977	66.80	1 887.2	137.1	750.7	1 989.8	51.23
1978	66.50	2 127.6	145.2	709.8	2 023.3	54.42
1979	98.30	2 628.8	152.5	935.7	1 749.2	61.01
1980	101.40	2 633.1	147.1	940.9	1 298.5	70.87

注：数据来自诸如 *American Metal Market*，*Metals Week* 以及美国商务部出版物，由加里·史密斯（Gary R.Smith）收集。

C = 12 个月的平均美国国内铜价（每磅美分）；

G = 年度国民生产总值（10 亿美元）；

I = 12 个月的平均工业生产指数；

L 12 个月的平均伦敦金属交易所铜价（20 英镑）；

H = 每年新房动工数（千单位）；

A = 12 个月的平均铝价（每磅美分）。

a. 根据这些数据，估计以下回归模型：

$$\ln C_t = \beta_1 + \beta_2 \ln I_t + \beta_3 \ln L_t + \beta_4 \ln H_t + \beta_5 \ln A_t + u_t$$

并解释所得结果。

b. 求出上述回归的残差和标准化残差并作图。你对这些残差中是否有自回归做些什么猜测？

c. 估计德宾-沃森 d 统计量并对数据中可能出现的自相关性质做出评论。

d. 做游程检验，看你的答案是否不同于刚才在(c)中所得到的。

e. 你怎样辨别AR(p)过程是否比 AR(1)过程更好地描述自相关？注：保留数据做进一步的分析。(见习题 12.28。)

12.27 给定表 12.8 中的数据：

表 12.8

Y，个人消费支出（10 亿 1958 年美元）	X，时间	$\hat{Y}$,估计的 Y*	$\hat{u}_1$，残差
281.4	1（=1956）	261.420 8	19.979 1
288.1	2	276.602 6	11.497 3
290.0	3	291.784 4	−1.784 4
307.3	4	306.966 1	0.333 8
316.1	5	322.147 9	−6.047 9
322.5	6	337.329 7	−14.829 7
338.4	7	352.511 5	−14.111 5
353.3	8	367.693 3	−14.393 3
373.7	9	382.875 1	−9.175 1
397.7	10	398.056 9	−0.356 9
418.1	11	413.238 6	4.861 3
430.1	12	428.420 6	1.679 5
452.7	13	443.602 2	9.097 7
469.1	14	458.784 0	10.315 9
476.9	15（=1970）	473.965 8	2.934 1

* 估计自回归 $Y_t = \beta_0 + \beta_1 X_t + u_t$。

a. 验证德宾-沃森 $d=0.414\,8$。

b. 干扰中有没有正的序列相关？

499 c. 如有，则分别利用：

ⅰ. 泰尔-纳加方法；

ⅱ. 德宾两步程序；

ⅲ. 科克伦-奥克特法估计 ρ。

d. 利用泰尔-纳加法变换数据并对变换后的数据做回归。

e. d 中估计的回归仍显示有自相关吗？如有，如何能将它除去？

12.28 参照习题 12.26 以及那里的表 12.7 中的数据。如果该题的结果表明序列相关，

500 a. 用科克伦-奥克特两步程序，估计可行 GLS 或广义差分回归，并比较你所得的结果。

b. 如果估计自 a 的科克伦-奥克特法的 ρ 值和从 d 统计量估计得出的结果相差较大，你将选择哪一个估计 ρ 的方法，为什么？

12.29 参照例 7.4。略去变量 X^2 和 X^3，再做回归并检查残差中的"序列"相关。如果发现有序列相关，你会怎样解释它？你会提出什么补救措施？

12.30 回到习题 7.21。先验地可以预期在这样的数据中有自相关。因此，建议你以一阶差分形式做对数实际货币供给对对数实际国民收入和长期利率的回归。计算此回归和重做原始形式的回归。一阶差分变换所依据的假定是否得到满足？如果不，这种变换很可能导致什么样的偏误？利用现有的数据加以说明。

12.31 **德宾-沃森 d 用于检验非线性问题。**继续考虑习题 12.19。将得自那个回归的残差按 X 值的递增次序排列。利用（12.6.5）所给的公式，从重新排列的残差估计 d。如果计算出来的 d 值表示自相关，这将意味着线性模型是不正确的。完整的模型应包含 X_i^2 和 X_i^3 项。你能给出这一程序的直觉理由吗？看看你的答案是否同 H. 泰尔所给的相一致。[10]

501 **12.32** 回到习题 11.22。求出残差并探明残差中是否有自相关。如果发现有序列相关，你会怎样变换数据？在本例中序列相关的意义是什么？

12.33 **蒙特卡罗实验。**参照表 12.1 和 12.2。利用那里给的 ε_t 和 X_t 数据，按下列模型产生 10 个 Y 值的一个样本：

$$Y_t = 3.0 + 0.5X_t + u_t$$

其中 $u_t = 0.9u_{t-1} + \varepsilon_t$。假定 $u_0 = 10$。

a. 估计这个方程并评论你的结果。

b. 现在假定 $u_0 = 17$。重复这个练习 10 次，再评论所得结果。

c. 保持上述结构不变。仅将 $\rho = 0.9$ 改为 $\rho = 0.3$。然后将你的结果同(b)中所给的相比较。

12.34 利用表 12.9 所给数据，估计模型：

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

其中 Y = 库存，X = 销售量，均以 10 亿美元计。

502 a. 估计上述回归。

b. 从估计的残差中探明是否有正的自相关：利用（i）德宾-沃森检验和（ii）(12.6.13) 所给的大样本正态性检验。

c. 如果 ρ 是正的，用贝伦布鲁特-韦布检验去检验假设 $\rho = 1$。

d. 如果你猜测自回归误差结构的阶数是 p，可用布劳殊-戈弗雷检验去证实这点。你会怎样选择这个阶数呢？

e. 根据此检验的结果，你会怎样转换数据得以把自回归除掉？说明你的全部计算。

f. 重复前面的步骤，但用以下模型：

$$\ln Y_t = \beta_1 + \beta_2 \ln X_t + u_t$$

g. 你在线性与对数线性两种设定之间如何取舍？说明你用的检验方法。

表 12.9　1950—1991 年美国制造业的库存与销售（10 亿美元）

年份	销售*	库存†	年份	销售*	库存†
1950	38 596	59 822	1970	108 352	178 594
1951	43 356	70 242	1971	117 023	188 991
1952	44 840	72 377	1972	131 227	203 227
1953	47 987	76 122	1973	153 881	234 406
1954	46 443	73 175	1974	178 201	287 144
1955	51 694	79 516	1975	182 412	288 992
1956	54 063	87 304	1976	204 386	318 345
1957	55 879	89 052	1977	229 786	350 706
1958	54 201	87 055	1978	260 755	400 929
1959	59 729	92 097	1979	298 328	452 636
1960	60 827	94 719	1980	328 112	510 124
1961	61 159	95 580	1981	356 909	547 169
1962	65 662	101 049	1982	348 771	575 486
1963	68 995	105 463	1983	370 501	591 858
1964	73 682	111 504	1984	411 427	651 527
1965	80 283	120 929	1985	423 940	665 837
1966	87 187	136 824	1986	431 786	664 654
1967	90 918	145 681	1987	459 107	711 745

1968	98 794	156 611	1988	496 334	767 387
1969	105 812	170 400	1989	522 344	813 018
			1990	540 788	835 985
			1991	533 838	828 184

* 年度数据为每月数字而不是经调整的季度数字的平均。

† 从 1982 年开始的经季度调整的期末数字和早期的数字是不可比的。

资料来源：*Economic Report of the President*，1993，Table B-53，p.408.

503 **12.35** 表12.10给出了美国1954—1981年间普通股即期（t 期）真实回报率（RR_t）、下期（$t+1$ 期）产出增长率（OG_{t+1}）和第 t 期通货膨胀率（Inf_t）的数据，都以百分数表示。

a. 将 RR_t 对通货膨胀率做回归。

b. 将 RR_t 对 OG_{t+1} 和 Inf_t 做回归。

c. 尤金·法马观察到："真实股票回报率与通货膨胀之间简单的负相关关系是荒谬的，因为它是两个结构性关系的结果：一个是股票即期真实回报率与预期产出增长率［由 OG_{t+1} 度量］之间的正相关关系，一个是预期产出增长率与即期通货膨胀率之间的负相关关系。"借用这种观点，评论上述两个回归结果。

d. 你预计在回归 a 和 b 中会出现自相关吗？为什么？若你认为会出现，则采取适当的修正措施并给出修正后的结果。

表 12.10　　美国 1965—1982 年间 RR、OG 与 Inf 数据

年份	RR	Growth	Inflation
1954	53.0	6.7	−0.4
1955	31.2	2.1	0.4
1956	3.7	1.8	2.9
1957	−13.8	−0.4	3.0
1958	41.7	6.0	1.7
1959	10.5	2.1	1.5
1960	−1.3	2.6	1.8
1961	26.1	5.8	0.8
1962	−10.5	4.0	1.8
1963	21.2	5.3	1.6
1964	15.5	6.0	1.0
1965	10.2	6.0	2.3
1966	−13.3	2.7	3.2

1967	21.3	4.6	2.7
1968	6.8	2.8	4.3
1969	−13.5	−0.2	5.0
1970	−0.4	3.4	4.4
1971	10.5	5.7	3.8
1972	15.4	5.8	3.6
1973	−22.6	−0.6	7.9
1974	−37.3	−1.2	10.8
1975	31.2	5.4	6.0
1976	19.1	5.5	4.7
1977	−13.1	5.0	5.9
1978	−1.3	2.8	7.9
1979	8.6	−0.3	9.8
1980	−22.2	2.6	10.2
1981	−12.2	−1.9	7.3

12.36 德宾 h 统计量：考虑如下工资决定模型：

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

其中 Y = 工资 = 真实时薪指数

X = 生产率 = 每小时产出指数

a. 利用表 12.4 中的数据，估计上述模型并解释你的结论。

b. 由于模型把滞后回归子作为一个回归元包括进来，所以不适合用德宾-沃森 d 来查明数据中是否存在序列相关。对这种所谓自回归模型，德宾曾提出所谓的 **h 统计量**（h statistic）来检验一阶自相关，其定义为[11]：

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\operatorname{var}(\hat{\beta}_3)]}}$$

其中 n 为样本容量，$\operatorname{var}(\hat{\beta}_3)$ 为滞后 Y_{t-1} 系数的方差，$\hat{\rho}$ 为一阶序列相关的估计值。

在大样本情形（技术上讲，渐近情形）下，德宾已证明，在 $\rho=0$ 的虚拟假设下，

$$h \sim N(0,\ 1)$$

即 h 统计量服从标准正态分布。我们从正态分布的性质知道，$|h|>1.96$ 的概率约为 5%。因此，若在一个应用中 $|h|>1.96$，那我们就可以拒绝 $\rho=0$ 的虚拟假设，即上述自回归模型中有一阶自相关的证据。

此检验可如下进行：第一步，用 OLS 估计上述模型（目前阶段不必担心任何估计问题）。第二步，记下此模型中的$\mathrm{var}(\hat{\beta}_3)$和例行计算出来的 d 统计量。第三步，利用 d 值得到$\hat{\rho}\approx(1-d/2)$。注意，有趣的是，尽管我们不能用 d 值来检验此模型中的序列相关，但我们可用它得到 ρ 的一个估计值。第四步，现在可以计算 h 统计量。第五步，若样本容量足够大，但计算的 $|h|$ 超过了 1.96，那我们就可以断定有存在一阶自相关的证据。当然，你可以选用你想选用的任何显著性水平。

在前面给出的自回归工资决定模型中应用 h 检验，并得出适当结论，然后与回归（12.5.1）中所给出的结论相比较。

12.37 **虚拟变量和自相关**。参照第 9 章中讨论的储蓄—收入回归。利用表 9.2 中给出的数据，并假定AR(1)模式，在考虑自相关的情况下，重新估计储蓄—收入回归。特别注意对虚拟变量的变换。将结论与第 9 章给出的结论相比较。

12.38 利用表12.4 中给出的工资—生产力数据，估计模型（12.9.8），并将结论与回归（12.9.9）中给出的结论相比较。你能得到什么结论？

【习题注释】

[1] Damodar Gujarati, "Labor's Share in Manufacturing Industries," *Industrial and Labor Relations Review*, vol.23, no.1, October 1969, pp.65–75.

[2] J.von Neumann, "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," *Annals of Mathematical Statistics*, vol.12, 1941, pp.367–395.

[3] 此表可见于 Johnston 的前引文献（第三版），第 559。

[4] G.Hildreth and J.Y.Lu, "Demand Relations with Autocorrelated Disturbances," Michigan State University, *Agricultural Experiment Station*, Tech.Bull.276, November 1960.

[5] D. Cochrane and G. H. Orcutt, "Applications of Least-Squares Regressions to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, vol.44, 1949, pp.32–61.

[6] 注意 $\hat{\rho}=\sum\hat{u}_t\hat{u}_{t-1}/\sum u_t^2$（为什么?）。尽管有偏，但 $\hat{\rho}$ 仍是真实 ρ 的一个一致估计量。

[7] J. Durbin, "Estimation of Parameters in Time-Series Regression Models," *Journal of the Royal Statistical Society*, series B, vol.22, 1960, pp.139–153.

[8] 资料来源：*Prices and Earnings in 1951 – 1969: An Econometric Assessment*, Department of Employment, Her Majesty's Stationery Office, 1971, Table C, p.37, Eq.63。

[9] Kenneth Wallis, "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations," *Econometrica*, vol.40, 1972, pp. 617–636. d_4 表可从 J.Johnston，前引文献，第 3 版，第 558 页找到。

[10] Henri Theil, *Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1978, pp.307–308.

[11] J. Durbin, "Testing for Serial Correlation in Least-squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Econometrica*, vol.38, pp.410－421.

附录 12A

12A.1 (12.1.11) 中误差项自相关的证明

由于 $v_t = u_t - u_{t-1}$，所以很容易证明 $E(v_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$，因为对每个 t 都有 $E(u) = 0$。现在，$\operatorname{var}(v_t) = \operatorname{var}(u_t - u_{t-1}) = \operatorname{var}(u_t) + \operatorname{var}(u_{t-1}) = 2\sigma^2$，因为每个 u_t 的方差都是 σ^2，且 u 独立分布。因此，v_t 是同方差性的。但

$$\operatorname{cov}(v_t, v_{t-1}) = E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] = -\sigma^2$$

则显然非零。因此，尽管 u 不存在自相关，但 v 都存在。

12A.2 方程 (12.2.3)、(12.2.4) 和 (12.2.5) 的证明

在AR(1)模式下，

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{1}$$

因此

$$E(u_t) = \rho E(u_{t-1}) + E(\varepsilon_t) = 0 \tag{2}$$

所以，

$$\operatorname{var}(u_t) = \rho^2 \operatorname{var}(u_{t-1}) + \operatorname{var}(\varepsilon_t) \tag{3}$$

505 因为 u 和 ε 都不存在自相关。

因为 $\operatorname{var}(u_t) = \operatorname{var}(u_{t-1}) = \sigma^2$，且 $\operatorname{var}(\varepsilon_t) = \sigma_\varepsilon^2$，所以得到：

$$\operatorname{var}(u_t) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{4}$$

现将(1)式两边同时乘以 u_{t-1} 并取期望则得到：

$$\operatorname{cov}(u_t, u_{t-1}) = E(u_t u_{t-1}) = E[\rho u_{t-1}^2 + u_{t-1}\varepsilon_t] = \rho E(u_{t-1}^2)$$

注意到 u_{t-1} 和 ε_t 之间的协方差为零(为什么?)，而且 $\operatorname{var}(u_t) = \operatorname{var}(u_{t-1}) = \sigma_\varepsilon^2/(1-\rho^2)$，我们有

$$\operatorname{cov}(u_t, u_{t-1}) = \rho \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \tag{5}$$

依此类推，

$$\operatorname{cov}(u_t, u_{t-2}) = \rho^2 \frac{\sigma_\varepsilon^2}{(1-\rho^2)}$$

$$\operatorname{cov}(u_t, u_{t-3}) = \rho^3 \frac{\sigma_\varepsilon^2}{(1-\rho^2)} \cdots\cdots$$

现在，相关系数就是协方差与方差之比，因此，

$$\operatorname{cor}(u_t,\ u_{t-1}) = \rho \quad \operatorname{cov}(u_t,\ u_{t-2}) = \rho^2 \cdots\cdots$$

【注释】

[1] 对此，参见 William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, N.J., 2000, Chap. 11, 和 Paul A. Rudd, *An Introduction to Classical Econometric Theory*, Oxford University Press, 2000, Chap. 19。

[2] Maurice G. Kendall and William R. Buckland, *A Dictionary of Statistical Terms*, Hafner Publishing Company, New York, 1971, p.8.

[3] Gerhard Tintner, *Econometrics*, John Wiley & Sons, New York, 1965.

[4] 作为一种惯例，我们将用下标 t 表示时间序列数据，而用通常的下标 i 表示横截面数据。

[5] 如果我们发现真正的问题是设定偏误而不是自相关，则如 7.7 节所表明，(12.1.3) 的参数的 OLS 估计量可能既有偏误，又非一致性的。

[6] 关于这点，参看 William H. Greene，前引文献，第 526 页。

[7] 我们在第 4 篇也将看到，尽管 Y 和 X 非平稳，可能发现 u 是平稳的。我们以后会解释其含义。

[8] 若 $S=0$，则得到$E(u_t^2)$。由于根据假定 $E(u_t)=0$，所以 $E(u_t^2)$表示误差项的方差，显然非零。(为什么？)

[9] 这个名称是不难解释的。按定义，u_t 与 u_{t-1}之间的（总体）相关系数是：

$$\rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{\operatorname{var}(u_t)}\sqrt{\operatorname{var}(u_{t-1})}}$$

$$= \frac{E(u_t u_{t-1})}{\operatorname{var}(u_{t-1})}$$

这是因为对每一 t，$E(u_t)=0$，并且因为我们保留了同方差性假定，而有$\operatorname{var}(u_t)=\operatorname{var}(u_{t-1})$。读者应能看出，$\rho$ 还是 u_t 对 u_{t-1}回归中的斜率系数。

[10] 注意，$r = \sum x_t x_{t+1} / \sum x_t^2$ 为 X_t 与 X_{t+1}(或 X_{t-1}，因为相关系数是对称的)之间的相关系数；$r^2 = \sum x_t x_{t+2} / \sum x_t^2$ 为 X_t 滞后两期之间的相关系数，等等。

[11] 证明见 Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, pp. 274－275。校正因子与第一次观测值（Y_1X_1）有关。关于这个问题，参看习题 12.18。

[12] 关于 $\hat{\beta}_2^{GLS}$是 BLUE 的正式证明，见克曼塔，同上。但繁琐的代数证明可利用矩阵符号而大为简化。参看 J. Johnston, *Econometric Methods*, 3d ed., McGraw-Hill, New York, 1984, pp. 291－293。

[13] 但要避免繁琐的代数运算，矩阵代数几乎是必需的。

[14] 见 Kmenta，前引文献，第 277－278 页。

[15] 参看 S. M. Goldfeld and R. E. Quandt, *Nonlinear Methods in Econometrics*, North

Holland Publishing Company, Amsterdam, 1972, p. 183。顺便指出，如果误差有正相关，R^2 趋向于过高的偏误，就是说它趋向于比没有自相关时的 R^2 大。

[16] 正式的证明，参看 Kmenta，前引文献，第 281 页。

[17] 即使干扰项 u_t 是同方差的和不相关的，其估计量 $\hat{u}_t$ 则是异方差的和自相关的。关于此问题，参看 G.S.Maddala, *Introduction to Econometrics*, Macmillan, 2d ed., New York, 1992, pp. 480－481。不过，可以证明，随着样本容量无限扩大，残差趋于收敛至其真实值 u_t。对此，可参见 E.Malinvaud, *Statistical Methods of Econometrics*, 2d ed., North-Holland Publishers, Amsterdam, 1970, p.88。

[18] Stanford Weisberg, *Applied Linear Regression*, John Wiley & Sons, New York, 1980, p.120.

[19] 其实，所谓的"**学生化**"**残差**（studentized residuals），乃是有一单位方差。但在实践中，标准化残差给出的图形一般看来都无异于学生化残差所给的。因此，我们使用标准化残差也无妨。关于这个问题，参看 Norman Draper and Harry Smith, *Applied Regression Analysis*, 3d ed., John Wiley & Sons, New York, 1998, pp.207－208。

[20] 在非参数检验中，我们不对观测值所来自的分布做任何假定。关于吉尔里检验，参看 R.C.Geary, "Relative Efficiency of Count of Sign Changes for Assessing Residual Autoregression in Least Squares Regression," *Biometrika*, vol.57, 1970, pp.123－127。

[21] J.Durbin and G.S.Watson, "Testing for Serial Correlation in Least-Squares Regression," *Biometrika*, vol.38, 1951, pp.159－171.

[22] 然而，费尔布拉第（R.W.Farebrother）已计算模型中没有截距项时的 d 值。参看他的"The Durbin-Watson Test for Serial Correlation When There Is No Intercept in the Regression," *Econometrica*, vol.48, 1980, pp.1553－1563。

[23] 更详尽的细节，参见 Gabor Korosi, Laszlo Matyas, and Istvan P. Szekey, *Practical Econometrics*, Avebury Press, England, 1992, pp.88－89。

[24] 参看本节稍后关于"准确"德宾-沃森检验的讨论。

[25] 细节见 Thomas B.Fomby, R.Carter Hill, and Stanley R.Johnson, *Advanced*, *Econometric Methods*, Springer Verlag, New York, 1984, pp.225－228。

[26] 更高深的讨论，参见 Ron C.Mittelhammer, George G. Judge, and Douglas J.Miller, *Econometric Foundations*, Cambridge University Press, New York, 2000, p.550。

[27] 参见 James Davidson, *Econometric Theory*, Blackwell Publishers, New York, 2000, p. 161。

[28] 前引文献，第 161 页。

[29] Fumio Hayashi, *Econometrics*, Princeton University Press, Princeton, N.J., 2000, p.45.

[30] 参见 L.G.Godfrey, "Testing Against General Autoregressive and Moving Average Error Models When the Regressor Include Lagged Dependent Variables," *Econometrica*, vol.46, 1978, pp. 1293－1302。以及 T.S.Breusch, "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, vol.17, 1978, pp.334－355。

[31] 比如，在回归 $Y_t = \beta_1 + \beta_2 X_t + u_t$ 中，误差项可表示为 $u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2}$，它表示了白噪音误差项的一个三期移动平均。

[32] 此检验基于第 8 章曾简要提到的拉格朗日乘数原理。

[33] 包含原回归元 X 的原因，是容许 X 可能不是严格非随机的这一事实。但如果它是严格随机的，则可从模型中省掉。对此，参见 Jeffrey M. Wooldridge, *Introducto-*

ry *Econometrics: A Modern Approach*, South-Western Publishing Co. 2000, p.386。

[34] Ron C. Mittelhammer et al.，前引文献，第 547 页。记住，**一个统计检验的功效等于 1 减去犯第Ⅱ类错误的概率，即减去接受一个错误假设的概率**。一个检验的最大功效为 1，最小功效为 0。一个检验的功效越接近于 0，该检验就越糟，越接近于 1，就越有效。其本质含义是，不存在单一个最有效的自相关检验。

[35] 一次观测的损失在大样本中也许关系不大，但在小样本中可能造成有分量的差异。若不按照指出的方法对第一次观测进行变换，误差方差就不是同方差的。对此，参见 Jeffrey Wooldridge，前引文献，第 388 页。关于第一次观测的重要性的某些蒙特卡罗结果，见 Russell Davidson and James G. MacKinnon, *Estimation an Inference in Econometrics*, Oxford University Press, New York, 1993, p.349, Table 10.1。

[36] Maddala 前引文献，第 232 页。

[37] 易于证明，令 $Y_t = \alpha_1 + \beta_1 t + \beta_2 X_t + u_t$。因此，$Y_{t-1} = \alpha + \beta_1(t-1) + \beta_2 X_{t-1} + u_{t-1}$，前者减去后者须得到 $\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t$，这就表明此方程中的截距项实际上就是原模型中趋势变量的系数。记住我们假定了 $\rho = 1$。

[38] 在习题 12.38 中，要求你对包括常数项的参数做这个模型。

[39] 水平值形式与一阶差分形式的 r^2 的比较略为复杂，作为对此的补充讨论，参见 Maddala，前引文献，第 6 章。

[40] 一阶差分回归中计算出来的 d 值是否能与原水平值回归中的 d 值做同样的解释并不清楚。但用游程检验可以看出，一阶差分回归的残差中没有自相关的证据。

[41] I. I. Berenblutt and G. I. Webb, "A New Test for Autocorrelated Errors in the Linear Regression Model," *Journal of the Royal Statistical Society*, Series B, vol.35, No.1, 1973, pp.33－50.

[42] 包含第一次观测后，ρ 的迭代值为：0.914 2，0.946 2，0.955 6，0.959 1，0.960 5，0.961 0，最后一个 ρ 值用于对数据进行变换，以得到广义差分方程。

[43] 若回归元如经济数据中极其常见地表现出时间趋势，则尤其如此。

[44] W. K. Newey, and K. West, "A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, vol. 55, 1987, pp.103－708.

[45] 你若能对付矩阵代数，这种方法在 Greene 的前引文献中有所讨论，第 462－463 页.

[46] Z. Griliches, and P. Rao, "Small Sample Properties of Several Two-stage Regression Methods in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, vol. 64, 1969, pp.253－272.

[47] 进一步讨论，参见 Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, 4th ed., 1998, pp.214－217。

[48] 见 Maddala 前引文献，第 321－322 页。

[49] Lois W. Sayrs, *Pooled Time Series Analysis*, Sage Publications, California, 1989, p.19.

[50] 参见 Jeffrey M. Wooldridge 前引文献，第 402－403 页。以及 A. K. Bera and C. M. Jarque, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence," *Economic Letters*, vol. 7, 1981, pp.313－318。

第 13 章　计量经济建模：模型设定和诊断检验

506 不能机械地做应用计量经济学，它需要理解、直觉和技巧。[1]

……通常我们在驾车通过一座桥梁时并不担心其结构的可靠性，因为我们合理地相信已经有人严格地检查过其工程原理和实践。经济学家做模型时也必须这样，否则的话，就必须奉送一句警告"使用导致坍塌概不负责"[2]。

经济学家多年来对"真理"的寻求曾给人一种观感：经济学家们就好像在一间黑房子里搜寻一只原本并不存在的黑猫，而计量经济学家还经常声称找到了一只。[3]

经典线性回归模型的假定之一（假定 9）是，分析中所使用的模型被"正确地"设定了：如果模型未被"正确"设定，那我们就遇到**模型设定误差**（model specification error）或**模型设定偏误**（model specification bias）的问题。由于寻找正确设定的模型就像寻找圣杯一样，所以我们在本章将缜密而又严格地考察这个假定。具体而言，我们要考虑如下问题：

1. 我们如何去发现一个"正确"的模型？换言之，在经验分析中选择一个模型的准则有哪些？
2. 我们在实践中容易遇到哪些类型的模型设定误差？
3. 设定误差的后果有哪些？
4. 我们如何侦察设定误差？换言之，我们可以使用哪些诊断工具？
5. 一旦侦察出设定误差，我们能采取哪些补救措施？有什么好处？

6. 我们如何评价几个表现不相上下的备选模型？

模型设定与评价的专题涉及范围广泛，在此领域中已经做过许多经验研究。除此之外，对这个问题还有一些哲学上的论争。尽管我们不能在一章的篇幅内充分说明这个专题，但我们希望能澄清模型设定和模型评价过程中所涉及的一些本质问题。

§13.1 模型选择准则

根据亨德里和理查德的观点，一个被选用于经验分析的模型应满足如下准则[4]：

1. 数据容纳性；即从模型做出预测必须有逻辑上的可能性。

2. 与理论一致；即必须有好的经济含义。比如，若米尔顿·弗里德曼的**永久收入假说**（permanent income hypothesis）成立，则在永久消费对永久收入的回归中，预期截距项的值应该为零。

3. 回归元的弱外生性；即解释变量或回归元必须与误差项不相关。

4. 表现出参数的不变性；即参数值的稳定性。否则，预测就很困难。弗里德曼曾指出："对一个假设（模型）有效性惟一重要的检验，就是其预测值与经验值的比较。"[5]在参数缺乏恒常性时，这种预测就不可靠。

5. 表现出数据的协调性；即从模型中估计的残差必须完全随机（从技术上讲必须是白噪音）。换言之，若模型适当，则此模型的残差必须是白噪音。否则，模型中就存在着某种形式的设定误差。稍后，我们将阐释设定误差的性质。

6. 模型有一定的包容性；即模型从能解释其结论的意义上讲应该包容或包括所有与之相竞争的模型。简言之，其他模型都不可能再改进我们所选定的模型。

508 列出一个"好"模型的准则是一方面，但实际上做出一个"好"模型则完全是另一回事儿，因为我们在实践中很可能会遇到下一节将要讨论的各种各样的设定误差。

§13.2 设定误差的类型

假定根据刚才列举的那些准则，我们得到一个我们认为是好的模型。为简明起见，令这个模型为：

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i} \tag{13.2.1}$$

其中 Y_i = 生产的总成本，X = 产量，方程（13.2.1）是教科书中常见的立

方总成本函数的例子。

但假使出于某种原因（比如说，不想花功夫去描绘散点图），研究者决定使用以下模型：

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \tag{13.2.2}$$

注意，我们改变了符号，以区别于真实模型。

由于（13.2.1）被认为是真实的，所以采用（13.2.2）就构成一种设定误差，即**漏掉一个有关变量**（omitting a relevant variable）（X_i^3）的误差。因此（13.2.2）中的误差项 u_{2i} 事实上是：

$$u_{2i} = u_{1i} + \beta_4 X_i^3 \tag{13.2.3}$$

我们将很快看到这一关系式的重要性。

现假定另一研究者用了下述模型：

$$Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + \lambda_4 X_i^3 + \lambda_5 X_i^4 + u_{3i} \tag{13.2.4}$$

如果（13.2.1）是"真实"的，则（13.2.4）也构成一种设定误差，现在是**包含了一个无需或无关变量**（including an unnecessary or irrelevant variable）的误差，意指在真实模型中 λ_5 为零，新误差项其实是：

$$\begin{aligned} u_{3i} &= u_{1i} - \lambda_5 X_i^4 \\ &= u_{1i} \quad \text{因为真模型中 } \lambda_5 = 0 \qquad \text{（为什么?）} \end{aligned} \tag{13.2.5}$$

再假定又一研究者拟定以下的模型：

$$\ln Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + u_{4i} \tag{13.2.6}$$

和真实模型（13.2.1）相比，（13.2.6）也构成一种设定偏误。其偏误在于使用了**错误的函数形式**（wrong functional form），在（13.2.1）中 Y 以线性形式出现，而在（13.2.6）中它以对数到线性的形式出现。

509 最后，考虑有研究者使用如下模型：

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4^* X_i^{*3} + u_i^* \tag{13.2.7}$$

其中 $Y_i^* = Y_i + \varepsilon_i$ 和 $X_i^* = X_i + w_i$，ε_i 和 w_i 均为测量误差。（13.2.7）所表明的是，研究者没有使用真正的 Y_i 和 X_i，却用了含有测量误差的替代变量 Y_i^* 和 X_i^*。因此，在（13.2.7）中研究者犯了**测量偏误的误差**（errors of measurement bias）。在应用研究中，数据不免受到近似计算、不完全覆盖以及数据缺落等误差的困扰。在社会科学中我们常用到的是第二手材料，通常无法得知第一手材料收集者是否造成了误差？以及误差的类型为何？

另一种设定误差的形式与随机误差 u_i（或 u_t）进入回归模型的方式有关系。比如考虑如下不含截距项的双变量回归模型：

$$Y_i = \beta X_i u_i \tag{13.2.8}$$

其中随机误差项以乘积的形式进入回归方程，并且 $\ln u_i$ 满足 CLRM 的假定，这与误差项以相加的形式进入如下模型是不同的：

$$Y_i = \alpha X_i + u_i \tag{13.2.9}$$

尽管这两个模型中的变量相同，我们已经分别用 β 和 α 来表示（13.2.8）和（13.2.9）中的斜率系数。现在问，若（13.2.8）是“正确”或“真实”的模型，那么所估计的 α 会给出真实 β 的一个无偏估计值，即 $E(\hat{\alpha})=\beta$ 吗？若不然，则对误差项不适当的随机设定将构成设定误差的另一个根源。

总之，在提出一个经验模型时，我们很可能会遇到如下一种或多种设定误差：

1. 漏掉一个有关变量；
2. 包含一个无需变量；
3. 采用错误函数形式；
4. 测量误差；
5. 对随机误差项不正确的设定。

在转而详细考察这些设定误差之前，先区分**模型设定误差**（model specification errors）和**模型误设误差**（model mis-specification errors）很有好处。我们以上讨论误差类型的前四种在本质上都属于模型设定误差，因为在我们脑
510 海中都有一个“真实”模型，只是处于某种原因我们没有估计这个正确的模型。在模型误设误差中，我们不知道真实模型是什么。这种情形下，我们或许联想到凯恩斯学派与货币主义学派之间的争论。在解释 GDP 的变化时，货币主义者认为货币是第一位的因素，而凯恩斯主义者则强调政府支出的作用。由此看来，这是两个不相上下的竞争模型。

接下来，我们将首先考虑模型设定误差，然后考察模型误设误差。

§ 13.3 模型设定误差的后果

不管设定误差的来源是什么，由此造成什么样的后果呢？为使讨论简单，我们将按三变量模型的框架回答此问题，并在本节只考虑此前讨论过的前两种设定误差形式，即（1）**模型拟合不足**（underfitting a model）即漏掉一个有关变量，和（2）**模型拟合过度**（overfitting a model）即包含了一个无需变量。尽管这里的讨论很容易就能推广到多于两个回归元的情形，可一旦超出三个变量的情形，代数运算就会很繁琐，而且矩阵代数也几乎不可缺少。

模型拟合不足（漏掉一个有关变量）

假如真实模型是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{13.3.1}$$

但出于某种原因我们拟合了如下模型：

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i \tag{13.3.2}$$

漏掉 X_3 的后果将是：

1. 如果漏掉的变量 X_3 与包含进来的变量 X_2 相关，也就是 r_{23}非零，则 $\hat{\alpha}_1$ 和 $\hat{\alpha}_2$ 是有偏误且非一致的。就是说，$E(\hat{\alpha}_1)$不等于 β_1，$E(\hat{\alpha}_2)$不等于 β_2，而且这种偏误不会随着样本容量的增大而消失。

2. 即使 X_2 与 X_3 不相关($r_{23}=0$)，尽管现在是无偏的，但仍有偏误。

3. 误差（干扰）的方差 σ^2 将被不正确地估计。

4. 习惯上计算的 $\hat{\alpha}_2$ 的方差($=\sigma^2/\sum x_{2i}^2$)，是真实估计量 $\hat{\beta}_2$ 的方差的一个有偏误的估计量。

5. 后果是，通常的置信区间和假设检验程序，对于所估计参数的统计显著性，容易导出误导性的结论。

6. 另外一个后果是，基于不正确模型做出的预测及预测（置信）区间都是不可靠的。

511 虽然对以上命题一一给出证明会使我们离题太远[8]，但我们还是在附录 13A 的 13A.1 节证明了：

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{32} \tag{13.3.3}$$

其中 b_{32}是漏去的变量 X_3 对包含进来的变量 X_2 回归的斜率($b_{32}=\sum x_{3i}x_{2i}/\sum x_{2i}^2$)。如（13.3.3）所示，除非 β_3 或 β_{32}或二者同时为零，否则 $\hat{\alpha}_2$ 就是有偏误的。我们排除了 β_3 为零的可能性，因为如果那样的话，我们就不存在设定误差的问题。系数 b_{32}在 X_2 与 X_3 无关时为零，这在大多数经济数据中都不太可能。

然而一般而言，偏误的程度将取决于偏误项 $\beta_3 b_{32}$。如果，比方说，β_3 是正的（即 X_3 对 Y 有正的影响），并且 b_{32}也是正的（即 X_2 与 X_3 正相关）。那么，总体而言，$\hat{\alpha}_2$ 将高估了真实的 β_2（即有正的偏误）。但这种结果并不足为奇，因为，X_2 不仅代表了其对 Y 的直接影响，还包括了其（通过 X_3）对 Y 的间接影响。简单地说，本来应归功于 X_3 的影响，由于 X_3 未“获准”进入模型，所以无从展示其效应，而记在 X_2 的头上。为了做一个简洁的说明，考虑第 7 章曾讨论过的一个例子。

说明性例子：再谈儿童死亡率一例

将儿童死亡率对人均 GNP 和妇女识字率做回归，我们得到方程 (7.6.2) 所示的回归结果，并给出这两个变量的偏斜率系数值分别为 −0.005 6和 −2.231 6。但如果我们去掉 FLR 变量，我们则得到方程 (7.7.2) 所示的结果。若我们把（7.6.2）视为正确模型，则（7.7.2）因漏掉了有关变量 FLR 而成为一个误设模型。现在你可以看到，在正确的模型中，PGNP 变量的系数为 −0.005 6，而在“不正确”的模型（7.7.2）中却为 −0.011 4。

从绝对值来看，与真实模型相比，现在 PGNP 对 CM 有更大的影响。但

如果我们把 FLR 对 PGNP 回归（将被排除的变量对包含进来的变量回归），此回归的斜率系数［方程（13.3.3）中的 b_{32}］为 0.002 56。[8] 这就表明，随着 PGNP 每提高一个单位，FLR 平均上升约 0.002 56 个单位。但若 FLR 上升这么多单位，则其对 CM 的影响为 $(-2.231\,6)(0.002\,56)=\hat{\beta}_3 b_{32}=-0.005\,43$。

因此，我们最后从（13.3.3）得到 $(\hat{\beta}_2+\hat{\beta}_3 b_{32})=[-0.005\,6+(-2.231\,6)(0.002\,56)]\approx-0.011\,1$，与从不正确模型（7.7.2）中得到的 PGNP 系数值大致相等。[9] 此例说明，PGNP 对 CM 的真实影响（$-0.005\,6$）远低于不正确模型（7.7.2）所给出的结果（$-0.011\,4$）。

512 现在来分析 $\hat{\alpha}_2$ 和 $\hat{\beta}_2$ 的方差：

$$\operatorname{var}(\hat{\alpha}_2)=\frac{\sigma^2}{\sum x_{2i}^2} \tag{13.3.4}$$

$$\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)}=\frac{\sigma^2}{\sum x_{2i}^2}\text{VIF} \tag{13.3.5}$$

其中 VIF（对共线性的一种度量）为第 10 章讨论过的方差膨胀因子 $[=1/(1-r_{23}^2)]$，而 r_{23} 为变量 X_2 和 X_3 之间的相关系数；我们在第 3 章和第 7 章已经熟悉了方程（13.3.4）和（13.3.5）。

由于公式（13.3.4）和（13.3.5）不同，所以 $\operatorname{var}(\hat{\alpha}_2)$ 一般不同于 $\operatorname{var}(\hat{\beta}_2)$。但我们知道，$\operatorname{var}(\hat{\beta}_2)$ 是无偏的。（为什么？）因此，$\operatorname{var}(\hat{\alpha}_2)$ 就有偏误，从而证实了前面第 4 点论断。由于 $0<r_{23}^2<1$，所以就现在的情形来说，$\operatorname{var}(\hat{\alpha}_2)$ 小于 $\operatorname{var}(\hat{\beta}_2)$。现在，我们面临着一个两难选择：尽管 $\hat{\alpha}_2$ 有偏，但其方差比无偏估计量 $\hat{\beta}_2$ 的方差小（当然，我们现在排除了 $r_{23}=0$ 的情形，因为在实践中回归元之间总是有些相关）。所以，我们这里遇到一个取舍的问题。[10]

但问题还没有结束，因为两个模型的 RSS 不一样，其自由度也不一样，所以从模型（13.3.2）估计的 σ^2 和从真实模型（13.3.1）估计的 σ^2 也就不相同。你或许记得，我们得到 σ^2 的一个估计值为 $\hat{\sigma}^2=\text{RSS}/\text{df}$，其大小取决于模型中包含的回归元个数和自由度（$=n-$待估计参数的个数）。* 如果我们现在在模型中增加变量，RSS 通常会减小（记住，随着模型中引入越来越多的变量，R^2 不断变大），但待估计参数增加也使自由度下降。净影响取决于 RSS 的下降是否足以抵消增加回归元所导致的自由度损失。很有可能，一个回归元对回归子具有强烈影响（比如它使 RSS 的减少比在模型中引入此变量所导致的自由度损失大得多），因此，包含这种变量不仅减少偏误，而且还会提高估计量的精确性（即减小标准误）。

另一方面，如果有关的那些变量对回归子只有微弱的影响，而且它们之间高度相关（即 VIF 很大），那么，我们虽然可能减小了模型中所包含变量系数的偏误，但也增大了其标准误（使得效率下降）。事实上，此时在偏误

* 原文漏掉减号。——译者注

和精确性之间的取舍将取决于各回归元的相对重要性。

513 为了给这种讨论做出个结论，让我们考虑 $r_{23}=0$ 的特殊情形，即 X_2 和 X_3 无关的情形。这时，b_{32}将等于零。(为什么?) 从而能从 (13.3.3) 看出，$\hat{\alpha}_2$ 现在是无偏的。[11] 而且，似乎从 (13.3.4) 和 (13.3.5) 看来，$\hat{\alpha}_2$ 和 $\hat{\beta}_2$ 的方差相等。那么，是否从模型中略去 X_3，尽管理论上它是个有关变量，也不致有什么害处呢? 一般来说，回答是否定的，因为这时从 (13.3.4) 估计出来的 $\mathrm{var}(\hat{\alpha}_2)$仍是有偏误的，致使我们的假设检验程序仍值得怀疑。[12] 再说，在大多数经济研究中，X_2 和 X_3 都会相关，从而产生了以上所讲的问题。**论点已十分清楚：一旦根据相关理论把模型建立起来，切忌从中再忽略掉一个变量。**

包含一个无关变量（模型拟合过度）

现在让我们假定：

$$Y_i=\beta_1+\beta_2X_{2i}+u_i \tag{13.3.6}$$

是真实模型，而我们拟合了以下模型：

$$Y_i=\alpha_1+\alpha_2X_{2i}+\alpha_3X_{3i}+v_i \tag{13.3.7}$$

从而导致了在模型中引入一个无需变量的设定误差。

这一假定误差将导致如下后果：

1. "不正确"模型中全部参数的 OLS 估计量都是无偏而又一致的，即 $E(\alpha_1)=\beta_1$，$E(\hat{\alpha}_2)=\beta_2$ 和 $E(\hat{\alpha}_3)=\beta_3=0$。

2. 误差方差 σ^2 的估计是正确的。

3. 通常的置信区间和假设检验程序仍然有效。

4. 然而，一般地说，诸 α 系数的估计量将是非有效的，也就是说，它们的方差一般都大于真实模型中 $\hat{\beta}$ 的方差。关于这些命题的部分证明，见于附录 13A 的第 13A.1 节。这里要考虑的问题是，这些 $\hat{\alpha}$ 的相对无效性，这是不难证明的。

我们从通常使用的 OLS 公式得知：

$$\mathrm{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_{2i}^2} \tag{13.3.8}$$

514 和

$$\mathrm{var}(\hat{\alpha}_2)=\frac{\sigma^2}{\sum x_{2i}^2(1-r_{23}^2)} \tag{13.3.9}$$

因此，
$$\frac{\mathrm{var}(\hat{\alpha}_2)}{\mathrm{var}(\hat{\beta}_2)}=\frac{1}{1-r_{23}^2} \tag{13.3.10}$$

由于 $0\leqslant r_{23}^2\leqslant 1$，推知 $\mathrm{var}(\hat{\alpha}_2)\geqslant \mathrm{var}(\hat{\beta}_2)$，也就是说，虽然平均而言，$\hat{\alpha}_2=\hat{\beta}_2$[即 $E(\hat{\alpha}_2)=\beta_2$]，但 $\hat{\alpha}_2$ 的方差一般都大于 $\hat{\beta}_2$ 的方差。

这一发现的含义是，包含无关变量 X_3 将使 $\hat{\alpha}_2$ 的方差不必要地增大，从而使 $\hat{\alpha}_2$ 的精度减小。这对 $\hat{\alpha}_1$ 也是成立的。

注意我们所考虑的两种设定偏误的不对称性。如果我们略去了一个有关变量，则留在模型中变量的系数，一般地说既偏误且非一致，误差的估计也是不正确的，从而通常的假设检验程序都是无效的。而另一方面，模型中含有无关变量，虽然仍能给出真实模型中系数的无偏且一致估计，而且通常的假设检验方法也仍然有效；但引入多余变量的惟一代价是，系数方差的估计值变大了，致使对参数进行概率推断的精度降低了。一个无益的结论似乎是：与其略掉有关变量，不如含有无关变量。但是这种哲理是不值得维护的，因为，增加无关变量将导致估计量的效率损失，并且还可以引发多重共线性问题，更不用说自由度的损失了。因此，

> 一般而言，最好的方法是，根据理论，仅仅包含那些直接影响因变量、而又不能由已被引进的其他变量来代替的解释变量。[13]

§13.4 设定误差的检验

知道设定误差的后果是一回事，而发现是否犯了这种错误则完全是另一回事。因为，没有人要成心去犯这种错误，出现设定误差常常是疏忽所致，
515 或由于基础理论薄弱，或由于缺乏检验模型的适当数据，所以我们想尽可能准确地制定模型而又无能为力。如戴维森曾指出：“由于经济学的非实验性质，所以我们一向对所观测到的数据的生成机制没有信心。对经济学中任何一个假设的检验，最终都取决于足以设定一个适当节俭的模型的附加假设，而这些假设既可能被证明是合理的，也可能没有被证明是合理的。”[14]

于是，实际问题不是为什么会造成这种错误（因为通常都会造成），而是如何发现这种错误。一旦发现了有设定误差，也就常常能找出补救的方法。例如，如果能证明模型中不恰当地漏掉了一个变量，明显的补救方法就是把那个变量加到分析中去，当然这里假定能获得有关的数据。

我们在本节中就是要讨论一些可用来侦察设定误差的检验方法。

侦察是否含有无需变量（对过度拟合模型的侦察）

假如，为了解释某一现象我们提出一个 k 变量模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{13.4.1}$$

然而，比方说，变量 X_k 是否真的属于模型（所应包含的变量），我们却无十分把握。一个简单的辨认方法，是用通常的 t 检验即 $t = \hat{\beta}_k/\mathrm{se}(\hat{\beta}_k)$ 去检验所估计 $\hat{\beta}_k$ 的显著性。又比方说，我们不能肯定 X_3 和 X_4 是否真的属于模

型。于是，我们很容易通过第 8 章所讲的 F 检验来判断。因此，侦察模型中是否出现了一个或多个无关变量并不困难。

但至关重要的是，记住在做这些显著性检验时，我们心目中得有一个具体的模型。尽管这个模型在某种程度上是试验性质的，但我们应把它看作保留假设或"真理"。于是，给定该模型，我们可通过平常的 t 和 F 检验去辨认一个或多个回归元是不是真正有关的变量。但要注意，切勿反复使用 t 和 F 检验来建立模型。就是说，我们不可以说，Y 之所以与 X_2 有关，只因为 $\hat{\beta}_2$ 是统计显著的。然后又因为 $\hat{\beta}_3$ 是统计显著的便把 X_3 包含在模型中，如此等等。这种建模策略被称为**自下而上的方法**（bottom-up approach）（从一个较小的模型开始，然后逐渐扩大模型），或者多少带些轻蔑口吻地称之为**数据开采**（data mining）方法、**回归捕捉**（regression fishing）方法、**数据琢磨**（data grubbing）方法、**数据窥探**（data snooping）方法、**数字斟酌**（number crunching）方法等。

516 数据开采的主要目标是在进行一些诊断检验之后提出一个"最好的"模型，即使最终选定的模型在如下意义上是一个"好"模型：其所有估计系数都具有"正确的"符号、基于 t 和 F 检验都是统计显著的、R^2 值足够高、德宾-沃森 d 统计量的值可以接受（约为 2）等。本专业的纯化论者很看不起数据开采的实践。用威廉·普尔（William Pool）的话说，"……寻求数据基础的经验规律而非经济理论的含义总是很危险的。"[15] "谴责"数据开采的原因之一如下：

在数据开采情况下的名义与真实显著性水平。不经心的研究者所面临的一种数据开采的危险是，诸如 1%、5%或 10%的常用显著性水平（α）并非真正的显著性水平。洛弗尔（Lovell）曾指出，如果有 c 个备用的回归元，根据数据开采情况，从中最后选出 k 个（$k \leqslant c$），则真实的显著性水平（α^*）和名义上的显著水平（α）有如下关系[16]：

$$\alpha^* = 1 - (1 - \alpha)^{c/k} \tag{13.4.2}$$

或近似地为：

$$\alpha^* \approx (c/k)\alpha \tag{13.4.3}$$

例如，取 $c = 15$，$k = 5$ 和 $\alpha = 5\%$，由（13.4.3），真实的显著性水平是 $(15/5)(5\%) = 15\%$。因此，如果一个研究者从数据开采中选择 15 个回归中的 5 个，而仅按名义的显著水平 5%报告这个浓缩模型的结果，并宣称这些结果在统计上是显著的，那么要别人接受这种结果，就是一桩有苦难言的事。因为我们知道，真实的显著性水平实际上是 15%。应该指出，若 $c = k$，则不存在数据开采的问题，真实和名义的显著性水平也就相同。当然，在实践中，研究者都是仅报告其最后结果而不透露此前是如何通过大量数据开采或**预检验**（pretesting）而得到这些结果的详情。[17]

尽管数据开采有一些明显的缺陷，但仍不断得到承认，特别是在应用计量经济学家中，纯粹主义者（即非数据开采）的建模方法完全抵挡不住如此

强烈的攻势。如查曼（Zaman）指出：

> 不幸的是，运用真实数据集的经验表明，这样一种（纯粹主义）方法既不可行又不理想。之所以说它不可行，是因为很少有经济理论只导致惟一的
> 517 一个模型。之所以说它不理想，是因为从数据中了解到的一个关键方面是数据支持和不支持什么类型的模型。即便原模型碰巧表现出优良的拟合性质，但说明和了解数据支持和不支持的模型类型也越来越重要。[18]

帕特森（Kerry Patterson）也表达出类似的观点，他指出：

> （数据开采）这种方法表明，经济理论与经验设定相互影响而不是各自为阵。[19]

为了避免陷入数据开采与纯粹主义关于建模思路的论争，我们可以借鉴彼得·肯尼迪（Peter Kennedy）的观点：

> （模型设定）需要对理论和数据的通盘考虑，设定搜索中所用到的检验程序应按照使数据开采成本最小的要求来设计。这方面的例子有：为样本外预测的检验预留下数据、调整显著性水平（罗维尔）和避免使用诸如最大化 R^2 之类值得怀疑的准则。[20]

如果我们从一个更开阔的视角来看待数据开采，把它看成一种寻求经验规律的过程，并能从这些经验规律中判断（现有）理论模型中是否存在错误和/或疏漏，那么，它的作用就太大了。再次引用肯尼迪的话："应用计量经济学家的艺术在于，容许数据驱动理论进展而又不致陷入太大的数据开采的危险。"[21]

对遗漏变量和不正确函数形式的检验

实际上，我们永远不能肯定用来做经验检验的模型是"真理，完全的真理，非真理莫属"。我们是根据理论或洞察力和先前的经验工作建立一个我们认为能抓住问题实质的模型，然后对模型进行经验检验。在获得结果之后，再按照上面讨论过的良好模型准则做事后调查分析。到了这一步我们才会知道所选模型是否适宜。在决定模型的适宜性时，我们着眼于结果的一些概括性特征，诸如 $\bar{R}^2$ 值，估计的 t 比率，估计的系数符号与事先预期的是否一致，德宾-沃森统计量，等等。如果这些诊断特征合理，我们就宣称所
518 选模型是现实的良好代表。同理，如果结果看来不够理想，或由于 $\bar{R}^2$ 值太低，或由于统计上显著或有正确符号的系数太少，或由于德宾-沃森 d 值太低，我们便开始担心模型的适宜性，并着手寻找补救方法。也许我们漏掉某个重要变量，或者用了错误的函数形式，或者没有先求时间序列的差分（以消除序列相关），如此等等。为了帮助我们确定模型不适宜性是否由这些问题的一种或几种所引起，不妨利用如下的一些方法。

残差分析。如在第 12 章中所看到的，对残差的分析曾是侦察自相关和

异方差性的一种良好的视觉鉴别法。然而，尤其对横截面数据而言，残差还可以用于分析模型的设定误差，诸如一个重要变量的遗漏或函数形式的误用。事实上，如果存在这种误差，残差图将会显示出明显不同的形状。

为了说明问题，让我们再考虑最先在第 7 章讨论过的立方总生产成本函数，假定真实总成本函数可表述如下，其中 Y＝总成本和 X＝产出：

$$Y_i=\beta_1+\beta_2X_i+\beta_3X_i^2+\beta_4X_i^3+u_i \tag{13.4.4}$$

但研究者拟合了以下的二次函数：

$$Y_i=\alpha_1+\alpha_2X_i+\alpha_3X_i^2+u_{2i} \tag{13.4.5}$$

而另一研究者则拟合下面的线性函数：

$$Y_i=\lambda_1+\lambda_2X_i+u_{3i} \tag{13.4.6}$$

虽然我们明知两位研究者都造成了设定误差，但为了教学的目的，不妨看看这些模型的残差估计值是怎样一种模样。（表 7.4 给出了这些成本—产出数据。）图 13.1 不言而喻：从左至右，逐渐接近真实模型的过程中，不仅残差（在绝对值上）变小，而且与错用模型联系在一起的突出的周期摆动也逐渐消失。

由此看到残差图分析的效用：如果有设定误差，残差图必定展现出明显的样式。

519

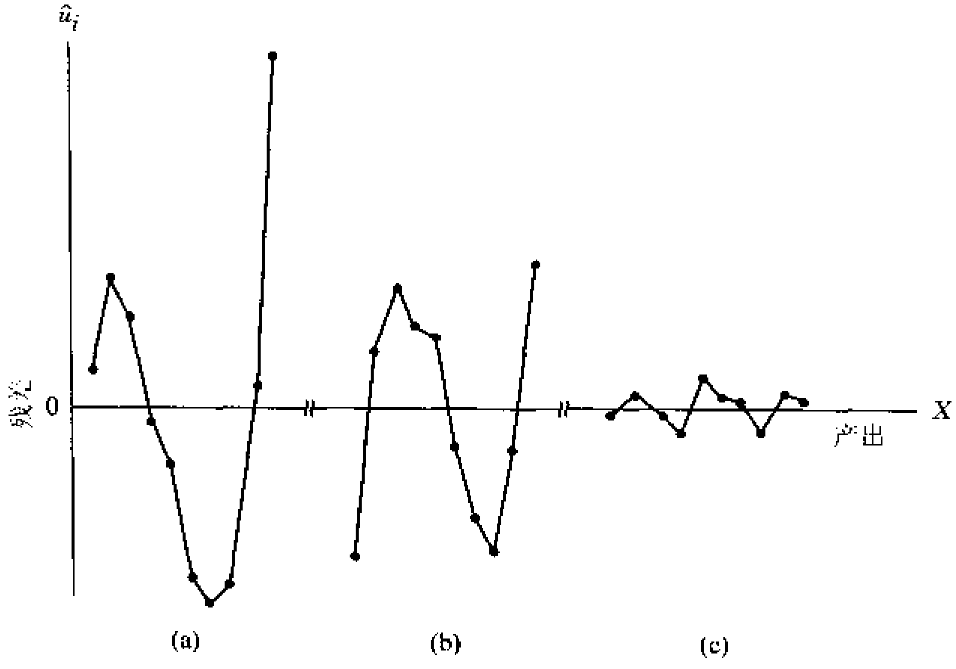


图 13.1　得自（a）线性、（b）二次及（c）立方总成本函数的残差 $\hat{u}_i$

再次使用德宾-沃森 d 统计量。如果我们分析一下表 13.1 中惯常算出的德宾-沃森 d 值，我们便看到，对线性成本函数，估计的 d 是 0.716，表明在估计的残差中有正的“相关”，因为，对 $n=10$ 和 $k'=1$，5%的临界 d 值

520 是 $d_L=0.879$ 和 $d_U=1.320$（而 0.716 低于 $d_L=0.879$）。同理，对二次成本函数，估算的 d 是 1.038，则 5%临界值是 $d_L=0.697$ 和 $d_U=1.641$，从而表明无定论。但若使用修改的 d 检验（见第 12 章），则因估算的 d 值小于 d_U，而可以说残差中有正"相关"。至于立方成本函数这一真实的设定，估计的 d 值表示残差中无任何正"相关"。[22]

表 13.1　　从线性、二次以及立方总成本函数估计的残差

观测号	线性模型的 $\hat{u}_i$*	二次模型的 $\hat{u}_i$†	立方模型的 $\hat{u}_i$**
1	6.600	23.900	−0.222
2	19.667	9.500	1.607
3	13.733	18.817	−0.915
4	−2.200	13.050	−4.426
5	−9.133	11.200	4.435
6	−26.067	−5.733	1.032
7	−32.000	−16.750	0.726
8	−28.933	−23.850	−4.119
9	4.133	−6.033	1.859
10	54.200	23.700	0.022

* $\hat{Y}_i=166.467+19.933X_i$　　$R^2=0.840\,9$

(19.021) (3.066)　　$\bar{R}^2=0.821\,0$

(8.752) (6.502)　　$d=0.716$

† $\hat{Y}_i=222.383-8.025\,0X_i+2.542X_i^2$　　$R^2=0.928\,4$

(23.488) (9.809) (0.869)　　$\bar{R}^2=0.907\,9$

(9.468) (−0.818) (2.925)　　$d=1.038$

** $\hat{Y}_i=141.767+63.478X_i-12.962X_i^2+0.939X_i^3$　　$R^2=0.998\,3$

(6.375) (4.778) (0.985 6) (0.059 2)　　$\bar{R}^2=0.997\,5$

(22.238) (13.285) (−13.151) (15.861)　　$d=2.70$

当我们拟合线性或二次模型时，所观测的正"相关"并非（一阶）序列相关，而是（模型）设定误差的一种度量。所看到的相关仅反映这样一个事实：本属于模型的某些变量被并入到误差项中去了，故而需从误差中分离出去作为解释变量而单独存在；如果我们从成本函数中排除 X_i^3，则如 (13.2.3) 所示，误设模型 (13.2.2) 中的误差项实际上是 $(u_{1i}+\beta_4X_i^3)$，因而，如果事实上 X_i^3 显著地影响着 Y，它必然展现出一种系统性的模式（例如正自相关）。

为了用德宾-沃森检验来侦察模型设定误差，我们如下进行：

1. 从假定的模型求得 OLS 残差。

2. 如果认为假定的模型因排除了一个有关的解释变量 Z（比方说）所以是误设的，就可将步骤 1 中所得的残差按 Z 值的递增次序排列。注：Z 变量可以是假定模型所含的 X 变量之一，或该变量的某一函数，如 X^2 或 X^3。

3. 从这样排列的残差，按通常的 d 公式，即：

$$d=\frac{\sum_{t=2}^{n}(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

计算 d 统计量。注意，下标 t 是这里（重排）的观测次数，不一定指时间序列数据。

4. 根据德宾-沃森表，如果 d 值是显著的，就可接受模型误设的假设。当这种情况出现时，补救措施也就寓于其中。

521 在我们的成本例子中，Z（$=X$）变量（产出）已经按从小到大的次序排列[23]，因此无需重新计算 d 统计量。正如我们已经看到的，对于线性和二次成本函数，d 统计量都表明了设定误差，如何补救是明显的：在线性成本函数中引进二次和三次项；在二次函数中引进三次项，简言之，使用立方成本模型。

拉姆齐的 RESET 检验。拉姆齐（Ramsey）曾提出称为 **RESET**（regression specification error test，意谓回归设定误差检验）的一般性设定误差检验。这里我们仅说明这种检验的最简单情形。为了建立概念，仍用我们的成本—产出例子并假定成本函数对产出是线性的：

$$Y_i=\lambda_1+\lambda_2X_i+u_{3i} \tag{13.4.6}$$

其中 Y = 总成本，X = 产出。如果用得自此回归的残差 $\hat{u}_i$ 对此模型的 Y_i 估
522 计值 $\hat{Y}_i$ 描图，就会得到一个如图 13.2 所示的图形。虽然 $\sum\hat{u}_i$ 和 $\sum\hat{u}_i\hat{Y}_i$ 都必然是零（为什么？参看第 3 章），图中的残差仍表明其均值系统地随 $\hat{Y}_i$ 而变化的模式。从而提示我们，如果以某种形式将 $\hat{Y}_i$ 当作回归元引入(13.4.6)，则应使 R^2 增大。而如果 R^2 的增大是统计上显著的（在第 8 章所讨论的 F 检验的基础上），就表明线性成本函数（13.4.6）是误设的。这就是 RESET 的基本思路。**RESET 的操作步骤如下：**

1. 从所选模型，例如（13.4.6），得到 Y_i 的估计值 $\hat{Y}_i$。

2. 将某种形式的 $\hat{Y}_i$ 作为增补的回归元引入，重做（13.4.6）。由图 13.2 我们看出 $\hat{u}_i$ 与 $\hat{Y}_i$ 之间有曲线关系，表明可引进 $\hat{Y}_i^2$ 和 $\hat{Y}_i^3$ 作为增补回归元，于是我们做回归：

$$Y_i=\beta_1+\beta_2X_i+\beta_3\hat{Y}_i^2+\beta_4\hat{Y}_i^3+u_i \tag{13.4.7}$$

3. 记来自（13.4.7）的 R^2 为 $R^2_{新}$；得自（13.4.6）的 R^2 为 $R^2_{旧}$，然后利用首次在（8.5.18）中引入的 F 检验，即：

$$F=\frac{(R^2_{新}-R^2_{旧})/新回归元的个数}{(1-R^2_{新})/(n-新模型中的参数个数)} \tag{8.5.18}$$

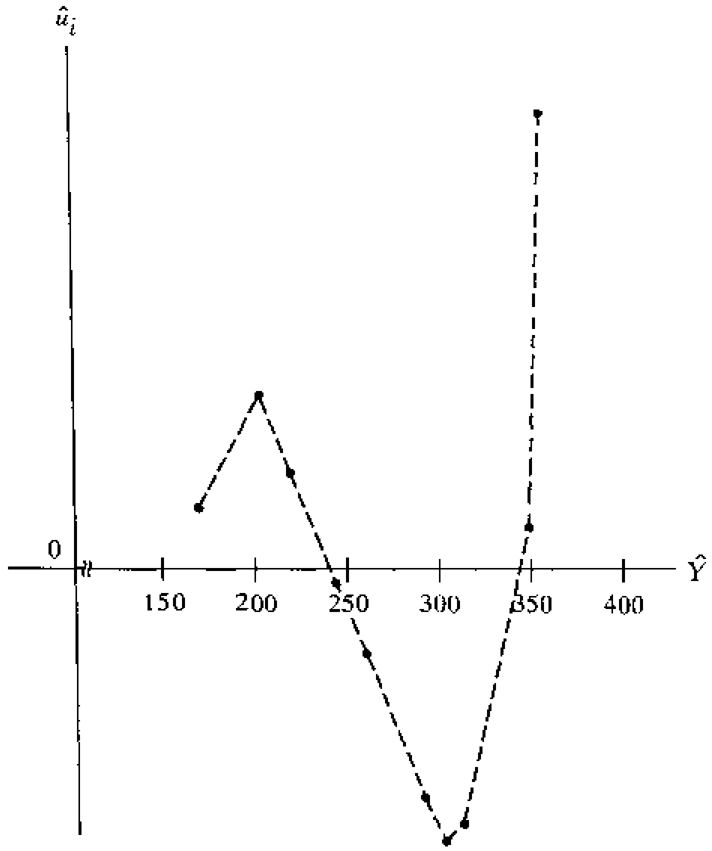


图 13.2　得自线性成本函数 $Y_i=\lambda_1+\lambda_2X_i+u_i$ 的残差与 Y 的估计值

判别由于（13.4.7）的使用，R^2 的增大是不是统计上显著的。

4．如果所计算的 F，比方说，在 5% 的水平上显著，就可接受模型（13.4.6）被误设的假设。

回到我们的说明性例子，我们有如下结果（括号中为标准误）：

$$\hat{Y}_i = 166.467 + 19.933X_i \tag{13.4.8}$$
$$(19.021)\quad(3.066)\qquad\qquad R^2=0.840\,9$$

$$\hat{Y}_i = 2\,140.722\,3 + 476.655\,7X_i - 0.091\,87\hat{Y}_i^2 + 0.000\,119\hat{Y}_i^3 \tag{13.4.9}$$
$$(132.004\,4)\quad(33.395\,1)\quad(0.006\,20)\quad(0.000\,007\,4)$$
$$R^2=0.998\,3$$

注：（13.4.9）中的 $\hat{Y}_i^2$ 和 $\hat{Y}_i^3$ 得自（13.4.8）。

现应用 F 检验求得：

$$F=\frac{(0.998\,3-0.840\,9)/2}{(1-0.998\,3)/(10-4)}=284.403\,5 \tag{13.4.10}$$

523 读者容易核实此 F 值是高度显著的，表明了模型（13.4.8）是误设的。当然，我们根据残差的视觉分析和德宾-沃森 d 值已得到过同样的结论。

RESET 的优点之一是，它不要求设定对立模型，故易于应用。但这问

时也是它的缺点，因为即使知道了模型误设，也不一定有助于另外选出一个更好的模型。

为增补变量的拉格朗日乘数（LM）检验。这是相对于拉姆齐的 RESET 检验的另一检验。为说明此检验。我们继续用前述说明性例子。

如果将线性成本函数（13.4.6）同立方成本函数（13.4.4）相比，前者就是后者的一个受约束形式（restricted version）（回顾我们在第 8 章中关于约束最小二乘的讨论）。受约束的回归（13.4.6）假定了平方和立方产出项的系数均为零。为检验此假定，LM 检验进行如下：

1. 用 OLS 法估计受约束回归（13.4.6）并求得残差 $\hat{u}_i$。

2. 如果无约束的回归（13.4.4）事实上是真实回归，则得自（13.4.6）的残差应与平方产出 X_i^2 和立方产出 X_i^3 有关。

3. 这就建议我们用在步骤 1 中得到的 $\hat{u}_i$ 去对全部回归元（包括受限回归中的回归元）做回归，这在本例中是指：

$$\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i \qquad (13.4.11)$$

其中 v 是具有通常性质的一个误差项。

4. 对于大样本，恩格尔曾证明，从（辅助）回归（13.4.11）估计出来的 R^2 的 n（样本大小）倍遵循自由度等于受约束回归中约束个数的 χ^2 分布，用符号表示，可写成：

$$nR^2 \underset{\text{asy}}{\sim} \chi^2_{\text{约束个数}} \qquad (13.4.12)$$

其中 asy 表示渐近地，即在大样本中。

524 5. 如果从（13.4.12）得到的 χ^2 值大于选定显著性水平的临界 χ^2 值，就拒绝受限回归；否则不拒绝。

对于我们的例子，回归结果如下：

$$\hat{Y}_i = 166.467 + 19.333X_i \qquad (13.4.13)$$

其中 Y 是总成本，而 X 是产出。关于此回归的标准误已见于表 13.1。

用从（13.4.13）得到的残差按方才讲的步骤 3 做回归时，得到如下结果：

$$\begin{aligned}
\hat{u}_i &= -24.7 + \quad 43.544\,3X_i - 12.961\,5X_i^2 + 0.939\,6X_i^3 \\
\text{se} &= \quad (6.375)\quad (4.779) \qquad (0.986) \qquad (0.059) \qquad\qquad (13.4.14) \\
&\qquad\qquad\qquad\qquad\qquad R^2 = 0.989\,6
\end{aligned}$$

虽然我们的样本只有 10 个，谈不上是大样本，但仅仅为了说明 LM 的操作方法，我们算出 $nR^2 = (10)(0.989\,6) = 9.896$。从 χ^2 表我们读出 2 个自由度的 1% 临界 χ^2 值是 9.21，因此，所测的 9.896 这个值是在 1% 水平上是统计显著的，从而我们的结论是拒绝约束回归（即线性成本函数）。我们得到与基于拉姆齐的 RESET 检验类似的结论。

§ 13.5　测量误差

我们一直在隐含地假定，我们对因变量 Y 和诸解释变量 X 的观测无任何误差。例如，在消费支出对家庭收入和财富的回归中，我们假定对这些变量的（观测）数据是“准确”的；它们不是由外推、内插或按任何系统的方式进位（如进位到最近似的百分之一美元，等等）而得到的猜测估计值（guess estimates）。可惜，这种理想情形由于种种原因，如非应答误差、报道误差和计算误差，实际上是找不到的。且不管什么原因，由于测量误差构成了又一类设定偏误，并带来下面注明的后果，所以它是一个潜在的麻烦问题。

因变量 Y 中的测量误差

考虑以下模型：

$$Y_i^* = \alpha + \beta X_i + u_i \tag{13.5.1}$$

其中 Y_i^* = 永久性消费支出[26]

X_i = 当前收入

u_i = 随机干扰项

525 由于 Y_i^* 不可直接观测，所以我们可能使用了这样的一个可观测变量 Y_i：

$$Y_i = Y_i^* + \epsilon_i \tag{13.5.2}$$

其中 ϵ_i 表示 Y_i^* 中的测量误差。于是我们估计的不是（13.5.1）而是：

$$\begin{aligned} Y_i &= \alpha + \beta X_i + u_i + \epsilon_i \\ &= \alpha + \beta X_i + (u_i + \epsilon_i) \\ &= \alpha + \beta X_i + v_i \end{aligned} \tag{13.5.3}$$

其中 $v_i = u_i + \epsilon_i$ 是一个合成误差项，包含着总体干扰项（也可称为方程误差项）和测量误差项。

为简单起见，假定 $E(u_i) = E(\epsilon_i) = 0$，$\mathrm{cov}(X_i, U_i) = 0$，$\mathrm{cov}(X_i, \epsilon_i) = 0$；就是说 Y_i^* 的测量误差与 X_i 不相关，和 $\mathrm{cov}(u_i, \epsilon_i) = 0$；就是说方程误差与测量误差不相关，有了这些假定就可以证明，从（13.5.1）或从（13.5.3）估计 β 都将给出真实 β 的一个无偏估计量（见习题 13.7）；就是说，因变量中的测量误差并不破坏 OLS 估计量的无偏性质。然而，从（13.5.1）和从（13.5.3）估计的 β 的方差和标准差将是不同的。这是因为，按照通常的公式（见第 3 章）我们得到：

$$\text{模型}(13.5.1): \mathrm{var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} \tag{13.5.4}$$

$$\text{模型(13.5.3)}:\operatorname{var}(\hat{\beta}) = \frac{\sigma_v^2}{\sum x_i^2} = \frac{\sigma_u^2 + \sigma_\varepsilon^2}{\sum x_i^2} \tag{13.5.5}$$

显然后者大于前者。[27]**因此，虽然因变量中的测量误差不影响参数估计及其方差的无偏性，但这时所估计的方差，却比没有这种测量误差时要大些。**

解释变量 X 中的测量误差

现假定模型不是（13.5.1）而是：

$$Y_i = \alpha + \beta X_i^* + u_i \tag{13.5.6}$$

其中 Y_i = 当前消费支出

X_i^* = 永久性收入

u_i = 干扰项（方程误差）

假令我们观测到的不是 X_i^*，而是 X_i：

$$X_i = X_i^* + w_i \tag{13.5.7}$$

其中 w_i 代表 X_i^* 中的测量误差，从而我们估计的不是（13.5.6），而是：

$$\begin{aligned} Y_i &= \alpha + \beta(X_i - w_i) + u_i \\ &= \alpha + \beta X_i + (u_i - \beta w_i) \\ &= \alpha + \beta X_i + z_i \end{aligned} \tag{13.5.8}$$

其中 $z_i = u_i - \beta w_i$，是方程与观测两种误差的一个混合。

现在即使我们假定 w_i 有零均值，序列独立且与 u_i 不相关，我们却不再能假定合成误差项独立于解释变量 X_i，因为[假定 $E(z_i)=0$]

$$\begin{aligned} \operatorname{cov}(z_i, X_i) &= E[z_i - E(z_i)][X_i - E(X_i)] \\ &= E(u_i - \beta w_i)(w_i) \qquad \text{利用(13.5.7)} \\ &= E(-\beta w_i^2) \\ &= -\beta\sigma_w^2 \end{aligned} \tag{13.5.9}$$

所以，（13.5.8）中的解释变量与误差项是相关的，从而违背了经典线性回归模型中的关键性假定：解释变量与随机干扰项无关。如果这一假定被破坏，则可以证明，OLS 估计量不仅是偏误的而且是非一致性的。就是，即使样本含量 n 无限增大，OLS 估计量仍有偏误。[28]

527 附录 13A 第 13A.3 节表明了对于模型（13.5.8）有

$$\operatorname{plim}\hat{\beta} = \beta\left[\frac{1}{1 + \sigma_w^2/\sigma_{X^*}^2}\right] \tag{13.5.10}$$

其中 σ_w^2 和 $\sigma_{X^*}^2$ 分别是 w_i 和 X^* 的方差，而 $\operatorname{plim}\hat{\beta}$ 指 β 的概率极限。

因为预料括号内的项会小于 1（为什么?），故（13.5.10）表明即使样

本含量无限增大，也不收敛于 β。实际上，若假定 β 为正，将低估 β，也就是它偏向于零。当然，如果 X 中没有测量误差（即 $\sigma_w^2=0$），将给出 β 的一个一致（性）估计量。

因此，当测量误差出现在解释变量中时，将使参数的一致性估计成为不可能，这就给我们提出一个严峻的问题。当然，如我们曾看到的，如果测量误差仅出现于因变量之中，估计量仍是无偏，并从而也是一致性的。如果测量误差出现在解释变量中，那怎么办？回答并不容易。一个极端情形是假定 σ_w^2 比 $\sigma_{X^*}^2$ 相对地小，以致为了一切实际的目的，我们都可以"假定没有"测量误差并照常进行 OLS 估计。当然，这里的困难在于 σ_w^2 和 $\sigma_{X^*}^2$ 是不易观察或测量的，因此也就无法判断其相对大小。

另一补救建议是利用这样的**工具或代理**变量：它们虽与原始 X 变量高度相关，却与方程和测量误差项（即 u_i 和 w_i）都不相关。如果能找到这样的代理变量，我们就能得到 β 的一个一致估计值。但这种工作说比做容易得多，实际上不容易找到一个好的代理变量；我们常常陷入一种埋怨天气不好而又无能为力的境况。另外，要弄清楚所选工具变量是否确实独立于误差项 u_i 和 w_i，也是不容易的。

文献中还有解决问题的其他建议。[29]但大多数都是针对某种给定情况而设计，并且以限制性很强的假定为基础。对于测量误差的问题，确实没有令人满意的答案，这就是为什么把数据观测得尽可能准确如此重要。

一个例子

528 我们构造一个例子以突出说明前面的论点，作为本节的结束。表 13.2 给出真实消费支出 Y^*，真实收入 X^*，观测消费支出 Y 及观测收入 X 的假想数据。此表还说明这些变量是怎样观测的。[30]

表 13.2　真实消费支出 Y^*，真实收入 X^*，观测消费支出及观测收入 X 的假想数据，所有数据均以美元计

Y^*	X^*	Y	X	ε	w	u
75.466 6	80.00	67.601 1	80.094 0	−7.865 5	0.094 0	2.466 6
74.980 1	100.00	75.443 8	91.572 1	0.463 6	−8.427 9	−10.019 9
102.824 2	120.00	109.695 6	112.140 6	6.871 4	2.140 6	5.824 2
125.765 1	140.00	129.415 9	145.596 9	3.650 9	5.596 9	16.765 1
106.503 5	160.00	104.238 8	168.557 9	−2.264 7	8.557 9	−14.496 5
131.431 8	180.00	125.831 9	171.479 3	−5.599 9	−8.520 7	−1.568 2
149.369 3	200.00	153.992 6	203.536 6	4.623 3	3.536 6	4.369 3
143.862 8	220.00	152.920 8	222.853 3	9.057 9	2.853 3	−13.137 2
177.521 8	240.00	176.334 4	232.987 9	−1.187 4	−7.012 0	8.521 8

182.274 8	260.00	174.525 2	261.181 3	−7.749 6	1.181 3	1.274 8

注：假定 X^* 数据是给定的，在推算其他变量时，我们做了如下假定：(1) $E(u_i) \quad E(\varepsilon_i)=E(w_i)=0$；(2) $\text{cov}(X,u)=\text{cov}(X,\varepsilon)=\text{cov}(u,\varepsilon)=\text{cov}(w,u)=\text{cov}(\varepsilon,w)=0$；(3) $\sigma_u^2=100$，$\sigma_w^2=36$，和 $\sigma_w^2=36$；(4) $Y_i^*=25+0.6X_i^*+u_i$，$Y_i=Y_i^*+\varepsilon_i$，和 $X_i=X_i^*+w_i$。

仅因变量 Y 有测量误差

根据所给数据，真实消费函数是：

$$\hat{Y}_i^* = 25.00 + 0.600\,0X_i^* \tag{13.5.11}$$
$$(10.477) \quad (0.058\,4)$$
$$t = (2.386\,1) \quad (10.276) \qquad R^2 = 0.929\,6$$

而与此相比，如果用 Y_i 代替 Y_i^*，则得到：

$$\hat{Y}_i = 25.00 + 0.600\,0X_i^* \tag{13.5.12}$$
$$(12.218) \quad (0.068\,1)$$
$$t = (2.046\,1)(8.811\,8) \qquad R^2 = 0.906\,6$$

如这些结果所表明，并借助于理论，估计的系数不变。因变量中出现测量误差的惟一影响是估计的系数标准误有变大的倾向[见(13.5.5)]，这可从(13.5.12)清楚地看到。顺便指出，(13.5.11)和(13.5.12)之所以有相同的回归系数，是因为样本的生成有意地配合了测量误差模型中的那些假定。

X 中的测量误差

已知真实回归是(13.5.11)。现假令我们不用 X_i^* 而用 X_i。(注：实际上 X_i^* 几乎不可观测。)回归结果如下：

$$\hat{Y}_i^* = 25.992 + 0.594\,2X_i \tag{13.5.13}$$
$$(11.081\,0) \quad (0.061\,7)$$
$$t = (2.345\,7) \quad (9.627\,0) \qquad R^2 = 0.920\,5$$

这些结果均与理论一致——当解释变量有测量误差时，所估系数是有偏误的。庆幸的是，本例中的偏误比较小——由(13.5.10)显见，偏误依赖于 $\sigma_w^2/\sigma_{X^*}^2$，而在数据的产生中我们假定了 $\sigma_w^2=36$ 和 $\sigma_{X^*}^2=3\,667$，从而导致较小的偏误因子，仅约为 0.98%（$=36/3\,667$）。

当 Y 和 X 都有测量误差时，也就是我们所做的不是 Y_i^* 对 X_i^* 的回归，而是 Y_i 对 X_i 的回归，会出现什么后果？这个问题留给读者去探讨（参见习题13.23）。

§ 13.6 对随机误差项不正确的设定

529 研究者所面临的一个常见问题是，对误差项 u_i 进入回归模型的设定。

由于误差项不能直接观测到，所以就不容易确定它进入模型的形式。为看出这一点，让我们回到（13.2.8）和（13.2.9）所给出的模型中。为便于说明，我们在那里已经假定了模型中不存在截距项。在这里，我们进一步假定（13.2.8）中的 u_i，使 $\ln u_i$ 满足通常的 OLS 假定。

如果我们假定（13.2.8）是"正确"模型，但我们估计的是（13.2.9），结果会怎么样呢？附录 13A 第 13A.4 节证明了，若 $\ln u_i \sim N(0, \sigma^2)$，则

$$u_i \sim \log \text{ 正态}\left[e^{\sigma^2/2}, e^{\sigma^2}(e^{\sigma^2}-1)\right] \tag{13.6.1}$$

结果：

$$E(\hat{\alpha}) = \beta e^{\sigma^2/2} \tag{13.6.2}$$

其中 e 为自然对数的底。

如你所见，$\hat{\alpha}$ 是一个有偏的估计量，因为其均值不等于真实的 β。

在有关非线性参数回归模型的章节中，我们将对随机误差项的设定做更多的探讨。

§13.7 嵌套与非嵌套模型

在进行设定检验时，区分**嵌套和非嵌套模型**（nested and non - nested models）很有好处。为说明这两者的差别，考虑以下的模型：

$$\text{模型 A:} \quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$
$$\text{模型 B:} \quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

我们说模型 B 被嵌套在模型 A 之中，是因为它是模型 A 的一个特殊情形：如果我们估计模型 A，然后检验假设 H_0：$\beta_4 = \beta_5 = 0$，并且不拒绝它（比方说基于 F 检验），那么模型 A 就简化为模型 B。若我们在模型 B 中增加变量 X_4，那么模型 A 在 $\beta_5 = 0$ 时就简化为模型 B；这里，我们只用 t 检验来检验 X_5 的系数为零的假设。

我们前面讨论过的设定误差检验和第 8 章中讨论过的约束 F 检验在本质上都属于这种嵌套假设检验，只是我们没有这么称呼而已。

530 现在考虑如下的模型：

$$\text{模型 C:} \quad Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$$
$$\text{模型 D:} \quad Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$$

其中 X 和 Z 各代表不同的变量。我们说模型 C 和 D 是非嵌套的，因为不能把一个作为另一个的特殊情形而推导出来。经济学与其他科学一样，解释同一现象会有多种争持不下的理论。例如货币主义者强调货币在解释 GDP 变化中的作用，而凯恩斯学派则用政府支出的变化去解释它们。

这里要指出，你可以使模型 C 和 D 包含相同的回归元，比如模型 D 中

可以包含 X_3，而模型 C 中可以包含 Z_2。可即便如此，它们仍是非嵌套模型，因为模型 C 没有包含 Z_3，而模型 D 没有包含 X_2。

即便进入模型的变量完全一样，函数形式也可能使两个模型成为非嵌套模型。比如，考虑模型：

$$\text{模型 E:}\quad Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$$

模型 D 和 E 仍是非嵌套模型，因为你不能把其中某个模型作为另一个模型的特殊情形而推导出来。

因为我们前面已经看过了对嵌套模型的检验（t 和 F 检验），所以在接下来的一节中，我们将讨论对非嵌套模型（即我们前面提到的模型误设误差）的某些检验。

§13.8 非嵌套假设的检验

根据哈维（Harvey）[32]，检验非嵌套假设的方法大体上分为两类：（1）**判别方法**（discrimination approach）：给定两个或多个相争持模型，我们根据某些拟合优度准则选择其一。（2）**辨识方法**（discerning approach）（为本书作者用词）：在考察一个模型时须顾及其他模型所提供的信息。下面扼要地解释这些方法。

判别方法

考虑以上模型 C 和 D。由于这两个模型具有相同的因变量，所以我们就可依据诸如我们曾讨论过的 R^2 或校正 R^2 之类的拟合优度准则，在两（或多）个模型之间做出选择。但必须牢记，在比较两（或多）个模型时，回归
531 子必须相同。除这些准则之外，还有其他的准则可以使用，其中包括**赤池信息准则**（AIC）、**施瓦茨信息准则**（SIC）和**马娄斯的 C_p 准则**（Mallow's C_p criterion）等。我们将在第 13.9 节讨论这些准则。多数现代统计软件包在其例行回归程序中都添加了这些准则中的一个或多个。在本章的第 11 节，我们将利用一个引申的例子来说明这些准则。基于这些准则中的一或多个，最终选择的模型具有最高 R^2 的值或最低的 AIC 或 SIC 值等。

辨识方法

非嵌套 F 检验或包容 F 检验。考虑前面介绍的模型 C 和 D，如何在这两个模型之间进行选择呢？为此，假设我们估计如下的嵌套或混合模型：

模型 F：$Y_i=\lambda_1+\lambda_2X_{2i}+\lambda_3X_{3i}+\lambda_4Z_{2i}+\lambda_5Z_{3i}+u_i$

注意模型F嵌套或包含了模型C和D。但C并不嵌套于D中，D也不嵌套于C中。因此它们属于非嵌套模型。

现在如果C是正确的，则$\lambda_4=\lambda_5=0$，而如果D是正确的，则$\lambda_2=\lambda_3=0$。用通常的F检验就可以做这个检验，非嵌套模型由此得名。

然而，这种检验程序却带来一些问题。首先，如果X与Z高度相关，则如在多重共线性一章中所看到的，很可能一或多个λ系数在统计上不显著，尽管我们有可能拒绝所有斜率系数同时为零的（联合）假设。就此情形，我们无法决定模型C抑或模型D是正确的模型。其次，还有另一个问题，假使我们选取模型C作为参考假设或模型，并发现它的所有系数都是显著的。我们把一个或两个Z变量加到模型中，并通过F检验发现它对解释平方和（ESS）的增补贡献是不显著的，因此我们就决定选择模型C。但假如我们反过来选择模型D作为参考假设。并发现它的所有系数也都是显著的，而当我们把一或两个X变量加到此模型并再次用F检验时，我们又会发现它对ESS的增补贡献也是不显著的，于是我们又会把模型D选做正确模型。因此，“参考假设的选择竟能决定模型选择的结果”[33]，尤其是在相互争持的诸回归元中有严重的多重共线性的情况下。最后，人为的嵌套模型F可能缺乏经济意义。

一个说明性例子：圣路易斯模型

532 为了明确名义GNP的变化，是由货币供给的变化来解释（货币主义），还是由政府支出的变化来解释（凯恩斯主义），我们考虑如下模型：

$$\begin{aligned}\dot{Y}_t &= \alpha+\beta_0\dot{M}_t+\beta_1\dot{M}_{t-1}+\beta_2\dot{M}_{t-2}+\beta_3\dot{M}_{t-3}+\beta_4\dot{M}_{t-4}+u_{1t}\\ &= \alpha+\sum_{i=0}^{4}\beta_i\dot{M}_{t-i}+u_{1t} \qquad (13.8.1)\end{aligned}$$

$$\begin{aligned}\dot{Y}_t &= \gamma+\lambda_0\dot{E}_t+\lambda_1\dot{E}_{t-1}+\lambda_2\dot{E}_{t-2}+\lambda_3\dot{E}_{t-3}+\lambda_4\dot{E}_{t-4}+u_{2t}\\ &= \gamma+\sum_{i=0}^{4}\lambda_i\dot{E}_{t-i}+u_{2t} \qquad (13.8.2)\end{aligned}$$

其中$\dot{Y}_t=t$时刻名义GNP的增长率

$\dot{M}_t=t$时刻货币供给量（指M_1）的增长率

$\dot{E}_t=t$时刻充分或高就业下政府支出的增长率

顺便指出（13.8.1）和（13.8.2）都属于**分布滞后模型**（distributed lag models）。这种模型是第17章要详加讨论的主题。眼下只需知道货币供给或政府支出的单位变化对GNP的影响是分布在一段期间而不是瞬时的。

因为不容易先验地在两个模型之间做出选择，故将两者糅合在一起，如下所示：

$$\dot{Y}_t = \text{常数项} + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{3t} \tag{13.8.3}$$

这个嵌套模型就是用以表达并估计著名的圣路易斯联邦储备银行（一个有货币学派倾向的银行）模型的一个形式。此模型给出对于美国在 1953 年第 1 季度至 1976 年第 4 季度期间的估计结果（括号中是 t 比率）[34]：

系数	估计值	系数	估计值	
β_0	0.40(2.96)	λ_0	0.08(2.26)	
β_1	0.41(5.26)	λ_1	0.06(2.52)	
β_2	0.25(2.14)	λ_2	0.00(0.02)	
β_3	0.06(0.71)	λ_3	−0.06(−2.20)	(13.8.4)
β_4	0.05(−0.37)	λ_4	−0.07(−1.83)	
$\sum_{i=0}^{4}\beta_i$	1.06(5.59)	$\sum_{i=0}^{4}\lambda_i$	0.03(0.40)	$R^2 = 0.40 \quad d = 1.78$

这些结果能表明一个模型胜于另一个模型吗？如果我们考虑 $\dot{M}$ 和 $\dot{E}$ 的单位变化对 $\dot{Y}$ 的累积效应，我们分别得到 $\sum_{i=0}^{4}\beta_i = 1.06$ 和 $\sum_{i=0}^{4}\lambda_i = 0.03$，前者是统计上显著的，而后者不是。这种比较会倾向于支持货币主义者的主张，即货币供给的变化决定着（名义）GNP 的变化。如何严格地评价这一主张，且留给读者作为一个习题。

533 **戴维森－麦金农 J 检验。**[35]由于刚才列出的非嵌套 F 检验程序中的种种问题，人们提出了另外的检验。其中之一是戴维森－麦金农 J 检验（Davidson-Mackinnon J test）。为说明此检验，假使我们要比较假设或模型 C 和 D。J 检验的步骤如下：

1. 估计模型 D 并由此得到 Y 的估计值 $\hat{Y}_i^D$。

2. 将步骤 1 中得到的估计值作为另一回归元增补到模型 C 中，并随即估计以下模型：

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 \hat{Y}_i^D + u_i \tag{13.8.5}$$

其中的值得自步骤 1。此模型是亨德里方法论中的**兼容性原则**（encompassing principle）之一例。

3. 用 t 检验对假设 $\alpha_4 = 0$ 进行检验。

4. 如果假设 $\alpha_4 = 0$ 不被拒绝，就可接受（即不拒绝）模型 C 为真模型。因为，(13.8.5) 所含的 $\hat{Y}_i^D$ 代表不为模型 C 所含有的变量影响，而这种影

响并没有增加模型 C 原有的解释能力。换句话说，D 模型不含有足以改进模型 C 的表现的任何额外信息，故模型 C 兼容了模型 D。类似地推理，如果虚拟假设被拒绝，则模型 C 不会是真模型（为什么?）。

5. 现在把假设或模型 C 和 D 的作用颠倒过来，先估计模型 C，并用由此得到的 Y 估计值作为回归元放到模型 D 中，重复步骤 4，以决定是否认为模型 D 胜过模型 C。更具体而言，我们估计如下模型：

$$Y_i=\beta_1+\beta_2 Z_{2i}+\beta_3 Z_{3i}+\beta_4 \hat{Y}_i^C+u_i \tag{13.8.6}$$

其中 $\hat{Y}_i^C$ 是得自模型 C 的 Y 的估计值。现在假设检验 $\beta_4=0$。如该假设不被拒绝，则选择模型 D 而不选 C。如假设 $\beta_4=0$ 被拒绝，则由于 D 没有改进 C 的表现（而 C 改进了 D 的表现——译者注），故选 C 而不选 D。

J 检验虽然直观上比较可取，却也遇到一些问题。由于（13.8.5）和（13.8.6）的两个检验是独立操作的，故有下述可能结局：

	假设：$\alpha_4=0$	
假设：$\beta_4=0$	**不拒绝**	**拒绝**
不拒绝	同时接受 C 和 D	接受 D 而拒绝 C
拒绝	接受 C 而拒绝 D	同时拒绝 C 和 D

534 如上表所示，如果 J 检验程序导致同时接受或同时拒绝两模型，我们就得不到明确的答案。当两模型均被拒绝时，任一模型都无助于对 Y 行为的解释。同理，若两模型均被接受，则有如克曼塔所说的：“显然数据还未充分到足以辨别两个假设［模型］的地步。”[36]

J 检验的另一个问题是，当我们用 t 统计量去检验模型（13.8.5）和（13.8.6）中估计的 Y 变量的显著性时，t 统计量只是渐近地即在大样本中遵从标准正态分布。因此，在小样本中，J 检验会过多地拒绝真假设或真模型，从而它不是（在统计意义上）很有功效的。

一个说明性例子

为说明 J 检验，考虑表 13.3 中的数据，该表给出美国 1970—1991 年期间私人人均消费支出（PPCE）和私人人均可支配收入（PDPI）数据，均以 1987 年美元计算。现考虑以下两个竞争模型：

$$\text{模型 A: } \mathrm{PPCE}_t=\alpha_1+\alpha_2\mathrm{PDPI}_t+\alpha_3\mathrm{PDPI}_{t-1}+u_t \tag{13.8.7}$$

$$\text{模型 B: } \mathrm{PPCE}_t=\beta_1+\beta_2\mathrm{PDPI}_t+\beta_3\mathrm{PPCE}_{t-1}+u_t \tag{13.8.8}$$

表 13.3　1970—1991 年美国私人人均消费支出与私人人均可支配收入，以 1987 年美元价值计算

年份	PPCE	PDPI	年份	PPCE	PDPI
1970	8 842	9 875	1981	10 770	12 156
1971	9 022	10 111	1982	10 782	12 146
1972	9 425	10 414	1983	11 179	12 349

1973	9 752	11 013	1984	11 617	13 029
1974	9 602	10 832	1985	12 015	13 258
1975	9 711	10 906	1986	12 336	13 552
1976	10 121	11 192	1987	12 568	13 545
1977	10 425	11 406	1988	12 903	13 890
1978	10 744	11 851	1989	13 029	14 005
1979	10 876	12 039	1990	13 044	14 068
1980	10 746	12 005	1991	12 824	13 886

资料来源：*Economic Report of the President*，1993，Table B-5，p.355.

535 模型 A 表示 PPCE 依赖于当前的和前期的 PDPI；这是所谓分布滞后模型之一例（参看第 17 章）。模型 B 设想 PPCE 依赖于当前的 PDPI 以及前期的 PPCE；此模型代表以自回归模型为名的一个模型（参看第 17 章）。模型中引进 PPCE 滞后值的理由是为了反映人们的消费惯性或习惯的持久性。

这两个模型各自的估计结果如下：

$$\text{模型 A: } \widehat{\text{PPCE}}_t = -1\,299.053\,6 + 0.920\,4\text{PDPI}_t + 0.093\,1\text{PDPI}_{t-1}$$
$$t = \quad (-4.037\,8)(6.017\,8) \qquad (0.630\,8) \tag{13.8.9}$$
$$R^2 = 0.988\,8 \qquad d = 0.809\,2$$

$$\text{模型 B: } \widehat{\text{PPCE}}_t = -841.856\,8 + 0.711\,7\text{PDPI}_t + 0.295\,4\text{PPCE}_{t-1}$$
$$t = \quad (-2.413\,7)(5.463\,4) \qquad (2.368\,1) \tag{13.8.10}$$
$$R^2 = 0.991\,2 \qquad d = 1.014\,4$$

如果按照判别方法，比如说根据最高 R^2 准则，在两者之间进行选择，我们就会选择（13.8.10）；然而，在（13.8.10）中，似乎两个解释变量个别而论都是统计上不显著的，而在（13.8.9）中仅仅当前的 PDPI 是统计上显著的（但须当心共线性的问题!）。

但从预测的目的考虑，两个估计的 R^2 值差别不大，选（13.8.10）而不选（13.8.9），未必适当。

为了应用 J 检验，且假定模型 A 是虚拟即维持假设，而模型 B 是对立或备择假设。按照上面探讨的 J 检验步骤，用得自模型（13.8.10）的 PPCE 估计值作为模型 A 中的一个新增回归元，并得到以下结果：

$$\widehat{\text{PPCE}}_t = 1\,322.795\,8 - 0.706\,1\text{PDPI}_t - 0.435\,7\text{PDPI}_{t-1} + 2.133\,5\,\widehat{\text{PPCE}}_t^B$$
$$t = \quad (1.589\,6)\,(-1.395\,8) \qquad (-2.192\,6) \qquad (3.314\,1) \tag{13.8.11}$$
$$R^2 = 0.993\,2 \qquad d = 1.711\,5$$

其中（13.8.11）右边的$\widehat{\text{PPCE}}_t^B$是得自模型 B 即（13.8.10）的 PPCE 的估计值。既然此变量的系数在统计上是显著的（在双尾 0.004 的水平上），按照

J 检验程序，我们必须拒绝模型 A 而接受模型 B。

再假定模型 B 是维持假设而模型 A 是备择假设，按照和前面完全一样的程序，我们得到如下结果：

$$
\begin{aligned}
\widehat{\mathrm{PPCE}}_t = & -6\,549.865\,9 + 5.117\,6\mathrm{PDPI}_t + 0.630\,2\mathrm{PPCE}_{t-1} - 4.677\,6\,\widehat{\mathrm{PPCE}}^A_t \\
t = & \quad (-2.497\,6)(2.542\,4) \qquad (3.414\,1) \qquad (-2.192\,6) \qquad (13.8.12) \\
& \qquad\qquad R^2 = 0.992\,0 \qquad d = 1.711\,5
\end{aligned}
$$

其中（13.8.12）右边的$\widehat{\mathrm{PPCE}}^A_t$是得自模型 A 即（13.8.9）的估计值。但在此回归中，右端的系数仍然是统计上显著的（在双尾 0.042 5 水平上）。这一结果表明，我们现在应拒绝模型 B，而接受模型 A！

以上的全部细节告诉我们，为了解释 1970—1991 年期间美国的私人人均消费支出行为，不见得哪个模型是特别有用的。

当然，我们仅仅考虑了两个相互媲美的模型，其实还可以比较多个模型。可以把 J 检验程序推广到多个模型的比较上，尽管分析上会立即变得复杂起来。

本例生动地表明为什么 CLRM 要假定分析中所用的回归模型是被正确地设定的。显然，在做一个模型时，全力注意被模型化的现象是极其重要的。

536 **模型选择的其他检验。**刚才讨论的 J 检验只是模型选择的许多检验中的一种，还有考克斯检验、JA 检验、P 检验、米松－理查德（Mizon-Richard）兼容性检验，以及这些检验的变种。显然，我们不能指望一一讨论这些专门的检验，有兴趣的读者可参阅散见于各个注释中所引的文献。

§ 13.9 模型选择准则

在本节，我们讨论几个已用于在互相竞争的模型之间做出选择和/或从预测的角度对模型进行比较的准则。我们在此区分**样本内**（in-sample）预测和**样本外**（out-of-sample）预测。样本内预测本质上告诉我们的是，所选择的模型在给定样本中对数据拟合得如何。样本外预测则考虑到一个拟合模型在给定回归元值情况下对回归子未来值的预测。

有几个准则用于这一目的。具体而言，我们讨论如下这些准则：（1）R^2，（2）校正 R^2（$=\bar{R}^2$），（3）赤池信息准则，（4）施瓦茨信息准则，（5）马娄斯的 C_p 准则，及（6）预测 χ^2 准则。所有这些准则都是为了最小化残差平方和（RRS）（或提高 R^2 的值）。但除了第一个准则之外，准则（2）、（3）、（4）和（5）都对包含回归元个数不断增加进行了惩罚。因此，在模型的拟合优度与其复杂性（由回归元个数来判断）之间有一种权衡取舍的关系。

R^2 准则

我们知道，对一个回归模型拟合优度的度量指标之一就是 R^2，其定义为：

$$R^2 = \mathrm{ESS/TSS} = 1 - \mathrm{RSS/TSS} \tag{13.9.1}$$

如此定义的 R^2 必然介于 0 和 1 之间。越接近 1，拟合得越好。但 R^2 有一些问题。首先，它度量的是样本内拟合优度，即度量了给定样本中所估计的 Y 值与其实际值有多么接近。它不能保证对样本外观测也能很好地预测。其次，在将两个或多个 R^2 进行比较时，因变量或回归子必须相同。最后，也是最重要的一点，当模型中添加越来越多的变量时，R^2 总不会变小。因此，通过单纯地向模型中添加更多的变量，玩"最大化 R^2"的游戏很诱人。当然，在模型中添加越来越多的变量的确能使 R^2 变大，但同时也使预测误差的方差变大。

校正 R^2 准则

537 作为对增加回归元来提高值的一种惩罚，亨利·泰尔提出我们在第 7 章中所研究的校正 R^2，记为 $\bar{R}^2$。记得：

$$R^2 = 1 - \frac{\mathrm{RSS}/(n-k)}{\mathrm{TSS}/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \tag{13.9.2}$$

你从这个公式可以看出，$\bar{R}^2 \leqslant R^2$，表明校正 R^2 是如何对增加更多的回归元进行惩罚的。我们在第 8 章曾指出，与 R^2 不同，校正 R^2 只有在所添加变量的 t 值的绝对值大于 1 时才会增加。比较而言，$\bar{R}^2$ 是一个比 R^2 更好的度量指标。但同样记住，为了能进行比较，被比较模型的回归子仍必须相同。

赤池信息准则

在 AIC 准则中，进一步对模型中增加回归元进行了惩罚，AIC 的定义为：

$$\mathrm{AIC} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{\mathrm{RSS}}{n} \tag{13.9.3}$$

其中 k 为回归元的个数（包括截距项），n 为观测次数。为了数学计算上的方便起见，把（13.9.3）写成：

$$\ln \text{AIC} = \frac{2k}{n} + \ln\left(\frac{\text{RSS}}{n}\right) \tag{13.9.4}$$

其中 ln AIC 为 AIC 的自然对数，$2k/n$ 为惩罚因子。有些教材和软件只以其对数定义 AIC，所以就没有必要再在 AIC 的前面加上 ln。正如你从这个公式中所见，与之相比，AIC 对添加更多回归元施加了更严厉的惩罚。在比较两个或多个模型时，具有最低 AIC 值的模型优先。AIC 的优越性之一在于，它不仅适用于样本内预测，还适用于预测一个回归模型在样本外的表现。此外，它对嵌套和非嵌套模型都适用，甚至还可以用于决定AR(p)模型的滞后长度。

施瓦茨信息准则

与 AIC 的思想类似，SIC 准则的定义为：

$$\text{SIC} = n^{k/n}\frac{\sum \hat{u}^2}{n} = n^{k/n}\frac{\text{RSS}}{n} \tag{13.9.5}$$

538 或以对数形式表示为：

$$\ln \text{SIC} = \frac{k}{n}\ln n + \ln\left(\frac{\text{RSS}}{n}\right) \tag{13.9.6}$$

其中［(k/n) $\ln n$］为惩罚因子。通过比较（13.9.6）和（13.9.4）明显可以看到，SIC 施加的惩罚比 AIC 更严厉。与 AIC 相似，SIC 的值越低的模型就越好。而且与 AIC 一样，SIC 可以用于比较一个模型在样本内或样本外的预测表现。

马娄斯的 C_p 准则

假设我们有一个包括截距在内 k 个回归元的模型。和平常一样，令 $\hat{\sigma}^2$ 为真实 σ^2 的估计量。但假设我们只选择 p（$p\leqslant k$）个回归元，并从使用这 p 个回归元的回归中得到 RSS。令 RSS_p 表示使用 p 个回归元的残差平方和。现在，马娄斯（C. P. Mallows）便提出了模型选择的如下准则，被称为 C_p 准则：

$$C_p = \frac{\text{RSS}_p}{\hat{\sigma}^2} - (n - 2p) \tag{13.9.7}$$

其中 n 为观测次数。

我们知道$E(\hat{\sigma}^2)$是真实 σ^2 的一个无偏估计量。现在，如果含有个回归元的模型拟合得足够充分的话，那么可以证明[38]：$E(\text{RSS}_p) = (n-p)\sigma^2$。结果近似有：

$$E(C_p) \approx \frac{(n-p)\sigma^2}{\sigma^2} - (n-2p) \approx p \tag{13.9.8}$$

在根据 C_p 准则选择一个模型时，我们想找到一个 C_p 值很低（约为 p）的模型。换句话说，根据节俭性原则，我们将选择一个含有 p 个回归元（$p<k$）并相当好地拟合数据的模型。

实践中，人们通常将计算得来的 C_p 对 p 进行描点。一个“充分的”模型将作为一个与 $C_p=p$ 线接近的点而出现，如图 13.3 所示。此图显示，模型 A 因比模型 B 更接近 $C_p=p$ 线而比模型 B 更受欢迎。

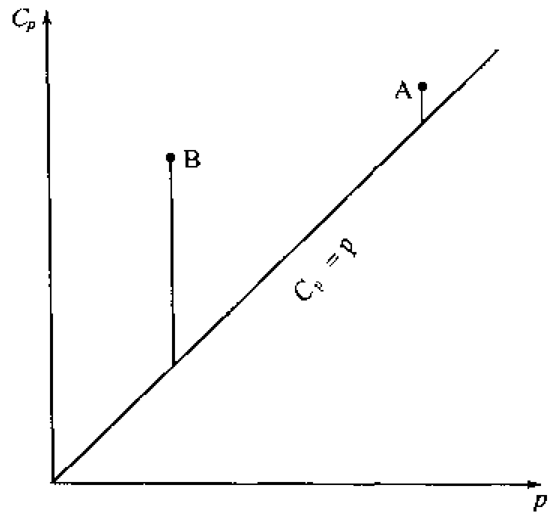


图 13.3　马娄斯的 C_p 描点图

对模型选择准则的一句忠告

我们已经讨论了几种模型选择的准则，但我们应该把这些准则看成是对我们在本章中讨论的各种设定检验的补充。以上讨论的某些准则只是纯粹描
539 述性的，或许没有什么很强的理论性。其中还有些易于受到数据开采的指控。可尽管如此，这些准则仍频繁地被实践者所使用，所以读者应该对此有所察觉。这些准则中没有哪一个肯定优于其他准则。[39] 大多数现代软件包现在都包括 R^2、校正 R^2、AIC 和 SIC。尽管马娄斯的 C_p 很容易根据定义计算出来，可软件仍没有例行地给出它。

用于预测的 χ^2

假使我们有一个基于 n 次观测的回归模型，并假使想利用它来预测回归子在另外 t 次观测中的（均）值。其他地方曾指出，把样本数据留存一部分，以看所估计的模型对未包含进样本的观测（后样本期间的观测）所做的

预测如何，这是一个好主意。

现在，预测 χ^2 检验的定义如下：

$$预测\ \chi^2 = \frac{\sum_{n+1}^{n+t} \hat{u}_i^2}{\hat{\sigma}^2} \tag{13.9.9}$$

其中 $\hat{u}_i^2$ 表示第 i 期（$= n+1,\ n+2,\ \cdots,\ n+t$）利用所拟合回归得到的参数和后样本期间回归元的值得出的预测误差。$\hat{\sigma}^2$ 是 σ^2 基于所拟合回归的通常 OLS 估计量。

540 如果我们假设参数值在样本期和后样本期保持不变，则可以证明 (13.9.9) 中给出的统计量服从自由度为 t 的 χ^2 分布，其中 t 表示留做预测的观测次数。如查伦扎（Charemza）和戴德曼（Deadman）所指出，预测 χ^2 检验具有弱统计功效，意味着它正确地拒绝一个错误的虚拟假设的概率很低，因此这个检验应该用作象征性的、而非决定性的检验。[40]

§ 13.10 计量经济建模的其他专题

正如本章的引言中所讲，计量经济建模和诊断检验的专题如此广泛而又具有发展前景，所以应该用一本专著来写这个专题。我们在上一节已经触及该领域的一些重大主题。在本节，我们考虑一些研究者可能会发现在实践中很有用的其他性质。具体而言，我们考虑如下专题：(1) **异常数据、杠杆数据和有影响力的数据**；(2) **递归最小二乘法**（recursive least squares）；(3) **邹至庄预测失灵检验**（Chow's prediction failure test）。不可避免，对每个专题的讨论都将十分简单。

异常数据、杠杆数据和有影响力的数据[41]

记得在最小化残差平方和时，OLS 对样本中的每个观测都赋予相等的权重。但由于异常数据、杠杆数据和有影响力的数据这三种特殊类型的数据点的出现，使每个观测对回归结果的影响可能不同。了解它们并掌握它们如何影响回归分析，对我们而言就很重要。

在回归的背景下，一个异常数据可定义为一个具有“很大残差”的观测。记得 $\hat{u}_i = (Y_i - \hat{Y}_i)$，即残差表示回归子的实际值与从回归模型估计出来的估计值之差（或正或负）。当我们说一个残差很大时，总是相对其他残差而言，而且这个很大的残差会因为它与估计的回归线之间垂直距离很大而立即吸引我们的注意。注意，在一个数据集中，可能有不止一个异常数据。我们在习题 11.22 中已经遇到过一个这样的例子，在那个例子中，要求你在由 20 个国家构成的样本中，将股票价格的百分比变化（Y）对消费者物价

指数的百分比变化（X）做回归。而对智利的那个观测就是一个异常数据。

541 如果一个数据点不成比例地远离绝大部分回归元值，那就认为它表现出（高度）杠杆性。一个杠杆数据点为什么会带来问题呢？因为它能把回归线向自己拉近，由此改变回归线的斜率，所以它也能带来麻烦。如果这种情况确实发生了，那我们就称这样一个杠杆数据点为一个有影响力的数据点。从样本中去掉这样一个数据点会显著地影响回归线。回到习题 11.22 中，你会发现，若你在包含智利数据的情况下将 Y 对 X 的回归，那么斜率系数就为正并且“在统计上高度显著”。但如果你去掉智利这个国家的观测数据，斜率系数实际上为零。因此，对智利的观测有杠杆作用，也是一个有影响力的观测。

为了进一步弄清楚异常数据、杠杆数据和有影响力数据的性质，考虑图 13.4。[42]

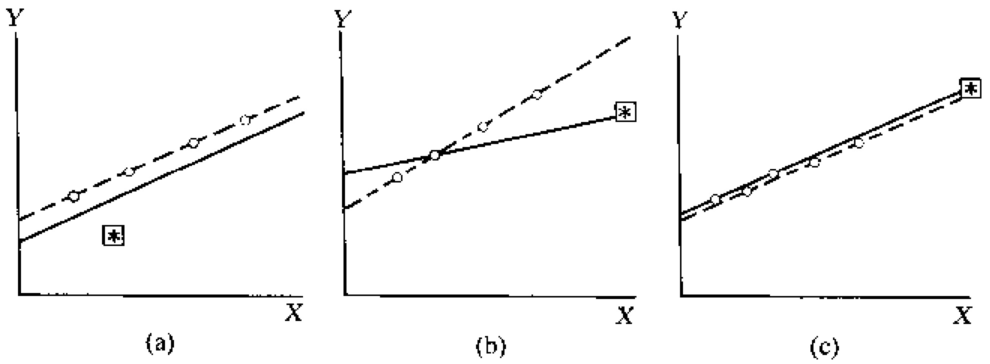


图 13.4

在每个子图中，实线表示对所有数据的 OLS 线，破折线表示去掉异常数据后的 OLS 线。在图（a）中，异常数据接近 X 的均值，具有较低的杠杆作用，对回归系数没有什么影响。在图（b）中，异常数据远离 X 的均值，具有很大的杠杆作用，并对回归系数产生明显影响。在图（c）中，异常数据具有很大的杠杆作用，但对回归系数的影响力很低，因为它基本上与其他观测在同一条直线上。

资料来源：节选自约翰·福克斯前引文献的第 268 页。

我们如何处理这种数据点呢？我们应该只是把它们去掉并只考虑其余的数据点吗？德雷珀（Draper）和史密斯认为：

> 不假思索地拒绝异常数据并不总是一个明智的选择。异常数据有时候提供了其他数据点不能提供的信息，因为它可能是因环境因素非同寻常的组合所致，而其中有些因素可能具有重要的意义，并要求进一步的研究而不是简单地拒绝。作为一个一般原则，只有在可追踪到诸如记录观测发生错误或仪器没有正确调整（在物理试验中）等原因时，才应该拒绝使用异常数据。否则就需要进行认真的研究。[43]

542 有哪些检验可用于侦察异常数据和杠杆数据点呢？文献中讨论了几种检验，但我们这里不讨论它们，因为这样会使我们离题太远。[44]诸如 Shazam 和 Microfit 之类的软件包都有侦察异常数据、杠杆数据和有影响力数据的例

行程序。

递归最小二乘法

我们在第 8 章考察了一个回归模型在涉及时间序列数据时的结构稳定性问题，并说明了在此如何使用邹至庄检验。明确地讲，你可能记得我们在那一章中讨论了美国在 1970—1995 年间的一个简单储蓄函数（储蓄作为收入的函数）。我们在那里还看到，储蓄—收入关系可能在 1982 年左右发生了变化。知道了这个结构转折点之后，我们就能用邹至庄检验来确定。

但如果我们不知道结构转折点又会怎么样呢？此时可以使用**递归最小二乘法**（RELS）。RELS 背后的基本思想很简单，并可以用储蓄—收入回归来解释。

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

其中 Y = 储蓄，X = 收入，样本期间为 1970—1995 年。（参见表 8.9 中的数据。）

假设我们首先使用了 1970—1974 年间的数据并估计了这个储蓄函数，得到 β_1 和 β_2 的估计值。然后，我们使用 1970—1975 年间的数据再次估计这个储蓄函数并得到这两个参数的估计值。然后我们再使用 1970—1976 年间的数据重新估计储蓄模型。以此类推继续增加 Y 和 X 的数据点直至用完全部样本。你可以想像到，每组回归都将给出 β_1 和 β_2 的一组新估计值。如果你把这些参数的估计值依次描点，将会看出估计参数是如何变化的。如果所考虑的模型在结构上是稳定的，这两个参数的估计值变化将很小，而且基本上是随机的。然而，如果参数的估计值变化明显，则意味着存在着结构转折。因此，RELS 成为时间序列数据中一个常用的例行程序，因为时间序列
543 是按照时间顺序排列的。当横截面数据按照某种"规格"或"规模"变量（如企业就业规模或资产规模等）排序时，递归最小二乘法也是一个有用的诊断工具。在习题 13.30 中，要求你在表 8.9 所给的储蓄数据中应用 RELS。

诸如 Shazam、Eviews 和 Microfit 之类的软件包现在都例行做递归最小二乘估计。RELS 还生成**递归残差**（recursive residuals），可作为几个诊断检验的基础。

邹至庄预测失灵检验

我们在第 8 章已经讨论过结构稳定性的邹至庄检验。邹至庄还证明了，他的检验方法略加修改后还可用于对回归模型预测功效的检验。我们再次回到美国 1970—1995 年间的收入—储蓄回归。

假设我们对 1970—1981 年期间估计了储蓄—收入回归，并得到基于

1970—1981 年期间的数据估计的截距和斜率系数 $\hat{\beta}_{1,70-81}$ 和 $\hat{\beta}_{2,70-81}$。现在我们利用 1982—1995 年期间收入的实际值和 1970—1981 年期间的截距和斜率值，来预测 1982—1995 年间每一年的储蓄值。这里的逻辑是，若参数值没有发生严重的结构变化，则基于前一期间系数估计值而估计出来的 1982—1995 年间的储蓄估计值，不应该与后一期间储蓄的实际值有很大不同。当然，若后一期间储蓄的实际值与预测值之间存在巨大差异，则整个数据期间储蓄—收入关系的稳定性就值得怀疑。

储蓄的实际值与估计值之间的差别是大是小可通过 F 检验检验如下：

$$F = \frac{\left(\sum \hat{u}_t^{*2} - \sum \hat{u}_t^2\right)/n_2}{\sum \hat{u}_t^2/(n_1 - k)} \tag{13.10.1}$$

其中 n_1 = 初始回归所基于的第一期间（1970—1981）中观测的次数，n_2 = 第二期间或预测期间的观测次数，$\sum \hat{u}_t^{*2}$ = 对所有观测（$n_1 + n_2$）估计出来的 RSS，$\sum \hat{u}_t^2$ = 对前 n_1 个观测估计出来的 RSS，k 为待估计参数的个数（此例中为 2）。若误差是独立同正态分布的，则（13.10.1）中给出的 F 统计量服从自由度分别为 n_2 和 n_1 的 F 分布。习题 13.31 要求你应用邹至庄预测失灵检验，以看储蓄—收入关系事实上是否发生了变化。顺便提一句，注意这个检验与前面讨论的预测 χ^2 检验的相似之处。

§13.11 一个总结性的例子：一个小时工资的决定模型

544 为了判定哪些因素决定了小时工资，考虑了如下模型

$$\begin{aligned}\text{Hwage}_i = {} & \beta_1 + \beta_2\text{Edu}_i + \beta_3\text{Gender}_i + \beta_4\text{Hispanic}_i + \beta_5\text{Lfexp}_i \\ & + \beta_6\text{Mastatus}_i + \beta_7\text{Race}_i + \beta_8\text{Region}_i + \beta_9\text{Union}_i + u_i\end{aligned} \tag{13.11.1}$$

其中

Hwage = 小时工资美元
Edu = 受教育年数
Gender = 女性取值 1，否则取值 0
Hispanic = 西班牙人取值 1，否则取值 0
Race = 既非白人又非西班牙人取值 1，否则取值 0
Lfexp = 潜在劳动力市场工作年限
Mastatus = 婚姻状况，已婚取值 1，否则取值 0
Region = 定居区域，南部取值 1，否则取值 0
Union = 加入工会情况，加入工会取值 1，否则取值 0

工资函数（13.11.1）的起源可追溯到雅克布·明瑟（Jacob Mincer）。[46] 如你所见，工资函数既包括了定量变量，又包括了定性或虚拟变量。根据经验，所有这些变量看起来都合乎逻辑。注意，种族变量分为三类：西班牙人、非西班牙白人和非西班牙非白人（主要是黑人或从非洲移民到美国的人）；因此，这里有两个种族虚拟变量。我们没有考虑的一组或参照组便是非西班牙白人。

1985 年采访 528 人得到的数据是美国人口普查局定期进行的当代人口调查（CPS）中的一部分。这些数据最早由伯恩特（Berndt）收集，并经阿瑟·戈德伯格改编。我们已在第 2 章讨论过这个来源，记住这些数据是横截面数据。

据经验，预计小时工资与受教育水平、工作经历、婚姻状况和工会状况正相关，而和西班牙籍、种族、性别和居住区域负相关；再次提请注意，所有的比较都是相对非西班牙白人而进行的。随便查阅一本劳动经济学方面的书，更多地了解小时工资的各种决定因素。[47]

我要我的学生利用这些数据估计（13.11.1）。回归结果在表 13.4 中给出。如你所见，（13.11.1）中所有变量都具有预期的符号，尽管不是所有变量都是个别统计显著的。约等于 0.282 6 的 R^2 值看来有些低，但这么低的
545 R^2 在大量观测的横截面数据中司空见惯。然而，由于计算出来约等于 25.56 的 F 值高度显著（因为其 p 值几乎为零），所以这个 R^2 在统计上是显著的。记住，F 统计量检验的是所有斜率系数同时为零，即所有解释变量的值联合起来对回归子也没有影响的假设。

注意到变量 Hispanic、Mastatus 和 Race 个别地看都是统计不显著的，但区域变量"勉强地"统计显著，于是我的一些学生又去掉前三个变量并得到
546 表 13.5 所示的回归结果。现在，在 5% 或更好的显著性水平上（即 p 值低于 5%），所有变量都是个别统计显著的。对各个系数的解释简单明了。比如，区域虚拟变量系数 −0.840 8 表明，保持所有其他变量不变，可能由于南方生活成本低或南方工会组织薄弱，使南方工人比其他地方工人平均每小时少挣 84 美分。类似地，保持所有其他因素不变，女工人平均比男工人每小时少挣 2.13 美元。仅从统计分析仍不能说明这个差距是否源于性别歧视。

表 13.4　　基于（13.11.1）的回归结果

因变量：HWAGE　　样本：1 528

变量	系数	标准误	t 统计量	概率
C	−4.182 714	1.275 908	−3.278 227	0.001 1
EDUCATION	0.937 130	0.082 625	11.341 94	0.000 0
GENDER	−2.140 661	0.391 546	−5.467 200	0.000 0
HISPANIC	−0.512 385	0.911 056	−0.562 408	0.574 1

LFEXP	0.098 486	0.017 494	5.629 597	0.000 0
MSTATUS	0.485 134	0.418 881	1.158 167	0.247 3
RACE	0.942 389	0.583 578	-1.614 849	0.107 0
REGION	-0.771 424	0.430 173	-1.793 287	0.073 5
UNION	1.468 088	0.512 735	2.863 248	0.004 4

R^2	0.282 693	因变量均值	9.047 538
校正 R^2	0.271 636	因变量标准差	5.144 082
回归标准误	4.390 177	赤池信息准则	5.813 515
残差平方和	10 003.03	施瓦茨信息准则	5.886 283
对数似然比	-1 525.768	F 统计量	25.567 45
德宾-沃森统计量	1.857 457	P（F 统计量）	0.000 000

表 13.5

因变量：HWAGE　　样本：1 528

变量	系数	标准误	t 统计量	概率
C	-4.289 796	1.258 229	-3.409 392	0.000 7
BDUCATION	0.953 006	0.082 184	11.595 96	0.000 0
GENDER	-2.134 171	0.391 740	-5.447 929	0.000 0
LFEXP	0.104 037	0.016 888	6.160 545	0.000 0
REGION	-0.840 832	0.427 621	-1.966 303	0.049 8
UNION	1.427 421	0.509 978	2.798 988	0.005 3

R^2	0.276 707	因变量均值	9.047 538
校正 R^2	0.269 779	因变量标准差	5.144 082
回归标准误	4.395 772	赤池信息准则	5.810 462
残差平方和	10 086.51	施瓦茨信息准则	5.858 974
对数似然比	-1 527.962	F 统计量	39.939 78
德宾-沃森统计量	1.858 629	P（F 统计量）	0.000 000

恰如所料，“短”回归（即去掉变量 Hispanic、Mastatus 和 Race 后的回归）比“长”回归（即包括所有变量的回归）经调整的 R^2 更低。但请注意赤池和施瓦茨统计量：它们在短回归中都比在长回归中要小，这就表明它们对模型中引入过多的回归元进行了惩罚。由于两个统计量如此接近，所以我们只须选择其中的一个，两个模型中的德宾-沃森 d 值都充分接近于 2，表明不存在某种“自相关”或设定误差。

由于回归（13.11.1）背后的数据在数据磁盘中给出了，所以你可能想对这些数据做个“实验”。性别和受教育虚拟变量之间或性别与婚姻状况虚

拟变量之间很可能存在某种交互作用。也可能小时工资与劳动市场上的工作经历之间的关系是非线性的，从而使得在回归模型中引入其平方项成为必要。这虽然听起来有些像数据开采，但我们也已经指出，数据开采在计量经济建模中也能起到一定的作用。当然，你在进行数据开采时，必须留意真实的显著性水平。

§13.12 向实务工作者进一言

我们在本章已经取得了很大进展。毫无疑问，建模既是科学又是艺术。一位实践中的研究者可能被这些理论上的精妙之处和一系列的诊断检验搞得晕头转向。但最好记住马丁·费尔德斯坦（Martin Feldstein）的忠告："应用计量经济学家与理论家一样，很快就会从实践中发现，一个有用的模型不是真实的或现实的模型，而是节俭的、比较合理的和富于信息的模型。"[48]

加拿大西蒙·弗雷泽（Simon Fraser）大学的彼得·肯尼迪给出如下"应用计量经济学的十大告诫"[49]：

1. 你应该使用常识和经济理论。

2. 你应该询问正确的问题（即实用性胜于数学上的优美）。

547 3. 你应该了解背景（不要做无知的统计分析）。

4. 你应该对数据进行审查。

5. 你不应该信奉复杂性。而应使用 **KISS 原则**，即保持尽可能简单（keep it stochastically simple）。

6. 你应该充分而又严格地看待结果。

7. 你应该当心数据开采的成本。

8. 你应该准备着妥协（不要信奉教科书中的方法）。

9. 你不应该把显著性和重要性相混淆（不要混淆统计显著性和实际显著性）。

10. 在出现敏感性时你应该坦白（即准备接受批评）。

你可能想详细阅读肯尼迪的文章，以体会他所热中的上述十大告诫的说服力。虽然有些建议听起来不可当真，但每一条可能都包含着苦涩的真理。

§13.13 要点与结论

1. CLRM 假定用于分析的计量经济模型是正确地设定的，这个假定有两层意义。一是，没有方程设定误差，二是，没有模型设定误差，本章主要考虑方程设定误差。

2. 本章讨论的方程设定误差包括（1）重要变量的遗漏，（2）多余变量

的引入，（3）错误函数形式的采用，（4）误差项 u_i 的非正确设定，以及（5）回归子和回归元中的测量误差。

3. 当模型漏掉真实的变量时，后果可能是很严重的，模型中所保留的变量的系数的 OLS 估计量不仅有偏误，而且不是一致性的。此外，这些系数的方差和标准误的估计都是不正确的，从而使通常的假设检验程序失效。

4. 模型含有无关变量的后果幸而不那么严重：有关和“无关”变量的系数的估计量仍然是无偏的，并且是一致性的。误差方差 σ^2 仍然被正确地估计。惟一的问题是所估计的方差趋之于过大，从而使参数的估计较不准确，即置信区间无必要地扩大。

5. 为了侦察方程设定误差，我们考虑了几种检验，诸如（1）残差分析，（2）德宾-沃森 d 统计量，（3）拉姆齐的 RESET 检验，以及（4）拉格朗日乘数检验。

548 6. 一类特殊的设定误差是回归子和回归元取值的测量误差。如果仅回归子中有测量误差，则 OLS 估计量是无偏的，且有一致性，但效率较低。如果回归元中有测量误差，则 OLS 估计量是偏误的，而且非一致性的。

7. 即使察觉到或猜测到有测量误差，如何补救常常是不容易的。工具或代理变量的使用虽然理论上诱人，但不很实际。因此，在实践中研究者应详细说明数据来源、收集的方法，所用的定义，等等。官方收集的数据常常附有多个注释，研究者应把这些告知读者。

8. 模型误设误差可能与方程设定误差一样严重。特别地，我们区分了嵌套和非嵌套模型。为了判别适当的模型，我们讨论了非嵌套或包容性的 F 检验和戴维森-麦金农 J 检验，并指出每个检验的局限性。

9. 在实践中选择一个经验模型时，研究者使用过一系列准则。我们讨论了其中的一些准则，比如赤池和施瓦茨信息准则、马娄斯的 C_p 准则和预测 χ^2 准则。我们讨论了这些准则的优缺点，并警告读者这些准则不是绝对的，对仔细地设定分析只能起到辅助作用。

10. 我们还讨论了一些附加的专题：（1）异常数据、杠杆数据和有影响力的数据；（2）递归最小二乘法；（3）邹至庄预测失灵检验。我们讨论了他们在应用研究中的作用。

11. 我们以彼得·肯尼迪“对应用计量经济学家的十大告诫”结束本章。这些告诫的关键所在，是要求研究者能考虑到计量经济学纯粹技术方面之外的东西。

习 题

问答题

13.1 参照方程（8.7.23）中所估计的子鸡需求函数，考虑到 13.1 节所讨论的良好模型的属性，你会说这个需求函数是“正确地”设

定的吗？

13.2 假定真模型是：

$$Y_i = \beta_1 X_i + u_i \tag{1}$$

但你没有去拟合这个过原点回归，却一贯地拟合了通常带有截距的模型：

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i \tag{2}$$

评述这一设定误差的后果。

549 **13.3** 继续做习题13.2，但假定真模型是（2），讨论拟合误设模型（1）的后果。

13.4 假令“真”模型是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \tag{1}$$

而我们加了一个“无关”变量 X_3 到模型中去（“无关”是指变量 X_3 的真系数 β_3 为零），并估计了：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i \tag{2}$$

a. 模型（2）的 R^2 和经校正的 R^2 会不会比模型（1）的大？

b. 从（2）得到的 β_1 和 β_2 的估计值是无偏的吗？

c. “无关”变量 X_3 的引入对 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 的方差有影响吗？

13.5 考虑如下“真实”（柯布-道格拉斯）生产函数：

$$\ln Y_i = \alpha_0 + \alpha_1 \ln L_{1i} + \alpha_2 \ln L_{2i} + \alpha_3 \ln K_i + u_i$$

其中 Y = 产出

L_1 = 生产性劳动力

L_2 = 非生产性劳动力

K = 资本

但若在经验调查中实际用的回归是：

$$\ln Y_i = \beta_0 + \beta_1 \ln L_{1i} + \beta_2 \ln K_i + u_i$$

假定你拥有有关变量的横截面数据，

a. 将会有 $E(\hat{\beta}) = \alpha_1$，和 $E(\hat{\beta}_2) = \alpha_3$ 吗？

b. 如果知道了 L_2 是生产函数中的一个无关变量，(a)中的答案能否成立？给出必要的推导。

13.6 参照方程(13.3.4)和(13.3.5)。你会看到，虽然有偏误，却比无偏的有更小的方差。你会怎样在偏误与较小方差之间做出权衡？提示：两种估计量的 MSE(均方误)可表达为：

$$\begin{aligned}\mathrm{MSE}(\hat{\alpha}_2) &= (\sigma^2/\textstyle\sum x_{2i}^2) + \beta_3^2 b_{32}^2 \\ &= \text{抽样方差} + \text{平方偏误} \\ \mathrm{MSE}(\hat{\beta}_2) &= \sigma^2/\textstyle\sum x_2^2(1 - r_{23}^2)\end{aligned}$$

关于 MSE,参看附录 A。

13.7 证明从(13.5.1)或(13.5.3)估计的 β 都是真 β 的一个无偏估计。

550 **13.8** 按照弗里德曼的永久收入假说,我们可写:

$$Y_i^* = \alpha + \beta X_i^* \qquad (1)$$

其中 Y_i^* = "永久"消费支出, X_i^* = "永久"收入。但我们观测到的不是"永久"变量,而是:

$$Y_i = Y_i^* + u_i$$
$$X_i = X_i^* + v_i$$

其中 Y_i 和 X_i 是可以观测或测量到的数量,而 u_i 和 v_i 分别是 Y_i^* 和 X_i^* 的测量误差。

利用可观测的数量,就可把消费函数写为:

$$\begin{aligned} Y_i &= \alpha + \beta(X_i - v_i) + u_i \\ &= \alpha + \beta X_i + (u_i - \beta v_i) \qquad (2) \end{aligned}$$

假定:(1) $E(u_i) = E(v_i) = 0$,(2) $\mathrm{var}(u_i) = \sigma_u^2$ 和 $\mathrm{var}(v_i) = \sigma_v^2$,(3) $\mathrm{cov}(Y_i^*, u_i) = 0$,$\mathrm{cov}(X_i^*, v_i) = 0$,和 (4) $\mathrm{cov}(u_i, X_i^*) = \mathrm{cov}(v_i, Y_i^*) = \mathrm{cov}(u_i, v_i) = 0$,证明在大样本中从 (2) 估计的 β 可表示为:

$$\mathrm{plim}\,(\hat{\beta}) = \frac{\beta}{1 + (\sigma_v^2/\sigma_{X^*}^2)}$$

a. 关于 $\hat{\beta}$ 的偏误性质,你能谈些什么?

b. 如果样本无限地加大,估计的 β 值会不会趋向于与真 β 相等?

13.9 **资本资产定价模型**。近代投资理论中的资本资产定价模型设定,一定时期内的证券(普通股)的平均回报率与证券的波动性,即所谓 β 系数(波动性是对风险的度量)有如下关系:

$$\bar{R}_i = \alpha_1 + \alpha_2 (\beta_i) + u_i \qquad (1)$$

其中 $\bar{R}_i$ = 证券 i 的平均回报率

β_i = 证券 i 的真 β 系数

u_i = 随机干扰项

真正的 β_i 不可直接观测而是按下式估算的:

$$r_{it} = \alpha_1 + \beta^* r_{mt} + e_t \qquad (2)$$

其中 r_{it} = 时间 t 时证券 i 的回报率

r_{mt} = 时间 t 时的市场回报率(指某个广为概括的市场指数,如工业证券的 S&P 指数的回报率)

e_t = 残差项

并且其中 β^* 是"真" β 系数的一个估计值。因此,我们实际上

估计的不是（1）而是：

$$R_i = \alpha_1 + \alpha_2 (\beta_i^*) + u_i \tag{3}$$

551 其中 β_i^* 是从回归（2）得到的。但因 β_i^* 是估计值，真 β 与 β^* 之间的关系可写为：

$$\beta_i^* = \beta_i + v_i \tag{4}$$

其中 v_i 可称为测量误差。

a. 这一测量误差对 α_2 的估计会有什么影响？

b. 从（3）估计的 α_2 会是真 α_2 的一个无偏估计吗？如果不是，它是 α_2 的一致性估计吗？如果不是，你有什么补救措施的建议？

13.10 考虑模型：

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \tag{1}$$

为了找出此模型是否因为漏掉变量 X_3 而成为一个误设的模型，你决定用模型（1）给出的残差仅仅对 X_3 一个变量做回归（注：在此回归中有一截距项）。然而，拉格朗日乘数检验要求你用（1）的残差兼对 X_2 和 X_3 及一常数项做回归。为什么你用的程序很可能是不适当的？[1]

13.11 考虑模型：

$$Y_i = \beta_1 + \beta_2 X_i^* + u_i$$

而实际上我们用以度量 X_i^* 的是这样的 X_i：

a. $X_i = X_i^* + 5$

b. $X_i = 3X_i^*$

c. $X_i = (X_i^* + \varepsilon_i)$，

其中 ε_i 是具有通常性质的一个纯随机项。这些测量误差对真 β_1 和 β_2 的估计将有什么影响？

13.12 参照回归方程（13.3.1）和（13.3.2），用类似于（13.3.3）的方法说明：

$$E(\hat{\alpha}_1) = \beta_1 + \beta_3 (X_3 - b_{32}\bar{X}_2)$$

其中 b_{32} 是遗漏变量 X_3 对所含变量 X_2 回归中的斜率系数。

13.13 深刻地评述利莫尔（Leamer）的下述观点：

我对超越统计学［指推断理论实际上来自数据］的兴趣起源于我对工作中的经济学家的观察。在经济专业中认为计量经
552 济理论无关重要者占有一个令人迷惑不解的大部分。计量经济理论与计量经济实践之间的鸿沟，足以引起职业者的不安。事实上，一种平静的均衡渗透着我们的期刊和「专业」会议，我们心安理得地把我们自己划分为一方面是孤独虔诚的牧师团——统计理论家，另一方面是一群顽固的罪过者——数据分析

者。牧师们被授权列出罪过的清单，并为他们所表现的特殊才能而受到敬畏。罪过者不被指望能避免罪过；但要求他们公开坦白他们的罪过。[2]

13.14 评价亨利·泰尔的如下陈述：

就这门艺术的现状而论，最切合实际的做法，是对置信系数和显著性作自由灵活的解释，如果置信区间和检验统计量是从“按习惯的回归策略得到的最后回归”计算出来的话。意思是说，一个 95% 的置信系数，也许实际上是一个 80% 的置信系数，而一个 1% 的显著性水平，也许实际上是一个 10% 的水平。[3]

13.15 关于 20 世纪50 年代和 60 年代初所采取的计量经济学方法论，布劳格（Blaug）做过如下的评述[4]：

大多数［经验研究］像是放下球网打网球：现代经济学家们过多地满足于显示真实世界符合于他们的预测，而无意去拒绝本可检验的预测，从而用容易的验证去替代困难的反驳［按照波普（Popper）的说法］。

你同意这种观点吗？不妨浏览布劳格的书，以更好地了解他的观点。

13.16 按照布劳格的意见，“没有证明的逻辑，但有证伪的逻辑。”[5]他这句话的意思是什么？

13.17 回到正文中讨论的圣路易斯模型，参照嵌套 F 检验所涉及的问题，严格评论回归（13.8.4）所展现的结果。

13.18 假设真模型是：

$$Y_i = \beta_1 + \beta_2 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

但你估计了：

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$

如果你利用 Y 在 $X = -3$、-2、-1、0、1、2、3 处的观测并估计了“不正确”的模型，这些估计值将出现什么偏误？[6]

13.19 为了看出 X_i^2 是否应属于模型 $Y_i = \beta_1 + \beta_2 X_i + u_i$，拉姆齐的 RESET 检验将估计这个线性模型，并从模型中得到 Y_i 的估计
553 值［即 $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$］，然后估计模型 $Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 \hat{Y}_i^2 + v_i$，并检验 α_3 的显著性。试证明：若最终表明在上述（RESET）方程中是统计显著的，就等同于直接估计如下模型：$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$。（提示：在 RESET 回归中代入 $\hat{Y}_i$。[7]）

13.20 说明如下命题正确与否的原因。[8]

a. 一个观测可能具有影响力但不是异常数据。

b. 一个观测可能是异常数据但不具有影响力。

c. 一个观测可能同时既是异常数据又具有影响力。

d. 如果在模型中 $Y_i=\beta_1+\beta_2X_i+\beta_3X_i^2+u_i$，发现 β_3 的估计值是统计显著的，那么，即便 β_2 的估计值在统计上不显著，我们也应该在模型中保留线性项 X_i。

e. 若你用 OLS 估计了模型 $Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i$ 或 $Y_i=\alpha_1+\beta_2x_{2i}+\beta_3x_{3i}+u_i$，则估计的回归线应该相同，其中 $x_{2i}=(X_{2i}-\bar{X}_2)$ 和 $x_{3i}=(X_{3i}-\bar{X}_3)$。

解答题

13.21 利用习题7.19 所给的对子鸡的需求数据。假使你被告知真实的需求函数是：

$$\ln Y_t=\beta_1+\beta_2\ln X_{2t}+\beta_3\ln X_{3t}+\beta_6\ln X_{6t}+u_t \tag{1}$$

而你有不同的看法并估计了以下的需求函数：

$$\ln Y_t=\alpha_1+\alpha_2\ln X_{2t}+\alpha_3\ln X_{3t}+v_t \tag{2}$$

其中 Y = 人均子鸡消费量（磅）

X_2 = 人均实际可支配收入

X_3 = 子鸡实际零售价格

X_6 = 子鸡替代品的复合实际价格

a. 假定前面给的需求函数（1）是真实的，做出设定误差的 RESET 和 LM 检验。

b. 假使我们发现（1）中 $\hat{\beta}_6$ 的在统计上不显著。这是否意味着用（2）去拟合数据就没有设定误差？

c. 如果我们发现 $\hat{\beta}_6$ 不显著，这是否意味着我们不应把替代品的价格作为一个变量引入到需求函数中来？

13.22 继续习题13.21。纯粹为了教学的目的，假定模型（2）是真实需求函数。

a. 如果现在我们估计了模型（1），这时我们犯了什么类型的设定错误？

b. 这种设定误差的理论性后果为何？用你掌握的数据做出说明。

13.23 假设真模型为：

$$Y_i^*=\beta_1+\beta_2X_i^*+u_i \tag{1}$$

554

而由于度量误差你估计了：

$$Y_i=\alpha_1+\alpha_2X_i+v_i \tag{2}$$

其中 $Y_i=Y_i^*+\varepsilon_i$，$X_i=X_i^*+w_i$，而 ε_i 和 w_i 则表示测量误差。

利用表 13.2 中给出的数据，罗列出我们估计（2）而没有估计真实模型（1）所导致的后果。

13.24 习题6.14曾要求你估计CES（常数替代弹性）生产函数中的劳动力与资本替代弹性，但该函数的真实性是以劳动力市场中的完全竞争为依据的。如若竞争不完全，则模型的正确形式是：

$$\log(V/L)=\log\beta_1+\beta_2\log W+\beta_3\log(1+1/E)$$

其中（V/L）= 单位劳动力的价值增值，L = 劳动力投入，W = 实际工资率，以及 E = 劳动力供给弹性。

a. 如果劳动市场实际上是不完全的，那么原先的CES替代弹性估计涉及哪一类设定误差？

b. 这种误差对替代弹性参数 β_2 产生什么理论上的后果？

c. 假定在习题6.14所列的工业中的劳动力供给弹性依次是2.0、1.8、2.5、2.3、1.9、2.1、1.7、2.7、2.2、2.1、2.9、2.8、3.2、2.9和3.1。把这些数据以及习题6.14所给的那些数据一同用来估计上述模型，并用设定误差理论评述你的结果。

13.25 蒙特卡罗实验[9]：10个人的每周永久收入为：200、220、240、260、280、300、320、340、380和400美元。永久消费（Y_i^*）和永久收入（X_i^*）的关系为：

$$Y_i^*=0.8X_i^* \qquad (1)$$

每个人的临时收入等于均值为0方差为1的正态总体（即标准正态变量）的随机抽样值 u_i 的100倍。假定消费中没有临时成分。就是说，观测消费和永久消费相同。

555 a. 从零均值和单位方差的正态总体中抽取10个随机数并得到10个观测收入 X_i（$=X_i^*+100u_i$）。

b. 利用（a）中得到的数据做永久（=观测）消费对观测收入的回归，并将你的结果同（1）所展示的结果相比较。先验地，截距应为零（为什么?）。你的结果是不是这样？为什么或为什么不？

c. 重复（a）100次，从而得到100个（b）那样的回归。然后用你的结果同真实回归（1）相比。你能得出什么一般性结论。

13.26 参照习题8.26按照那里给的变量定义，考虑以下两个解释 Y 的模型：

$$\text{模型 A:}\quad Y_t=\alpha_1+\alpha_2X_{3t}+\alpha_3X_{4t}+\alpha_4X_{6t}+u_t$$
$$\text{模型 B:}\quad Y_t=\beta_1+\beta_2X_{2t}+\beta_3X_{5t}+\beta_4X_{6t}+u_t$$

用嵌套 F 检验，你将怎样在这两个模型之间进行选择？

13.27 继续习题13.26，用 J 检验，你会怎样在这两个模型之间做出决定？

13.28 回到习题7.19，那是关于美国对子鸡的需求问题。在那里给出

了 5 个模型。

a. 模型 1 和模型 2 有什么差别？如果模型 2 是正确的，而你估计了模型 1，那么你所犯的是什么类型的错误？你会用哪一种检验——方程设定误差（检验）或模型误设误差（检验）？说明必要的计算。

b. 你会怎样在模型 1 和模型 5 之间做出选择？你会用哪些检验，为什么？

13.29 参照表8.9，它给出 1970—1995 年间个人储蓄（Y）和个人收入（X）的数据。现考虑下述模型：

$$\text{模型 A：}\quad Y_t = \alpha_1 + \alpha_2 X_t + \alpha_3 X_{t-1} + u_t$$
$$\text{模型 B：}\quad Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

你怎样在这两个模型之间做出选择？明确地陈述你用的检验程序及全部计算。假使有人争辩说利率变量属于此储蓄函数，你会怎样对此做检验？收集 3 个月的国库券利率作为利率的代理变量，并说明你的答案。

13.30 利用习题 13.29 中的数据。为了熟悉递归最小二乘法，对 1970—1981、1970—1985、1970—1990 和 1970—1995 年期间估计储蓄函数。评论储蓄函数中估计系数的稳定性。

13.31 继续习题13.30。假设你针对 1970—1981 年估计了储蓄函数。利用如此估计得到的参数和 1982—1995 年间个人可支配收入的数据，对后一期间估计预测储蓄，并利用邹至庄预测失灵检验，说明它是否拒绝储蓄函数在这两个期间没有发生变化的假设。

13.32 在 k 变量回归模型中遗漏一个变量。参照方程（13.3.3），它给出了从模型 $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ 中漏掉变量 X_3 的偏误。对此可做如下推广：在一个 k 变量模型 $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$ 中，假设我们遗漏了变量 X_k，则可以证明，所包含变量 X_j 的斜率系数的偏误为：

556

$$E(\hat{\beta}_j) = \beta_j + \beta_k b_{kj} \qquad j = 2, 3, \cdots, (k-1)$$

其中 b_{kj} 为被排除变量对模型中包含的所有其他解释变量的辅助回归中 X_j 的（偏）斜率系数。[10]

回到习题 13.21。当我们从模型中排除掉变量 $\ln X_6$ 时，求出方程（1）中系数偏误的大小。这种排除严重吗？给出必要的计算。

【习题注释】

[1] 参看 Maddala，前引文献，第 477 页。

[2] Edward E. Leamer, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, John Wiley & Sons, New York, 1978, p.vi.

[3] Henry Theil, *Principles of Econometrics*, John Wiley & Sons, New York,

1971, pp.605－606.

[4] M. Blaug, *The Methodology of Economics. Or How Economists Explain*, Cambridge University Press, New York, 1980, p.256.

[5] 前引文献，第 14 页。

[6] 节选自 G.A.F., *Linear Regression Analysis*, John Wiley&Sons, New York, 1977, p.176。

[7] 节选自 Kerry Peterson 的前引文献，第 184－185 页。

[8] 节选自 Norman R. Draper 和 Harry Smith 的前引文献，第 606－607 页。

[9] 摘自 Christopher Dougherty, *Introduction to Econometrics*, Oxford University Press, New York, 1992, pp. 253－256。

[10] 这还可以推广到从模型中排除不止一个有关 X 变量的情形。对此，可参见 Chandan Mukherjee 等人的前引文献，第 215 页。

附录 13A

13A.1 $E(b_{12})=\beta_2+\beta_3 b_{32}$的证明[方程(13.3.3)]

三变量总体回归模型的离差形式可以写成：

$$y_i=\beta_2 x_{2i}+\beta_3 x_{3i}+(u_i-\bar{u}) \tag{1}$$

首先将两边同时乘以 x_2 并求和，然后又将两边同时乘以 x_3 并求和，通常的正规方程就是：

$$\sum y_i x_{2i}=\beta_2\sum x_{2i}^2+\beta_3\sum x_{3i}x_{2i}+\sum x_{2i}(u_i-\bar{u}) \tag{2}$$

$$\sum y_i x_{3i}=\beta_2\sum x_{2i}x_{3i}+\beta_3\sum x_{3i}^2+\sum x_{3i}(u_i-\bar{u}) \tag{3}$$

将（2）两边同时除以$\sum x_{2i}^2$，我们得到：

$$\frac{\sum y_i x_{2i}}{\sum x_{2i}^2}=\beta_2+\beta_3\frac{\sum x_{2i}x_{3i}}{\sum x_{2i}^2}+\frac{\sum x_{2i}(u_i-\bar{u})}{\sum x_{2i}^2} \tag{4}$$

现在回想：

$$b_{12}=\frac{\sum y_i x_{2i}}{\sum x_{2i}^2}$$

$$b_{32}=\frac{\sum x_{2i}x_{3i}}{\sum x_{2i}^2}$$

方程（4）便可写成

$$b_{12}=\beta_2+\beta_3 b_{32}+\frac{\sum x_{2i}(u_i-u)}{\sum x_{2i}^2} \tag{5}$$

557 对（5）两边同时取期望，我们最后得到：

$$E(b_{12})=\beta_2+\beta_3 b_{32}$$

其中用到(a)对于给定样本，b_{32}为已知量；(b)β_2 和 β_3 都是常数；和(c)u_i 与 X_{2i}(X_{3i}也一样)不相关。

13A.2 含有无关变量的后果：无偏性质

对真模型（13.3.6）我们有：

$$\hat{\beta}_2=\frac{\sum y x_2}{\sum x_2^2} \tag{1}$$

并且我们知道它是无偏的。

对模型(13.3.7) 我们得到：

$$\hat{\alpha}_2=\frac{(\sum y x_2)(\sum x_3^2)-(\sum y x_3)(\sum x_2 x_3)}{\sum x_2^2\sum x_3^2-(\sum x_2 x_3)^2} \tag{2}$$

现在离差形式的真模型是：

$$y_i=\beta_2 x_2+(u_i-\bar{u}) \tag{3}$$

用(3) 代入(2) 中并化简，得：

$$E(\hat{\alpha}_2)=\beta_2\frac{(\sum x_2^2)(\sum x_3^2)-(\sum x_2 x_3)^2}{\sum x_2^2\sum x_3^2-(\sum x_2 x_3)^2}=\beta_2 \tag{4}$$

即 $\hat{\alpha}_2$ 仍是无偏的。

我们还得到：

$$\hat{\alpha}_3=\frac{(\sum y x_3)(\sum x_2^2)-(\sum y x_2)(\sum x_2 x_3)}{\sum x_2^2\sum x_3^2-(\sum x_2 x_3)^2} \tag{5}$$

用(3) 代入(5) 中并化简得：

$$E(\hat{\alpha}_3)=\beta_2\frac{(\sum x_2 x_3)(\sum x_2^2)-(\sum x_2 x_3)(\sum x_2^2)}{\sum x_2^2\sum x_3^2-(\sum x_2 x_3)^2}=0 \tag{6}$$

由于真模型中没有 x_3，这便是它在真模型中的值。

13A.3 (13.5.10) 的证明

558 我们有：

$$Y = \alpha + \beta X_i^* + u_i \tag{1}$$

$$X_i = X_i^* + w_i \tag{2}$$

因此我们有离差形式：

$$y_i = \beta x_i^* + (u_i - u) \tag{3}$$

$$x_i = x_i^* + (w_i - \bar{w}) \tag{4}$$

现在，如果我们使用：

$$Y_i = \alpha + \beta X_i + u_i \tag{5}$$

就得到：

$$\begin{aligned}\hat{\beta} &= \frac{\sum xy}{\sum x^2} = \frac{\sum[\beta x^* + (u - \bar{u})][x^* + (w - \bar{w})]}{\sum[x^* + (w - w)]^2} \qquad \text{利用(3)和(4)} \\ &= \frac{\beta \sum x^{*2} + \beta \sum x^*(w - \bar{w}) + \sum x^*(u - \bar{u}) + \sum(u - \bar{u})(w - \bar{w})}{\sum x^{*2} + 2\sum x^*(w - \bar{w}) + \sum(w - \bar{w})^2}\end{aligned}$$

由于两个变量之比的期望值不等于它们的期望值之比，我们不能取这个表达式的期望值（注：期望值运算子 E 是一个线性运算子）。因此，我们先用 n 除分子和分母的每一项，然后取概率极限，即取：

$$\hat{\beta} = \frac{(1/n)[\beta \sum x^{*2} + \beta \sum x^*(w - w) + \sum x^*(u - u) + \sum(u - u)(w - \bar{w})]}{(1/n)[\sum x^{*2} + 2\sum x^*(w - w) + \sum(w - \bar{w})^2]}$$

的 plim（关于 plim 的细节，见附录 A）。因两个变量之比的概率极限就是它们的概率极限之比，故对每项取概率极限得：

$$\text{plim}\hat{\beta} = \frac{\beta\sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_w^2}$$

559 其中 $\sigma_{X^*}^2$ 和 σ_w^2 是随样本无限增大时 X^* 和 w 的方差。这里我们利用了这样一个事实，即随着样本的无限增大，误差 u 和 w 之间以及它们和真 X^* 之间都无相关关系。由以上表达式，我们最后得到：

$$\text{plim}\hat{\beta} = \beta\left[\frac{1}{1 + (\sigma_w^2/\sigma_{X^*}^2)}\right]$$

这正是我们要证明的结论。

13A.4 方程（13.6.2）的证明

由于此模型中没有截距项，所以根据过原点回归的公式得到（的估计值如下：

$$\hat{\alpha} = \frac{\sum X_i Y_i}{\sum X_i^2} \tag{1}$$

代入真模型（13.2.8）中的 Y，我们得到

$$\hat{\alpha} = \frac{\sum X_i(\beta X_i u_i)}{\sum X_i^2} = \beta \frac{\sum X_i^2 u_i}{\sum X_i^2} \tag{2}$$

统计理论表明，若 $u_i \sim N(0,\sigma^2)$，则

$$u_i = \log \text{正态}[e^{\sigma^2/2}, e^{\sigma^2}(e^{\sigma^2}-1)] \tag{3}$$

因此，

$$E(\hat{\alpha}) = \beta E\left(\frac{\sum X_i^2 u_i}{\sum X_i^2}\right) = \beta\left(E\frac{X_1^2 u_1 + X_2^2 u_2 + \cdots + X_n^2 u_n}{\sum X_i^2}\right)$$

$$= \beta e^{\sigma^2/2}\left(\frac{\sum X_i^2}{\sum X_i^2}\right) = \beta e^{\sigma^2/2}$$

其中使用到 X 都是非随机的和每个 u_i 的期望值都是 $e^{\sigma^2/2}$ 的事实。

由于 $E(\hat{\alpha}) \neq \beta$，所以 $\hat{\alpha}$ 是 β 的一个有偏估计量。

【注释】

[1] Keith Cuthbertson, Stephen G. Hall, and Mark P. Taylor, *Applied Econometrics Techniques*, Michigan University Press, 1992, p.X.

[2] David F. Hendry, *Dynamic Econometrics*, Oxford University Press, U.K., 1995, p.68.

[3] Peter Kennedy, *A Guide to Econometrics*, 3d ed., The MIT Press, Cambridge, Mass., 1992, p.82.

[4] D. F. Hendry and J.F.Richard, "The Econometric Analysis of Economic Time Series," *International Statistical Review*, vol. 51, 1983, pp.3 – 33.

[5] Milton Friedman, "The Methodology of Positive Economics," *in Essays in Positive Economics*, University of Chicago Press, Chicago, 1953, p.7.

[6] 但请参见习题 13.32。

[7] 代数上的处理可参看 Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, pp.391-399。熟悉矩阵代数的读者可参考 J. Johnston, *Econometrics Methods*, 4th ed., McGraw-Hill, New York, 1997, pp.119 – 121。

[8] 回归结果为：

$$\widehat{\text{FLR}} = 47.597\,1 + 0.002\,56\text{PGNP}$$
$$\text{se} = (3.555\,3) \qquad (0.001\,1) \qquad r^2 = 0.072\,1$$

[9] 注意，在真实模型中，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 都是其真实值的无偏估计值。

[10] 为了避开偏误和有效性之间的取舍关系，可以选择最小化均方误 MSE，因为它同时考虑了偏误和有效性。关于 MSE，可参见统计学方面的附录，即附录 A。也可参见习题 13.6。

[11] 但注意 $\hat{\alpha}_1$ 仍是有偏误的，这可如下直观看出：我们知道 $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3$，而 $\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2\bar{X}_2$，即便 $\hat{\alpha}_2 = \hat{\beta}_2$，这两个截距估计量也不同。

[12] 详细分析可参见 Adrian C. Darnell, *A Dictionary of Econometrics*, Edward

Elgar Publisher, 1994, pp.371－372.

[13] Michael D. Intriligator, *Econometric Models, Techniques and Applications*, Prentice Hall, Englewood Cliffs, N.J., 1978, p.189. 回顾简单性原则。

[14] James Davidson, *Econometric Theory*, Blackwell Publishers, Oxford, U. K., 2000, p.153.

[15] William Pool, "Is Inflation Too Low," *the Cato Journal*, vol. 18, no. 3, Winter 1999, p.456.

[16] M. Lovell, "Data Mining," *Review of Economics and Statistics*, vol. 65, 1983, pp.1－12.

[17] 对预检验的详尽讨论及其可能导致的偏误，参见 Wallace, T. D., "Pretest Estimation in Regression: A Survey," *American Journal of Agricultural Economics*, vol. 59, 1977, pp.431－443。

[18] Asad Zaman, *Statistical Foundations for Econometric Techniques*, Academic Press, New York, 1996, p.226.

[19] Kerry Patterson, *An Introduction to Applied Econometrics*, St. Martin's Press, New York, 2000, p.10.

[20] Peter Kennedy, "Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics," unpublished manuscript.

[21] 肯尼迪的前引文献，第 13 页。

[22] 在本例中 $d=2$ 这个值表示无设定误差。(为什么?)

[23] 如果按 X_i^2 或 X_i^3 排列，也没有关系。因为 X_i^2 和 X_i^3 都是已经排列好的 X_i 的（增）函数。

[24] J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society*, series B, vol. 31, 1969, pp.350－371.

[25] R. F. Engle, "A General Approach to Lagrangian Multiplier Model Diagnostics," *Journal of Econometrics*, vol. 20, 1982, pp.83－104.

[26] 这一谚语出自 Milton Friedman。参见习题 13.8。

[27] 但注意，此方差仍是无偏的，因在所述条件下合成误差项 $v_i=u_i+\varepsilon_i$ 仍满足最小二乘法的基本假定。

[28] 如附录 A 所示，$\hat{\beta}$ 是 β 的一致性估计量，如果随着 n 无限增大，$\hat{\beta}$ 的抽样分布最终收缩到 β。这可技术性地表述为 $\operatorname{plim}_{n\to\infty}\hat{\beta}=\beta$。又如附录 A 所指出，一致性是个大样本性质。当一个估计量的有限或小样本性质（如无偏性）不能确定时，一致性常被用来研究该估计量的性态。

[29] 参看 Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, *Advanced Econometric Methods*, Springer Verlag, New York, 1984, pp.273－277。还参看 Kennedy 的前引文献，第 138－140 页，有关加权回归以及工具变量的讨论。

[30] 感谢 Kenneth J. White 对此例的构造，参看他们 *Computer Handbook Using SHAZAM*，用以配合 Damodar Gujarati, *Basic Econometrics*, September 1985, pp. 117－121 的使用。

[31] 更一般地，可使用我们在第 8 章曾简要讨论过的似然比检验、瓦尔德检验和拉格朗日乘数检验。

[32] Andrew Harvey, *The Econometric Analysis of Time Series*, 2d ed., The MIT Press, Cambridge, Mass., 1990, Chap.5.

[33] Thomas B. Fomby, R. Carter Hill, and Stanley R.Johnson, *Advanced Econometric Methods*, Springer-Verlag, New York, 1984, p.416.

[34] 参见 Keith M.Carlson, "Does the St.Louis Equation Now Believe in Fiscal Policy?" *Review*, *Federal Reserve Bank of St.Louis*, vol.60, no.2, February 1978, p.17, Table IV。

[35] R.Davidson and J. G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypothesis," *Econometrica*, vol. 49, 1981, pp.781 - 793.

[36] Jan Kmenta 前引文献，第 597 页。

[37] 也可参见 Badi H. Baltagi, *Econometrics*, Springer, New York, 1998, pp.209 - 222。

[38] Norman D. Drapper and Harry Smith, *Applied Regression Analysis*, 3d ed., John Wiley & Sons, New York, 1998, p.332. 参见此书中某些关于 C_p 的例子。

[39] 对这个专题一个有益的讨论可参见 Francis X. Diebold, *Elements of Forecasting*, 2d ed., South Western Publishing, 2001, pp.83 - 89。总体上，迪博尔德建议的是 SIC 准则。

[40] Wojciech W. Charemza and Derek F. Deadman, *New Directions in Econometric Practice: A General to Specific Modelling, Cointegration and Vector Autoregression*, 2d ed., Edward Elgar Publishers, 1997, p.30. 也可参见第 250 - 252 页他们对各种模型选择准则的见解。

[41] 接下来的讨论受到如下这本书的影响 Chandan Mukherjee, Howard White, and Marc Wyuts, *Econometrics and Data Analysis for Developing Countries*, Routledge, New York, 1998, pp.137 - 148。

[42] 节选自 John Fox, *Applied Regression Analysis*, *Linear Models*, *and Related Methods*, Sage Publications, California, 1997, p.268。

[43] Norman R. Draper 和 Harry Smith 的前引文献，第 76 页。

[44] 这里有一些容易获得的文献来源：Alvin C. Rencher, *Linear Models in Statistics*, John Wiley & Sons, New York, 2000, pp.219 - 224; A.C. Atkinson, *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford University Press, New York, 1985, Chap.3; Ashis Sen and Muni Srivastava, *Regression Analysis: Theory, Methods, and Applications*, Springer-Verlag, New York, 1990, Chap.8; 和 John Fox 的前引文献第 11 章。

[45] 详细情况参见 Jack Johnston and John DiNardo, *Econometric Methods*, 4th ed., McGraw-Hill, New York, 1997, pp.117 - 121。

[46] 参见 J. Mincer, *School*, *Experience and Earnings*, Columbia University Press, New York, 1974。

[47] 比如参见 George Borjas, *Labor Economics*, 2d ed., McGraw-Hill, New York, 2000。

[48] Martin S.Feldstein, "Inflation, Tax Rules and Investment: Some Econometric Evidence," *Econometrica*, vol.30, 1982, p.829.

[49] Peter Kennedy, op.cit., pp.17 - 18.

第3篇

计量经济学专题

561 在第 1 篇里我们介绍了经典线性回归模型及其全部假定。在第 2 篇里我们详细分析了一个或多个假定不被满足时所产生的后果，以及可能的处理方法。在第 3 篇里我们转而研究一些有选择性，然而常常用得到的知识。我们特别研究这些专题：（1）非线性于参数的回归模型，（2）定性响应回归模型，（3）综列数据回归模型，（4）动态经济模型。

第 14 章考虑本质非线性于参数的模型。利用方便的软件包，估计此类模型不再是个巨大的挑战。尽管所涉及的数学可能会吓倒一些读者，但是非线性于参数的回归模型可以从直观上得到解释。本章运用一些适当的例子，显示了如何估计和解释此类模型。

第 15 章考虑解释变量本质上是定性的回归模型。因此本章补充了第 9 章的内容，在第 9 章中我们已经讨论过解释变量本质上是定性的回归模型。本章的基本精神是提出了回归子属于“是”或“不是”一类的模型。因为 OLS 在估计此类模型时存在一些问题，所以几个替代方法就应运而生。本章考虑了两个替代方法，即 **logit 模型**和 **probit 模型**。本章还讨论了定性响应模型的几个变化模型，即 **Tobit 模型**和**泊松（Poisson）回归模型**，并简单地讨论了定性响应模型的几个扩展，例如有序 probit、有序 logit 和多项式 logit。

562 第 16 章讨论了综列数据回归模型。此类模型联合了时间序列和横截面观测。尽管通过联合这种观测我们增加了样本容量，综列数据回归模型还是遇到了若干估计方面的挑战。本章仅仅考虑了此类模型的基本要义，并引导读者去查阅更深入研究的资料。

在第 17 章中，我们考虑了含有解释变量的现期和过期或滞后值的模型，还考虑了含有因变量的滞后值作为解释变量的模型，这两种模型分别称为**分布滞后模型**和**自回归模型**。虽然这类模型在经验计量经济学中非常有用，却由于它们违背了经典回归模型的一个或多个假定而带来了特殊的估计问题。我们将在考伊克（Koyck）适应性预期（AE）和部分调节等模型的构架中来讨论这些特殊问题。我们还注意到理性预期（RE）学派的倡导者针对 AE 模型的批评。

第 14 章 非线性回归模型

563 本书的重点是关于线性回归模型的，即线性于参数的模型和（或）能够变换以使其可以线性于参数的模型。但是，有时由于理论或经验的原因，我们不得不考虑非线性于参数的模型。[1]本章我们来看一看此类模型并研究它们的特性。

§14.1 本质上的线性和非线性回归模型

在第 2 章开始讨论线性回归模型的时候，我们陈述过本书所关心的基本上是线性于参数的模型。它们可以线性于变量，也可以非线性于变量。如果你参照过表 2.3，你将会看到一个线性于参数和变量的模型，它是一个线性回归模型，而且一个线性于参数但是非线性于变量的模型也是线性回归模型。从另一方面说，如果一个模型非线性于参数，那么它就是非线性（非线性于参数）回归模型，而不论该模型的变量线性与否。

564 然而，这里必须小心，有些模型可能看起来非线性于参数，而本质上却是线性的，因为通过合适的变换它们可以变成线性于参数的回归模型。但是，如果此类模型不能线性化于参数，则它们就被称作**本质非线性回归模**

型。从现在起，当我们谈到非线性回归模型时，我们的意思是它是本质非线性的。为了简单起见，我们把它称为 NLRM。

为了弄清这两者之间的区别，我们来回顾一下习题 2.6 和 2.7。在习题 2.6 中，模型 a，b，c，和 c 都是线性回归模型，因为它们都非线性于参数。模型 d 是个混合物，因为 β_2 是线性的，而 β_1 是非线性的。但是如果我们令 $\alpha=\ln\beta_1$，则该模型线性于 α 和 β_2。

在习题 2.7 中，模型 d 和 e 是本质非线性的，因为没有一个方法将之线性化。模型 c 显然是一个线性回归模型。至于模型 a 和 b 呢？把 a 的两边同时取对数，可以得到，它是线性于参数的。因此，模型 a 本质上是线性回归模型。模型 b 是逻辑（概率）分布函数的一个例子，我们将在第 15 章研究这个问题。从表面上看，这好像是一个非线性回归模型。但一个简单的数学技巧将使其转变为名义上的线性回归模型。

$$\ln\left(\frac{1-Y_i}{Y_i}\right)=\beta_1+\beta_2X_i+u_i \tag{14.1.1}$$

因此，模型 b 是本质线性的。我们将会在下一章考察类似于（14.1.1）的模型的应用。

现在来考虑著名的**柯布-道格拉斯（C－D）生产函数**。令 Y = 产出，X_2 = 劳动投入，X_3 = 资本投入，我们把该生产函数写成三种不同的形式：

$$Y_i=\beta_1X_{2i}^{\beta_2}X_{3i}^{\beta_3}e^{u_i} \tag{14.1.2}$$

565 或者，

$$\ln Y_i=\alpha+\beta_2\ln X_{2i}+\beta_3\ln X_{3i}+u_i \tag{14.1.2a}$$

其中 $\alpha=\ln\beta_1$。因此这种形式的 C－D 函数是本质线性的。

现在来考虑这个版本的 C－D 函数：

$$Y_i=\beta_1X_{2i}^{\beta_2}X_{3i}^{\beta_3}u_i \tag{14.1.3}$$

或者，

$$\ln Y_i=\alpha+\beta_2\ln X_{2i}+\beta_3\ln X_{3i}+\ln u_i \tag{14.1.3a}$$

其中 $\alpha=\ln\beta_i$。这个模型也是线性于参数的。

但是现在再来考虑如下版本的 C－D 函数：

$$Y_i=\beta_1X_{2i}^{\beta_2}X_{3i}^{\beta_3}+u_i \tag{14.1.4}$$

正如我们刚才所提到的，版本（14.1.2a）和（14.1.3a）C－D 函数是本质线性（于参数）回归模型，但是无法将（14.1.4）作变换以使变换后的模型线性于参数。[2]因此，（14.1.4）本质上是非线性回归模型。

另外一个众所周知的但却是本质非线性的函数是**常替代弹性（CES）生产函数**，它是柯布-道格拉斯生产函数的一个特例。CES 生产函数采用如下形式：

$$Y_i=A\left[\delta K_i^{-\beta}+(1-\delta)L_i^{\ -\beta}\right]^{-1/\beta} \tag{14.1.5}$$

其中 Y = 产出，K = 资本投入，L = 劳动投入，A = 规模参数，δ = 分布函数（$0<\delta<1$），β = 替代参数（$\beta\geqslant-1$）。[3]无论你把该生产函数中的随机误

差项变换为什么形式，都不可能使它变成线性（对参数而言）回归模型。它本质上是非线性回归模型。

§14.2 线性和非线性回归模型的估计

为了弄清估计线性和非线性回归模型的区别，我们来考虑以下的两个模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{14.2.1}$$

$$Y_i = \beta_1 e^{\beta_2 X_i} + u_i \tag{14.2.2}$$

现在你已经知道了（14.2.1）是个线性回归模型，同时（14.2.2）是一个非线性回归模型。回归（14.2.2）被称为**指数回归模型**，并且常常被用来测量变量的增长，如人口、GDP，或者货币供给。

566 假使我们考虑用 OLS 来估计这两个模型的参数。在 OLS 中，我们将残差平方和最小化，对于模型（14.2.1）来说，残差平方和为：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{14.2.3}$$

这里按照惯例 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 是真实诸 β 的估计量。将上面的表达式对这两个未知量进行微分，我们得到（3.1.4）和（3.1.5）所示的正规方程。同时解这些方程，可以得到由方程（3.1.6）和（3.1.7）给出的 OLS 估计量。仔细观察你会发现在这些方程中，未知量（诸 β）在左手侧而已知量（X 和 Y）在右手侧。结果，根据所得到的数据，我们得到了这两个未知量的显示解。

现在来考虑如果我们尽量将（14.2.2）中的 RSS 最小化会发生什么。正如附录 14A 中 14A.1 部分所示，对应于（3.1.4）和（3.1.5）的正规方程如下所示：

$$\sum Y_i e^{\beta_2 X_i} = \beta_1 e^{2\beta_2 X_i} \tag{14.2.4}$$

$$\sum Y_i X_i e^{\hat{\beta}_2 X_i} = \hat{\beta}_1 \sum X_i e^{2\hat{\beta}_2 X_i} \tag{14.2.5}$$

与线性回归模型的例子中的正规方程不同，非线性回归模型的正规方程的左手侧和右手侧都有未知量（诸 β）。因此，我们不能根据已知量来得出未知量的显示解。换句话说，未知量要用它们本身和数据来表达。因此，尽管我们能用最小二乘法来估计非线性回归模型的参数，但仍然不能得到未知量的显示解。顺便指出，OLS 适用于一个非线性回归模型时叫做**非线性最小二乘法**（NLLS）。那么，这个问题的解决办法如何？我们看一下节。

§14.3 估计非线性回归模型：试错法

为了做好准备，让我们来考虑一个具体的例子。表 14.1 中的数据是关

于美国一家重要的信托基金支付给其投资顾问的管理其资产的费用。支付的费用取决于该基金的净资产值。正如你所见，基金的净资产值越高，咨询费越低，这一点我们可以从图 14.1 清楚地看出。

表 14.1　支付的咨询费和资产规模

	费用，%	资产*
1	0.520	0.5
2	0.508	5.0
3	0.484	10
4	0.46	15
5	0.439 8	20
6	0.423 8	25
7	0.411 5	30
8	0.402	35
9	0.394 4	40
10	0.388	45
11	0.382 5	55
12	0.373 8	60

* 资产代表净资产值，单位为 10 亿美元。

561

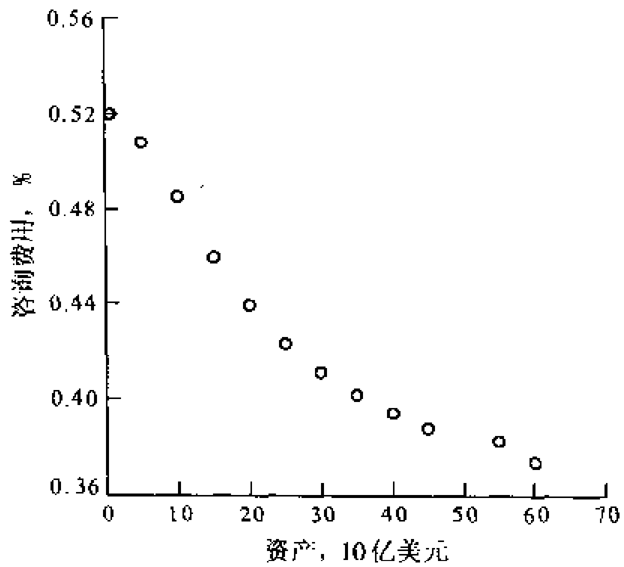


图 14.1　咨询费和基金资产的关系

567 为了弄清（14.2.2）中的指数回归模型如何拟合表 14.1 中的数据，我们可以用试错法来进行。假定最初 $\beta_1 = 0.45$，$\beta_2 = 0.01$。这些数据纯粹是猜测的，

有时候可以根据以前的经历或者经验来猜测，而有时候仅仅是为了使其拟合线性回归模型，尽管它可能并不适当。暂时不要担心这些取值是如何得到的。

既然知道了 β_1 和 β_2 的值，我们就可以把(14.2.2)写作：

$$u_i = Y_i - \beta_1 e^{\beta_2 X_i} = Y_i - 0.45e^{0.01X_i} \tag{14.3.1}$$

568 因此，

$$\sum u_i^2 = \sum (Y_i - 0.45e^{0.01X_i})^2 \tag{14.3.2}$$

因为 Y，X，β_1 和 β_2 是已知的，我们可以很容易地找出（14.3.2）中的误差平方和。[4] 记住在普通最小二乘法中，我们的目标是找到使误差平方和尽可能小的未知参数的值。如果从模型估计出的 Y 值与实际 Y 值尽可能接近，我们就达到了目标。根据给定的数值，我们可以得出 $\sum u_i^2 = 0.304\,4$。但我们如何知道这是们所能得到的尽可能小的误差平方和呢？如果选取 β_1 和 β_2 的值分别为 0.50 和 −0.01 会怎样呢？重复刚才所拟定的程序，我们发现现在将得到 $\sum u_i^2 = 0.007\,3$。显然，此时误差平方和比前面所得到的等于 0.304 4 的误差平方和更小。但我们如何知道是否达到了尽可能小的误差平方和呢？因为一旦将诸 β 取另一组值，我们又将得到另一个误差平方和。

如你所见，这种试错法或者说**迭代**过程提供了一种工具。如果一个人有无限的时间和耐心，试错法最终可以得出能保证误差平方和最小的和的值。但是你可能会问，我们如何从（$\beta_1 = 0.45$；$\beta_2 = 0.01$）到（$\beta_1 = 0.50$；$\beta_2 = -0.1$）呢？显然，我们需要某种算法规则，它要在我们停止前告诉我们如何从未知量的一组值到另一组值。幸运的是，这种算法规则是存在的，我们将在下一部分讨论。

§ 14.4　估计非线性回归模型的方法

估计 NLRM 有几种方法，或者说算法规则，即：（1）直接搜索或者试错，（2）直接最优化，和（3）迭代线性化。[5]

直接搜索或试错法或不用求导的方法

569 在前面一部分我们讲述了如何使用这种方法。尽管直观上该方法很有吸引力，因为它不像另一种方法需要使用微积分，但是它并没有被广泛使用。首先，如果一个 NLRM 涉及几个参数，那么这种方法就会变得很难处理，并且需要复杂的计算因而成本高昂。举例来说，如果一个 NLRM 包含 5 个参数，所考虑的每个参数有 25 个可选值，那么你必须计算 $(25)^5 = 9\,765\,625$ 次误差平方和。其次，你不能保证你所选取的最后一组参数能提供绝对最小

的误差平方和。用微积分的术语来讲，你可以得到局部最小值，但不是绝对最小值。实际上，没有方法可以保证完全最小。

直接最优化

在直接最优化方法中，我们将误差平方和对每一个系数或参数进行微分，然后令得到的方程等于零，同时解所得到的正规方程。在方程(14.2.4)和(14.2.5)中，我们已经可以看出这一点。但是，正如你从这些方程中所能看出的，我们不能得出这些方程的显示解或分析解。因此就需要某种迭代方法。其中有种方法称作**最速下降法**。我们不打算讨论该方法的技术细节，尽管它们在某种程度上会有所涉及，但读者可以从参考书中了解这些细节。像试错法一样，最速下降法也涉及选取未知参数的初始试验值的问题，但它需要比漫无目的方法或试错法更有系统地进行。后者的一个缺点是它可能会极端缓慢地与参数的最终值一致。

迭代线性化方法

在这种方法中，我们将关于参数的初始值的非线性化方程线性化，然后用OLS来估计线性化方程，并且调整最初选取的参数值。这些经过调整的参数值可用来再次线性化该模型，然后我们再一次用OLS进行估计，重新调整估计值。继续这个过程直到从最后一组迭代中得到的估计值没有实质性变化。线性化非线性方程的主要技巧就是微积分中的**泰勒（Taylor）级数展开**。该方法的基本细节可参见附录14A中的14A.2部分。有两种方法将利用泰勒级数展开来估计NLRM进行了系统化，即**高斯-牛顿（Gauss Newton）迭代法**和**牛顿-拉夫森（Newton-Raphson）迭代法**。由于这些方法中的一种或两种现在已经并入几个计算机软件包中，而且讨论它们的技术细节远远超出了本书的范围，因此这里不需要详细阐述。[6]下一部分我们讨论使用这些方法的一些例子。

§14.5 说明性的例子

570 **例14.1 信托基金咨询费**

参照表14.1所给出的数据和NLRM(14.2.2)。利用Eviews 4非线性回归方法（它使用线性化方法）[7]，我们得到下面的回归结果、系数、它们的标准误和 t 值，如下表所示：

变量	系数	标准误	t 值	p 值
截距	0.508 9	0.007 4	68.224 6	0.000 0
资产	−0.005 9	0.000 48	−12.315 0	0.000 0

$$R^2 = 0.938\,5 \qquad d = 0.349\,3 \tag{14.5.1}$$

根据这些结果，我们可以把估计模型写成：

$$\overline{\text{Fee}}_i = 0.508\,9\text{Asset}^{-0.005\,9}$$

在我们讨论这些结果以前，必须注意你不必提供参数的初始值以开始线性化过程，Eviews 可以自己完成这个工作。为了得到（14.5.1）所示的结果需要 5 次 Eviews 迭代。然而，你也能提供你自己的初始值以开始这个过程。为了说明，我们选取 β_1 和 β_2 的值分别为 $\beta_1 = 0.45$ 和 $\beta_2 = 0.01$。我们得到和（14.5.1）一样的结果，但它需要 8 次迭代。注意到这一点是很重要的，即如果你的初始值与最终值相差不远，那么就需要更少的迭代次数。在某些例子中，你能够简单地通过将回归子对回归元进行 OLS 回归而忽视非线性化来选取参数的初始值。例如，利用表 14.1 中的数据，如果你对资产的费用进行回归，则 β_1 和 β_2 的 OLS 估计值分别为 $\beta_1 = 0.502\,8$ 和 $\beta_2 =$ 0.002，它们更接近（14.5.1）所给出的最终值。（关于技术细节，参见附录 14A，14A.3。）

现在再来讨论 NLLS 估计量的特性。你可能会回想起，在具有正态分布的误差项的线性回归模型的例子中，我们能够用 t、F 或者 χ^2 检验来提出准确的推断程序（如假设检验），而不论所选的是小样本还是大样本。不幸的是，这不是 NLRM 的例子，即使它具有正态分布的误差项。**NLLS 估计量不是正态分布的，不是无偏的，并且没有最小方差，而不论所选取的是有限样本，小样本还是大样本。**因此，我们不能用 t 检验（来检验单个系数的重要性）或者 F 检验（来检验估计回归的总体重要性），因为不能根据估计的残值来得出误差方差的无偏估计。而且，残值（实际 Y 值和根据 NLRM 估计的 Y 的估计值之间的差额）的总和不必一定等于零，ESS 和 RSS 的和不必一定等于 TSS，因此 R^2 = ESS /RSS 对于该模型来说可能并不是一个有意义的描述性统计量。但是，我们可以把 R^2 计算为：

$$R^2 = 1 - \frac{\sum \hat{u}_i^{\,2}}{\sum (Y_i - \bar{Y})^2} \tag{14.5.2}$$

其中 Y = 回归子，$\hat{u}_i - Y_i - \hat{Y}_i$，这里的 $\hat{Y}_i$ 是从拟合的 NLRM 得出的 Y 的估计值。

571 因此，关于非线性回归中的回归系数的推断通常是以大样本理论为基础。大样本理论告诉我们，当样本规模很大时，对于具有正态分布的误差项的非线性回归模型的最小和最大似然估计量是接近正态分布和无偏的，而且

几乎具有最小方差。当误差项不是正态分布时，也能运用大样本理论。[8]

简单来说，NLRM 中所有的推断程序都是大样本的，或者是渐近的。再回到例 14.1，(14.5.1) 中所给出的估计量 t 只有在大样本的背景下才是有意义的。从这种意义上来讲，我们可以说方程（14.5.1）所示的估计系数是独立且统计显著的。当然，本例中我们的样本是相当小的。

再回到方程（14.5.1)，我们如何计算出 Y（费用）相对于 X（资本规模）的变化率呢？运用基本的推导规则，读者能看出 Y 相对 X 的变化率为：

$$\frac{\mathrm{d}Y}{\mathrm{d}X}=\beta_1\beta_2e^{\beta_2X}=(-0.005\,9)(0.508\,9)e^{-0.005\,9X} \tag{14.5.3}$$

正如你所见，费用的变化率依赖于资产值。例如，如果 $X=20$（百万），从 (14.5.3) 中我们能看出交纳的管理费用的预期变化率大约是 $-0.003\,1\%$。当然，答案将随着计算中所使用的 X 值的变化而变化。根据从（14.5.2）所算出的 R^2 来判断，0.938 5 的值暗示我们所选取的 NLRM 与表 14.1 中的数据拟合得很好。0.349 3 的德宾-沃森估计值暗示存在自相关性或者可能有模型设定误差。显然有程序可用来解决这些问题以及 NLRM 中的异方差性问题，这里我们就不再讨论这个专题了。感兴趣的读者可以查阅参考书。

例 14.2 墨西哥经济的柯布-道格拉斯产量

参照习题 14.9 中给出的数据，即关于 1955—1974 年间墨西哥经济的一些数据。我们可以看出（14.1.4）所给出的 NLRM 是否与这些数据相拟合，注意 Y= 产量，X_2= 劳动投入，X_3= 资本投入。运用 Eviews 4，经过两次迭代，我们可以得到如下回归结果。

变量	系数	标准误	t 值	p 值
截距	0.529 2	0.271 2	1.951 1	0.067 7
劳动	0.181 0	0.141 2	1.281 4	0.217 3
资本	0.882 7	0.070 8	12.465 8	0.000 0

$$R^2=0.994\,2 \qquad d=0.289\,9$$

572 因此，估计的柯布-道格拉斯产量是：

$$\mathrm{GDP}_t=0.529\,2\,劳动_t^{0.181\,0}\,资本_t^{0.882\,7} \tag{14.5.2}$$

渐近地说明，这个方程显示出在该模型中仅仅只有资本投入系数是重要的。习题 14.9 将会要求你把这些结果同根据（14.1.2）中给出的倍增的柯布-道格拉斯生产函数得出的结果相比较。

例 14.3 美国人口增长，1970—1999

习题 14.8 中的表给出了 1970—1999 年间美国的总人口。下面这种类型的逻辑增长模型通常被用来测量人口增长：

$$Y_t = \frac{\beta_1}{1 + e^{(\beta_2 + \beta_3 t)}} + u_t \tag{14.5.4}$$

其中 Y＝人口；t＝时间，按年月的顺序来度量；诸 β 是参数。注意关于这个模型的一个有趣的事情。尽管只有两个变量，即人口和时间，但是却有三个未知参量，这一点显示出在 NLRM 中参数可以比变量多。

样本：1970—1999

包括的观察值：30

一次迭代后趋向收敛

	系数	标准误	t 值	p 值
β_1	1 432.738	508.011 3	2.820 2	0.008 9
β_2	1.798 6	0.412 4	4.361 3	0.000 2
β_3	−0.011 7	0.000 8	−14.065 8	0.000 0

$R^2 = 0.999\ 7 \qquad d = 0.334\ 5$

因此，估计模型是：

$$\hat{Y}_t = \frac{1\ 432.739}{1 + e^{1.798\ 6 - 0.011\ 7t}} \tag{14.5.5}$$

因为我们有一个合理的大样本，因此所有估计的系数都是渐近统计显著的。低德宾-沃森统计量意味着误差项可能是自相关的。习题 14.8 要求你将上面的模型与半对数模型：$\ln Y_t = \beta_1 + \beta_2$时间$+ u_t$ 进行比较，并计算出这两个模型中隐含的人口增长率。

§ 14.6 要点与结论

573 本章中所讨论的要点可以概括为：

1. 尽管线性回归模型在理论和实际中占主导地位，但有时非线性回归模型还是有用的。

2. 线性回归模型中需要使用的数学是相当简单的，因为人们可以得到该模型的系数的显示解或分析解。我们很好地建立了此类模型推断的小样本和大样本理论。

3. 与此相对照，对于本质非线性回归模型，我们不能得到其准确的参数值。它们只能从数学上估计。

4. 有几种获得 NLRM 估计的方法，如：（1）试错法，（2）非线性最小二乘法，和（3）运用泰勒级数展开来线性化。

5. 计算机软件包现在已经形成了永久性的惯例，例如高斯-牛顿，

牛顿-拉夫森和马奎德（Marquard）。它们都是迭代惯例。

6. 在有限样本中，NLLS 估计量并不具有理想的特性。但是在大样本中，它们具有这种特性。因此，在小样本中，NLLS 的结果必须详细说明。

7. 自相关性、异方差性和模型设定问题可能会对 NLRM 造成麻烦，就如同它们会对线性回归模型造成麻烦那样。

8. 我们用几个例子来说明了 NLRM。因为能利用非常容易使用的软件包，NLRM 估计不再是秘密。因此，每当由于理论或者实际的原因需要使用它们时，读者不应当回避这种模型。实际上，如果参考习题 12.10，你将从方程（1）中看出有一个需要使用 NLRM 估计的本质非线性回归模型。

习题

问答题

14.1 本质线性和非线性回归模型意味着什么？举几个例子。

14.2 因为柯布-道格拉斯生产函数中的误差项可以是倍增(multiplicatively)或者递增的（additively），那么在这两者之间你如何做决定呢？

14.3 OLS 和非线性最小二乘估计之间的区别是什么？

14.4 饱和蒸汽中压力和温度间的关系可以表示为[1]：

$$Y = \beta_1 (10)^{\beta_2 t/(\gamma + t)} + u_t$$

574 其中 Y = 压力，t = 温度。运用非线性最小二乘法求出这个模型的正规方程。

14.5 判断下面的陈述是对的还是错的。给出你的理由。

a. 即使假定是正态分布的，NLLS 回归中的统计推断也不能建立在通常的 t，F 和 χ^2 检验的基础上。

b. 判定系数（R^2）对于 NLRM 来说并不是一个特别有意义的数字。

14.6 如何线性化本章所讨论的 CES 生产函数？列出必要的步骤。

14.7 描述时间变量的行为的模型被称为增长模型。此类模型可以用在很多不同的领域，如经济学、生物学、植物学、生态学和人口统计学。增长模型可以采取不同的形式，包括线性的和非线性的。考虑下面的模型，这里 Y 是我们想测量其增长的变量；t 是时间，按年月顺序度量；U_t 是随机误差项。

a. $Y_t = \beta_1 + \beta_2 t + u_t$

b. $\ln Y_t = \beta_1 + \beta_2 t + u_t$

c. 逻辑增长模型：$Y_t = \dfrac{\beta_1}{1 + \beta_2 e^{-\beta_3 t}} + u_t$

d. 冈珀茨（Gompertz）增长模型：$Y_t = \beta_1 e^{-\beta_2 e^{\beta_3 t}} + u_t$

通过考虑 Y 相对于时间的增长来找出这些模型的特性。

解答题

14.8 表14.2 中的数据给出了 1970—1999 年间美国的人口，单位：百万。拟合习题 14.7 给出的增长模型，并判断哪个模型拟合得更好，解释模型的参数。

表 14.2

观察值	美国人口	时间	观察值	美国人口	时间
1970	205.052	1	1985	238.466	16
1971	207.661	2	1986	240.651	17
1972	209.896	3	1987	242.804	18
1973	211.909	4	1988	245.021	19
1974	213.854	5	1989	247.342	20
1975	215.973	6	1990	249.948	21
1976	218.035	7	1991	252.639	22
1977	220.239	8	1992	255.374	23
1978	222.585	9	1993	258.083	24
1979	225.055	10	1994	260.599	25
1980	227.726	11	1995	263.044	26
1981	229.966	12	1996	265.463	27
1982	232.188	13	1997	268.008	28
1983	234.307	14	1998	270.561	29
1984	236.348	15	1999	273.131	30

资料来源：*Economic Report of the President*，2000.

14.9 表14.3 给出了 1955—1974 年间墨西哥的实际 GDP，劳动和实际固定资产的数据。考察方程（14.1.4）所给出的附加的柯布-道格拉斯生产函数是否与这些数据相拟合。将你的结果与从（14.1.2）所给出的倍增的柯布-道格拉斯生产函数而得出的结果相比较。哪一个拟合得更好？

575

表 14.3 **墨西哥经济的生产函数数据**

观察值	GDP	劳动	资本	观察值	GDP	劳动	资本
1955	114 043	8 310	182 113	1965	212 323	11 746	315 715
1956	120 410	8 529	193 749	1966	226 977	11 521	337 642
1957	129 187	8 738	205 192	1967	241 194	11 540	363 599

1958	134 705	8 952	215 130	1968	260 881	12 066	391 847
1959	139 960	9 171	225 021	1969	277 498	12 297	422 382
1960	150 511	9 569	237 026	1970	296 530	12 955	455 049
1961	157 897	9 527	248 897	1971	306 712	13 338	484 677
1962	165 286	9 662	260 661	1972	329 030	13 738	520 553
1963	178 491	10 334	275 466	1973	354 057	15 924	561 531
1964	199 457	10 981	295 378	1974	374 977	14 154	609 825

注：GDP 以 1960 年比索计算，百万；
劳动以千人为单位；
资本以 1960 年比索计算，百万。

资料来源：Victor J.Elias, *Sources of Growth: A Study of Seven Latin American Economies*, International Center for Economic Growth, ICS Press, San Francisco, 1992, Tables E－5, E－12, E－14.

【习题注释】

［1］根据 Draper 和 Smith，前引文献，第 554 页改编而来。

附录 14A

14A.1 方程（14.2.4）和（14.2.5）的推导

将方程（14.2.2）写成：

$$u_i = Y_i - \beta_1 e^{\beta_2 X_i} \tag{1}$$

因此，

$$\sum u_i^2 = \sum (Y_i - \beta_1 e^{\beta_2 X_i})^2 \tag{2}$$

因而误差平方和是 β_1 和 β_2 的函数，因为 Y 和 X 的值是已知的。因此，为了最小化误差平方和，我们必须将它对这两个未知数进行微分，即：

$$\frac{\partial \sum u_i^2}{\partial \beta_1} = 2\sum (Y_i - \beta_1 e^{\beta_2 X_i})(-1 e^{\beta_2 X_i}) \tag{3}$$

$$\frac{\partial \sum u_i^2}{\partial \beta_2} = 2\sum (Y_i - \beta_1 e^{\beta_2 X_i})(-\beta_1 e^{\beta_2 X_i} X_i) \tag{4}$$

576 根据一阶最优化条件，令上面的方程等于 0 并解之。我们就可以得到方程（14.2.4）和（14.2.5）。注意在微分误差平方和时，我们使用了链式规则。

14A.2 线性化方法

熟悉微积分的同学将会回忆起**泰勒定理**，该定理提出任意连续的且具有连续的 n 阶导数的函数$f(X)$在 $X=X_0$ 这一点附近都可以近似等于如下所示的一个多项式函数和余数之和，即：

$$f(X)=\frac{f(X_0)}{0!}+\frac{f'(X_0)(X-X_0)}{1!}+\frac{f''(X_0)(X-X_0)^2}{2!}+\cdots+\frac{f^n(X_0)(X-X_0)^n}{n!}+R \qquad (1)$$

其中$f'(X_0)$是 $f(X)$的一阶导数在 $X=X_0$ 处的值，$f''(X_0)$是 $f(X)$的二阶导数在 $X=X_0$ 处的值，以此类推。$n!$（读作 n 的阶乘）代表$n(n-1)(n-2)\cdots 1$，习惯上 $0!=1$，并且 R 代表余数。如果取 $n=1$，我们将会得到一个线性近似值；取 $n=2$，我们就会得到一个二次多项式近似值。如你所见，多项式的次数越高，近似值就越接近初始函数。(1) 中所给出的级数被称为在 $X=X_0$ 附近的泰勒级数展开。举例来说，考虑这个函数：

$$Y=f(X)=\alpha_1+\alpha_2X+\alpha_3X^2+\alpha_4X^3$$

假设我们想在这一点附近将其近似。现在我们得到：

$$f(0)=\alpha_1 \quad f'(0)=\alpha_2 \quad f''(0)=2\alpha_3 \qquad f'''(0)=6\alpha_4$$

因此我们可以得到下面的近似值：

$$一阶：Y=\alpha_1+\frac{f'(0)}{1!}=\alpha_1+\alpha_2X+余项(=\alpha_3X^2+\alpha_4X^3)$$

$$二阶：Y=f(0)+\frac{f'(0)}{1!}X+\frac{f''(0)}{2!}X^2$$

$$=\alpha_1+\alpha_2X+\alpha_3X^2+余项(=\alpha_4X^3)$$

$$三阶：Y=\alpha_1+\alpha_2X+\alpha_3X^2+\alpha_4X^3$$

577 三阶近似值准确重现了初始方程。

泰勒级数近似值的目标通常是为了选取一个次数更低的多项式以使余数项无关紧要。它通常通过线性函数和降低高阶项来近似非线性函数。

泰勒级数近似值可以很容易地展开为包含多于一个 X 的函数。例如，考虑下面的函数：

$$Y=f(X,Z) \qquad (2)$$

并且假设我们想在 $X=a$ 和 $Z=b$ 这一点附近将其展开。泰勒定理显示：

$$f(x,z)=f(a,b)+f_x(a,b)(x-a)+f_z(a,b)f(z-b)+\frac{1}{2!}[f_{xx}(a,b)(x-a)^2-2f_{xz}(a,b)(x-a)(z-b)+f_{zz}(a,b)(z-b)^2]+\cdots \qquad (3)$$

其中 f_x = 该函数对 X 的偏导，f_{xx} = 该函数对 X 的二阶偏导，对于变量 Z 来说与此类似。如果想得到该函数的线性近似，我们就可以使用（3）中的两项，如果想得到一个二次多项式或者二次近似值，我们就可以使用（3）中的三项，依此类推。

14A.3 对（14.2.2）中指数函数的线性近似

我们所考虑的函数是：

$$Y = f(\beta_1, \beta_2) = \beta_1 e^{\beta_2 X} \tag{1}$$

注：为了便于操作，我们降低了观察下标。

记住在这个函数中未知数是 β 系数。让我们在 $\beta_1 = \beta_1^*$ 和 $\beta_2 = \beta_2^*$ 处线性化该函数，这里加的是给定的固定值。为了将之线性化，我们按照如下方法进行：

$$Y = f(\beta_1, \beta_2) = f(\beta_1^*, \beta_2^*) + f_{\beta_1}(\beta_1^*, \beta_2^*)(\beta_1 - \beta_1^*) + f_{\beta_2}(\beta_1^*, \beta_2^*)(\beta_2 - \beta_2^*) \tag{2}$$

其中 f_{β_1} 和 f_{β_2} 是函数（1）关于这两个未知数的偏导，并且其将在（假定的）未知参数的值处取值。注意在上面的表达式中，我们只用了一阶偏导，因为我们在线性化该函数。现在假定 $\beta_1^* = 0.45$ 和 $\beta_2^* = 0.01$，这是真实系数的纯猜测估计值。现在，根据标准的微分规则：

$$f(\beta_1^* = 0.45, \beta_2^* = 0.01) = 0.45e^{0.01X_i} \tag{3}$$

$$f_{\beta_1} = e^{\beta_2 X_i} \quad 和 \quad f_{\beta_2} = \beta_1 X_i e^{\beta_2 X_i}$$

578 计算出在给定的未知数值处这些偏导的值并代入（2），我们得到：

$$Y_i = 0.45e^{0.01X_i} + e^{0.01X_i}(\beta_1 - 0.45) + (0.45)X_i e^{0.01X_i}(\beta_2 - 0.01) \tag{4}$$

可以将其写为：

$$(Y_i - 0.45e^{0.01X_i}) = e^{0.01X_i}\alpha_1 + 0.45X_i e^{0.01X_i}\alpha_2 \tag{5}$$

其中

$$\alpha_1 = (\beta_1 - 0.45), \ \alpha_2 = (\beta_2 - 0.01) \tag{6}$$

现在让 $Y_i^* = (Y_i - 0.45e^{0.01X_i})$，$X_1 = e^{0.01X_i}$，$X_{2i} = 0.45X_i e^{0.01X_i}$ 应用这些定义和误差项 u_i，我们可以把(5)写为：

$$Y_i^* = \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_i$$

你瞧，现在我们得到了一个线性回归模型。因为根据这些数据我们能轻易地算出 Y_i^*，X_{1i} 和 X_{2i} 的值，所以能通过 OLS 来估计（7）并得出 α_1 和 α_2 的值。然后，从（6），我们可以得到：

$$\beta_1 = \hat{\alpha}_1 + 0.45 \text{ 和 } \beta_2 = \hat{\alpha}_2 + 0.01 \tag{8}$$

将这些值分别称之为 β_1^{**} 和 β_2^{**}。利用这些修正值，我们可以开始进行（2）给出的迭代程序，从而得到另一组系数的值。按照这种方法继续进行迭代（或者说线性化），直至系数的值没有重大变化。例 14.1 用了 5 次迭代，而墨西哥柯布-道格拉斯的例子则用了 32 次迭代。但这些迭代背后隐藏的逻辑恰恰是我们解释过了的程序。

对于信托基金的费用结构来说，（6）所给出的 X_1，X_2，Y^* 如下表所示；基础的数据由表 14.1 给出。根据这些值，可以得出（7）的回归结果是：

因变量：Y^*

方法：最小二乘法

变量	系数	标准误	t 统计量	概率
X_1	0.039 775	0.006 229	6.385 6	0.000 1
X_2	0.001 303	0.000 157	8.309 5	0.000 0

$R^2 = 0.948\ 378$　　德宾-沃森 $d = 0.583\ 37$

现在利用（8），读者可以证明：

$$\beta_1^{**} = 0.489\ 7 \quad \text{和} \quad \beta_2^{**} = 0.011\ 3 \tag{9}$$

579

表 14.4

Y^*	X_1	X_2
0.022 49	0.452 25	0.226 12
0.032 38	0.473 07	2.365 3
0.031 58	0.497 32	4.973 2
0.029 64	0.522 82	7.842 3
0.030 43	0.549 63	10.992 6
0.034 39	0.577 81	14.445 2
0.041 09	0.607 43	18.223 0
0.049 65	0.638 58	22.350 3
0.059 23	0.671 32	26.852 8
0.069 18	0.705 74	31.758 3
0.094 02	0.779 96	42.898 0
0.099 39	0.819 95	49.197 2

将这些数字与这两个参数的初始猜测值分别为 0.45 和 0.01 时进行比较。运用 (9) 给出的新的估计，你可以再次开始迭代程序，直至其“收敛”，即最后一轮估计与它前一轮的估计相差不远。当然，如果你初始的猜测值与最终值比较接近，那么你将需要更少的迭代次数。也请注意我们仅仅只用了泰勒级数展开式中的线性项。如果你打算用展开式中的二次项或者更高次数项，也许你将更快达到最终值。但许多应用表明线性近似是非常好的。

【注释】

[1] 在第 4 章我们曾提到，在具有正态分布的误差项的假设条件下，OLS 估计量不但是 BLUE 的，而且在所有的估计量中是 BUE（最优无偏估计），而不论其线性与否。但如果我们撤掉正态分布的假设条件（正如戴维森和麦金农所提到的），则很可能得到非线性的和（或）有偏的估计量，并且这个统计量可能比 OLS 估计量更有效。参阅 Russell Davidson 和 G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, p.161。

[2] 如果你对该模型做对数变换，它将不再适用，因为 $\ln(A+B)\neq\ln A+\ln B$。

[3] 关于 CES 生产函数的特性，参阅 Michael D. Intriligator, Ronald Bodkin and Cheng Hsiao, *Econometric Models, Techniques, and Applications*, 2d ed., Prentice Hall, 1996, pp.294－295。

[4] 注意我们称 $\sum u_i^2$ 为误差平方和，而不是通常的残差平方和，因为我们假定参数值为已知。

[5] 下面的讨论源于：Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 4th ed., McGraw-Hill, 1998, Chap.10; Norman R. Draper and Harry Smith, *Applied Regression Analysis*, 3d ed., John Wiley & Sons, 1998, Chap.24; Arthur S. Goldberger, *A Course in Econometrics*, Harvard University Press, 1991, Chap.29; Russell Davidson and James MacKinnon, op.cit., pp.201－207; John Fox, *Applied Regression Analysis, Linear Models, and Related Methods*, Sage Publications, 1997, pp.393－400; and Ronald Gallant, *Nonlinear Statistical Models*, John Wiley and Sons, 1987.

[6] 有一种有时会用到的方法（即马奎特方法），它是最速下降法和线性化（或泰勒级数）法的折中。感兴趣的读者可以参阅参考书以了解该方法的细节。

[7] Eviews 提供了三种选择：二次爬山法（quadratic hill climbing）、牛顿-拉夫森方法，和 Berndt-Hall-Hall-Hausman。缺省选择是二次爬山法，它从牛顿-拉夫森方法变化而来。

[8] John Neter, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman, *Applied Regression Analysis*, 3d ed., Irwin, 1996, pp.548－549.

第 15 章　定性响应回归模型

580 到目前为止，在我们考虑的所有回归模型中都隐含地假定了回归子、因变量或响应变量（response variable）Y 是定量的，而解释变量是定量的、定性的（或虚拟变量）或二者兼有之。事实上，关于虚拟变量在第 9 章中，我们看到了虚拟回归元是如何被引入到回归模型中的，以及它们在特定情况下所扮演的角色。

本章中我们将考虑几种模型，其中回归子本身在本质上是定性的。尽管在社会科学和医学研究的各领域中的应用不断增加，定性响应回归模型仍然引起了令人感兴趣的估计和解释的挑战。本章中我们仅仅接触这个领域中的一部分主题，而将细节留给更多专门的书籍。[1]

§ 15.1　定性响应模型的性质

581 假如我们想研究成年男子的劳动力参与（LFP）的决定问题。因为一个成人或者在劳动力行列中，或者不在，所以 LFP 是一个是或不是的决定。这样，响应变量或回归子只能取两个值，即如果这个人在劳动力行列中，则取值 1；如果他或她不在其中，则取值 0。换言之，回归子是一个**二分变量**

(dichotomous variable)。劳动力经济学研究表明，LFP 的决定是失业率、平均工资率、教育和家庭收入等的一个函数。

另一个例子，考虑美国总统的选举。假设有两个政党——民主党和共和党。这里的因变量是两个政党间的投票选举。假定令 $Y=1$，若投票给民主党候选人；令 $Y=0$，若投票给共和党候选人。耶鲁大学的经济学家费尔(Ray Fair) 和几个政治科学家[2]在这个专题上做了大量的研究。在投票选择中使用的一些变量包括 GDP 增长率、失业率和通货膨胀率、候选人是否参加再竞选等等。这里重要的是回归子是一个定性变量。

回归子在性质上是定性的其他例子可以想出好几个。例如，一个家庭或者拥有一所住宅或者不拥有，有残疾保险或者没有，夫妻两人都在工作或者只一人在工作等。类似地，某种药物在医治一种疾病中是否有效，一家厂商决定是否宣布一种股利，一位参议员是否对减税提案投赞成票，总统是否对一法案行使否决权，等等。

我们不必将响应变量仅仅局限于是或否二分类型。回到总统选举的例子中，假定这里有三个政党——民主党、共和党和独立党派。这里的响应变量是三分 (trichotomous) 的。一般地，我们可以拥有一个**多分响应变量**（或多类型响应变量)。

我们计划先考虑二分回归子，然后再分析这个基本模型的各种扩展。但是，在我们这样做之前，把回归子区分为定性和定量两种回归模型，注意其基本差别是很重要的。

在一个 Y 是定量的模型中，我们的目标是对给定回归元的值来估计回归子的期望值，或均值。用第 2 章的话来说，我们所需要的是 $E(Y_i \mid X_{1i}, X_{2i}, \cdots, X_{ki})$，其中诸 X 是回归元，既有定量的也有定性的。在 Y 为定性的模型中，我们的目标是找出诸如向民主党候选人投票，或者拥有住宅，或者属于一个工会，或者参加一项运动等等这些事情发生的概率。因此，定性响应模型通常被称为概率模型。

在本章的余下部分，我们将寻求以下问题的答案：

1. 我们怎样估计定性响应模型？能否简单地用平常的 OLS 方法进行估计？

582 2. 有特殊推断问题吗？换言之，假设检验程序是否与我们已经学过的有所不同？

3. 若一个回归子是定性的，我们怎样对这种模型的拟合优度进行量度？在这种模型中计算惯用的 R^2 有什么价值吗？

4. 一旦超出二分回归子的情况，我们怎样估计和解释多分回归模型呢？还有，我们该怎样处理这样的模型：它的回归子是序数的，也就是一个有序类型变量，比如说学校教育（不到 8 年、8～11 年、12 年以及 13 年或以上)；或者它的回归子是名义的，即没有内在的顺序，比如说人种（黑人、白人、西班牙人、亚洲人和其他)？

5. 对于这些现象我们该如何建模：比如每年看内科医生的次数、给定年份中一个厂商获得专利的个数、一年中大学教授发表论文的篇数、五分钟

内接到电话的次数或者五分钟内通过某个收费站的汽车辆数？这些被称为**计数数据**或者**稀有事件数据**的现象是泊松（概率）过程的一些例子。

在本章中我们将仅在初级水平上为其中一些问题提供答案，有些专题相当高级，并且在数学和统计学上需要比本书所假设的更多的背景知识。各有关章末注所引用的参考文献可以提供更多的细节。

开始学习定性响应模型之前，我们首先考虑二分响应回归模型。这里有三种方法可以为二分响应变量建立一个概率模型：

1. **线性概率**模型（LPM）
2. **logit** 模型
3. **probit** 模型

由于 LPM 相对简单而且能用 OLS 进行估计，因此我们将首先考虑 LPM，另外两个模型将在随后的部分中讨论。

§15.2 线性概率模型

为了建立概念，考虑下面的回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{15.2.1}$$

其中 X = 家庭收入；$Y = 1$，如果该家庭拥有住宅；$Y = 0$，如果该家庭不拥有住宅。

模型（15.2.1）看似一个典型的线性回归模型，但由于回归子是二值的，或二分的，因此这个模型被称为**线性概率模型**（LPM）。这是因为 Y_i 在给定 X_i 下的条件期望 $E(Y_i|X_i)$ 可解释为在给定 X_i 下事件（家庭拥有住
583 宅）将发生的条件概率，即 $\Pr(Y_i = 1|X_i)$。这样在我们这个例子里，$E(Y_i|X_i)$ 给出了一个家庭拥有自己的住宅且其收入是某给定的数额 X_i 的概率。

把（15.2.1）这样的模型命名为 LPM 的理由可从下面看出：假定 $E(u_i) = 0$，如同平常那样（为了得到无偏估计量），我们得到：

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i \tag{15.2.2}$$

现在，令 P_i = "$Y_i = 1$"（即事件发生）的概率，而 $1 - P_i$ = "$Y_i = 0$"（即事件不发生）的概率，则变量 Y_i 有如下的（概率）分布：

Y_i	概率
0	$1 - P_i$
1	P_i
总和	1

即 Y_i 遵从**贝努里**（Bernoulli）**概率分布**。

因此，由数学期望的定义，我们有：

$$E(Y_i)=0(1-P_i)+1(P_i)=P_i \tag{15.2.3}$$

比较（15.2.2）和（15.2.3），我们得到：

$$E(Y_i \mid X_i)=\beta_1+\beta_2 X_i=P_i \tag{15.2.4}$$

即模型（15.2.1）的条件期望事实上可以用的条件概率来解释。通常，一个贝努里随机变量的期望是随机变量等于 1 的概率。从上面可以知道，如果有 n 次独立试验，每次成功的概率为 np，失败的概率为（$1-p$），而且 X 代表成功的次数，那么我们就称 X 服从二项式分布。二项式分布的均值为 np，方差为 $np(1-p)$。成功这个词已在行文中加以定义了。

既然概率必须介于 0 和 1 之间，我们就有了一个约束条件：

$$0 \leqslant E(Y_i \mid X_i) \leqslant 1 \tag{15.2.5}$$

即条件期望（或条件概率）必须介于 0 和 1 之间。

从前面的讨论来看，OLS 似乎容易适用于二分因变量回归模型。那么，这里可能就没有什么新东西了。不幸的是，情况并非如此，因为 LPM 提出
584 了如下几个问题：

干扰项 u_i 的非正态性

虽然 OLS 并不要求干扰项 u_i 一定是正态分布的，但为了统计推断的目的，我们仍然假定这些干扰服从正态分布。[3]然而，由于干扰项 u_i 和 Y_i 一样，在 LMP 中只取两个值，u_i 的正态性假定便不成立。为了看清这点，我们把（15.2.1）写成：

$$u_i=Y_i-\beta_1-\beta_2 X_i \tag{15.2.6}$$

u_i 的概率分布为

	u_i	概率
当 $Y_i=1$ 时	$1-\beta_1-\beta_2 X_i$	P_i
当 $Y_i=0$ 时	$-\beta_1-\beta_2 X_i$	$(1-P_i)$

（15.2.7）

显然，我们不可能假定 u_i 是正态分布的；它们遵循贝努里分布。

但是不能进行正态性假定也许并不像它所表现的那么重要，因为我们知道 OLS 的点估计仍然保持无偏性（回忆一下，如果我们的目的是点估计，正态性假定就无关重要）。此外，当样本无限增大时，统计理论表明 OLS 估计量一般都趋于正态分布。[4]因此，在大样本中，LPM 的统计推断可沿用正态性假定下的通常 OLS 过程。

干扰项 u_i 的异方差性

即使 $E(u_i)=0$ 和 $\operatorname{cov}(u_i,u_j)=0$ 对 $i\neq j$（即无序列相关性），我们却不能声称 LPM 中的干扰项 u_i 是同方差性的。然而，这并不令人吃惊。正如统
585 计理论所表明的，对于一个贝努里分布，理论均值和方差分别为 p 和 $p(1-p)$，其中 p 是成功（即事件发生）的概率，从而表明方差是均值的函数。因此误差是异方差性的。

对于（15.2.7）所给的误差项的分布，运用方差的定义，读者应该可以证明（见习题 15.10）：

$$\operatorname{var}(u_i)=P_i(1-P_i) \tag{15.2.8}$$

即 LPM 中误差项是异方差性的。既然 $P_i=E(Y_i\mid X_i)=\beta_1+\beta_2X_i$，$u_i$ 的方差最终依赖于 X，从而它不是同方差性的。

我们知道，当异方差出现时，OLS 估计虽然无偏，却不是有效的，就是说，OLS 估计量不具有最小方差性质。然而，正如非线性问题一样，异方差性问题也不是一种不能克服的障碍。在第 11 章中，我们讨论过处理异方差性问题的几种方法。由于 u_i 的方差依赖于$E(Y_i\mid X_i)$，解决异方差性问题的一个方法是进行数据变换，将模型（15.2.1）的两边除以：

$$\sqrt{E(Y_i\mid X_i)[1-E(Y_i\mid X_i)]}=\sqrt{P_i(1-P_i)}=\text{say}\sqrt{W_i}$$

即，

$$\frac{Y_i}{\sqrt{w_i}}=\frac{\beta_i}{\sqrt{w_i}}+\beta_2\frac{X_i}{\sqrt{w_i}}+\frac{u_i}{\sqrt{w_i}} \tag{15.2.9}$$

你可以容易地证明，（15.2.9）中被变换的误差项是同方差的。因此，在估计（15.2.1）之后，我们现在可以用 OLS 对（15.2.9）进行估计了。而 OLS 只是 w_i 以为权数的加权最小二乘（WLS）。

在理论上，我们刚才所描述的情景是美好的。但是在实际中，真实的 $E(Y_i\mid X_i)$是未知的，因此权重 w_i 也是未知的。为了估计 w_i，我们可以用下面的二步法[5]：

步骤 1. 对（15.2.1）进行 OLS 回归，暂时撇开异方差性问题。于是得到 $\hat{Y}_i=$ 真 $E(Y_i\mid x_i)$的估计值。再由此求 w_i 的估计值$\hat{w}_i=\hat{Y}_i(1-\hat{Y}_i)$。

586 **步骤 2.** 用估计值 $\hat{w}_i$ 去做如同（15.2.9）的数据变换，然后用 OLS（即加权最小二乘）估计变换后的方程。

我们马上将在例子中对这一方法进行解释。但我们首先需要讨论 LPM 中的另一个问题。

0 $E(Y, X)$ 1 不被满足

由于线性概率模型中的 $E(Y_i|X)$ 度量着在给定 X 下事件 Y 发生的条件概率，它必须落在 0 与 1 之间。虽然先验上这是正确的，但无法保证 $E(Y_i|X_i)$ 的估计量 $\hat{Y}_i$ 一定能满足这一约束条件。这是 LPM 的 OLS 估计的真正问题所在。有两种解决问题的方法。一种是用平常的 OLS 方法估计 LPM，看估计的 i 是否位于 0 与 1 之间，如果有些 i 小于 0（即是负的），则取其为零；如有某些大于 1，则取其为 1。另一种方法是设计一种估计技术，以保证所估的条件概率 i 必定落在 0 与 1 之间，稍后讨论的对数单位和概率单位模型将能保证所估概率确实落在 0 到 1 这个逻辑界限内。

可疑的拟合优度：R^2 值

在二分响应模型中，习惯计算的 R^2 的价值是有限的。为看出其中道理，考虑以下图形，对于给定的 X，Y 不是 0 就是 1。因此，所有的 Y 值必定要么落在 X 轴上，要么落在 $Y=1$ 的一条直线上。因此，一般地说，不能期望有任何 LPM 能很好地拟合这样的散点，不管是无约束 LPM（图 15.1a），抑或是断尾或受约束 LPM（图 15.1b）。后者指用一种限制 i 不超越逻辑带域“0－1”的方法去估计 LPM。结果是对这样的模型，按惯例算出的 R^2 很可能比 1 小很多。在大多数实例中，R^2 介于 0.2 与 0.6 之间。对这种模型，只有当实际的散点非常密集在点 A 和 B 处时（图 15.1c），R^2 才会高，比方说高于 0.8，因为这时容易通过 A 和 B 两点的连接而把直线的位置固定下来。这时，预测的 Y_i 值将非常靠近 0 或 1。

为了这些缘故，奥尔德里奇（John Aldrich）和纳尔逊（Forrest Nelson）争辩说：“在有定性应因量的模型中应避免使用判定系数作为一种摘要统计量。”[6]

LPM：一个数值例子

我们用一个数值例子来说明前节中关于 LPM 的一些问题。表 15.1 给出 40 个家庭的住宅所有权 Y（1＝拥有住宅，0＝不拥有住宅）和家庭收入 X（千美元）的虚构数据。根据这些数据，用 OLS 估计的 LPM 如下：

$$\hat{Y}_i = -0.945\,7 + 0.102\,1X_i$$

$$(0.122\,8) \quad (0.008\,2) \qquad (15.2.10)$$

$$t = (-7.698\,4) \quad (12.515) \qquad R^2 = 0.804\,8$$

表 15.1　拥有住宅（$Y=1$ 如果拥有住宅，否则为 0）以及收入 X（千美元）的假想数据

家庭	Y	X	家庭	Y	X
1	0	8	21	1	22
2	1	16	22	1	16
3	1	18	23	0	12
4	0	11	24	0	11
5	0	12	25	1	16
6	1	19	26	0	11
7	1	20	27	1	20
8	0	13	28	1	18
9	0	9	29	0	11
10	0	10	30	0	10
11	1	17	31	1	17
12	1	18	32	0	13
13	0	14	33	1	21
14	1	20	34	1	20
15	0	6	35	0	11
16	1	19	36	0	8
17	1	16	37	1	17
18	0	10	38	1	16
19	0	8	39	0	7
20	1	18	40	1	17

首先我们来解释这一回归。截距值 $-0.945\,7$ 给出零收入的家庭拥有自己的住房的“概率”。由于此值是负的，而概率又不可能是负值，我们就把该值当作零看待，这样做在本例中是说得过去的。[7] 斜率值 0.102 1 意味着收入每增加 1 单位（本例中是 1 000 美元），平均拥有住宅的概率增加 0.102 1或约 10%。当然，对某一给定的收入水平，我们可从（15.2.10）估计出拥有住宅的实际概率。例如，对于 $X=12$(12 000 美元)，估计拥有住宅的概率是：

$$(\hat{Y}_i \mid X=12) = -0.945\,7+12(0.102\,1)$$
$$= \quad 0.279\,5$$

就是说，收入为 12 000 美元的家庭拥有住宅的概率约为 28%，表 15.2 展示了对应于表中所列各种收入水平的估计概率，该表中最值得注意的特点是有 6 个估计值为负值，并有 6 个值大于 1。这清楚地表明了，尽管 $E(Y_i \mid X)$ 是正的且小于 1，而其估计值 $\hat{Y}_i$ 却不一定是正的或小于 1。这就是为什么当因变量是二分变量时不宜使用 LPM 的一个理由。

即使所估计的 Y_i 全部是正值且小于 1，LPM 仍受异方差性问题的困扰，这可从（15.2.8）容易看出。这样一来，我们就不能信赖（15.12.10）

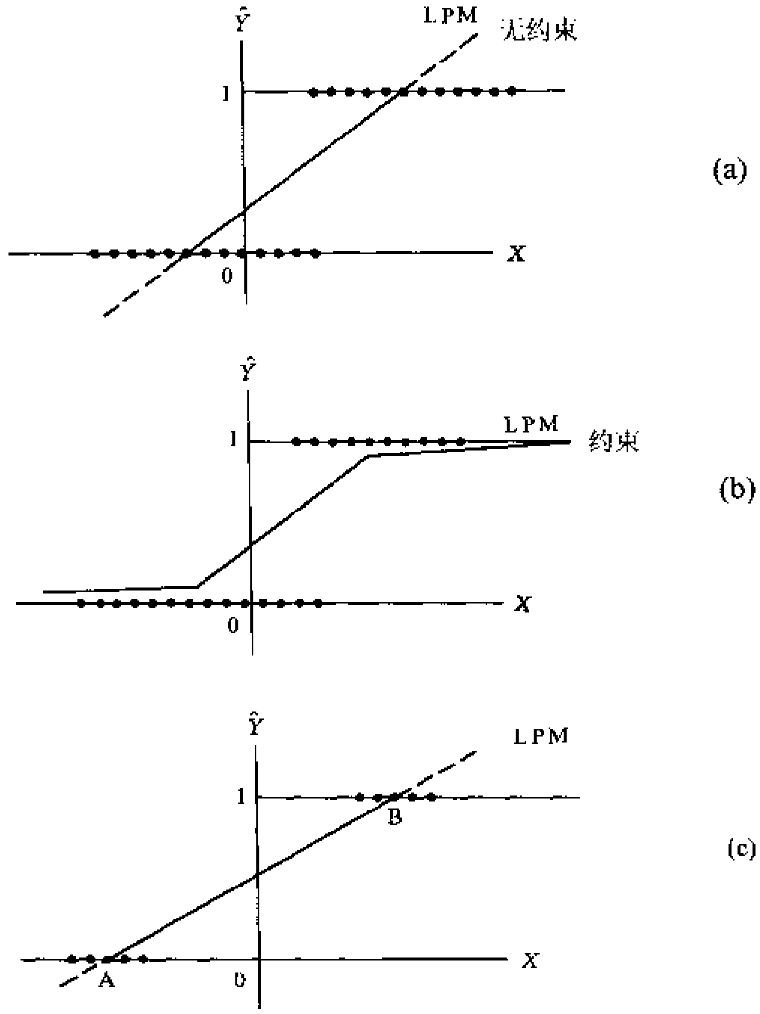


图 15.1　线性概率模型

所报道的估计标准误。(为什么?) 但是我们可用先前讨论过的加权最小二乘法以获得这些标准误的更为有效的估计值。应用WLS时所必需的权重 $\hat{w}_i$ 也列在表 15.2 中。但注意，由于某些 Y_i 是负的，某些 Y_i 大于 1，对于这些 Y_i 来说，$\hat{w}_i$ 将是负的。因此，我们在 WLS 中不能使用这些观测值（为什么?)，从而在本例中把观测值的个数从 40 减少到 28。[8] 删去这些观测值的 WLS 回归将是：

$$\frac{\hat{Y}_i}{\sqrt{w_i}} = -1.245\,6\frac{1}{\sqrt{w_i}} + 0.119\,6\frac{X_i}{\sqrt{w_i}} \tag{15.2.11}$$

$$(0.120\,6) \qquad (0.006\,9)$$

$$t = (-10.332) \qquad (17.454) \qquad R^2 = 0.921\,4$$

这些结果表明，和（15.12.10）相比，估计的标准误变小了，从而估计的 t 比率（在绝对值上）变大了。然而，我们承认这一结果不免有些难处，因为在估计（15.12.11）时，我们被迫放弃了 12 个观测值。而且，由于 w_i 是估

计值，严格地说，通常的统计假设检验程序仅在大样本中有效（参看第11章）。

表 15.2　　住宅所有权例子的实际 Y、估计 Y 以及权重 w_i

Y_i	$\hat{Y}_i$	$\hat{w}_i^{\ddagger}$	$\sqrt{\hat{w}_i}$	Y_i	$\hat{Y}_i$	$\hat{w}_i^{\ddagger}$	$\sqrt{\hat{w}_i}$
0	-0.129*			1	1.301+		
1	0.688	0.214 6	0.463 3	1	0.688	0.214 7	0.463 3
1	0.893	0.095 6	0.309 1	0	0.280	0.201 6	0.499 0
0	0.178	0.146 3	0.382 5	0	0.178	0.146 3	0.382 5
0	0.280	0.201 6	0.449 0	1	0.688	0.214 7	0.463 3
1	0.995	0.004 98	0.070 5	0	0.178	0.146 3	0.382 5
1	1.098+			1	1.097+		
0	0.382	0.236 1	0.485 9	1	0.893	0.095 6	0.309 1
0	-0.026 5*			0	0.178	0.146 3	0.382 5
0	0.076	0.070 2	0.265 0	0	0.076	0.070 2	0.265 0
1	0.791	0.165 3	0.406 6	1	0.791	0.165 3	0.405 5
1	0.893	0.095 6	0.309 1	0	0.382	0.236 1	0.485 9
0	0.484	0.249 7	0.499 7	1	1.199+		
1	1.097+			1	1.097+		
0	-0.333*			0	0.178	0.146 3	0.382 5
1	0.995	0.004 98	0.070 5	0	-0.129*		
1	0.688	0.214 7	0.463 3	1	0.791	0.165 3	0.406 6
0	0.076	0.070 2	0.265 0	1	0.688	0.214 7	0.463 3
0	-0.129*			0	-0.231*		
1	0.893	0.095 6	0.309 1	1	0.791	0.165 3	0.406 6

* 当作零，以避免负的概率。

\+ 当作1，以避免概率值大于1。

‡ $\hat{Y}_i(1-\hat{Y}_i)$。

§ 15.3　LPM 的应用

在有了方便的计算机软件包可用来估计（即将讨论的）对数单位和概率单位模型之前，LPM由于它的简单性，曾被相当广泛的使用。现在我们来阐明它的一些应用。

590 **例15.1**

科恩-雷-勒曼（Cohen-Rea-Lerman）研究[9]

在为美国劳工部做的一项研究工作中，科恩（Cohen）、雷（Rea）和勒曼（Lerman）意欲把各类劳工的“劳动力参与”当作一些社会经济一人口统计变量的函数来分析。在所有的回归中因变量都是一个虚拟变量：如果一个人参与劳动队伍，就取值1；如果他或她不参与，就取值0。在表15.3中我们复制了几个虚拟应变量回归中的一个。

在解释计算结果前，注意这些特点：上述回归是用OLS估计的。为了对异方差性做校正，作者们曾在他们的某些回归中使用前述的两步法，但发现这样得到的标准误和未经异方差校正的没多大区别。这也许纯粹由于样本较大，即约为25 000所致。由于样本含量很大，即使误差项取的是二分值，所估的 t 值仍可用于检验这个通常OLS程序中的统计显著性。R^2 的估计值0.175看来相当低，但鉴于这是个大样本结果，根据8.5节所给的 F 检验，此 R^2 仍是显著的。最后，注意作者是怎样把定量和定性变量融合在一起以及他们是怎样考虑交互作用的。

转到对所得结果的解释，我们看到每一斜率系数都给出对应于解释变量的一个给定单位变化，事件发生的条件概率的变化率。比如说，附着于变量“65岁及以上”的系数 $-0.275\,3$ 表示在保持其他因素不变情况下，该年龄组的妇女参与劳动（力）队伍的概率（同年龄为22～54岁的基底类别妇女相比）要低出约27%。类似地，附着于变量“受教育年数在16年及以上”的系数0.306 1表示，在保持其他因素不变的情况下，受这样多教育的妇女参与劳动队伍的概率（同基底类别，即受教育年数少于5年的妇女相比）要高出约31%。

现在考虑婚姻状况和年龄的**交互作用项**。表中数据表明，从未结婚的女人（和基底类相比）的劳动力参与概率要高出差不多29%，而年龄为65岁及以上的妇女（仍同基底类相比），劳动力参与概率则要低出约28%。但从未结婚且年龄为65岁或以上的妇女的劳动力参与概率和基底类相比，却低出20%。这意味着年龄为65岁及以上且从未结婚的女人很可能比65岁及以上而属于已婚或其他类的妇女更多地参与劳动力。

仿照以上的程序，读者不难解释表15.3中的其余系数。从给定的这些信息中还容易求得各种类别的劳动参与的条件概率。比方说，如果我们想求已婚（其他）、年龄为22～54岁、受教育年数为12～15年、失业率为2.5%～3.4%、就业改变为3.5%～6.49%、相对就业机会为74%及以上以及FILOW为7 500美元及以上的（参与）概率就得到：

$$0.436\,8+0.152\,3+0.223\,1-0.021\,3+0.030\,1+0.057\,1-0.245\,5 \\ =0.632\,6$$

换句话说，有上述特征的妇女，其劳动力参与概率估计约为63%。

表 15.3 劳动力参与

居住在最大的 96 个标准大都市统计区（SMSA）的 22 岁及以上的妇女的回归（因变量：在 1966 年参与或不参与劳动力队伍）

解释变量	系数	t 比率
常数	0.436 8	15.4
婚姻状况		
已婚，配偶存在	—	—
已婚，其他	0.152 3	13.8
从未结婚	0.291 5	22.0
年龄		
22～54 岁	—	—
55～64 岁	−0.059 4	−5.7
65 岁及以上	−0.275 3	−9.0
受教育年数		
0～4	—	—
5～8	0.125 5	5.8
9～11	0.170 4	7.9
12～15	0.223 1	10.6
16 年及以上	0.306 1	−13.3
失业率（1966），%		
低于 2.5	—	—
2.5～3.4	−0.021 3	−1.6
3.5～4.0	−0.026 9	−2.0
4.1～5.0	−0.029 1	−2.2
5.1 及以上	−0.031 1	−2.4
就业变化（1965～1966），%		
低于 3.5	—	—
3.5～6.49	0.030 1	3.2
6.5 及以上	0.052 9	5.1
相对就业机会，%		
低于 62	—	—
62～73.9	0.038 1	3.2
74 及以上	0.0571	3.2
FILOW，美元		
低于 1 500 及负值	—	—
1 500～7 499	−0.145 1	−15.4
7 500 及以上	−0.245 5	−24.4

交互作用（婚姻状况及年龄）			
婚姻状况	年龄		
其他	55～64 岁	−0.040 6	−2.1
其他	65 岁及以上	−0.139 1	−7.4
从未结婚	55～64 岁	−0.110 4	−3.3
从未结婚	65 岁及以上	−0.204 5	−6.4
交互作用（年龄与完成的受教育年数）			
年龄	受教育年数		
65 岁及以上	5～8	−0.088 5	−2.8
65 岁及以上	9～11	−0.084 8	−2.4
65 岁及以上	12～15	−0.128 8	−4.0
65 岁及以上	16 年及以上	−0.162 8	−3.6
	$R^2=0.175$		
观测值的个数＝25 153			

注：—表示基底或省略类别。

FILOW：家庭收入减去本人工薪收入。

资料来源：Malcolm S. Cohen, Samuel A. Rea, Jr., and Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, U.S. Department of Labor, 1970, Table F 6, pp.212 213.

592

例 15.2　对债券评级的预测

根据 1961—1966 年期间 200 种 Aa（优质）和 Baa（中等质量）债券的时间序列与横截面混合数据，约瑟夫·卡佩莱利（Joseph Cappelleri）估计了如下的债券评级预测模型。[10]

$$Y_i=\beta_1+\beta_2X_{2i}^2+\beta_3X_{3i}+\beta_4X_{4i}+\beta_5X_{5i}+u_i$$

其中 Y_i ＝1，如果债券评级为 Aa 穆迪评级）

＝0，如果债券评级为 Baa 穆迪评级）

X_2 ＝负债资本比率，杠杆作用的一种衡量

$$=\frac{\text{长期债券的美元价值}}{\text{总资产的美元价值}}\times 100$$

X_3 ＝利润率

$$=\frac{\text{税后收入的美元价值}}{\text{净总资产的美元价值}}\times 100$$

X_4＝利润率的标准差，利润率变异性的一种度量

X_5＝净总资产（千美元），规模的一种度量

可以先验地预期 β_2 和 β_4 是负的（为什么？），而 β_3 和 β_5 是正的。

在对异方差性和一阶自相关做校正后，卡佩莱利得到以下结果[11]：

$$\begin{aligned}\hat{Y}_i=\ &0.686\,0+0.017\,9X_{2i}^2+0.048\,6X_{3i}+0.057\,2X_{4i}+0.378(E-7)X_5\\ &(0.177\,5)(0.002\,4)\quad(0.048\,6)\quad(0.017\,8)\quad(0.039)(E-8)\end{aligned}\tag{15.3.1}$$

$R^2=\ 0.693\,3$

注：0.378　E－7 表示 0.000 000 037 8，下同。

除 X_4 的系数外，所有系数都有正确的符号。为什么利润率的变异性有一个正的系数，我们让金融专业的学生去寻觅其中的道理。因为人们会预料，其他条件不变，利润的变异性越大，得到 Aa 的穆迪评级的可能性越小。

对回归的解释是直截了当的。例如，附着于 X_3 的 0.048 6 表示其他条件相同时，利润率每增加一个百分点，平均而言，将导致债券获得 Aa 评级的概率增大约 0.05。类似地，平方杠杆比率每提高一个单位，债券被划为 Aa 等级的概率将降低 0.02。

593 **例15.3 预测债券违约**

为了预测市政府对市政债券违约，鲁宾费尔德（Daniel Rubinfeld）曾研究 1930 年马萨诸塞州的 35 个市政府的一个样本，其中确实有几宗违约的。他选用 LPM 模型并估计如下[12]：

$$\hat{P} = \underset{(0.29)}{1.96} - \underset{(0.009)}{0.029}\ \text{TAX} - \underset{(2.13)}{4.86}\ \text{INT} + \underset{(0.028)}{0.063}\ \text{AV} + \underset{(0.003)}{0.007}\ \text{DAV} - \underset{(0.88)}{0.48}\ \text{WELF} \tag{15.3.2}$$

$$R^2 = 0.36$$

其中 $P=0$，如果市政府违约，否则等于 1；TAX = 1929、1930、1931 年 3 年的平均税率；INT = 1930 年分摊做利息支付的当年预算的百分比；AV = 1925—1930 年财产估价的百分比增长；DAV = 1930 年直接净负债总额对财产估价总额的比率；及 WELF = 1930 年预算中分摊做慈善事业、养老金和士兵福利的百分比。

(15.3.2) 的含义也是相当清楚的。例如，当其他情况不变时，税率每增加千分之一，违约的概率将提高 0.03 或 3%。R^2 值相当低，但如前面看到的，一般地说，R^2 总是比较低而且在判断模型的拟合优度中作用是有限的。

§ 15.4 LPM 以外的其他方法

如我们已看到的，LPM 受到一些问题的困扰，诸如 (1) u_i 的非正态性，(2) u_i 的异方差性，(3) $\hat{Y}_i$ 落在 0－1 区域的范围之外，以及 (4) R^2 值一般地说比较低。但这些困难是可以克服的。例如，我们可用 WLS 去解决异方差性问题或增大样本含量以减轻非正态性问题。通过受约束最小二乘法或数学规划技术，还可迫使所估的概率落入 0－1 区间。

但即使这样做，LPM 的根本问题还在于其在逻辑上不是一个很有吸引力的模型，因为它假定了 $P_i = E(Y=1 \mid X)$ 随 X 而线性地增加，即 X 的边际或增补效应一直保持不变。例如，在住房所有权一例中，我们求出，X 每增加一单位（1 000 美元），拥有住房的概率一律增加 0.10，且不问收入水
594 平是 8 000 美元，10 000 美元，18 000 美元还是 22 000 美元。这显然是不现

实的。事实上，人们会预料对有非线性关系：收入很低的家庭将不会拥有一所住房，但收入充分高，比如说超过 X^* 的家庭很可能拥有自己的住房。超过 X^* 的任何收入增加将不会对拥有房子的概率有什么影响。因此，在收入分布的两端，X 的一个小小的增加实质上将不影响拥有住房的概率。

因此，我们所需要的是具有如下二分性质的一个（概率）模型：(1) 随着 X_i 增加，$P_i = E(Y=1 \mid X)$ 也增加，但永远不超出 0－1 这个区间；(2) P_i 和 X_i 之间的关系是非线性的，即“随着 X_i 变小，概率趋于零的速度越来越慢，而随着 X_i 变得很大，概率趋于 1 的速度也越来越慢”[13]。

从几何图形看，我们所要的（概率）模型有点像图 15.2 那样。注意，在此模型中，概率位于 0 与 1 之间并且随着 X 非线性地变化。

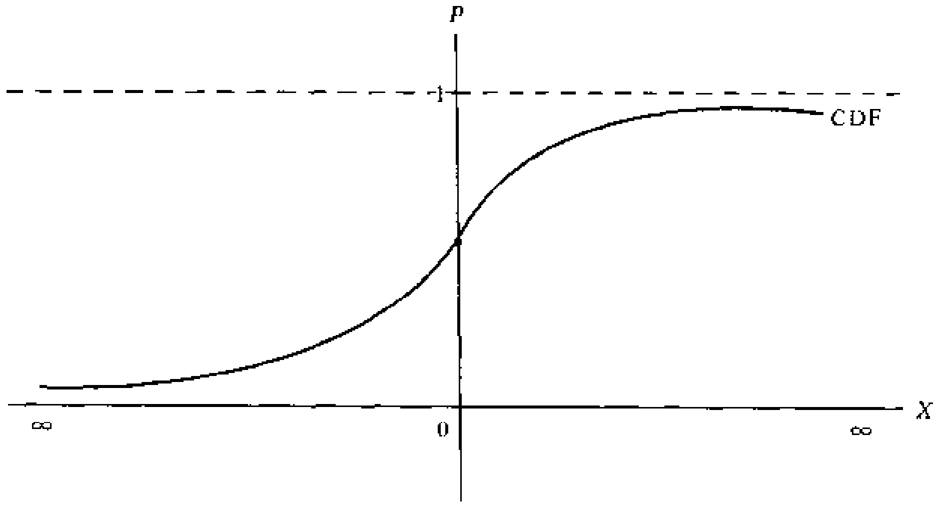


图 15.2　一个累积分布函数

读者必将发觉图中的 S 形曲线很像是一个随机变量的**累积分布函数**（CDF）。[14]因此，当回归中的响应变量是取 0－1 值的二分变量时，容易用 CDF 去建立回归模型。现在的实际问题是用哪一个 CDF？因为，尽管所有
595 的 CDF 都是 S 形，但对每一随机变量有惟一的一个 CDF。由于历史和实际两方面的原因，通常选择用以代表 0－1 响应模型的 CDF 是 (1) 逻辑和 (2) 正态两模型，前者给出 **logit** 模型，而后者给出 **probit**（或 **normit**）模型。

虽然对对数单位或概率单位模型做详尽的讨论会超出本书的范围，我们将多少非正式地去说明怎样估计并怎样解释这些模型。

§ 15.5　LOGIT 模型

我们继续用住房所有权的例子说明对数单位模型的基本概念。回顾在解释住房所有权对收入的线性关系时，LPM 曾是：

$$P_i = E(Y=1 \mid X_i) = \beta_1 + \beta_2 X_i \tag{15.5.1}$$

其中 X 为收入，而 $Y=1$ 表示家庭拥有住房，但现在考虑如下的住宅所有权表达式：

$$P_i = E(Y=1 \mid X_i) = \frac{1}{1+e^{-(\beta_1+\beta_2 X_i)}} \tag{15.5.2}$$

为了易于阐明，我们把（15.5.2）写成：

$$P_i = \frac{1}{1+e^{-Z_i}} = \frac{e^z}{1+e^z} \tag{15.5.3}$$

其中，$Z_i = \beta_1 + \beta_2 X_i$。

方程（15.5.3）代表一个以（累积）**逻辑分布函数**（logistic distribution function）为名的模型。[15]

容易证实，随着从 $-\infty$ 变到 $+\infty$，P_i 从 0 变到 1，而且 P_i 对 Z_i（从而对 X_i）有非线性关系，这样就满足了上述两点要求。[16]但看来似乎为了满足这些要求，却造成了估计问题。从（15.5.2）可以清楚地看出，P_i 不仅对 X 非线性，而且对诸 β 也是非线性的。这就意味着我们不能用熟知的 OLS 程序去估计参数。[17]不过这个问题的形式成分多于真实成分。因为，(15.5.2)是本质上线性的，对此可参看下面的说明。

596 如果拥有住房的概率 P_i 由（15.5.3）给出，则不拥有住房的概率 $(1-P_i)$是：

$$1-P_i = \frac{1}{1+e^{Z_i}} \tag{15.5.4}$$

因此，我们可以得到：

$$\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i} \tag{15.5.5}$$

现在 $P_i/(1-P_i)$ 就是拥有住房的**机会比率**（odds ratio）——一个家庭将拥有住房的概率对不拥有住房的概率之比。比方说，如果 $P_i=0.8$，则有利于拥有住房的机会是 4 比 1。

现在如果取（15.5.5）的自然对数，我们就得到一个非常有意思的结果，就是：

$$\begin{aligned} L_i &= \ln\left(\frac{P_i}{1-P_i}\right) = Z_i \\ &= \beta_1 + \beta_2 X_i \end{aligned} \tag{15.5.6}$$

即机会比率的对数 L_i 不仅对 X_i 为线性，而且（从估计的观点看）对参数也是线性的。[18] L 被称为 logit，从而像（15.5.6）这样的模型取名为 **logit 模型**。

注意 logit 模型的这些特点：

1. 随着 P 从 0 变到 1（就是，随着 Z 从 $-\infty$ 变到 $+\infty$），对数单位 L 从

$-\infty$变到$+\infty$。就是说，虽然概率（必须）落在0与1之间，但logit并不受此约束。

2. 虽然 L 对 X 为线性，但概率本身却不然。这一性质和概率随 X 而线性增大的LPM模型（15.5.1）形成了对比。[19]

3. 虽然在上述模型中仅含有单个 X 变量或回归元，但是我们可以根据所依据的理论，增加所需的回归元。

4. 若 L，即logit，是正的，这就意味着当回归元的值增加时，回归子等于1（意味着一些有利事件的发生）的机会也增大。若 L 为负，随着 X
597 值的增加，回归子等于1的机会将减小。换言之，当机会比率由1减到0时，logit会变负并且在幅度（即绝对值）上会越来越大；当机会比率由1增大至无穷时，logit为正并且越来越大。[20]

5.（15.5.6）中给出的logit模型较为正式的解释：斜率 β_2 给出 X 每单位变化的 L 的变化，就是说，它告知人们随着收入变化一个单位，比如1 000美元，有利于拥有住宅的对数—机会比率是怎样变化的。截距 β_1 是当收入为零时的有利于拥有住宅的对数—机会比率的值。像大多数的截距所做的解释那样，这种解释不一定有什么实在的意义。

6. 对给定的某个收入水平，比方说 X^*，我们其实想估计的并不是有利于拥有住房的机会比，而是拥有住房本身的概率。不过，一旦有了 $\beta_1+\beta_2$ 的估计值，这是容易直接从（15.5.3）中求出的。但这又提出了最重要的问题：首先是怎样估计 β_1 和 β_2? 答案将在下节中给出。

7. 好比LPM假定 P_i 和 X_i 有线性关系，logit模型假定机会比率的对数与 X_i 有线性关系。

§15.6 LOGIT模型的估计

为达到估计的目的，我们把（15.5.6）写成下式：

$$L_i=\ln\left(\frac{P_i}{1-P_i}\right)=\beta_1+\beta_2X_i+u_i \tag{15.6.1}$$

我们稍后将讨论随机干扰项的性质。

为了估计此模型，除 X_i 外我们还需要回归子（或logit）L_i 的数值。这取决于我们用于分析的数据类型。我们要区分两种数据类型：（1）个体（或微观）水平上的数据，和（2）群组数据或重复观测数据。

个体水平上的数据

如果像表15.1中的情况，我们拥有个体家庭的数据，那么（15.6.1）

的 OLS 估计就是不可行的。这很容易明白。就表 15.1 中的数据而言，$P_i=1$，若某家庭拥有住宅；$P_i=0$，若家庭不拥有住宅。但如果我们将这些值直接代入 logitL_i，就会得到：

$$L_i=\ln\left(\frac{1}{0}\right)\qquad \text{若家庭拥有住宅}$$

$$L_i=\ln\left(\frac{0}{1}\right)\qquad \text{若家庭不拥有住宅}$$

显然这些表达式是没有意义的。因此，如果拥有微观或个体水平上的数据，我们就无法运用标准 OLS 过程。在这种情况下，我们可能不得不求助于最
561 大似然方法对参数进行估计。尽管这个方法的初级部分在第 4 章的附录中讨论过，但考虑到有些读者愿意对此有更多的了解，我们将在附录 15A，15A.1 节中对其在目前上下文中的应用加以讨论。[21]诸如 Microfit、Eviews、Limdep、Shazam、PcGive 和 Minitab 这样的软件包对于估计个体水平上的 logit 模型有着固定的过程。在本章的后面我们将对 ML 方法的使用举例说明。

群组或重复观测数据

现在考虑一下表 15.4 所给出的数据。根据收入水平以及每个收入水平下拥有住宅的家庭个数，这张表给出了几个家庭的群组或重复观测数据。对应于每个收入水平 X_i，都有 N_i 个家庭，n_i 表示其中拥有住宅的家庭个数（$n_i \leqslant N_i$）。因此，如果我们计算：

$$\hat{P}_i=\frac{n_i}{N_i} \tag{15.6.2}$$

即相对频率，我们就能将它作为对应于每个 X_i 的真实 P_i 的一个估计值。如果相当大，$\hat{P}_i$ 将是 P_i 的良好估计值。[22]利用 $\hat{P}_i$，可以得到估计的 logit 如下：

$$\hat{L}_i=\ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)=\hat{\beta}_1+\hat{\beta}_2X_i \tag{15.6.3}$$

599 如果在每一 X_i 处的观测个数 N_i 都足够大，那么这将是真 logitL_i 的一个相当好的估计。

表 15.4　关于 X_i（收入）、N_i（收入为 X_i 的家庭个数）、和 n_i（N_i 中拥有住宅的家庭个数）的假想数据

X（千美元）	N_i	n_i
6	40	8
8	50	12
10	60	18

13	80	28
15	100	45
20	70	36
25	65	39
30	50	33
35	40	30
40	25	20

总之，给定了像表15.4那样的群组或重复观测数据，就能获得用于估计模型（15.6.1）的logit这个因变量的数据。但这时能对（15.6.3）应用OLS并按平常的方式估计参数吗？答案是肯定的，而又不完全如此，因为我们还没谈到随机干扰项的性质。可以证明，如果 N_i 相当大而且在一个给定收入组 X_i 中的每一次观测都可视为一个独立分布的二项式变量，那么，

$$u_i \sim N\left[0, \frac{1}{N_i P_i(1-P_i)}\right] \tag{15.6.4}$$

就是说，u_i 服从均值为零、方差为 $1/[N_i P_i(1-P_i)]$ 的正态分布。[23]

因此，如同LPM的情形那样，logit模型的干扰项也是异方差性的。这样一来，我们必须使用加权最小二乘法而不是普通最小二乘法。但是为了实际的目的，我们将用 $\hat{P}_i$ 代替未知的 P_i 并用：

$$\hat{\sigma}^2 = \frac{1}{N_i \hat{P}_i(1-\hat{P}_i)} \tag{15.6.5}$$

作为 σ^2 的估计量。

现在我们来叙述估计logit回归（15.6.1）的各个步骤：

1. 对每一收入水平 X_i，计算拥有住房的估计概率 $\hat{P}_i = n_i/N_i$。
2. 对每一 X_i 求logit[24]：

$$\hat{L}_i = \ln[\hat{P}_i/(1-\hat{P}_i)]$$

3. 为解决异方差性的问题，将（15.6.1）变换如下[25]：

$$\sqrt{w_i}L_i = \beta_1\sqrt{w_i} + \beta_2\sqrt{w_i}X_i + \sqrt{w_i}u_i \tag{15.6.6}$$

600 我们把它写为：

$$L_i^* = \beta_1\sqrt{w_i} + \beta_2 X_i^* + v_i \tag{15.6.7}$$

其中权重 $w_i = N_i\hat{P}_i(1-\hat{P}_i)$；$L_i^*$ = 变换的或加权的 L_i；X_i^* = 变换的或加权的 X_i；v_i = 变换的误差项。记住原始的误差方差是 $\hat{\sigma}_u^2 = 1/[N_i P_i(1-P_i)]$，容易验证变换后的误差项 v_i 是同方差性的。

4. 用OLS去估计（15.6.6），应记得WLS就是对于变换后数据的OLS。注意（15.6.6）不含截距项（为什么？）。因此还须用过原点回归程序估计（15.6.6）。

5. 按照平常的 OLS 方式建立置信区间和（或）检验假设。但要记住，严格地说，仅当样本合理地大时，所做的结论才是有效的。（为什么？）因此，对于小样本，如何解释估计的结果，需要审慎。

§ 15.7 群组 LOGIT 模型：一个数值例子

为了说明刚才讨论过的理论，我们将使用表 15.4 中的数据。既然表中的数据是群组数据，那么以这些数据为基础的 logit 模型被称为群组 logit 模型，简称 glogit。所需的原始数据以及应用 glogit 所必须的其他相关计算列在了表 15.5 中。以表 15.5 中的数据为基础的加权最小二乘回归（15.6.7）的计算结果如下：注意在（15.6.7）中没有截距项，因此过原点回归的程序在这里是适用的。

$$
\begin{aligned}
\hat{L}_i^* &= -1.594\,74\sqrt{w_i} + 0.078\,62X_i^* \\
se &= \quad (0.110\,46) \qquad (0.005\,39) \\
t &= (-14.436\,19) \qquad (14.566\,75) \qquad R^2 = 0.964\,2
\end{aligned}
\tag{15.7.1}
$$

R^2 是实测和估计的 L_i^* 之间的平方相关系数。L_i^* 和 X_i^* 是加权的 L_i 和 X_i，如（15.6.6）中所示。

Logit 模型的估计值的解释

我们该如何解释（15.7.1）？这里有不同的方法，一些是直观的，一些则不是。

Logit 的解释。如（15.7.1）所示，估计的斜率系数表明，当加权收入增加一个单位（1 000 美元），赞成拥有住宅的机会的加权对数值会上升 0.08 个单位。然而，这种机械的解释是不会很有吸引力的。

602 **机会比率的解释**。 记得 $L_i = \ln[P_i/(1-P_i)]$。因此，取估计的 logit 的反对数，我们将得到 $P_i/(1-P_i)$，即机会比率。这样，对（15.7.1）取反对数，得到：

$$
\begin{aligned}
\frac{\hat{P}_i}{1-\hat{P}_i} &= e^{-1.594\,74\sqrt{w_i} + 0.078\,62X_i^*} \\
&= e^{-1.594\,74\sqrt{w_i}} \cdot e^{0.078\,62X_i^*}
\end{aligned}
\tag{15.7.2}
$$

使用计算机你能够容易地算出 $e^{0.078\,62} = 1.081\,7$。这意味着加权收入每增加一个单位，同意拥有住宅的加权机会比率增加 1.081 7 或者大约增加 8.17%。**一般地说，如果你取第 j 个斜率系数的反对数，再从中减去 1 并乘以 100，你将得到对应于第 j 个回归元每增加 1 单位的机会比率的百分比变化。**

表 15.5　　估计所有权的对数单位模型的数据

X (千美元) (1)	N_i (2)	n_i (3)	$\hat{P}_i$ (4)=(3)÷(2)	$1-\hat{P}_i$ (5)	$\frac{\hat{P}_i}{1-\hat{P}_i}$ (6)	$\hat{L}_i=\ln(\frac{\hat{P}_i}{1-\hat{P}_i})$ (7)	$N_i\hat{P}_i(1-\hat{P}_i)=w_i$ (8)	$\sqrt{w_i}=\sqrt{N_i\hat{P}_i(1-\hat{P}_i)}$ (9)=$\sqrt{(8)}$	$\hat{L}_i^*=\hat{L}_i\sqrt{w_i}$ (10)=(7)(9)	$\hat{X}_i^*=X_i\sqrt{w_i}$ (11)=(1)(9)
6	40	8	0.20	0.80	0.25	−1.386 3	6.40	2.529 8	−3.507 1	15.178 8
8	50	12	0.24	0.76	0.32	−1.152 6	9.12	3.019 9	−3.480 7	24.159 2
10	60	18	0.30	0.70	0.43	−0.847 2	12.60	3.549 6	−3.007 2	35.496 0
13	80	28	0.35	0.65	0.54	−0.619 0	18.20	4.266 1	−2.640 7	55.459 3
15	100	45	0.45	0.55	0.82	−0.200 7	24.75	4.974 9	−0.998 5	74.623 5
20	70	36	0.51	0.49	1.04	0.040 0	17.49	4.182 5	0.167 3	83.650 6
25	65	39	0.60	0.40	1.50	0.405 4	15.60	3.949 7	1.601 2	98.742 5
30	50	33	0.66	0.34	1.94	0.663 3	11.20	3.349 6	2.221 8	100.488 0
35	40	30	0.75	0.25	3.0	1.098 6	7.50	2.738 6	3.008 6	95.840 5
40	25	20	0.80	0.20	4.0	1.386 3	4.00	2.000	2.772 6	80.000 0

附带提一下，如果想对没有加权的 logit 进行分析，你所要做的就是将 $\hat{L}_i^*$ 除以$\sqrt{w_i}$。表 15.6 给出了加权和没加权估计 logit 的每个观测值和其他一些数据，稍候我们将进行讨论。

概率的计算。由于一些人对 logit 和机会比率的语言不熟悉，所以我们通常计算出某个收入水平下拥有住宅的概率。假定我们要计算 $X=20$ (20 000美元) 下的这种概率。将 $X=20$ 代入 (15.7.1) 中，得到 $\hat{L}_i^*=-0.093\ 11$，然后除以 $\sqrt{w_i}=4.266\ 1$ (见表 15.5)，我们得到 $\hat{L}_i=-0.022\ 26$。因此，在收入水平为 20 000 美元时，有：

$$-0.022\ 26=\ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right)$$

表 15.6　　L^*、X^*、$\hat{L}_i^*$、概率和概率的变化[+]

L^*	X^*	$\hat{L}_i^*$	Logit	概率 $\hat{P}$	概率的变化[+]
−3.507 10	15.178 8	−2.840 96	−1.122 99	0.245 45	0.014 56
−3.480 70	24.159 20	−2.916 48	0.965 75	0.275 72	0.015 70
−3.480 70	35.496 00	−2.869 88	−0.808 50	0.308 21	0.016 76
−2.640 70	55.459 30	−2.442 93	−0.572 63	0.360 63	0.018 13
−0.998 50	74.623 50	−2.066 52	−0.415 38	0.397 62	0.018 83
0.167 30	83.650 60	−0.093 11	−0.022 26	0.494 43	0.019 65
1.601 20	98.742 50	1.464 72	0.379 84	0.591 66	0.018 99
2.221 18	100.488 00	2.558 96	0.763 96	0.682 21	0.017 04
3.008 60	95.840 50	3.167 94	1.156 77	0.760 74	0.014 31
2.772 60	80.000 00	3.100 38	1.550 19	0.824 94	0.011 35

L^* 和 X^* 来自于表 15.5。$\hat{L}_i^*$ 表示估计的 L^*。Logit 表示没加权的 logit。概率是拥有住宅的估计概率。概率的变化是收入每变化一单位的变化。

[+] 由 $\hat{\beta}_2\hat{P}(1-\hat{P})=0.078\ 62\hat{P}(1-\hat{P})$ 计算得到。

因此，

$$\frac{\hat{P}_i}{1-\hat{P}_i}=e^{-0.022\ 26}=1.022\ 5$$

解得：

$$\hat{P}_i=\frac{e^{-0.022\ 26}}{1+e^{-0.022\ 26}}$$

读者能够看到，概率的估计值为 0.494 4。也就是说，20 000 美元的工资条件下，某个家庭拥有住宅的概率大约为 49%。表 15.6 中显示了各种工资水平下计算出的概率。如表中所示，拥有住宅的概率随收入的增加而增加，但不是线性的，因为使用的不是 LPM 模型。

概率变化率的计算。正如你从表 15.6 中可以推断出的，拥有住宅的概率取决于收入水平。我们怎样才能计算出收入变化所引起的概率变化率呢？如章末注 19 所说，这不仅取决于估计斜率系数，而且还取决于测量概率的变化时所处的概率水平，后者当然取决于计算概率时的收入水平。

举个例子，假定我们要度量 20 000 美元的收入水平下拥有住宅概率的变化。于是，依据章末注 19，在水平 20（千美元）上收入每增加一单位，概率的变化是：$\hat{\beta}(1-\hat{P})\hat{P}=0.078\,62(0.505\,6)(0.494\,4)=0.019\,65$。

当收入水平为 40 000 美元时，概率的变化为 0.011 35。这作为一道习题留给读者。表 15.6 显示了每个收入水平下拥有住宅的概率变化，这些概率也被画在了图 15.3 中。

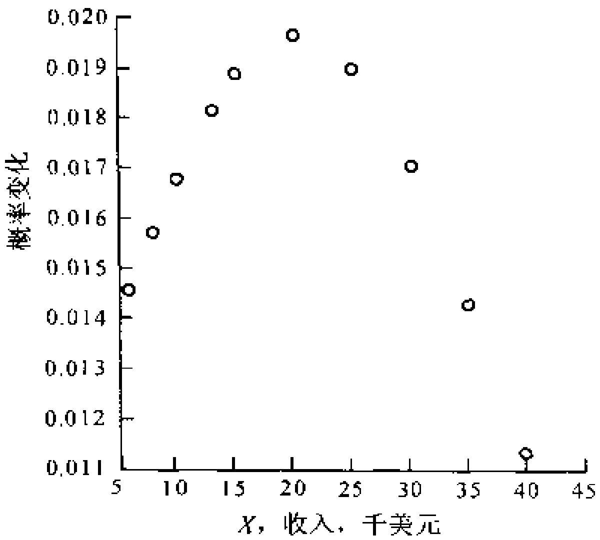


图 15.3　与收入有关的概率变化

604 为了总结 glogit 模型的讨论，以房屋的所有权为例，在 OLS 或者无加权回归的基础上我们给出了结果：

$$
\begin{aligned}
\hat{L}_i &= -1.658\,7+0.079\,2X_i \\
\text{se} &= \quad (0.095\,8)(0.004\,1) \qquad\qquad (15.7.3)\\
t &= (-17.32)\quad(19.11) \qquad r^2=0.978\,6
\end{aligned}
$$

这个回归与（15.7.1）所给出的加权最小二乘回归之间的区别，我们留给读者去比较。

§15.8　非群组或个体数据的 LOGIT 模型

为了做好准备，考虑表 15.7 中所给的数据。若某学生中级微观经济学的期末成绩为 A，则 $Y=1$；如果成绩为 B 或 C，则 $Y=0$。斯佩克特
605 （Specter）和马切奥（Mazzeo）使用平均学分绩（GPA）、TUCE、和个性化

教学系统（PSI）作为成绩的预测元。这里 logit 模型可以写成：

$$L_i = \left(\frac{\hat{P}_i}{1 - P_i}\right) = \beta_1 + \beta_2 \mathrm{GPA}_i + \beta_3 \mathrm{TUCE}_i + \beta_4 \mathrm{PSI}_i + u_i \qquad (15.8.1)$$

表 15.7　　关于个性化教学体系对课程成绩的影响数据

观测值	GPA 级	TUCE 级	PSI	评级	字母评级	观测值	GPA 级	TUCE 级	PSI	评级	字母评级
1	2.66	20	0	0	C	17	2.75	25	0	0	C
2	2.89	22	0	0	B	18	2.83	19	0	0	C
3	3.28	24	0	0	B	19	3.12	23	1	0	B
4	2.92	12	0	0	B	20	3.16	25	1	1	A
5	4.00	21	0	1	A	21	2.06	22	1	0	C
6	2.86	17	0	0	B	22	3.62	28	1	1	A
7	2.76	17	0	0	B	23	2.89	14	1	0	C
8	2.87	21	0	0	B	24	3.51	26	1	0	B
9	3.03	25	0	0	C	25	3.54	24	1	1	A
10	3.92	29		1	A	26	2.83	27	1	1	A
11	2.63	20	0	0	C	27	3.39	17	1	1	A
12	3.32	23	0	0	B	28	2.67	24	1	0	B
13	3.57	23	0	0	B	29	3.65	21	1	1	A
14	3.26	25	0	1	A	30	4.00	23	1	1	A
15	3.53	26	0	0	B	31	3.10	21	1	0	C
16	2.74	19	0	0	B	32	2.39	19	1	1	A

注：评级 Y = 1，如果期末成绩为 A

　　　　= 0，如果期末成绩为 B 或 C

TUCE = 为测试进来时的宏观经济学水平而举行的期初考试分数

PSI = 1，如果采用新的教学方法

　　= 0，如果不采用新的教学方法

GPA = 进来时的平均学分绩

资料来源：L. Spector and M. Mazzeo, "Probit Analysis and Economic Education," *Journal of Economic Education*, vol.11, 1980, pp.37 – 44.

如 15.6 节中所指出的那样，我们不能简单的令 $P_i=1$（若家庭拥有住宅）和 0（若家庭不拥有住宅）。这种情况下 OLS 和加权最小二乘法都是无能为力的。我们必须使用最大似然法的非线性估计过程。这个方法的细节将会在附录 15A 的 15A.1 节给出。因为大多数现代统计软件包通常都可以估计非群组数据的 logit 模型，所以我们将展出使用表 15.7 中数据的模型（15.8.1）的结果，并说明如何解释这些结果。数字结果以列表的形式写在表 15.8 中，而且是使用 Eviews 4 而得到的。在解释这些结果之前，先谈些一般性认识是适合的：

1. 由于我们使用的是最大似然法，它一般来说是一个大样本的方法，因此估计的标准误是渐近性的。

2. 我们使用（标准正态）Z 统计量而不用 t 统计量来对系数的统计显著性进行评估。从而，推断是以正态表为基础的。记住，如果样本非常大，t 分布收敛于正态分布。

3. 如前面提到的，对于二分回归子模型，拟合优度 R^2 的普通度量没有多大意义。类似 R^2 的所谓拟 R^2 有众多方法可供选择。[26] Eviews 提供了一种这样的方法，即 McFadden R^2，记作 R^2_{McF}。在我们的例子中，R^2_{McF}的值为 0.374 0。[27] 和 R^2 一样，R^2_{McF}也在 0 和 1 之间。另一个相对简单的拟合优度方法是计数 R^2。它的定义如下：

606

$$\text{计数 } R^2=\frac{\text{正确预测的次数}}{\text{总观测次数}} \tag{15.8.2}$$

表 15.8　(15.8.1) 的回归结果

因变量：成绩

方法：ML－二分 logit

5 次迭代后获得收敛

变量	系数	标准误	Z 统计量	概率
C	−13.021 3	4.931	−2.640 5	0.008 2
GPA	2.826 1	1.262 9	2.237 7	0.025 2
TUCE	0.095 1	0.141 5	0.672 23	0.501 4
PSI	2.378 6	1.064 5	2.234 5	0.025 5
McFadden $R^2=0.374\ 0$　　LR 统计量 (3 df) = 15.404 19				

由于 logit 模型中回归子取值为 1 或 0，如果所预测的概率大于 0.5，那么我们把它归类到 1，但如果它小于 0.5，则将其归类到 0。然后，我们数出正确预测次数并用（15.7.2）计算出 R^2。稍后我们将举例说明。

但是，需要注意的是，在二分回归子模型中拟合优度是次重要的。回归系数的期望符号以及它们统计上和（或）实际上的显著性才是首要的。

4. 为了检验所有斜率系数同时为零的虚拟假设，线性回归模型中 F 检验的等价物是**似然比**（LR）**统计量**。给定虚拟假设，LR 统计量服从自由度为解释变量个数的 χ^2 分布。本例中自由度为 3。（注：计算自由度时不包括截距项。）

现在我们来解释一下（15.8.1）中所给出的回归结果。等式中每个斜率系数都是一个偏斜率系数。它度量了某个回归元的值变动一个单位（其他回归元不变）所引起的 logit 估计值的变化。于是，GPA 系数值2.826 1意味着，其他变量保持不变，如果 GPA 增加一个单位，估计 logit 平均增加约 2.83 个单位，同时也表明两者之间正相关。如你所看到的，尽管 TUCE 的影响在统计上不显著，但其他所有回归元对 logit 均有正效应。这样，所有回归元一起对于期末成绩就有一个显著性的影响，如 LR 统计为 15.40，其 P 值大约是 0.001 5，相当的小。

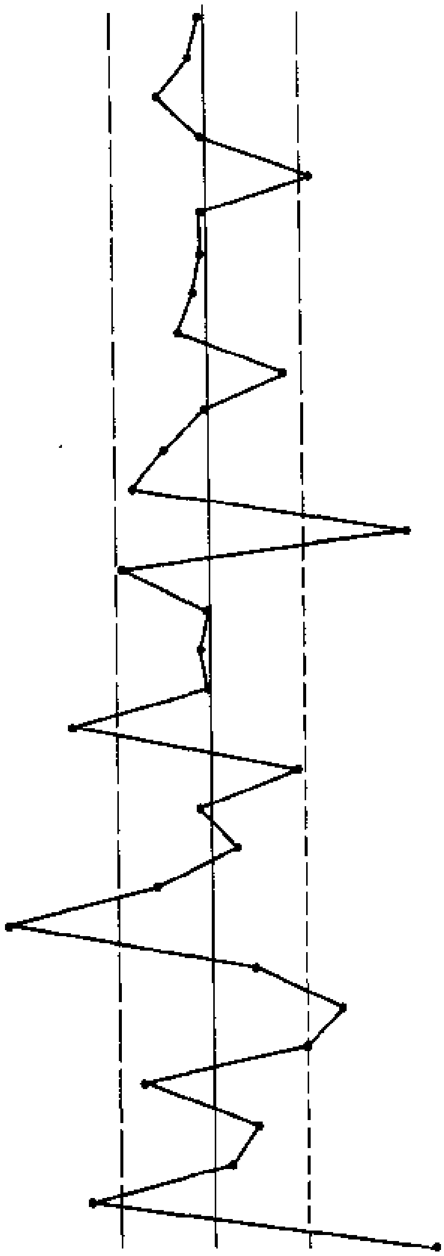
前面提到过，通过对各种斜率系数取反对数从而得到的机会比率，是一个更有意义的解释。那么，如果对 PSI 的系数 2.378 6 取反对数，你将得到 10.789 7（$\approx e^{2.378\,6}$）。这表明接受新的授课方法的学生得到 A 的可能性比没有接受的学生高出了 10 倍多，当然要其他回归元保持不变。

假如我们要计算出某个学生获得 A 的实际概率，考虑一下表 15.7 中的 10 号学生。将这个学生的实际数据代入到表 15.8 所给出的估计 logit 模型，
607 读者可以检查，这个学生的估计 logit 值为 0.817 8。使用等式（15.5.2），读者可以很容易地算出估计概率为 0.693 51。既然该生期末成绩实际上为 A，而且我们的 logit 模型对这个得到 A 的同学给以概率 1，那么估计概率 0.693 51 虽然不正好是 1，但接近 1。

回忆一下先前计数 R^2 的定义。表 15.9 给出了我们的说明性例子中回归子的实际值和估计值。从这张表中我们可以看出，32 次观测中有 6 个不正确的预测值（第 14、19、24、26、31 和 32 号学生）。因此，计数 R^2 的值为 $26/32=0.812\,5$，而 McFadden R^2 的值为 0.374 0。尽管这两个值没有直接的可比性，但是它们为我们提供了有关大小顺序的一些思路。此外，回归子为二分变量模型，拟合优度的重要性不可夸大。

表 15.9 **以表 15.8 中的回归为基础的实际值和拟合值**

观测	实际值	拟合值	残差	残差描点图
1	0	0.026 58	−0.026 58	
2	0	0.059 50	−0.059 50	
3	0	0.187 26	−0.187 26	
4	0	0.025 90	−0.025 90	
5	1	0.569 89	0.430 11	
6	0	0.034 86	−0.034 86	
7	0	0.026 50	−0.026 50	
8	0	0.051 56	−0.051 56	
9	0	0.111 13	−0.111 13	
10	1	0.693 51	0.306 49	
11	0	0.024 47	−0.024 47	
12	0	0.190 00	−0.190 00	
13	0	0.322 24	−0.322 24	
*14	1	0.193 21	0.806 79	
15	0	0.360 99	−0.360 99	
16	0	0.030 18	−0.030 18	
17	0	0.053 63	−0.053 63	
18	0	0.038 59	−0.038 59	
*19	0	0.589 87	−0.589 87	
20	1	0.660 79	0.339 21	
21	0	0.061 38	−0.061 38	
22	1	0.904 85	0.095 15	
23	0	0.241 77	−0.241 77	
*24	0	0.852 09	−0.852 09	
25	1	0.838 29	0.161 71	
*26	1	0.481 13	0.518 87	
27	1	0.635 42	0.364 58	
28	0	0.307 22	−0.307 22	
29	1	0.841 70	0.158 30	
30	1	0.945 34	0.054 66	
*31	0	0.529 12	−0.529 12	
*32	1	0.111 03	0.888 97	



* 指不正确的预测。

§ 15.9 Probit 模型

608 我们曾指出，为了解释二分因变量的行为，我们有必要使用适当选择的

CDF。logit 模型使用累积逻辑函数，如（15.5.2）所示。但这并非惟一可使用的 CDF。在某些应用中人们发现正态 CDF 是有用的。来自正态 CDF[28] 的估计模型通常称为 **probit 模型**，虽然有时也称 **normit 模型**。原则上，我们可以用正态 CDF 代替（15.5.2）中的逻辑 CDF 并按照第 16.5 节那样做。但我们将不采用这一路线，而是按照以麦克法登（McFadden）[29]所推出的以效用理论或行为理性选择剖视为依据的 probit 模型来陈述。

为了阐明使用概率单位模型的动机，假定在我们的住房所有权一例中，第 i 个家庭对是否拥有住房的决定，依赖于一种不可观测的效用指数（也被称为潜在变量）I_i，而后者又按照这样一种方式——指数 I_i 的值越大，家庭拥有住宅的概率就越大——取决于一个或几个解释变量，比方说收入 X_i。我们把指数 I_i 描述为：

$$I_i = \beta_1 + \beta_2 X_i \tag{15.9.1}$$

其中 X_i 表示第 i 个家庭的收入。

然而（不可观测的）I_i 是怎样同拥有住房的实际决定发生关系的呢？如前，令 $Y=1$，如果家庭拥有住房，否则 $Y=0$。现在，一个合理的假定是，对每一家庭都有这样一个指数临界值或门槛值，且记为 I_i^*：如果 I_i 超过 I_i^*，该家庭将拥有住房，否则不拥有。门槛值 I_i^* 和 I_i 一样是不可观测的，但如果我们假定每个都是有相同均值和方差的正态变量，就不但有可能估计（15.9.1）所给指数的参数，而且有可能获得有关不可观测的指数本身的某些信息。下面给出这方面的计算。

给定正态性假定，$I_i^* \leqslant I_i$ 的概率可由标准化正态 CDF 算出[30]：

$$P_i = P(Y=1 \mid X) = P(I_i^* \leqslant I_i) = P(Z_i \leqslant \beta_1 + \beta_2 X_i) = F(\beta_1 + \beta_2 X_i) \tag{15.9.2}$$

609 其中 $P(Y=1/X)$ 表示给定解释变量 X 的值时一个事件出现的概率；Z 是标准正态变量，即 $Z \sim N(0,\sigma^2)$。F 是标准正态 CDF，在此处可显示为：

$$\begin{aligned} F(I_i) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-z^2/2} \mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-z^2/2} \mathrm{d}z \end{aligned} \tag{15.9.3}$$

因为 P_i 代表事件发生的概率，在这里是拥有住宅的概率，故可由标准正态曲线下 $-\infty$ 到 I_i 所围的面积测出，如图 15.4a 所示。

现在为了获得关于效用指数 I_i 以及 β_1 和 β_2 的信息，我们取(15.9.2)的反函数，得到：

$$\begin{aligned} I_i &= F^{-1}(I_i) = F^{-1}(P_i) \\ &= \beta_1 + \beta_2 X_i \end{aligned} \tag{15.9.4}$$

其中 F^{-1} 是正态 CDF 的反函数。所有这一切的涵义均可从图 15.4 看清楚。在此图的 a 格，我们（从纵坐标）对给定的 $I_i^* \leqslant I_i$ 读出拥有住房的（累积）概率，而在 b 格里，我们（从横坐标）对给定的值 P_i 读出 I_i 值，后者不过

是前者的逆过程。

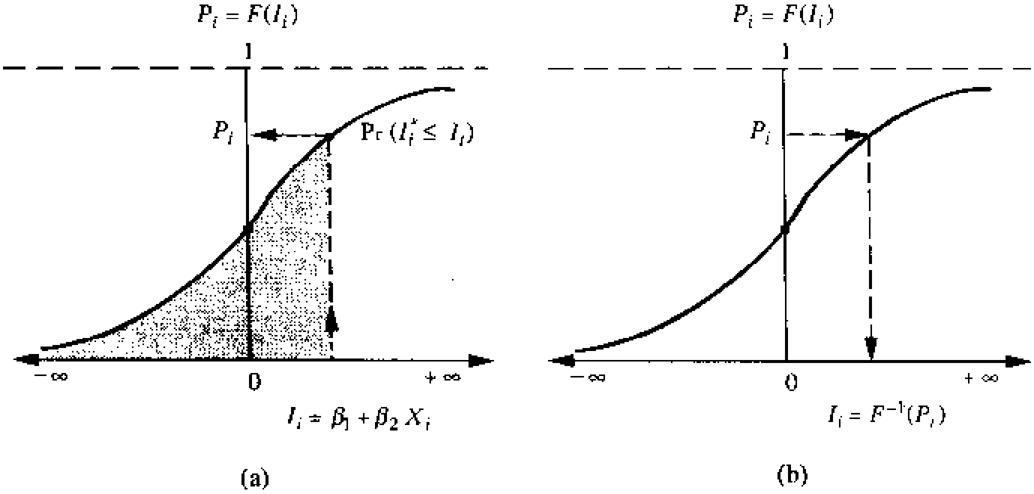


图 15.4　PROBIT 模型：(a) 给定 I_i，从纵坐标读取 P_i，(b) 给定 P_i，从横坐标读取 I_i

但是，实际上我们怎样能既获得指数 I_i 又对 β_1 和 β_2 进行估计呢？在 logit 模型的情形下，答案取决于我们拥有群组数据还是非群组数据。我们分别考虑这两种情况。

使用群组数据的 probit 估计：gprobit

610　我们将使用 glogit 曾用过的同样数据，由表 15.4 给出。既然已经得到了各个收入水平下拥有住宅的相对频率 P_i（概率的经验度量），如表 5.5 所示，那么我们就能用它来求得表 15.10（或图 15.5）中所示的正态 CDF 中的 I_i。

一旦我们有了 I_i 的估计值，正如稍后我们要指出的，对 β_1 和 β_2 的估计就相对简单了。顺便提一下，在 probit 分析的用语中，观测不到的效用指数又称为**正态等效利差**（n.i.d.），或简称 **normit**。由于每当 $P_i<0.5$ 时，n.i.d 或 I_i 将为负数，故实践中将 n.i.d. 加上 5 并将其结果称为一个 probit。

表 15.10　对标准正态 CDF 的指数 I_i 的估计

$\hat{P}_i$	$I_i=F^{-1}(\hat{P}_i)$
0.20	−0.841 6
0.24	−0.706 3
0.30	−0.524 4
0.35	−0.385 3
0.45	−0.125 7
0.51	0.025 1

0.60	0.253 3
0.66	0.412 5
0.75	0.674 5
0.80	0.841 6

注:(1)$\hat{P}_i$ 来自表 15.5;(2)I_i 来自对标准正态 CDF 的估计。

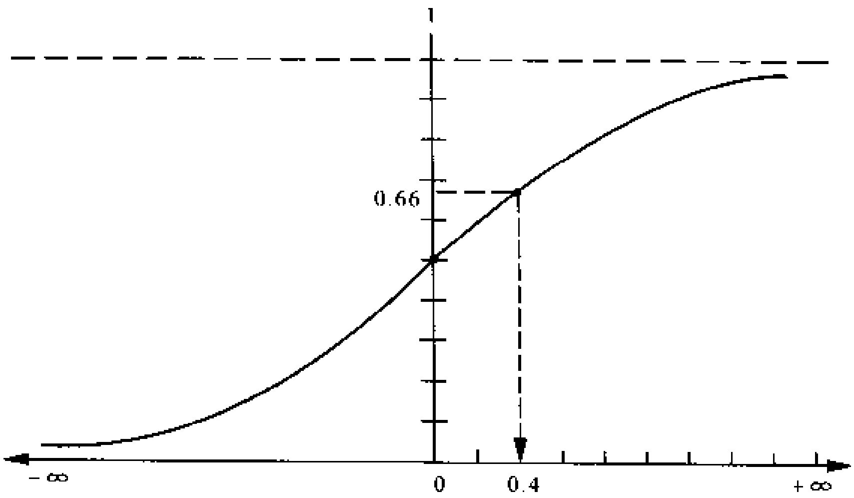


图 15.5 正态 CDF

611 **gprobit 的举例说明:住房的例子**

让我们继续使用住房的例子。对于这个例子我们已经得到了 glogit 模型的结果。对于同样的数据,群组 probit(gprobit)结果如下:

使用表 15.10 中给出的 n.i.d.(= I),回归结果如表 15.11 所示。[31] 基于 probits(= n.i.d. + 5)的回归结果如表 15.12 所示。

除了截距项,这些结果与前面表中所给出的完全相同。但这并不令人吃惊。(为什么?)

表 15.11

因变量:I

变量	系数	标准误	t 统计量	概率
C	−1.016 6	0.057 2	−17.747 3	1.039 7E−07
收入	0.048 46	0.002 47	19.558 5	4.854 7E−08
	R^2 = 0.979 51	德宾-沃森统计量 = 0.913 84		

表 15.12

因变量：I

变量	系数	标准误	t 统计量	概率
C	3.983 3	0.057 28	69.533 6	2.037 37E-12
收入	0.048 46	0.002 47	19.558 5	4.854 7E-08

$R^2=0.979\ 5$　　德宾-沃森统计量 = 0.913 8

注：这些结果没有对异方差性进行校正（参见习题 15.12）。

表 15.11 中 probit 估计值的解释。我们如何解释前面的结果呢？假定我们想弄清楚 X（以千美元计的收入）的一单位变动对 $Y=1$（即某家庭购买住宅）的概率的影响。为此，回顾一下方程（15.9.2）。我们要求函数对 X（关于收入的概率变化率）的导数。求导的结果是：

$$\frac{\mathrm{d}P_i}{\mathrm{d}X_i}=f(\beta_1+\beta_2X_i)\beta_2 \tag{15.9.5}$$ [32]

612 其中，$f(\beta_1+\beta_2X_i)$是在 $\beta_1+\beta_2X_i$ 处取值的标准正态概率密度函数。如你将意识到的，这个值的计算将取决于 X 变量的特定值。我们从表 15.5 中取 X 的一个值，比如说，$X=6$（千美元）。使用表 15.11 中给出的参数估计值，这样我们就需要求正态密度函数的值 $f[-1.016\ 6+0.048\ 46(6)]=f(-0.725\ 48)$。参照正态分布表，你将发现对于 $Z=-0.725\ 48$，正态密度约为 0.306 6。[33]现在将这个值乘以斜率系数的估计值，我们得到 0.014 85。这意味着从 6 000 美元收入水平开始，如果收入上升 1 000 美元，一个家庭购买住宅的可能性将上升约 1.4%。（与表 15.6 中的结果比较一下。）

在前面的讨论中，通过比较 LPM 和 logit 模型可以看出，使用 probit 模型计算概率的变化有点烦琐。

假定你想从拟合的 gprobit 模型中找出估计概率，而不去计算概率的变化，这很容易做到。利用表 15.11 中的数据并将表 15.5 中 X 值插入，读者可以算出n.i.d. 的估计值（保留两位小数），如下所示：

X	6	8	10	13	15	20	25	30	35	40
n.i.d. 估计值	-0.72	-0.63	-0.53	-0.39	-0.29	-0.05	0.19	0.43	0.68	0.92

现在诸如 Minitab 这样的统计软件包可以容易地计算出与各个 n.i.d. 相应的（累积）概率。例如，对应于一个 n.i.d. 值 -0.63，估计概率为

0.264 7；对应于一个 n.i.d. 值 0.43，估计概率为0.669 1。如果将这些估计值与表 15.5 中的实测值进行比较，你会发现两者相当的接近，这表明拟合模型十分好。我们刚才所做的已经形象地画在图 15.4 中了。

非群组或个体数据的 probit 模型

让我们再回到表 15.7。表中提供了 32 个个体的中级微观经济学考试的期末成绩。其中考试与变量 GPA、TUCE 和 PSI 有关。logit 回归的结果由表 15.8 给出。让我们看看 probit 结果像什么样。注意到如同个体数据的 logit 模型的情况那样，我们不得不使用一个以最大似然法为基础的非线性估计过程。由 Eviews 4 计算出的回归结果列在表 15.13 中。

613 **表 15.13** **因变量：成绩**

方法：ML－二分 probit

5 次迭代后获得收敛

变量	系数	标准误	t 统计量	概率
C	−7.452 3	2.542 4	−2.931 1	0.003 3
GPA	1.625 8	0.693 8	2.343 0	0.019 1
TUCE	0.051 7	0.083 8	0.616 6	0.537 4
PSI	1.426 3	0.595 0	2.397 0	0.016 5

LR 统计量（3 df）＝15.545 8　　McFadden $R^2=0.377\ 4$

概率（LR 统计量）＝0.001 4

“从定性上讲”，probit 模型的结果和 logit 模型的结果具有可比性，其中 logit 模型的 GPA 和 PSI 都个别地具有统计显著性。所有系数联合起来也具有统计显著性，因为 LR 统计值为 15.545 8，并且 p 值为 0.001 4。由于下节将讨论到的理由，我们不能对 logit 和 probit 回归系数直接进行比较。

为了比较的目的，我们对表 15.14 中的成绩数据给出基于将使用以成绩数据为基础的线性概率模型的结果。定性地看，LPM 的结果再次与 logit 和 probit 模型的结果是类似的，因为，GPA 和 PSI 都个别地具有统计显著性，而 TUCE 不具有。而且，由于 F 值 6.645 6 在统计上是显著的（因为它的 p 值只有 0.001 5），故解释变量合在一起看对成绩有显著的影响。

表 15.14　　因变量：成绩

变量	系数	标准误	t 统计量	概率
C	-1.498 0	0.523 8	-2.859 4	0.007 9
GPA	0.463 8	0.161 9	2.864 0	0.007 8
TUCE	0.010 4	0.019 4	0.538 6	0.594 3
PSI	0.378 5	0.139 1	2.720 0	0.011 0

$R^2=0.415\,9$　　德宾-沃森 $d=2.346\,4$　　F 统计量 $=6.645\,6$

在各种回归模型中某一个回归元的值变化一个单位的边际效应

在线性回归模型中，斜率系数所度量的是，在所有其他变量都保持不变的情况下，一个回归元一单位的变化所引起的回归子的平均变化。

LPM 中，斜率系数直接度量了由一个回归元一单位变化而导致一个事件出现变化的概率，其中所有其他变量的效应保持不变。

614 在 logit 模型中，变量的斜率系数显示了在保持所有其他的变量不变的情况下，伴随着变量的一单位变化而产生的机会比率的对数变化。但正如前面提到的，对于 logit 模型来说，一个事件发生的概率的变化率是由 $\beta_j P_i(1-P_i)$ 给出的，其中是 β_j 指第 j 个回归元的（偏回归）系数。但在计算时，包含在分析中的所有变量均须考虑。

正如我们早先看到的，在 probit 模型中，概率的变化率是相当复杂的，而且由 $\beta_j f(Z_i)$ 给出。这里 $f(Z_i)$ 是标准正态变量的密度，并且 $Z_i=\beta_1+\beta_2 X_{2i}+\cdots+\beta_k X_{ki}$，即分析中所使用的回归模型。

因此，在 logit 和 probit 模型中，所有回归子都涉及概率变化的计算，而在 LPM 中只有第 j 个回归子才涉及。这种区别可能就是 LPM 模型在早期受欢迎的一个原因。

§ 15.10　LOGIT 和 PROBIT 模型

尽管对于我们成绩的例子来说，LPM、logit 和 probit 模型定性地给出了相似的结果，但由于早先所提到的 LPM 的问题，我们仍然把注意力局限在 logit 和 probit 模型。在 logit 和 probit 之间，哪个模型更可取呢？在大多数应用中，这两个模型十分类似，主要的区别在于逻辑分布是否有稍微平坦的尾部，这一点从图 15.6 可以看出。也就是说，logit 的条件概率比 probit 的以

更慢的速度趋近于 0 或 1。这一点，从表 15.15 中可以清楚地看出。因此，没有必然的原因去选择一个模型而放弃另一个模型。由于 logit 模型使用相对简单的数学，因而在实际中许多研究者选择了它。

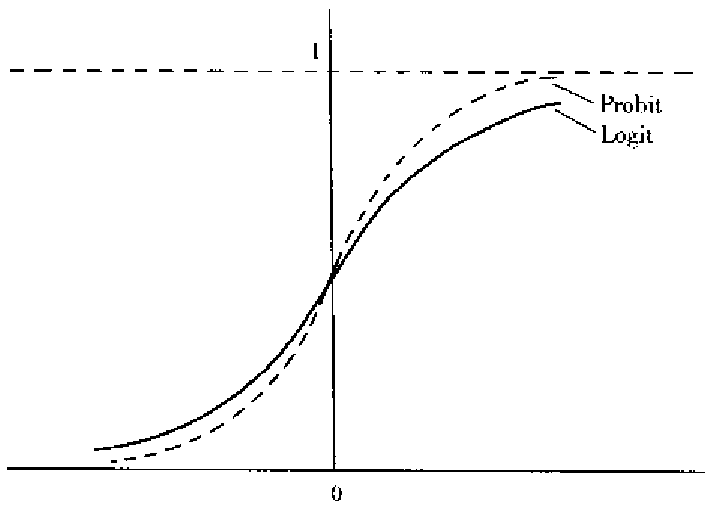


图 15.6 LOGIT 和 PROBIT 的累积分布

615 **表 15.15** 累积概率函数的值

Z	累积正态 $P_1(Z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z}e^{-s^2/2}ds$	累积逻辑 $P_2(Z)=\frac{1}{1+e^{-Z}}$
−3.0	0.001 3	0.047 4
−2.0	0.022 8	0.119 2
−1.5	0.066 8	0.182 4
−1.0	0.158 7	0.268 9
−0.5	0.308 5	0.377 5
0	0.500 0	0.500 0
0.5	0.691 5	0.622 5
1.0	0.841 3	0.731 1
1.5	0.933 2	0.817 6
2.0	0.977 2	0.880 8
3.0	0.998 7	0.952 6

尽管这两个模型是类似的，但是在解释由这两个模型估计出的系数时必须小心。例如，对于我们成绩的例子来说，用 probit 模型得出的 1.625 8 的

GPA 系数和用 logit 模型得出的 2.826 1 的 GPA 系数并不能直接进行比较。原因在于，尽管标准逻辑分布（logit 的基础）和标准正态分布（probit 的基础）的均值都等于 0，但它们的方差是不同的：对于标准正态分布来说是 1（正如我们已知的），而对于标准逻辑分布来说是 $\pi^2/3$，此处 $\pi \approx 22/7$。因此，如果你把 probit 系数乘以大约 1.81（近似等于 $\pi\sqrt{3}$），你就会近似得到 logit 系数。对于我们的例子而言，GPA 的 probit 系数是 1.625 8，将它乘以 1.81，我们得到 2.94，它很接近 logit 系数。相反，如果把 logit 乘以 0.55（=1/1.81），你将会得到 probit 系数。但是雨宫建议将 logit 的估计值乘以 0.625 以得到相对于 probit 估计值的更好的一个估计值。[34] 相反，将 probit 系数乘以 1.6（=1/0.625）将给出相应的 logit 估计。

顺便提一下，雨宫也列出了 LPM 模型和 logit 模型的系数的关系，如下所示：

$$\beta_{\text{LPM}} = 0.25\beta_{\text{logit}} \qquad \text{去掉截距}$$

并且

$$\beta_{\text{LPM}} = 0.25\beta_{\text{logit}} + 0.5 \qquad \text{加上截距}$$

我们留给读者去发现这些近似值是否支持我们成绩的例子。

§ 15.11 TOBIT 模型

probit 模型的一个扩展就是 tobit 模型，它最先由詹姆斯·托宾（卓越的诺贝尔经济学家）提出。为了解释这个模型，我们继续讨论房屋所有权的例子。在 probit 模型中，我们所关心的是估计拥有一间房屋所有权的概率，它是一些社会经济变量的函数。而在 tobit 模型中，我们的兴趣是弄清楚一个人或一个家庭在房屋上的花费，它跟社会经济变量有关。现在我们就面临着一个困境，即：如果消费者不买房，显然我们没有这类消费者的买房费用的数据；只有当消费者确实买了房我们才有此类数据。

因此，消费者被分成两个组，一组由 n_1 个消费者组成，我们有关于他们的回归元（即收入、债券利息率、家庭中的人数等等）和回归子（在购房上的费用）的信息，另一组由 n_2 个消费者组成，关于他们我们仅有回归元的信息而无回归子的信息。一个仅对某些观测有回归元的信息的样本叫做**截取样本**。[35] 因此，tobit 模型也被认为是截取回归模型。由于在回归元取值上的限制，一些作者称此类模型为**限值因变量模型**。

统计上，我们可以把 tobit 模型表达为：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + u_i \qquad \text{如果 RHS}>0 \\ &= 0 \qquad \text{其他} \end{aligned} \tag{15.11.1}$$

其中 RHS = 右手侧。注：附加变量 X 可轻易地加进这个模型。

能不能只用 n_1 个观测值来估计回归（15.11.1）而不管其余的 n_2 个观测值呢？答案是否定的，因为从 n_1 个观测值得到的参数的 OLS 估计值是有偏误的，并且是非一致的。也就是说，它们是渐近有偏误的。[36]

为了看到这一点，考虑图 15.7。该图表明，如果 Y 观测不到（由于截取），则所有这些观测（$=n_2$）（用叉号标出）将全部落在水平轴上。如果 Y 被观测到，则观测点（$=n_1$）（用点标出）将落在 X—Y 平面上。直观上
617 就能看清楚，如果仅根据 n_1 个观测来估计回归，所得到的截距和斜率系数和把全部 n_1+n_2 个观测都考虑进来而得到的有所不同。

那么要怎样去估计像（15.11.1）这样的 tobit（或截取）回归模型呢？实际使用的技巧包括最大似然法，它的应用相当复杂并且超出了本书论述的范围，但读者能从参考书中获得更多关于 ML 法的信息。[37]

赫克曼（James Heckman）提出了一个 ML 法的替代方法，它相对来说要简单一些。[38] 该替代方法由一个两步骤的估计程序组成。第一步，我们首先估计消费者拥有房屋的概率，这可以利用 probit 模型来完成。第二步，通过给模型（15.11.1）加入一个由 probit 估计导出的变量［称作**反米尔斯比**
618 （Inverse Mills Ratio）或**危害率**］来对其进行估计。关于它的具体操作可以参考赫克曼的文章。赫克曼程序给出了（15.11.1）的参数的一致性估计，但它们并不像 ML 估计那样有效。因为大多数现代统计软件包都有 ML 方法，故应用这些软件包比赫克曼两步程序更受欢迎。

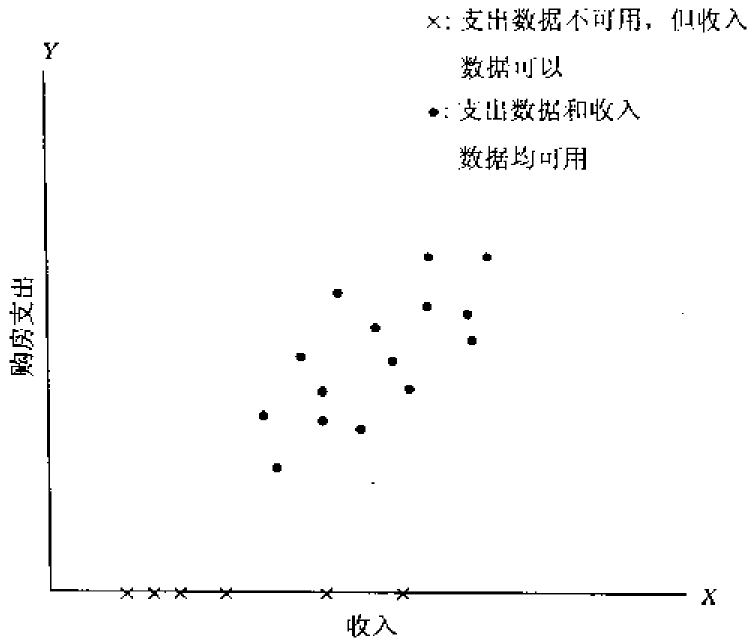


图 15.7　相对于收入消费者计划在购房上的花费

Tobit 模型的举例说明：费尔的婚外恋模型[39]

在一篇有趣且具有理论创新性的文章中，费尔收集了 601 个第一次结婚

的男人和女人的样本，并分析了他们对一个关于婚外恋问题的反应。[40]该研究中的变量定义如下：

Y = 在过去 0，1，2，3，4～10（记为 7）年的婚外恋数目

$Z_1 - 0$（对于女人），1（对于男人）

Z_2 = 结婚的年数

Z_4 = 孩子，如果没有孩子，则取值 0；如果有孩子，则取值 1

Z_5 = 宗教信仰，分 1～5 个等级，1 为不信教

Z_6 = 受教育年数：初等教育 = 9；高等教育 = 12，Ph.D. 或其他 = 20

Z_7 = 职业，按“Hollingshead”等级，1～7

Z_8 = 婚姻的自我评价，1 = 很不幸福，5 = 非常幸福

总共 601 个应答者，451 人没有婚外恋，150 人有一次或一次以上的婚外恋。

根据图 15.7，如果我们在纵轴上标出事件的次数并且在横轴上标出比方说受教育年数，则有 451 个观测点分布在横轴上。因此，我们将得到一个截取样本，并且 tobit 模型可以适用。（注意 Y 的观测值是整数而它的估计值不一定是。——译者注）

表 15.16 给出了用（不适合的）OLS 和（适合的）ML 方法得出的上述模型的估计结果。如你所见，OLS 包括 451 个没有婚外恋的人和 150 个有一次或一次以上婚外恋的人。ML 法对此进行了明显的考虑（指对有和没有婚外恋给予不同的对待。——译者注）但 OLS 法则没有，因此这两个估计是有区别的。由于已经讲过的原因，人们可以信赖 ML 而忽视 OLS 估计。这两个模型的系数可以和任何其他的回归模型的系数一样进行解释。Z_8 的负系数（婚姻幸福程度）意味着婚姻的幸福程度越高，则婚外恋发生的频率越低，这也许这不是什么惊人的发现。

附带提一下，注意到如果我们感兴趣的是婚外恋发生的概率而不是此类
619 事件的次数，则我们可以使用 probit 模型，规定 $Y=0$（对于没有婚外恋的人）和 $Y=1$（对于有婚外恋的人），结果如表 15.17 所示。运用 probit 模型的知识，读者应该可以自己解释该表所示的 probit 结果。

619 **表 15.16** **婚外恋的 OLS 和 TOBIT 估计**

解释变量	OLS 估计	Tobit 估计
截距	5.872 0（5.162 2）*	7.608 4（1.947 9）'
Z_1	0.054 0（0.179 9）	0.945 7（0.889 8）
Z_2	−0.050 9（−2.253 6）	−0.192 6（−2.379 9）
Z_3	0.169 4（4.110 9）	0.533 1（3.636 8）
Z_4	−0.142 6（−0.407 2）	1.019 1（0.796 5）
Z_5	−0.477 6（−4.274 7）	−1.699 0（−4.190 6）
Z_6	−0.013 7（−0.214 3）	0.025 3（0.111 3）

Z_7	0.104 9 (1.180 3)	0.212 9 (0.663 1)
Z_8	−0.711 8 (−5.931 9)	−2.273 2 (−5.472 4)
R^2	0.131 7	0.151 5

* 括号里的数据是 t 值

+ 括号里的数据是 Z（标准正态）值

注：总共有 601 个观测值，其中对于因变量（婚外恋的次数）有 451 个值为零，其他 150 个为非零值。

表 15.17 因变量：Y^*

方法：ML－二分 probit

样本：1－601

涉及的观测值：601

5 次迭代后获得收敛

变量	系数	标准误	t 统计量	概率
C	0.779 402	0.512 549	1.520 638	0.128 4
Z_1	0.173 457	0.137 991	1.257 015	0.208 7
Z_2	−0.024 584	0.010 418	−2.359 844	0.018 3
Z_3	0.054 343	0.018 809	2.889 278	0.003 9
Z_4	0.216 644	0.165 168	1.311 657	0.189 6
Z_5	−0.185 468	0.051 626	−3.592 551	0.000 3
Z_6	0.011 262	0.029 517	0.381 556	0.702 8
Z_7	0.013 669	0.041 404	0.330 129	0.741 3
Z_8	−0.271 791	0.053 475	· 5.082 608	0.000 0

平均解释变量	0.249 584	S.D. 解释变量	0.433 133
回归的 S.E.	0.410 279	赤池信息准则	1.045 584
残差平方和	99.650 88	施瓦茨准则	1.111 453
对数似然性	−305.198 0	汉南－奎恩准则	1.071 224
受限制对数似然性	−337.688 5	平均对数似然性	−0.507 817
LR 统计量（8 df）	64.981 07	麦克法登 R^2	0.096 215
概率（LR 统计量）	4.87E−11		
解释变量为零的观测次数	451	总观测次数	601
解释变量为 1 的观测次数	150		

§15.12 对计数数据建模：泊松回归模型

620 在许多现象中回归子是计数型的，例如一个家庭每年的假期数，一个企业每年所取得的专利次数，每年看牙医或医生的次数，每周去杂货店的次数，每年收到的违章停车票或超速警告票数，在给定期间内住院的天数，在一段时间（如5分钟）通过一个收费站的车辆数等等。每个例子中潜在的变量都是离散的，只能取有限的数值。有时计数数据也可能是稀少或不经常发生的事件，例如一周内被闪电击中的次数，2周内不止一次中了彩票，或4周内有两次或两次以上心脏病突发。对此类现象如何建模呢？

正如贝努里分布被选择用来对线性概率模型中是/否的决定来建模一样，专门适合计数数据的概率分布就是泊松概率分布。下面给出了泊松分布的PDF[41]：

$$f(Y_i)=\frac{\mu^{Y}e^{-\mu}}{Y!}\qquad Y=0,1,2,\ldots \tag{15.12.1}$$

其中$f(Y)$指变量Y取非负整数值的概率，且$Y!$（读作Y的阶乘）代表$Y!=Y\times(Y-1)\times(Y-2)\times2\times1$。可证明：

$$E(Y)=\mu \tag{15.12.2}$$

$$\operatorname{var}(Y)=\mu \tag{15.12.3}$$

注意泊松分布的一个有趣的性质：它的方差和均值是一样的。

泊松分布模型可以写成：

$$Y_i=E(Y_i)+u_i=\mu_i+u_i \tag{15.12.4}$$

其中诸Y如同均值为μ_i的泊松随机变量那样独立分布，而每个μ_i表达如下：

$$\mu_i=E(Y_i)=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\ldots+\beta_kX_{ki} \tag{15.12.5}$$

其中诸X是可能会影响均值的变量。例如，如果计数变量是指在给定的一
621 年内参观纽约天主教艺术博物馆的次数，则这个次数将依赖于其他变量，如消费者的收入、门票价格、离博物馆的距离以及停车费。

出于估计的目的，我们将模型写成：

$$Y_i=\frac{\mu^{Y}e^{-\mu}}{Y!}+u_i \tag{15.12.6}$$

这里用了（15.12.5）代替u_i。如你所见，得出的回归模型对于参数是非线性的，使我们必须用到上一章所论讨的非线性回归估计。我们来考虑一个具体的例子，看看它是如何计算的。

一个说明的例子：老年人跌倒次数的研究

此处所使用的数据是由内特（Netere）等人收集的。[42]这些数据是关于65岁或65岁以上的100个人的。研究的目的是记录这群人的跌倒次数（$=Y$），它分别与性别（$X_2=0$，对于女性；$=1$，对于男性）、平衡指数（X_3）和力量指数（X_4）有关。平衡指数越高，对象越稳定，而且力量指数越高，对象越强壮。为了查明教育或教育加有氧运动是否对跌倒次数有影响，作者引入了一个附加变量（X_1），即所谓的干预变量；$X_1=0$，如果只有教育，$X_1=1$，如果教育加上有氧运动。对象被随机地分配给两个干预方法。

利用 Eviews 4，可得出表 15.18 的结果。

表 15.18

因变量：Y

样本：1～100

7 次迭代后获得收敛

$Y=\mathrm{EXP}\,(C(0)+C(1)*X1+C(2)*X2+C(3)*X3+C(4)*X4)$

	系数	标准误	t 统计量	概率
C (0)	0.370 20	0.345 9	1.070 1	0.287 3
C (1)	−1.100 36	0.170 5	−6.452 5	0.000 0
C (2)	−0.021 94	0.110 5	−0.198 5	0.843 0
C (3)	0.010 66	0.002 7	3.948 3	0.000 1
C (4)	0.009 27	0.004 14	2.238 0	0.027 5

$R^2=0.485\,7$　　校正 $R^2=0.464\,0$

对数似然 $=-197.209\,6$　　德宾-沃森 统计量 $=1.735\,8$

注：EXP（ ）表示（ ）中表达式为 e（自然对数的底）次方。

622 **对结果的解释**

记住我们所得到的表 15.18 的结果是对第 i 个人而言的，其估计的均值是 μ_i。也就是说，我们所估计的是：

$$\hat{\mu}_i=e^{0.370\,2-1.100\,366X_{1i}-0.021\,94X_{2i}+0.010\,6X_{3i}+0.009\,27X_{4i}} \qquad (15.12.7)$$

为了找出第 i 个对象的实际均值，我们需要对该对象的不同 X 变量取值。举例来说，第 99 个对象有这些变量值：$Y=4$，$X_1=0$，$X_2=1$，$X_3=50$ 和 $X_4=56$。将这些值代入（15.12.7）可得到第 99 个对象的估计的均值为 $\hat{\mu}_{99}=3.353\,8$。该对象的实际 Y 值是 4。

现在如果想查明与第 99 个对象相类似的某一对象每年跌倒少于 5 次的概率，那么我们可得到如下的结果：

$$P(Y<5)=P(Y=0)+P(Y=1)+P(Y=2)+P(Y=3)+P(Y=4)$$

$$=\frac{(3.353\,8)^0e^{-3.353\,8}}{0!}+\frac{(3.353\,8)^1e^{-3.353\,8}}{1!}+$$

$$\frac{(3.3538)^2e^{-3.3538}}{2!}+\frac{(3.3538)^3e^{-3.3538}}{3!}$$

$$+\frac{(3.3538)^4e^{-3.3538}}{4!}$$

$$=0.7491$$

我们也能知道回归元对 Y 的均值的边际或偏效应，如下所示。根据说明性的例子，假定我们想弄清楚力量指数（X_4）增加一单位对 Y 的均值的影响。既然：

$$\mu=e^{C_0+C_1X_{1i}+C_2X_{2i}+C_3X_{3i}+C_4X_{4i}} \tag{15.12.8}$$

我们需要求$\partial\mu/\partial X_4$。利用微积分的链式法则，显而易见它等于：

$$\frac{\partial\mu}{\partial X_4}=C_4e^{C_0+C_1X_{1i}+C_2X_{2i}+C_3X_{3i}+C_4X_{4i}}=C_4\mu \tag{15.12.9}$$

也就是说，均值相对于回归元的变化率等于该回归元的系数乘以均值。当然，均值取决于该模型中所有回归元的取值。这一点和我们前面讨论过的 logit 和 probit 模型相类似，在这两个模型中变量的边际贡献也取决于该模型所有变量的取值。

回到个别系数的统计显著性，可以看出截距和变量 X_2 个别地具有统计显著性。但请注意表中所给出的标准误是渐近的，因此，t 值应渐近地加以解释。正如前面提到的，一般而言，所有的非线性迭代估计过程的结果仅仅在大样本上才具有真实性。

在结束对于泊松回归模型的讨论前，请注意该模型做了限制性的假定，即泊松过程的均值和方差是相同的，而且某事件在任一时点上发生的概率是不变的。

§15.13 定性响应回归模型的其他专题

623 正如在开始所提到的，定性响应回归模型的问题是相当多的。本章介绍的是这方面的一些基本模型。对于想进一步研究这个主题的读者，我们在下文中将简要讨论这个领域的一些其他的模型。这里我们不深入探讨，因为这样做将远远超出本书的范围。

序数 logit 和 probit 模型

在双变量 logit 和 probit 模型中，我们感兴趣的是为是或否的响应变量而

建模。但通常响因变量或回归子有多于两个的结果，并且这些结果的性质常常是序数的。也就是说，它们不能表达为一段区间。通常，在调查型研究中响应是在 Likert 型上度量的，例如“坚决同意”，“某种程度上同意”或“坚决不同意”。或者教育调查的响应可能会是“低于高中教育”，“高中教育”，“大学”或“专家程度”。这些响应常常被标号为 0（低于高中教育），1（高中教育），2（大学），3（研究生）。这些就是序数计数法，因为种类之间存在着明显的等级，但不能说 2（大学教育）是 1（高中教育）的 2 倍或者 3（研究生教育）是 1（高中教育）的 3 倍。

为了研究诸如上面所提到的现象，我们可以扩展双变量 logit 和 probit 模型来把多重等级的类型考虑进去。当我们必须使用多阶段和逻辑概率分布来考虑各种各样的等级类型时，会应用到很多算术。对于所涉及的数学基础及其应用，读者可以查阅早先引用的格林和曼达拉的文献。在相对直观的水平上，读者可以查阅里奥（Liao）的专著。[43]诸如 Limdep，Eviews 和 Shazam 的软件为序数 logit 和 probit 模型的估计提供了计算程序。

多项 logit 和 probit 模型

在序数 probit 和 logit 模型中，响应变量有两个以上的次序或等级类型。但有些情况下回归不是无序的。例如上班所采用的交通方式的选择。选择可能会是自行车、摩托车、汽车、公共汽车或火车。尽管这些是分类的响应，但此处没有等级或次序；就其性质而言，它们基本上是名义上的。再举一例，考虑职业的分类，如不熟练的，半熟练的和很熟练的。这里仍然没有次序。与此类似，职业选择（如自我雇佣、为私人企业工作、为当地政府工作，以及为联邦政府工作等）从性质上讲也基本上是名义的。

624 多项 logit 和 probit 模型的技巧可用来研究这种名义种类。而且它要求很少的数学。前面引用的参考书提到了这些技巧的要点。如果在某些专门的例子中需要用到此类模型，则早先引用的统计软件包可以为它们提供工具。

持续时间模型

考虑如下这些问题：（1）什么决定了失业的持续期间？（2）什么决定了白炽灯的寿命？（3）什么因素决定了罢工的持续期间？（4）什么决定了 HIV 呈阳性的病人的存活时间？

诸如此类的问题是持续期间模型的主题，它们被普遍称之为**存活分析**或**从时间到事件数据分析**。上面所引用的每个例子中，关键变量都是时间的长度，要把它作为一个随机变量而建模。其数学问题再次涉及合适的概率分布的 CDF 和 PDF。尽管技术细节很繁琐，但是关于这个主题已有了一些容易

理解的著作。[44]诸如 Stata 和 Limdep 的统计软件可以容易地用于估计此类模型。这些软件包提供了一些演算好的例子来帮助使用此类模型的研究者。

§15.14 要点与结论

1. 定性响应回归模型是指在该类模型中，响应变量或回归子是非定量的或者说由一个区间度量的。

2. 最简单的定性响应回归模型是二分模型，在该类模型中，回归子为是/否或存在/不存在类型。

3. 最简单的二分响应模型是线性概率模型，在该类模型中，二分响应因变量通过利用标准的 OLS 方法来对相关解释变量进行回归。这里简单性不是一种美德，因为 LPM 饱受几个估计问题的折磨。即使某些估计问题可以克服，但 LPM 最基本的缺点是它假定某些事件发生的概率是随着回归元的水平而线性增加的。如果我们使用 logit 和 probit 模型，则可以避免这个限制性的假定。

4. 在 logit 模型中，因变量是机会比率的对数，它是回归元的线性函数。logit 模型的概率是逻辑分布的。如果有分组形式的数据，则可以用 OLS 来估计 logit 模型的参数，但要我们充分考虑误差项的异方差性，如果只能在个体或微观的水平上得到数据，则要求使用参数的非线性估计程序。

625 5. 如果我们选择正态分布作为合适的概率分布，则可以使用 probit 模型。该模型在数学上有点复杂，因为要涉及积分。但出于实际的考虑，logit 和 probit 模型都能得出类似的结果。因而实际上怎样选择就取决于计算的简单性，这并不是个难题，因为现在很容易得到复杂的统计软件包。

6. 如果响应变量是计数型的，则在实际应用中所经常使用的模型是泊松回归模型，它建立在泊松概率分布的基础上。

7. 与 probit 模型紧密相关的是 tobit 模型，它也被称作截取回归模型。在该模型中，响应变量仅仅在某些条件满足时才可观测到。例如，只有当一个人开始决定购买汽车时，购车费用的问题才是有意义的。但是，马德拉注意到 tobit 模型"仅仅当潜在变量［如隐藏在现象背后的基本变量］原则上可取负值并且观测到的零值是截取和不可观测的结果时才可以应用"[45]。

8. 二分响应回归模型有许多扩展。包括序数 probit 和 logit 以及名义 probit 和 logit 模型。隐藏在这些模型背后的哲理和简单 logit 以及 probit 模型一样，尽管数学变得更加复杂。

9. 最后，我们简要考虑了所谓的持续时间模型，在该模型中，现象（如失业或生病）的持续时间取决于几个因素。在此类模型中，持续时间的长度成为研究者感兴趣的变量。

习题

问答题

15.1 参照表 15.2 所给数据，若 $\hat{Y}_i$ 为负，则令它等于 0.01；若 $\hat{Y}_i$ 大于 1，则令它等于 0.99。重新计算权值并用 WLS 估计 LPM。将你的结果同（15.2.11）所给的相比较，并做出评议。

15.2 对于表15.1 所给住房所有权数据，logit 模型的最大似然估计如下：

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = -493.54 \quad + \quad 32.96\ 收入$$
$$t = \qquad (-0.000\,008)\ (0.000\,008)$$

评议这些结果，但注意所有高于 16（千美元）的收入都对应于 $Y=1$，而所有低于 16 的收入都对应于 $Y=0$。对于这种情形，你会先验地有什么预期？

15.3 费希尔[1]把 762 个家庭对耐用品 Y 的购买（$Y=1$，若购买；$Y=0$，若不购买）当作若干变量的函数来研究，得到如下结果：

表 15.3 支付的咨询费和资产规模

解释变量	系数	标准误
常数	0.141 1	—
1957 年可支配收入，X_1	0.025 1	0.011 8
(可支配收入 $=X_1)^2$，X_2	−0.000 4	0.000 4
支票存款，X_3	−0.005 1	0.010 8
储蓄存款，X_4	0.001 3	0.004 7
U.S. 储蓄债券，X_5	−0.007 9	0.006 7
住房情况：租赁，X_6	0.046 9	0.093 7
住房情况：拥有，X_7	0.013 6	0.071 2
每月租金，X_8	−0.754 0	1.098 3
每月抵押付款，X_9	−0.980 9	0.516 2
个人非分期债务，X_{10}	−0.036 7	0.032 6
年龄，X_{11}	0.004 6	0.008 4
年龄的平方，X_{12}	−0.000 1	0.000 1
婚姻状况，X_{13}（1＝已婚）	0.176 0	0.050 1
儿女人数，X_{14}	0.039 8	0.035 8

(儿女人数 $= X_{14})^2$, X_{15}	-0.0036	0.007 2
购买计划, X_{16} (1 = 计划; 0 = 无计划)	0.176 0	0.038 4
	$R^2 = 0.1336$	

注：所有金融变量均以千美元计。

住房情况：租赁（1，若租赁；否则为 0）；

住房情况：拥有（1，若拥有；否则为 0）。

资料来源：Janet A. Fisher, "An Analysis of Consumer Good Expenditure," *The Review of Economics and Statistics*, vol.64, no.1, Table 1, 1962, p.67.

a. 对方程的拟合做一般性评论。

b. 你会怎样解释附着于支票存款变量的系数 -0.0051? 你怎样说明该变量带有负号的合理性?

c. 引进年龄平方和儿女个数平方的理由是什么? 为什么两者都带有负号?

d. 假定除收入变量外，其余变量都取零值，试求收入为 20 000 美元的家庭购买耐用品的条件概率。

e. 给定 $X_1 = 15\,000$ 美元，$X_3 = 3\,000$ 美元，$X_4 = 5\,000$ 美元，$X_6 = 0$，$X_7 = 1$，$X_8 = 500$ 美元，$X_9 = 300$ 美元，$X_{10} = 0$，$X_{11} = 35$，$X_{13} = 1$，$X_{14} = 2$，$X_{16} = 0$，估计购置耐用品的条件概率。

15.4 表15.3 给出的劳动力参与回归的 R^2 值 0.175，是相当低的一个数值。你能检验这个数值的统计显著性吗? 你用哪一种检验? 为什么? 试对这类模型的 R^2 值做一般性的评论。

15.5 对回归（15.7.1）中各个不同的收入水平，估计购置住房的概率。将这些概率对收入描图并对所得到的关系进行评议。

627 **15.6**[2] 说明 probit 回归表 15.11 中的截距等于 $-\mu_x/\sigma_x$，而斜率等于 $1/\sigma_x$，其中 μ_x 和 σ_x 是 X 的均值和标准差。

15.7 根据 54 个标准大都市统计地区（SMSA）的资料，德马里斯（Demaris）估计出以下的 logit 模型，以说明高谋杀率与低谋杀率的区别[3]：

$$\begin{aligned} \ln\hat{O}_i &= \quad 1.1387 + 0.0014P_i + 0.0561C_i - 0.4050R_i \\ \text{se} &= \qquad\qquad (0.0009)\quad(0.0227)\quad(0.1568) \end{aligned}$$

其中 O = 高谋杀率的机会，P = 1980 年以千计的人口，C = 从 1970 年到 1980 年的人口增长率，R = 阅读智商，而 se 为渐近标准误。

a. 你怎样解释各个系数?

b. 哪些系数是个别地统计上显著的?

c. 阅读智商的一单位增加对出现较高谋杀率的机会有什么影响?

d. 人口增长率增加一个百分点，对出现较高谋杀率的机会有什

么影响?

15.8 比较并评论 WLS 回归(15.7.3)和 OLS 回归(15.7.1)。

解答题

15.9 从荷兰中央统计局 1980 年对家庭预算的调查,克拉默(J.S.Cramer)得到了 2 820 个家庭样本的如下 logit 模型。(这个结果是使用最大似然法经过三次迭代得到的。[4])logit 模型的目的是将汽车拥有权作为(对数)收入的函数。汽车拥有权是一个二分变量:$Y=1$,如果一个家庭拥有一辆汽车,否则为零。

$$\hat{L}_i = -2.772\,31 + 0.347\,582 \ln \text{Income}$$
$$t = (-3.35) \qquad (4.05)$$
$$\chi^2(1\text{df}) = 16.681 \ (p\text{ 值} = 0.000\,0)$$

其中,$\hat{L}_i$ = 估计的 logit,lnIncome 是收入的对数。χ^2 度量着模型的拟合优度。

a. 解释所估计的 logit 模型。

b. 你怎样从估计的 logit 模型中得到拥有汽车概率的表达式?

c. 一个收入为 20 000 美元的家庭拥有一辆汽车的概率是多少?

d. 对估计的 logit 模型的统计显著性做一评论。

15.10 建立方程(15.2.8)。

15.11 Bowen 和 Bok 使用 logit 模型对招收一切中学毕业的新生和只招收黑人新生的高校进行了的一项关于毕业率的重要研究,得到了表 15.19 中的结果。[5]

a. 对于招收一切新生和只招收黑人新生的毕业率,你能做出什么一般性的结论?

b. 机会比率是两个机会之比。比较两组招收一切新生的情况,一组是 SAT 的得分超过 1 299,另一组是 SAT 的得分小于 1 000(基准组)。1.393 的机会比率意味着第一组中学生的毕业机率比第二组中的高出了 39%。表中列出的各种机会比率是否和先验的期望率一致?

c. 对于估计参数的统计显著性你能说些什么?估计模型的总体显著性又能说些什么?

表 15.19　预测毕业率的逻辑回归模型,1989 年入校

变量	招收一切新生			只招收黑人		
	参数估计	标准误	机会比率	参数估计	标准误	机会比率
截距	0.957	0.052	—	0.455	0.112	—
女性	**0.280**	0.031	1.323	**0.265**	0.101	1.303

黑人	**−0.513**	0.056	0.599			
拉丁美洲人	**−0.350**	0.080	0.705			
亚洲人	**0.122**	0.055	1.130			
其他人种	**−0.330**	0.104	0.719			
SAT＞1299	**0.331**	0.059	1.393	0.128	0.248	1.137
SAT 1200～1299	**0.253**	0.055	1.288	0.232	0.179	1.261
SAT 1100～1199	**0.350**	0.053	1.420	0.308	0.149	1.361
SAT 1000～1099	**0.192**	0.054	1.211	0.141	0.136	1.151
SAT 的数据不具备	**−0.330**	0.127	0.719	0.048	0.349	1.050
中学等级的前 10%	**0.342**	0.036	1.407	**0.315**	0.117	1.370
中学等级的数据不具备	−0.065	0.046	0.937	−0.065	0.148	0.937
高的社会经济地位（SES）	**0.283**	0.036	1.327	**0.557**	0.175	1.746
低的 SES	**−0.385**	0.079	0.680	**−0.305**	0.143	0.737
SES 不具备	**0.110**	0.050	1.116	0.303	0.172	1.031
SEL−1	**1.092**	0.058	2.979	**0.712**	0.161	2.038
SEL−2	**0.193**	0.036	1.212	**0.280**	0.119	1.323
女子大学	**−0.299**	0.069	0.742	0.158	0.269	1.171
观测次数	32 524			2 354		
−2log 似然函数						
有限制的	31 553			2 667		
无限制的	30 160			2 569		
χ^2	1 393 有 18d.f.			98 有 14d.f.		

注：黑体系数在 0.05 水平上是显著的；其他系数则不是。模型中省略的组别有白人、男性、SAT＜1 000、中学等级的后 90%、中等的 SES、SEL−3 以及男女兼收的大学。毕业率是六年制的、首所学校的毕业率，如附录表 D.3.1 中注释里所定义。学院选择组别如附录表 D.3.1 中注释里所定义。关于社会经济地位（SES）的定义参见附录 B。

SEL−1＝平均 SAT 得分在 1 300 以及以上的学院

SEL−2＝平均 SAT 得分在 1 150 到 1 299 之间的学院

SEL−3＝平均 SAT 得分低于 1 150 的学院

资料来源：Bowen and Bok, op.cit., p.381.

15.12 在表15.11 所给出的 logit 模型中干扰项有如下的方差：

$$\sigma_u^2 = \frac{P_i(1-P_i)}{N_i f_i^2}$$

其中 f_i 是对应于 $F^{-1}(P_i)$ 的标准密度函数。

a. 给定上述方差，你是怎样转换表 15.10 中的模型，以使得转换后的方差项变成同方差的呢？

b. 利用表 15.10 的数据做此数据转换。

c. 用转换后的数据估计 probit 模型并与先前的回归结果做比较。

15.13 由于作为拟合优度的度量，R^2 并不特别适合于二分因变量模型，一个取代 R^2 的量是下述的 χ^2 检验统计量：

$$\chi^2 = \sum_{i=1}^{G} \frac{N_i(\hat{P}_i - P_i^*)^2}{P_i^*(1 - P_i^*)}$$

其中 N_i = 第 i 组中的观测数

$\hat{P}_i$ = 事件发生的实际概率（$= n_i/N_i$）

P_i^* = 估计的概率

G = 组数（即所观测的 χ_i 的水平的个数，例如在表 15.4 中是 10）

可以证明对于大样本，χ^2 将遵循（$G-k$）个自由度的 χ^2 分布，其中 k 是估计模型中的参数个数（$k<G$）。

试用上述 χ^2 变量检验回归（15.7.1），并对这样得到的拟合优度进行评议，再将它同所报道的 R^2 值做比较。

15.14 表15.20 给出对菊花蚜虫喷洒不同浓度的鱼藤酮得到的结果数据，以每束约 50 支菊花计算。做出一个适当的模型，把死亡概率表达为 X 对数（即剂量的对数）的函数，并对所做的结果进行评议。并计算习题 15.13 所讨论的 χ^2 适度检验。

表 15.20 鱼藤酮对菊花蚜虫的毒性研究

630

浓度 毫克/升 X	Log（X）	总 数 N_i	死亡数 n_i	$\hat{P}_i = n_i/N_i$
2.6	0.415 0	50	6	0.120
3.8	0.579 7	48	16	0.333
5.1	0.707 6	46	24	0.522
7.7	0.886 5	49	42	0.857
10.2	1.008 6	50	44	0.880

资料来源：D.J.Fennet，*Probit Analysis*，Cambridge University Press，London，1964.

15.15 要求攻读研究生计划的13 名申请人有如下的 GRE 数量和词汇分数。其中有 6 名学生获得入学准许。

表 15.21

	GRE 能力测试分数		准人研究生计划（准 = 1，不准 = 0）
学生编号	数量，Q	词语，V	
1	760	550	1
2	600	350	0
3	720	320	0

4	710	630	1
5	530	430	0
6	650	570	0
7	800	500	1
8	650	680	1
9	520	660	0
10	800	250	0
11	670	480	0
12	670	520	1
13	780	710	1

资料来源：Donald F. Morrison, *Applied Linear Statistical Methods*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1983, p.279 (adapted).

a. 根据 GRE 数量和词语分数用 LPM 模型预测入学概率。

b. 这个模型能令人满意吗？如果不，你有什么其他的建议？

15.16 为了研究对6 盒两升装软饮料赠送价格折扣券的效果，蒙哥马利（Douglas Montgomery）和佩克（Elizabeth Peck）收集了下表中的数据。一个含有 5 500 个消费者的样本被随机地划分到 11 种不同的折扣类别中，如表所示，每类有 500 人。响应变量是消费者是否在一个月内将折扣券兑换。

a. 把兑换率看作因变量，价格折扣看作解释变量，看对数单位模型是否适合数据。

b. 看概率单位模型和对数单位模型是否拟合得一样好。

c. 如果价格折扣是 17 美分，预测的兑换率是多少？

d. 为使折扣券的兑换率达到 70%，价格折扣估计是多少？

631

表 15.22

价格折扣 X, ¢	样本大小 N_i	兑换折扣券个数 n_i
5	500	100
7	500	122
9	500	147
11	500	176
13	500	211
15	500	244
17	500	277
19	500	310
21	500	343
23	500	372
25	500	391

资料来源：Douglas C. Montgomery and Elizabeth A. Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, 1982, p.243 (notation changed).

15.17 为了发现谁是和谁不是银行存户（支票存户、储蓄存户等等），卡斯基（John Caskey）和彼得森（Andrew Peterson）利用美国家庭数据分别对 1977 年和 1989 年估计了一个概率模型。表 15.23 中所给的斜率系数衡量着相应的回归元的每单位变化对拥有银行账户的概率的潜在影响。这些边际影响是在模型所含回归元的均值处计算的。

a. 1977 年，婚姻情况对拥有银行账户的影响是什么？1989 年又是什么？这些结果有经济意义吗？

b. 为什么在 1977 年和 1989 年，代表少数民族的变量都有负的系数？

c. 你能对（幼年）子女数变量带有负号做出合理的解释吗？

d. 表中所给的 χ^2 统计量表征什么？（提示：习题 15.13。）

632 **表 15.23** **因变量为拥有银行存户的 PROBIT 回归**

	1977 年数据		1989 年数据	
	系数	隐含斜率	系数	隐含斜率
常数	−1.06		−2.20	
	(3.3)*		(6.8)*	
收入（以 1991 年美元计，千美元）	0.030	0.002	0.025	0.002
	(6.9)		(6.8)	
已婚	0.127	0.008	0.235	0.023
	(0.8)		(1.7)	
子女数	−0.131	−0.009	−0.084	0.008
	(3.6)		(2.0)	
户主（HH）年龄	0.006	0.000 4	0.021	0.002
	(1.7)		(6.3)	
HH 的教育水平	0.121	0.008	0.128	0.012
	(7.4)		(7.7)	
男 HH	−0.078	−0.005	−0.144	−0.011
	(0.5)		(0.9)	
少数民族	−0.750	−0.050	−0.600	−0.058
	(6.8)		(6.5)	
就业	0.186	0.012	0.402	0.039
	(1.6)		3.6	
住房所有者	0.520	0.035	0.522	0.051
	(4.7)		(5.3)	
对数似然性	−430.7		−526.0	
χ^2 统计量	480		602	

(H_0:除常数外全部系数等于零)		
观测个(次)数	2 025	2 091
样本中正确预测的百分率	91	90

* 括号中的数字是 t 统计量。

资料来源:John P. Caskey and Andrew Peterson, "Who Has a Bank Account and Who Doesn't: 1977 and 1989," Research Working Paper 93 - 10, Federal Reserve Bank of Kansas City, October 1993.

15.18 蒙特卡罗研究:为了帮助理解 probit 模型,威廉·贝克(William Becker)和唐纳德·瓦尔德门(Donald Waldman)假定以下的一个关系式[6]:

$$E(Y|X)=-1+3X$$

于是,令 $Y_i=-1+3X+\varepsilon_i$,其中假定 ε_i 为标准正态(即均值为0,方差为1)变量。由此产生如表15.24所列的含有35个观测值的样本。

a. 由此表的 Y 和 X 数据,你能估计一个 LPM 吗?记住真模型是$E(Y|X)=-1+3X$。

b. 给定 $X=0.48$,估计$E(Y|X=0.48)$并将此估计值与真$E(Y|X=0.48)$比较。注意 $\bar{X}=0.48$。

c. 利用表15.24中给出的 Y^* 和 X 数据,估计一个 probit 模型。你可以利用任一统计软件包。作者们所估计的 probit 模型如下:

$$\hat{Y}_i^*=-0.969+2.764X_i$$

试求:$P(Y^*=1|X=0.48)$也就是求 $P(Y_1>0|X=0.48)$。看你的答案是否和作者们的答案0.64一致。

d. 此表所给 X 值的样本标准差是0.31。如果 X 偏离 X 均值上方一个标准差,预测的概率变化是多少?也就是问,$P(Y^*=1|X=0.79)$是什么?作者的答案是0.25。

表 15.24 若 $Y>0$ 由模型 $Y=-1+3X+\varepsilon$ 和 $Y^*=1$ 产生的假想数据组

Y	Y^*	X	Y	Y^*	X
−0.378 6	0	0.29	−0.375 3	0	0.56
1.197 4	1	0.59	1.970 1	1	0.61
−0.464 8	0	0.14	−0.405 4	0	0.17
1.140 0	1	0.81	2.441 6	1	0.89
0.318 8	1	0.35	0.815 0	1	0.65

2.201 3	1	1.00	−0.122 3	0	0.23
2.447 3	1	0.80	0.142 8	1	0.26
0.115 3	1	0.40	−0.668 1	0	0.64
0.411 0	1	0.07	1.828 6	1	0.67
2.695 0	1	0.87	−0.645 9	0	0.26
2.200 9	1	0.98	2.978 4	1	0.63
0.638 9	1	0.28	−2.332 6	0	0.09
4.319 2	1	0.99	0.805 6	1	0.54
−1.990 6	0	0.04	−0.898 3	0	0.74
−0.902 1	0	0.37	−0.235 5	0	0.17
0.943 3	1	0.94	1.142 9	1	0.57
−3.223 5	0	0.04	−0.296 5	0	0.18
0.169 0	1	0.07			

资料来源：William E. Becker and Donald M. Waldman, "A Graphical Interpretation of Probit Coefficients," *Journal of Economic Education*, Fall 1989, Table 1, P.373.

【习题注释】

[1] "An Analysis of Consumer Good Expenditure," *The Review of Economics and Statistics*, vol.64, no.1, 1962, pp.64−71.

[2] 选做题。

[3] Demaris, op.cit., p.46.

[4] J.S.Cramer, *An Introduction to the Logit Model for Economist*, 2d ed., published and distributed by Timberlake Consultants Ltd., 2001, p.33. 这些结果是从 Timberlake Consultants 出版的统计软件包 PcGive 10 的第 51 页复制的。

[5] William G.Bowen and Derek Bok, *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*, Princeton University Press, Princeton, N.J., 1998, p.381.

[6] William E.Becker and Donald M.Waldman, "A Graphical Interpretation of Probit Coefficients," Journal of Economic Education, vol.20, no.4, Fall 1989, pp.371−378.

附录 15A

15A.1 个体（非群组）数据的 LOGIT 和 PROBIT 模型的最大似然估计[1]

正文中假定，给定个体的收入 X，我们对估计一个个体拥有住宅的概率感兴趣。我们还假定这个概率可以由逻辑函数（15.5.2）表达出来。为了

方便起见，复制如下：

$$P_i = \frac{1}{1+e^{-(\beta_1+\beta_2 X_i)}} \tag{1}$$

我们不实际观测 P_i，而只是观测结果 $Y=1$，如果个体拥有住宅；和 $Y=0$，如果个体不拥有住宅。

因为每个 Y_i 都是一个贝努里随机变量，所以我们可以写成：

$$\Pr(Y_i=1)=P_i \tag{2}$$

$$\Pr(Y_i=0)=1-P_i \tag{3}$$

假定我们有一个 n 次观测值的随机样本。令 $f_i(Y_i)$ 表示 $Y_i=1$ 或 0 的概率，观测的 n 个 Y 值的联合概率，即 $f(Y_1,Y_2,\cdots,Y_n)$ 为：

$$f(Y_1,Y_2,\cdots,Y_n)=\prod_1^n f_i(Y_i)=\prod_1^n P_i^{Y_i}(1-P_i)^{1-Y_i} \tag{4}$$

其中 Π 是乘积符号。注意，因为每个 Y_i 是独立的，而且有着相同的（逻辑）密度函数，所以我们可以将联合密度函数写成个体密度函数的乘积。等式（4）中的联合概率就是著名的**似然函数**（LF）。

等式（4）在使用上有一点不便。但如果我们对它取自然对数，就可得到所谓的**对数似然函数**（LLF）：

$$\begin{aligned}\ln f(Y_1,Y_2,\cdots,Y_n) &= \sum_1^n [Y_i\ln P_i+(1-Y_i)\ln(1-P_i)] \\ &= \sum_1^n [Y_i\ln P_i - Y_i\ln(1-P_i)+\ln(1-P_i)] \qquad (5) \\ &= \sum_1^n \left[Y_i\ln\left(\frac{P_i}{1-P_i}\right)\right]+\sum_1^n \ln(1-P_i)\end{aligned}$$

从（1）可以容易的证明：

$$(1-P_i)=\frac{1}{1+e^{\beta_1+\beta_2 X_i}} \tag{6}$$

和

$$\ln\left(\frac{P_i}{1-P_i}\right)=\beta_1+\beta_2 X_i \tag{7}$$

使用(6)和(7)，我们将 LLF(5)写成：

$$\ln f(Y_1,Y_2,\cdots,Y_n)=\sum_1^n Y_i(\beta_1+\beta_2 X_i)-\sum_1^n \ln[1+e^{(\beta_1+\beta_2 X_i)}] \tag{8}$$

从（8）可以看出，因为 x_i 已知，所以对数似然函数是参数 β_1 和 β_2 的函数。

在 ML 中我们的目标是使 LF（或 LLF）最大化，即通过使观测的诸 Y 的概率尽可能大（最大），从而获得未知参数的值。为了这个目的，我们就每个未知数对（8）求偏微分，令其表达式为零并解出这些表达式。然后应

用最大化的二阶条件证明，事实上我们所得到的参数的值可以使 LF 最大化。

635 因此，你必须分别用 β_1 和 β_2 对（8）求微分，然后按上面所说的进行。你很快就会意识到，结果表达式对参数来说有很高的非线性，不能得到确切的解。这就是为什么我们将不得不使用前面章节所讨论的非线性估计方法中的一种以得到数值解。一旦得到 β_1 和 β_2 的数值解，我们就可以容易地估计（1）了。

除了在（1）中我们使用正态 CDF 而不是逻辑 CDF，probit 模型的 ML 过程与 logit 模型中的 ML 过程是相似的。结果表达式变得相当的复杂，但总的思路是一致的。所以我们就到此为止。

【附录注释】

[1] 接下来的讨论在很大程度上依据于 John Neter, Michael H. Kutner, Christopher J. Nachsteim, and William Wasserman, *Applied Linear Statistical Models*, 4th ed., Irwin, 1996, pp.573－574。

【注释】

[1] 在入门的水平上，读者可能会发现以下的资料非常有用。Daniel A. Powers and Yu Xie, *Statistical Methods for Categorical Data Analysis*, Academic Press, 2000; John H. Aldrich and Forrest Nelson, *Linear Probability, Logit, and Probit Models*, Sage Publications, 1984; Tim Futing Liao, *Interpreting Probability Models: Logit, Probit and Other Generalized Linear Models*, Sage Publications, 1994. 文献的一个非常综合的评论可以参见 G.S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, 1983。

[2] 例如，参见 Ray Fair, "Econometrics and Presidential Elections," *Journal of Economic Perspective*, Summer 1996, pp.89－102, 以及 Machael S. Lewis-Beck, *Economics and Elections: The Major Western Democracies*, University of Michigan Press, Ann Arbor, 1980。

[3] 回忆一下，我们已经建议应该用合适的正态检验，例如 JB 检验，来检测某个应用中的正态假定。

[4] 证明的依据是中心极限定理，可参阅 E. Malinvaud, *Statistical Methods of Econometrics*, Rand McNally, Chicago, 1966, pp.195－197。如果诸回归元被认为是随机的并且是正态分布的，那么，即使干扰项是非正态分布的，F 和 t 检验仍可使用。而且记住，当样本大小无限增大时，二项分布收敛于正态分布。

[5] 为表明这一程序的正确性，参看 Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, New York, 1964, pp.249－250。基本上它是一个大样本的方法，有关异方差的章节中，我们在可行的或估计的广义最小二乘法的标题下讨论过。

[6] Aldrich and Nelson, op. cit., p.15. 在含有虚拟回归子的模型中测量拟合优度的其他方法参见 T. Amemiya, "Qualitative Response Models," *Journal of Economic Literature*, vol.19, 1981, pp.331－354。

[7] 不妨把较大的负值粗略地解释为：零收入的家庭拥有自己的住宅近乎不可能。

[8] 为了避免自由度的损失，当 Y_i 的估计值为负时，可令 $\hat{Y}_i = 0.01$；而当 Y_i

的估计值大于或等于 1 时，令 $\hat{Y}_i=0.99$，参看习题 15.1。

[9] Malcolm S. Cohen, Samuel A. Rea, Jr., and Robert I. Lerman, *A Micro Model of Labor Supply*, BLS Staff Paper 4, U.S. Department of Labor, 1970.

[10] Joseph Cappelleri, "Predicting a Bond Rating," unpublished term paper, C.U.N.Y. 本文所用模型是下文中所用模型的一个修改：Thomas F. Pogue and Robert M. Soldofsky, "What Is in a Bond Rating?" *Journal of Financial and Quantitative Analysis*, June 1969, pp. 201–228。

[11] 在进行异方差校正前，一些概率的估计值是负值或大于 1；这时，为便于计算权 w_i，将这些估计值相应地调整成 0.01 和 0.99。

[12] D. Rubinfeld, "An Econometric Analysis of the Market for General Municipal Bonds," unpublished doctoral dissertation, MIT, 1972. 此例的结果重载于 Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 2d ed., McGraw-Hill, New York, 1981, p. 279。

[13] John Aldrich and Forrest Nelson, op. cit., p. 26.

[14] 如同附录 A 中所讨论的，随机变量 X 的 CDF 是指它取值小于或等于 x_0 的 X 概率，这里 x_0 是 X 的某个特定的数值。简言之，X 的 CDF，记为 $F(X)$，是指 $F(X=x_0)=P(X\leqslant x_0)$。

[15] 逻辑模型曾被广泛应用于分析增长现象，诸如人口、GNP、货币供给等等。关于 logit 与 probit 模型的理论和实际的细节，可参看 J. S. Kramer, *The Logit Model for Economists*, Edward Arnold Publishers, London, 1991; 和 G. S. Maddala, op. cit.。

[16] 注意，随着 $Z_i\to+\infty$，e^{-Z_i} 趋于零；而随着 $Z_i\to-\infty$，e^{-Z_i} 无限地增大，其中 $e=2.718\ 28$。

[17] 当然，还可使用第 14 章中讨论的非线性估计技术。亦可参见 15.8 节。

[18] 记得 OLS 的线性假定并不要求 X 变量一定是线性的。这样，作为模型的回归元还可以有 X^2、X^3，等等。对于我们的目的，关键在于模型对参数而言是线性的。

[19] 利用微积分知识，可以证明 $dP/dX=\beta_2P(1-P)$，这表明概率对 X 的变化率不仅涉及 β_2，而且涉及用来测度概率变化的概率水平（第 15.7 节有更多的讨论）。顺便指出，当 $P=0.5$ 时 X_i 的一单位变化对 P 的影响最大，而当 P 接近 0 或 1 时影响最小。

[20] 这一点归功于 David Garson。

[21] 关于应用于对数单位模型的最大似然法的较简单的讨论，参看 John Aldrich and Forrest Nelson, op. cit., pp. 49–54。还参看 Alfred Demaris, *Logit Modeling: Practical Applications*, Sage Publications, Newbury Park, Calif., 1992。

[22] 记得由初等统计学，一个事件的概率是其相对频率在样本含量无限地变大时的极限。

[23] 如初等概率论所说的，成功（这里指拥有住宅）的比例 $\hat{P}_i$ 遵循均值等于真 P_i、方差等于 $P_i(1-P_i)/N_i$ 的二项式分布；并且随着 N_i 无限地增大，此二项式分布追近于正态分布。(15.6.4) 所给 u_i 的分布性质来自这一基本理论。关于细节，参看 Henry Theil, "On the Relationships Involving Qualitative Variables," *American Journal of Sociology*, vol. 76, July 1970, pp. 103–154。

[24] 因 $\hat{P}_i = n_i / N_i$，故 L_i 又可表示为$\hat{L}_i = \ln n_i / (N_i - n_i)$顺便指出，为了避免 $\hat{P}_i$ 取值 0 或 1，实践中可把 $\hat{L}_i$ 计算为$\hat{L}_i = \ln\left(n_i + \frac{1}{2}\right) / \left(N_i - n_i + \frac{1}{2}\right) = \ln(\hat{P}_i + 1/2N_i)/(1 - \hat{P}_i + 1/2N_i)$。作为一种经验之谈，我们建议对每一 X_i 值 N_i 不可小于 5，关于更多的细节，参看 D.R.Cox, *Analysis of Binary Data*, Methuen, London, 1970, p.33。

[25] 如果我们估计（15.6.1）而不考虑异方差性，则由第 11 章得知，虽然估计量是无偏的，却不是有效的。

[26] 一个易于理解的讨论可参见 J.Scott Long, *Regression Models for Categorical and Limited Dependent Variables*, Sage Publications, Newbury Park, California, 1997, pp.102－113。

[27] 技术上，这个被定义为：$1 - (\mathrm{LIF}_{ur}/\mathrm{LLF}_r)$，其中 LLF_{ur}是模型中包含所有回归元的无限制对数似然函数；LLF_r 是模型中仅含有截距项的受限制的对数似然函数。概念上，LLF_{ur}和 LLF_r 分别等价于线性回归模型中的 RSS 和 TSS。

[28] 参见附录 A 中的正态 CDF 的讨论。简言之，如果一个变量 X 服从均值为 μ、方差为 σ^2 的正态分布，那么它的 PDF 为：

$$f(X) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

它的 CDF 为

$$F(X) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

其中 X_0 是 X 的一个特定的值。

[29] D.McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in P.Zarembka (ed.), *Frontier in Econometrics*, Academic Press, New York, 1973.

[30] 拥有均值为零和单位（＝1）方差的正态分布就是大家所熟知的标准或标准化正态变量（参见附录 A）。

[31] 以下结果均未对异方差性进行校正。关于校正异方差的合适的程序参见习题 15.12。

[32] 我们使用导数的链条规则：

$$\frac{dP_i}{dX_i} = \frac{dF(t)}{dt} \cdot \frac{dt}{dX}$$

其中 $t = \beta_1 + \beta_2 X_i$。

[33]注意，标准正态 Z 能在 $-\infty$到 $+\infty$之间变动，但密度函数$f(Z)$一直为正。

[34] T.Amemiya, "Qualitative Response Model: A Survey," *Journal of Economic Literature*, vol.19, 1981, pp.481－536。

[35] 截取样本应该和断尾样本区分开来，后者只有在回归元可以观测到的时候回归子的信息才可得到。这里我们不讨论这一问题，但有兴趣的读者可以参考 William H.Greene, *Econometric Analysis*, Prentice Hall, 4th ed., Englewood Cliffs, N.J., Chap.19。直观的讨论可以参见 Peter Kennedy, *A Guide to Econometrics*, The MIT Press, Cambridge, Mass., 4th ed., 1998, Chap.16。

[36] 如果我们只考虑 n_1 个观测值而忽略其他的，这样就会产生偏差，那么就不能保证为所需的零。而且没有我们就不能保证 OLS 估计是无偏的。这种偏差可以

很容易的从附录3A中的方程（4）和（5）中看到。

[37] 参见 Greene, op.cit. 一个专业性不强的讨论可以从 Richard Breen, *Regression Models: Censored, Sample Selected or Truncated Data*, Sage Publications, Newbury Park, California, 1996。

[38] J.J.Heckman, "Sample Selection Bias as a Specification Error," *Econometrica*, vol.47, pp.153－161.

[39] Ray Fair, "A Theory of Extramarital Affairs," *Journal of Political Economy*, vol.86, 1978, pp.45－61. 这篇文章和数据可以查看网址http://fairmodel.econ.yale.edu/rayfair/pdf/1978DAT.ZIP。

[40]《今日心理学》(*Psychdogy Today*) 发表了有关性的101个问题的调查，它要求读者将他们的答案寄来。1970年7月，以收集的约2 000个电子形式的回答为基础，对杂志调查的结果进行了讨论。费尔从这些回答中抽取了601个作为样本。

[41] 对于这个分布的细节可参见任何统计的标准书籍。

[42] John Neter, Michael H.Kutner, Christopher J.Nachtsheim, and William Wasserman, *Applied Regression Models*, Irwin, 3d ed., Chicago, 1996. 数据可以从书中所附的数据盘中得到，并可参考习题14.28。

[43] Tim Futing Liao, op.cit.

[44] 例如，参见 David W.Hosmer, Jr., 和 Stanley Lemeshow, *Applied Survival Analysis*, John Wiley & Sons, New York, 1999。

[45] G.S.Maddala, *Introduction to Econometrics*, 2d ed., Macmillan, New York, 1992, p.342.

第 16 章 综列数据回归模型

636 在第 1 章里，我们简要讨论了一般可用于经验分析的数据类型，它们被称为**时间序列数据**、**横截面数据**和**综列数据**。在时间序列数据中，我们观测一段时期内一个或多个变量的值（比如，几个季度或几年的 GDP)。在横截面数据中，一个或多个变量的值是由同一时点上几个样本单元或实体的数据组成（比如，犯罪率是在给定的年份中由美国的 50 个州的数据组成)。而综列数据则是调查经历一段时间的同样的横截面数据。简言之，综列数据具有空间和时间两种特性。

表 1.1 中我们已经看到了这种数据类型的一个例子。表中列出了 1990 和 1991 年美国 50 个州鸡蛋产量和价格的数据。对于任一给定的年份，鸡蛋产量和价格的数据代表着一个横截面样本。对于任意给定的州，表中给出了关于鸡蛋产量和价格的两个时间序列的观测值。这样，关于鸡蛋产量和价格我们总共有了(50 × 2) = 100 次(混合)观测值。

综列数据还有其他一些名称，诸如**混合数据**（时间序列和横截面观测值的混合)，**时间序列和横截面数据的联合**，**纵列数据**（对一个变量或一组观测对象在一段时间内的研究)，**历史事件分析**（比如，一段时间内通过一系列状态或条件下的运动对观测物进行研究)，**群队分析**（例如，对某商学院的 1 965 名毕业生就业路径的跟踪研究)。尽管这些名字有一些微小的差别，但是它们在本质上都包含了横截面单元在一段时期的活动。因此，从一般的

意义上而言，我们将使用综列数据这个术语去涵盖一个或多个这样的术语。而且，**我们将基于这样的综列数据的回归模型称之为综列数据模型**。

637 现在综列数据正越来越多的被经济研究所使用。一些著名的综列数据组有：

1. **动态收入综列研究（PSID）**。由密歇根大学的社会研究所于 1968 年开始主持，研究所每年收集大约 5 000 个家庭的各种社会经济和人口统计变量。

2. **收入和参与项目调查（SIPP）**。这个类似 PSID 的调查由商业部人口普查局主持，每年将对被访问者就其经济状况进行 4 次调查。

还有许多其他的调查被不同的政府机构所采用。

一开始给大家一个警告。综列数据回归的主题非常多，而且其中所涉及的一些数学和统计学的知识十分复杂。我们只希望能够触及综列数据回归模型的一些基本东西，而把细节留给参考书。[1]但是要事先提醒大家，其中一些参考书非常专业。幸运的是，诸如 Limdep，PcGive，SAS，STATA，Shazam，Eviews 以及其他一些好用的软件包可以使得实际使用综列数据回归模型的任务变得十分容易。

§ 16.1 为什么使用综列数据？

比起横截面或时间序列数据，综列数据有哪些优点呢？巴尔凯（Baltagi）将综列数据的优点列在下面[2]：

1. 既然综列数据与一定时期内的个人、企业、州、国家等有关，那么这些单元中一定存在着差异性。正如稍后我们将看到的，综列数据估计的技巧能够通过对特定个体变量的考虑而将这种差异性非常清晰地加以研究。我们将用个体这个词从一般意义上表示诸如个人、企业、州、国家等微观单元。

2. 通过时间序列和横截面数据的混合，综列数据提供“更加有价值的数据，变量之间增加了多变性和减少了共线性，并且提高了自由度和有效性”。

638 3. 通过对重复横截面数据的研究，综列数据更适用于变化中的动态的研究。也就是说失业、工作的转变以及劳动力的流动性更适于用综列数据进行研究。

4. 综列数据能够更好地检测和度量单纯使用横截面数据或时间序列数据无法观测到的影响。例如，如果我们在联邦和（或）州的最低工资中加入最低工资增长的连续波动，那么就可以更好地研究最低工资法对就业和收入的影响。

5. 综列数据能够使我们对更加复杂的行为模型进行研究。比如，比起纯粹的横截面数据或时间序列数据，综列数据能够更好地处理诸如经济规模和技术变化这些现象。

6. 通过使数据适用于数千个单元，综列数据能够将偏差降到最低，而这种偏差可能是由于我们将个人或企业情况加总到更大的集合体中所产生的。

简言之，综列数据能够在很多方面丰富经验分析，而这些是仅仅使用横截面或时间数据无法做到的。但这并不表示综列数据模型就不存在任何问题，在学习一些理论和讨论一个例子后我们将讨论这些问题。

§16.2 综列数据：一个解释性的例子

为了以后的讨论，我们先来考虑一个具体的例子。表 16.1 所给出的数据取自由格伦费尔德（Y.Grunfeld）提出的一个著名的投资理论研究。[3]

格伦费尔德对于弄清企业的实际价值（X_2）和实际资本存量（X_3）如何决定实际总投资（Y）这个问题很感兴趣。尽管最初的研究包括几个公司，但出于例证的目的，我们获得了 4 个公司的数据，即通用电气（GE）、通用汽车（GM）、美国钢铁（US）以及西屋（Westing house）电气公司。每个公司关于前面所说的 3 个变量的数据来自于 1935—1954 年。这样就用了 4 个横截面单元以及 20 个时间期间。因此，总的来说我们有 80 个观测值。一个先验的结论是 Y 与 X_2、X_3 正相关。

原则上我们能够进行 4 个时间序列回归，每个公司一个，或者进行 20 个横截面回归，每年一次回归，尽管在后一种情况下我们将不得不担心自由度。[4]

640 混合或者说合并所有 80 次观测值，我们就能够如下写出格伦费尔德投资函数：

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it}$$
$$i = 1,\ 2,\ 3,\ 4$$
$$t = 1,\ 2,\ \cdots,\ 20 \tag{16.2.1}$$

其中 i 表示第 i 个横截面单元，t 表示第 t 年。作为一种惯例，我们将用 i 表示横截面标识符，用 t 表示时间标识符。假设 N 个横截面单元和 T 个时期均有一个最大值，如果每个横截面单元都有相同数目的时间序列观测，那么这样的综列（数据）被称为**平衡综列**。在本例中，当样本中的每个公司有 20 次观测时，我们就得到了一个平衡综列。如果各个综列单元的观测次数不相同，我们就称之为**非平衡综列**。本章中我们将主要考虑平衡综列。

首先假设 X 是非随机的，而且误差项遵从经典假设，即 $E(u_{it}) \sim N(0, \sigma^2)$。

仔细留意公式中下标 2 和 3，它们的意义是很明显的。

我们该怎样估计（16.2.1）呢？答案就在下面。

639 **表 16.1** **4 个公司的投资数据，1935—1954**

观测值	I	F_{-1}	C_{-1}	观测值	I	F_{-1}	C_{-1}
	GE				US		
1935	33.1	1 170.6	97.8	1935	209.9	1 362.4	53.8
1936	45.0	2 015.8	104.4	1936	355.3	1 807.1	50.5
1937	77.2	2 803.3	118.0	1937	469.9	2 673.3	118.1
1938	44.6	2 039.7	156.2	1938	262.3	1 801.9	260.2
1939	48.1	2 256.2	172.6	1939	230.4	1 957.3	312.7
1940	74.4	2 132.2	186.6	1940	361.6	2 202.9	254.2
1941	113.0	1 834.1	220.9	1941	472.8	2 380.5	261.4
1942	91.9	1 588.0	287.8	1942	445.6	2 168.6	298.7
1943	61.3	1 749.4	319.9	1943	361.6	1 985.1	301.8
1944	56.8	1 687.2	321.3	1944	288.2	1 813.9	279.1
1945	93.6	2 007.7	319.6	1945	258.7	1 850.2	213.8
1946	159.9	2 208.3	346.0	1946	420.3	2 067.7	232.6
1947	147.2	1 656.7	456.4	1947	420.5	1 796.7	264.8
1948	146.3	1 604.4	543.4	1948	494.5	1 625.8	306.9
1949	98.3	1 431.8	618.3	1949	405.1	1 667.0	351.1
1950	93.5	1 610.5	647.4	1950	418.8	1 677.4	357.8
1951	135.2	1 819.4	671.3	1951	588.2	2 289.5	341.1
1952	157.3	2 079.7	726.1	1952	645.2	2 159.4	444.2
1953	179.5	2 371.6	800.3	1953	641.0	2 031.3	623.6
1954	189.6	2 759.9	888.9	1954	459.3	2 115.5	669.7
	GM				WEST		
1935	317.6	3 078.5	2.8	1935	12.93	191.5	1.8
1936	391.8	4 661.7	52.6	1936	25.90	516.0	0.8
1937	410.6	5 387.1	156.9	1937	35.05	729.0	7.4
1938	257.7	2 792.2	209.2	1938	22.89	560.4	18.1
1939	330.8	4 313.2	203.4	1939	18.84	519.9	23.5
1940	461.2	4 643.9	207.2	1940	28.57	628.5	26.5
1941	512.0	4 551.2	255.2	1941	48.51	537.1	36.2
1942	448.0	3 244.1	303.7	1942	43.34	561.2	60.8
1943	499.6	4 053.7	264.1	1943	37.02	617.2	84.4
1944	547.5	4 379.3	201.6	1944	37.81	626.7	91.2
1945	561.2	4 840.9	265.0	1945	39.27	737.2	92.4

1946	688.1	4 900.0	402.2	1946	53.46	760.5	86.0
1947	568.9	3 526.5	761.5	1947	55.56	581.4	111.1
1948	529.2	3 245.7	922.4	1948	49.56	662.3	130.6
1949	555.1	3 700.2	1 020.1	1949	32.04	583.8	141.8
1950	642.9	3 755.6	1 099.0	1950	32.24	635.2	136.7
1951	755.9	4 833.0	1 207.7	1951	54.38	732.8	129.7
1952	891.2	4 924.9	1 430.5	1952	71.78	864.1	145.5
1953	1 304.4	6 241.7	1 777.3	1953	90.08	1 193.5	174.8
1954	1 486.7	5 593.6	2 226.3	1954	68.60	1 188.9	213.5

注：$Y = I =$ 总投资 = 厂房与设备的增添，加上维护和维修，经 P_1 平缩后的百万美元计

$X_2 = F =$ 厂商价值 = 普通和优先股的 12 月 31 日价格（或当年 12 月 31 日与次年 1 月 31 日平均价格）乘以未清偿普通股与优先股数，加上借款的 12 月 31 日总账面价值，经 P_2 平缩的百万美元计

$X_3 = C =$ 厂房设备存量 = 经 P_1 平缩的厂房设备经添增的累积和，减去经 P_3 平缩的折旧扣除

$P_1 =$ 生产者耐用设备综合价格平缩因子（1947 = 100）

$P_2 =$ GNP 综合价格平缩因子（1947 = 100）

$P_3 =$ 折旧费用平缩因子 = 金属与金属产品批发价格指数的 10 年移动平均（1947 = 100）

资料来源：Reproduced from H.D.Vinod and Aman Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, New York, 1981, pp.259 - 261.

§16.3 综列数据回归模型的估计：固定效应方法

对（16.2.1）的估计取决于我们对截距、斜率和误差项 u_{it} 的假定。以下有几种可能[5]：

1. 假定通过时间和空间的截距和斜率是不变的，误差项在时间和个体上存在差异。
2. 斜率不变而截距随个体而变化。
3. 斜率不变但截距随时间和个体而变化。
4. 所有系数（截距和斜率系数）均随个体而变化。
5. 截距和斜率随个体和时间而变化。

641 正如你所看到的，相继的每种情形将越来越多的复杂性（并且可能更加现实）引入对综列数据回归模型的估计中，比如说（16.2.1）。当然，如果在模型中增加回归元也会增加模型的复杂性，因为回归元之间存在着共线性的可能。

要想深入地探讨先前所说的各种可能性需要单独的一本书，而且市场上

已经有好几本这样的书了。[6]

1. 所有系数都不随时间和个体而变化

最简单的方法（也可能是自然的方法）就是不考虑空间和时间的混合数据，只是对通常的 OLS 回归进行估计。即将每个公司的 20 次观测值逐一堆置，于是模型中的每个变量总共可以得到 80 次观测值。OLS 结果如下：

$$
\begin{aligned}
\hat{Y} &= -63.3041 + 0.1101X_2 + 0.3034X_3 \\
\text{se} &= (29.6124) \quad (0.0137) \quad (0.0493) \\
t &= (-2.1376) \quad (8.0188) \quad (6.1545) \qquad (16.3.1) \\
R^2 &= 0.7565 \qquad \text{德宾-沃森估计值} = 0.2187 \\
& \qquad n = 80 \qquad \text{df} = 77
\end{aligned}
$$

如果检查**混合回归**的结果并且运用常见的准则，你将会看到所有系数都具有独立的统计显著性，斜率系数具有所期望的正号，而且 R^2 的值相当高。正如期望的一样，Y 与 X_2 和 X_3 正相关。美中不足的是，德宾-沃森统计的估计值十分低，这表明数据里可能存在着自相关。当然，正如我们所知，低的德宾-沃森值也可以归咎于设定误差。比如说，需要估计的模型假设 GE、GM、US 以及西屋电气的截距值是相同的。它还假定两个 X 变量的斜率系数对于四个公司是完全一样的。很明显，这些假定相当严格。因此，尽管很简单，但混合回归方程（16.1.2）可能扭曲了这四个公司 Y 和 X 之间关系的真实情况。我们需要做的就是找出一些办法将四个公司的特性分别考虑进来。接下来将解释怎样做。

2. 斜率系数不变而截距随个体而变化：固定效应或最小二乘虚拟变量回归模型

642 将每个公司或每个横截面单元的个体性考虑进来的一种方法是，假定截距随每个公司变化但斜率系数仍然不变。为了看到这一点，我们将模型（16.2.1）写成：

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \qquad (16.3.2)$$

注意到我们将下标 i 加到了截距项上，这就表明四个公司的截距是不同的。这种差异可能是由于每个公司的特性所引起的，比如管理者风格或管理者哲学。

在文献中，模型（16.3.2）就是为人熟知的**固定效应**（回归）模型（FEM）。固定效应这个词归因于每个个体的截距不随时间变化这一事实，也就是非时间变异，尽管截距可能随个体（这里指的是四个公司）的不同而变化。如果我们将截距项写成 β_{1it}，那么这表明每个公司或个体的截距是时间变异的。值得注意的是，由模型（16.3.2）所给出的 FEM 假定斜率系数

的回归元不会随个体或时间的变化而改变。

我们该如何实际的考虑公司间（固定效应）截距的变化呢？通过第 9 章中所学的虚拟变量方法我们可以容易地做到，尤其是**级差截距虚拟变量**。因此，(16.3.2) 可以写成：

$$Y_{it} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \qquad (16.3.3)$$

其中，如果观测值属于 GM，则 $D_{2i}=1$，否则为 0；如果观测值属于 US，则 $D_{3i}=1$，否则为 0；如果观测值属于 WEST，则 $D_{4i}=1$，否则为 0。由于我们用的是四个公司，因此只需用 3 个虚拟变量。换言之，α_1 代表 GE 的截距，而 α_2、α_3、α_4 以及级差截距系数就能够说明 GM、US 和 WEST 的截距相对于 GE 的截距有多大的不同。总而言之，GE 变成了一个供比较的公司。当然，你可以自由地选择任何一家公司作为比较公司。

顺便说一下，如果你想得到每个公司确切的截距值，就必须在开始的时候引入四个虚拟变量进行回归，也就是加入共同的截距。如果你不这么做，就会掉进虚拟变量陷阱。

648 由于我们使用虚拟变量对固定效应进行估计，在文献中，模型(16.3.3)就是大家知道的**最小二乘虚拟变量(LSDV)模型**。所以固定效应这个词和 LSDV 可以互换使用。过去 LSDV 模型也被称为**协方差模型**，X_2 和 X_3 就是所谓的协方差。

基于(16.3.3)的结果如下：

$$\begin{aligned}\hat{Y}_{it} = &-245.7924 + 161.5722D_{2i} + 339.6328D_{3i} + 186.5666D_{4i} + 0.1079X_{2i} + 0.3461X_{3i}\\ \text{se} = &\quad(35.8112)\quad(46.4563)\quad(23.9863)\quad(31.5068)\quad(0.0175)\quad(0.0266)\\ t = &\quad(-6.8635)\quad(3.4779)\quad(14.1594)\quad(5.9214)\quad(6.1653)\quad(12.9821)\\ &R^2 = 0.9345 \quad d = 1.1076 \quad \text{df} = 74\end{aligned} \qquad (16.3.4)$$

比较一下这个回归和 (16.3.1)。在 (16.3.4) 中，所有估计系数均具有独立的高显著性，而 t 系数估计量的 p 值相当得小。四个公司的截距值具有统计上的差异：GE 为 −245.792 4、GM 为 −84.220 (= −245.792 4 + 161.572 2)、US 为 93.877 4 (= −245.792 4 + 339.632 8)、WEST 为 −59.225 8(= −245.792 4 + 186.566 6)。这些截距上的差异可能由每个公司独特的性质所引起，比如管理风格或管理者能力的差异。

哪个模型更好呢，(16.3.1) 抑或 (16.3.4)？答案很明显。估计系数统计上的显著性、R^2 的值明显增加以及德宾－沃森 d 值的显著提高这些事实均表明，(16.3.1) 出现了设定错误。但是，R^2 值的增加并不意外，因为模型 (16.3.4) 中含有更多的变量。

我们也能够为这两个模型提供一个正式的测试。与 (16.3.4) 相比，由于给所有公司强加了一个相同的截距，模型 (16.3.1) 是一个受约束的模型。因此，我们可以用第 8 章讨论过的**受约束 F 检验**。运用公式(8.7.10)，读者可以容易地进行检测。在本例中，F 值为：

$$F = \frac{(R_{\text{UR}}^2 - R_{\text{R}}^2)/3}{(1 - R_{\text{UR}}^2)/74} = \frac{(0.9345 - 0.7565)/3}{(1 - 0.9345)/74} = 66.9980 \qquad (16.3.5)$$

其中，受约束 R^2 值来自于（16.3.1），不受约束 R^2 值来自（16.3.4）。因为（16.3.1）假定 GE、GM、US 和 WEST 的截距相同，所以约束数为 3。

很清楚，66.998 0（分子自由度为 3，分母自由度为 74）这个 F 值高度显著，因此，受约束回归（16.3.1）是无效的。

644 **时间效应**。正如用虚拟变量来考虑个体（公司）的影响，在一定意义上我们也可以考虑格伦费尔德投资函数随时间变化而引起的时间效应。而投资函数的变化是由诸如技术变迁、政府法规和（或）税收政策变化这些因素以及诸如战争或其他冲突这样的外部效应所引起的。如果我们每年引入一个时间虚拟变量，时间效应很容易被考虑进来。既然我们有从 1935 到 1954 年间 20 年的数据，那么就应该引入 19 个时间虚拟变量（为什么?）。模型（16.3.3）写成：

$$
\begin{aligned}
Y_{it} = \lambda_0 + \lambda_1 \mathrm{DUM35} + \lambda_2 \mathrm{DUM36} + \cdots + \lambda_{19} \mathrm{DUM53} + \\
\beta_2 X_{2it} + \beta_3 X_{3it} + u_{it} \qquad (16.3.6)
\end{aligned}
$$

其中，如果观测来自 1935 年，DUM35 的值就为 1，否则为 0；其他以此类推。我们将 1954 年作为基准年，它的截距值为 λ_0。（为什么?）

我们没有给出（16.3.6）的回归结果，因为没有一个单独的时间虚拟变量具有独立的统计显著性。（16.3.6）的 R^2 值为 0.769 7，而（16.3.1）的值为 0.756 5，仅仅增加了 0.013 2。运用受约束 F 检验可以看到，增加值是不显著的，这可能说明年份或时间效应不显著，也可能表明投资函数并未随时间发生很大的变化。以上作为留给读者的一道练习题。

我们已经看到单个公司的影响具有统计显著性，但单个年份的影响却没有。这是不是由于没有同时考虑个体和时间的影响而造成模型的设定错误呢？让我们来考虑这种可能性吧。

3. 斜率系数不变而截距随个体和时间而变化

为了考虑这种可能性，我们将（16.3.4）和（16.3.6）合并，如下所示：

$$
\begin{aligned}
Y_{it} = \alpha_1 + \alpha_2 D_{\mathrm{GM}_i} + \alpha_3 D_{\mathrm{US}_i} + \alpha_4 D_{\mathrm{WEST}_i} + \lambda_0 + \lambda_1 \mathrm{DUM35} + \cdots \\
+ \lambda_{19} \mathrm{DUM53} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_{it} \qquad (16.3.7)
\end{aligned}
$$

当我们进行回归时，我们发现公司虚拟变量和 X 的系数都具有独立的统计显著性。但是，没有一个时间虚拟变量是这样的。因此，我们再回到（16.3.4）。

得出的总结是可能存在个别公司效应而无时间效应。换言之，除了截距之外，四个公司的投资函数是完全相同的。在我们已经考虑过的所有情形中，变量 X 对 Y 有很强的影响。

4. 所有系数都随个体而变化

645 这里，我们假定所有个体单元或横截面单元的截距和斜率系数是不同的。也就是说，GE、GM、US 和 WEST 的投资函数都不相同。我们可以容易地将 LSDV 模型扩展来考虑这种情形。重新考虑（16.3.4）。在那里我们使用相加的方式引入个体虚拟变量。但是关于虚拟变量，在第 9 章中我们表明了斜率虚拟变量之间的交互作用，或者说级差是如何说明斜率系数之间的差异。为了在格伦费尔德投资函数中做到这一点，我们所必须要做的就是将每个公司虚拟变量与每个 X 变量相乘（这将在（16.3.4）中增加 6 个变量）。这样我们估计的模型如下：

$$
\begin{aligned}
Y_{it} = {} & \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \beta_2 X_{2it} + \beta_3 X_{3it} + \gamma_1(D_{2i}X_{2it}) + \gamma_2(D_{2i}X_{3it}) \\
& + \gamma_3(D_{3i}X_{2it}) + \gamma_4(D_{3i}X_{3it}) + \gamma_5(D_{4i}X_{2it}) + \gamma_6(D_{4i}X_{3it}) + u_{it} \qquad (16.3.8)
\end{aligned}
$$

你将注意到，γ 表示级差斜率系数，正如 α_2、α_3 和 α_4 表示级差截距。如果一个或多个 γ 系数具有统计显著性，这就是告诉我们一个或多个斜率系数与基准组不同。例如 β_2 和 γ_1 具有统计显著性，在这种情况下，$(\beta_2 + \gamma_1)$ 就给出了通用汽车的 X_2 的斜率系数值。这表明对于通用汽车，X_2 的斜率系数与通用电气这个比较公司是有差异的。

若所有的级差截距和所有的级差斜率系数都具有统计显著性，我们就可以得出这样一个结论，即通用汽车、美国钢铁和西屋电气的投资函数都是有别于通用电气的。如果事实确实如此，那么估计混合回归模型（16.3.1）就几乎没有意义了。

让我们来检验一下（16.3.8）的回归结果。为了便于阅读，（16.3.8）的回归结果被放在了表 16.2 这样的列项表格中。

表 16.2　　回归结果(16.3.8)

变量	系数	标准误	t 值	p 值
截距	−9.956 3	76.351 8	−0.130 4	0.896 6
D_{2i}	−139.510 4	109.280 8	−1.276 6	0.206 1
D_{3i}	−40.121 7	129.234 3	−0.310 4	0.757 2
D_{4i}	9.375 9	93.117 2	0.100 6	0.920 1
X_{2i}	0.092 6	0.042 4	2.184 4	0.032 4
X_{3i}	0.151 6	0.062 5	2.425 0	0.018 0
$D_{2i}X_{2i}$	0.092 6	0.042 4	2.184 4	0.032 4
$D_{2i}X_{3i}$	0.219 8	0.068 2	3.219 0	0.002 0
$D_{3i}X_{2i}$	0.144 8	0.064 6	2.240 9	0.028 3
$D_{3i}X_{3i}$	0.257 0	0.120 4	2.133 3	0.036 5
$D_{4i}X_{2i}$	0.026 5	0.111 4	0.238 4	0.812 2

$D_{4i}X_{3i}$	−0.060 0	0.378 5	−0.158 4	0.874 5

$R^2 = 0.951\ 1 \qquad d = 1.089\ 6$

正如这些结果所表明的，Y 与 X_2 和 X_3 显著相关。但是，几个级差斜率系数具有统计显著性。例如对于 GE，X_2 的斜率系数为 0.090 2，而 GM 则为 0.182 8(0.090 2+0.092 6)。有趣的是，没有一个级差截距在统计上是显著的。

646 总而言之，看上去四个公司的投资函数都是不同的。这可能表示四个公司的数据是没有"可混合性"的，这就能够单独对每个公司的投资函数进行估计（见习题 16.13）。需要提醒的是，综列数据回归模型不适用于单个公司的情况，尽管时间序列数据和横截面数据都是可以得到的。

使用固定效应或者 LSDV 模型时的一个注意事项。尽管 LSDV 模型易于使用，但它仍有一些问题需要记在心中。

第一，如果像在模型（16.3.7）中那样引入过多的虚拟变量，那么你将面临自由度的问题。在（16.3.7）情况下，我们有 80 次观测值，却仅仅有 55 个自由度——我们为三个公司虚拟变量失去了 3 个自由度、为 19 个年份虚拟变量失去了 19 个自由度、为 2 个斜率系数失去了 2 个自由度以及为一个共同截距失去了 1 个自由度。

第二，模型中有如此多的变量，因而总会存在多重共线性的可能性，而这种可能性会给准确估计一个或多个参数带来困难。

第三，假定在 FEM（16.3.1）中我们也可以包括诸如性别、肤色和种族这些变量。但这些变量也是时间恒定的，因为个人的性别、肤色或种族是不随时间而变化的。因此，LSDV 方法是不能够鉴别这种时间恒定变量的影响的。

第四，我们必须小心地考虑误差项 u_{it}。目前为止，我们所得到的结果都是建立在误差项遵从经典假设这一假定之上的，即 $u_{it} \sim N(0, \delta^2)$。既然下标 i 和 t 分别指的是横截面数据和时间序列数据，那么对于 u_{it} 的经典假设就必须做出修正。以下是几种可能性。

1. 我们可以假定对于所有横截面单元误差项的方差是相同的，或假定误差项的方差是异方差的。

2. 我们可以假定对于每个个体在时间上没有自相关。于是，比如说，我们就能假定通用汽车的投资函数的误差项是非自相关的。或者我们假定 AR(1)模式是自相关的。

3. 对于某个给定的时间，通用汽车的误差项可能与美国钢铁或者西屋电气的误差项相关。[7]或者，我们可以假定不存在这样的相关性。

647 4. 我们能够想到误差项的其他一些排列和组合。正如你将很快意识到的，对于一个或者多个这些可能性的考虑将会使分析大大的复杂化。空间和数学上的需要不允许我们考虑所有可能性。关于各种可能性的少许容易理解的讨论在代尔曼（Dielman）、塞斯（Sayrs）和克曼塔的书中可以找到。[8]不

过，如果我们求助于即将要讨论的随机效应模型，其中一些问题可能得以缓解。

§16.4 综列数据回归模型的估计：随机效应方法

尽管可以直接应用固定效应模型或者 LSDV 模型，但如果有好几个横截面单元，由于自由度的原因，建立的模型将是代价高昂的。此外，正如克曼塔所写道的：

> 与协方差（即 LSDV）模型相联系的一个明显的问题是，引入虚拟变量（并随之造成自由度数目的减少）是否是确实必要的。一个合理的基本协方差模型是这样一个特定的回归模型，模型内不能包含不随时间变化（并且可能是另一些确实随时间变化但对于所有横截面单元具有相同值）的相关解释变量，而且包含虚拟变量是对我们的无知的一种掩饰（重点强调）。[9]

如果虚拟变量确实代表了对于（真实）模型知识的一种缺乏，那么为什么不通过干扰项 u_{it}来表达这种无知呢？这里有一个精确的方法，支持它的人称之为**误差组成模型（ECM）或随机效应模型（REM）**。

基本的想法是从（16.3.2）开始的：

$$Y_{it}=\beta_{1i}+\beta_2X_{2it}+\beta_3X_{3it}+u_{it} \tag{16.4.1}$$

我们假定 β_{1i}是一个均值为 β_1（这里没有下标 i）的随机变量，而不再将其看成一个固定值。那么，单个公司的截距值就可表示为：

$$\beta_{1i}=\beta_1+\varepsilon_i \qquad i=1,2,3,\cdots,N \tag{16.4.2}$$

其中，ε_i 是一个均值为零、方差 σ_ε^2 的随机误差项。

实质上要说的是，例中的四个公司是从更大的这种公司集合中提取出来的，而这些公司的截距都有一个相同的均值（$=\beta_1$），并且每个公司截距值的个别差异都反映在误差项 ε_i 中。

将（16.4.2）代入（16.4.1）中，我们得到：

$$\begin{aligned}Y_{it}&=\beta_1+\beta_2X_{2it}+\beta_3X_{3it}+\varepsilon_i+u_{it}\\&=\beta_1+\beta_2X_{2it}+\beta_3X_{3it}+w_{it}\end{aligned} \tag{16.4.3}$$

648 其中

$$w_{it}=\varepsilon_i+u_{it} \tag{16.4.4}$$

合成的误差项 w_{it}包括两个部分：ε_i 和 u_{it}。前者是横截面或特定个体误差部分，后者则是时间序列和横截面混合误差部分。

误差组成模型这个术语正是因为合成误差项 w_{it} 由两个（或多个）误差部分组成而得名。

ECM 通常的假定是：

$$
\begin{aligned}
&\varepsilon_i \sim N(0,\sigma_\varepsilon^2) \\
&u_{it} \sim N(0,\sigma_u^2) \\
&E(\varepsilon_i u_{it})=0 \quad E(\varepsilon_i\varepsilon_j)=0 \quad (i\neq j) \\
&E(u_{it}u_{is})=E(u_{it}u_{jt})=E(u_{it}u_{js})=0 \qquad (i\neq j;\ t\neq s)
\end{aligned}
\tag{16.4.5}
$$

即个体之间的误差部分是不相关的，并且兼考虑横截面和时间序列单元的误差部分也没有自相关。

注意区分 FEM 和 ECM 之间的不同。在 FEM 中，每个横截面单元都有自己的（固定的）截距值，N 个横截面单元就有 N 个这样的值。另一方面，在 ECM 中截距 β_1 代表所有（横截面）截距的平均值，而误差部分 ε_i 则表示单个截距对这个平均值的（随机）偏离。不过，要记住 ε_i 不是直接可观测的，也就是众所周知的**不可观测的或潜在的、变化的**。

可从（16.4.5）中所做的假定推出结果如下：

$$E(w_{it})=0 \tag{16.4.6}$$

$$\operatorname{var}(w_{it})=\sigma_\varepsilon^2+\sigma_u^2 \tag{16.4.7}$$

现在如果 $\sigma_\varepsilon^2=0$，那么模型（16.2.1）和（16.4.3）之间就没有不同了。在这种情况下，我们可以简单地将（横截面和时间序列）观测值进行混合，然后像我们在（16.3.1）中那样进行混合回归。

如（16.4.7）所示，误差项 w_{it} 是同方差性的。但是，它可以表明 w_{it} 和 $w_{is}(t\neq s)$ 是相关的。这也就是说，一个给定横截面单元的误差项在不同的两个时点上是相关的。相关系数 $\operatorname{corr}(w_{it},w_{is})$ 表示如下：

$$\operatorname{corr}(w_{it},w_{is})=\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_u^2} \tag{16.4.8}$$

649 注意到上述相关系数有两个特性。第一，从（16.4.8）中可以清楚地看到，对于任意给定横截面单元，两个不同时间上的误差项的相关系数值保持不变，而无论这两个时间段相距有多远。这与我们在第 12 章中所讨论的一阶［AR(1)］模式形成了强烈的对比。在一阶模式中我们发现时间段之间的相关性随时间而下降。第二，对于所有横截面单元，（16.4.8）所给出的相关性的结构保持不变，即对于所有个体它是同一的。

如果我们不考虑这个相关性的结构并且用 OLS 对（16.4.3）进行估计，得出的估计量将是低效的。这里最合适的方法就是广义最小二乘法。

由于 GLS 在数学上的复杂性，我们将不会在当前的上下文中进行讨论。[10]既然现在大多数的统计软件包可以方便地估计 ECM（和 FEM 一样），那么我们将仅仅给出这个投资例子的结果。不过在这样做之前需要注意，我们能够容易地将一个随机误差部分扩展到（16.4.4）中从而考虑到时间的变动（见习题 16.6）。表 16.3 中列出了格伦费尔德投资函数的 ECM 估计结果。

这个回归中有几点需要注意。第一，如果你将四个公司的随机效应值相加，结果将会（正如它所应该的那样）为零。（为什么?）第二，随机误差部分的平均值 ε_i 为 -73.035 3，即共同截距值。GE 的随机效应值 -169.928 2 告诉我们 GE 的随机误差部分对共同截距值有多大的偏离。类似的解释可以应用于随机效应的其他三个值。第三，R^2 值可以通过变换的 GLS 回归得到。

如果将表 16.3 中 ECM 模型的结果与 FEM 的结果进行比较，你将看到除了表 16.2 中我们允许两个变量的斜率系数可以随横截面单元而变化之外，大体上两个 X 变量的系数值似乎差别不大。

表 16.3　　　　格伦费尔德投资函数的 EMC 估计

变量	系数	标准误	t 统计量	p 值
截距	-73.035 3	83.949 5	-0.869 9	0.387 0
X_2	0.107 6	0.016 8	6.401 6	0.000 0
X_3	0.345 7	0.016 8	13.023 5	0.000 0
随机效应				
GE	-169.928 2			
GM	-9.507 8			
USS	165.561 3			
WEST	13.874 75			
		$R^2=0.932\ 3$ (GLS)		

§16.5　固定效应与随机效应模型的比较

650 研究者所面临的挑战是，哪个模型更好，FEM 还是 ECM? 这个问题的答案取决于对个体或特定横截面、误差部分 ε_i 和回归元 X 之间可能的相关性所做的假定。

若假定 ε_i 与 X 不相关，那么 ECM 可能合适一些；若 ε_i 与 X 是相关的，FEM 则可能更适用一些。为什么要预计独立的误差部分和一个或多个回归元之间的相关性呢？考虑一个例子。假设有一个含有大量个体的随机样本，我们相对他们的工资或收入函数建立模型。假定收入是教育、工作经验等的函数。现在如果我们用 ε_i 代表内在的能力、家庭背景等，那么当我们建立包括 ε_i 的收入函数模型时，它就很有可能与教育相关，因为内在的能力和家庭背景常常是教育的决定性因素。正像伍尔德里奇（Wooldridge）所主张的，“在很多应用中，适合于使用综列数据的全部原因是允许不能被观测到的效应（即 ε_i）可以与解释变量相关。”[11]

ECM 的基本假设是，ε_i 是一个从非常大的总体中提取的随机变量。但

是有时候情况并非如此。比如假定我们想对美国50个州的犯罪率进行研究。显然，在这种情况下，50个州是一个随机样本这样的假设是不可靠的。

记住这两种方法之间的基本区别后，关于FEM和ECM之间的选择我们还有什么可以说的吗？这里，贾奇所做的判断也许是有帮助的[12]：

1. 若 T（时间序列数据的数目）较大而 N（横截面单元的数量）较小，那么通过FEM和ECM估计的参数值之间很可能没什么差别。于是，这时的选择依据就基于计算上的便利了。从这个理由来看，FEM可能更加可取。

2. 当 N 较大而 T 较小时，两种方法的估计值会有显著的差异。回忆一下，在ECM中，$\beta_{1i}=\beta_1+\epsilon_i$，这里 ϵ_i 是横截面随机部分，而在FEM中我们认为 β_{1i} 固定不变而非随机变量。在后一种情况下，对于样本中被观测的横截面单元的统计推断是有条件的。如果确信我们样本中个体或横截面单元不是从一个较大的样本中随机取出的，那么FEM是合适的。但如果样本中的横截面单元被看作随机抽取，那么ECM是合适的，因为在这种情况下统计推断是无条件的。

3. 如果个别的误差部分 ϵ_i 与一个或多个回归元是相关的，那么ECM估计量是有偏差的，而从FEM中获得的估计量则是无偏的。

651 4. 如果 N 较大而 T 较小，并且坚持ECM的基本假设，那么ECM估计量比FEM估计量更有效。[13]

是否有一种正式的检验可以帮助我们在FEM和ECM之间做出选择？有，豪斯曼（Hausman）于1978年提出了一个检验。[14]我们将不讨论这个检验的细节，因为它们超出了本书的范围。[15]豪斯曼检验的基本虚拟假设是FEM和ECM估计量没有实质上的差异。豪斯曼提出的检验统计量有一个渐近的 χ^2 分布。若虚拟假设被拒绝了，结论就是ECM不合适而使用FEM可能会更好。这种情况下，样本中的 ϵ_i 的统计推断将是有条件的。

抛开豪斯曼检验，记住约翰逊和迪那多（DiNardo）所提出的警告是很重要的。在决定选择固定效应或随机效应模型时，他们提出："……没有一个简单的规则可以帮助研究者航行通过固定效应这块巨石以及误差量度和动态选择这块暗礁。尽管在横截面数据方面它们有了改善，但是综列数据不是解决计量经济学家所有问题的灵丹妙药。"[16]

§16.6 综列数据回归：一些结论性的意见

正如开始时所特别提到的，建立综列数据模型的主题是庞大而且复杂的。我们仅仅只是触及它的皮毛。在我们没有讨论的主题中，可能会碰到以下的情况。

1. 综列数据的假设检验。

2. ECM 中的异方差性和回归性。

3. 非平衡综列数据。

4. 动态综列模型，其中回归子(Y_{it})的滞后值作为一个解释变量而出现。

5. 涉及综列数据的联立方程。

6. 定性因变量和综列数据。

这些主题的一个或多个能够在本章中所引用的参考书中找到，更多地了解这一主题会促使读者去参考它们。这些参考书也引用了几个商业和经济的不同领域中的实证研究，而这些研究使用的是综列数据回归模型。建议初学者去读一些应用实例，从而对研究者在实际中如何运用这样的模型有一个感性的认识。

§16.7 要点与结论

652 1. 综列回归模型以综列数据为基础。综列数据由同一个横截面或个体单元在几个时间段上的观测组成。

2. 使用综列数据有几个好处。第一，它们大大增加了样本大小。第二，通过重复研究横截面观测值，综列数据更加适合研究变化的动态。第三，综列数据使我们能够研究更为复杂的行为模型。

3. 尽管有这些确实的优点，综列数据也提出了几个估计和推断的问题。由于数据涉及横截面和时间两个维度，因而困扰横截面数据（例如异方差性）和时间序列数据（例如自相关）的问题就需要提及。还有一些其他问题，比如说在同一时点上个体单元的交叉相关。

4. 有几个估计的技巧来解决一个或多个这些问题。两个最为著名的是（1）固定效应模型（FEM），（2）随机效应模型（REM）或误差组成模型(ECM)。

5. 在 FEM 中，回归模型的截距允许在个体间存在差异，也就是说承认每个个体或横截面单元有一些自己的特性这一事实。为了考虑不同的截距，我们可以使用虚拟变量。FEM 中所使用的虚拟变量就是著名的最小二乘虚拟变量（LSDV）模型。FEM 适用于个体的特定截距与一个或多个回归元相关的情形。LSDV 的一个缺点就是，当横截面单元的数目 N 非常大时，我们将不得不引入 N 个虚拟变量（但禁止了共同截距项），从而消耗了许多自由度。

6. ECM 是 FEM 的一个替换物。ECM 中假定个体单元的截距是从一个有着不变均值的非常大的总体中随机抽取得到。那么，单个截距可以表示为对这个不变均值的偏离。ECM 相对 FEM 而言的一个优点就是在自由度上是经济的，因为我们不必估计 N 个横截面截距，我们只需估计截距的均值以及方差。ECM 使用与每个横截面单元的随机截距与回归元不相关的情形。

7. 豪斯曼检验可用于决定 FEM 和 ECM 的选择。

8. 尽管综列数据回归在应用研究中越来越受欢迎，综列数据也越来越容易得到，但是它也不是适用于任何情形。在每种情况下都必须做出一些实际的判断。

习 题

问答题

16.1 (a) 横截面数据、(b) 时间序列数据和 (c) 综列数据的特点是什么？

658 **16.2** 一个固定效应模型 (FEM) 表示什么意思？既然综列数据同时有时间和空间两个维度，那么，FEM 是怎样考虑这两个维度的？

16.3 一个误差组成模型 (ECM) 表示什么意思？它与 FEM 有什么不同？什么时候 ECM 是适用的？什么时候 FEM 是适用的？

16.4 FEM、最小二乘虚拟变量 (LSDV) 模型和协方差模型三者之间有区别吗？

16.5 什么时候综列数据回归模型是适用的？举一些例子。

16.6 怎样扩展模型 (16.4.4) 以考虑一个时间误差部分？这时方程 (16.3.6)、(16.3.7) 和 (16.3.8) 会有什么情况发生？

16.7 考虑表1.1 中给出的鸡蛋产量和价格数据。这里哪一个模型是适用的，FEM 还是 ECM？为什么？

16.8 (16.3.4) 的回归结果中，四个公司的固定效应截距是什么？这些效应是否具有统计上的差异？

16.9 对于本章中讨论的投资例子，表 16.3 给出了基于 ECM 的结果。如果将 (16.3.4) 中所给出的结果与之比较，你可以得出什么样的一般性结论？

16.10 基于密歇根收入动态研究，豪斯曼试图估计一个工资或收入模型。这个模型将 629 名高校毕业生作为一个样本，对这些毕业生做一个为期 6 年的调查，从而得到 3 774 次观测值。这个研究中因变量是工资的对数，解释变量是年龄（分成几个年龄组）、前一年的失业情况、前一年健康不佳的情况、自我雇佣的情况、居住地的方位（南 = 1，其他为 0），居住地的区域（农村 = 1，其他为 0）。豪斯曼同时使用了 FEM 和 ECM。表 16.4 中给出了结果（括号里是标准差）：

a. 结果是否具有经济学意义？

b. 两个模型所产生的结果是否存在着巨大的差异？如果有，什么可以解释这些差异？

c. 基于表中所给出的数据，如果可能的话，你会选择哪个模型？

表 16.4 工资等式(因变量:工资的对数*)

变量	固定效应		随机效应	
1. 年龄段 1(20~35)	0.055 7	(0.004 2)	0.039 3	(0.003 3)
2. 年龄段 2(35~45)	0.035 1	(0.005 1)	0.009 2	(0.003 6)
3. 年龄段 3(45~55)	0.020 9	(0.005 5)	−0.000 7	(0.004 2)
4. 年龄段 4(55~65)	0.020 9	(0.007 8)	−0.009 7	(0.006 0)
5. 年龄段 5(65 岁以上)	−0.017 1	(0.015 5)	−0.042 3	(0.012 1)
6. 前一年的失业率	−0.004 2	(0.015 3)	−0.027 7	(0.015 1)
7. 前一年的健康欠佳	−0.020 4	(0.022 1)	−0.025 0	(0.021 5)
8. 自我雇佣	−0.219 0	(0.029 7)	−0.267 0	(0.026 3)
9. 南部	−0.156 9	(0.065 6)	−0.032 4	(0.033 3)
10. 农村	−0.010 1	(0.031 7)	−0.121 5	(0.023 7)
11. 常数	-	-	0.849 9	(0.043 3)
s^2	0.056 7		0.069 4	
自由度	3 135		3 763	

* 3 774 观测值;括号里是标准误。

复制于 Cheng Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986, p.42. *Original*。

资料来源:J.A.Hausman, "Specification Tests in Econometrics," *Econometrica*, vol.46, 1978, pp. 1251－1271.

654

解答题

16.11 参考表1.1 中的数据。

a. 令 Y = 鸡蛋产量(百万计)和 X = 鸡蛋价格(美分每打)。分别在 1990 和 1991 年对模型:$Y_i = \beta_1 + \beta_2 X_i + \mu_i$ 进行估计。

b. 混合这两年的观测值并估计混合回归模型。在混合数据时你要做什么样的假设?

c. 使用固定效应模型,将两年进行区分,求出回归结果。

d. 区分 50 个州后,你能够使用固定效应模型吗?为什么可以或者不可以?

e. 将州效应和年效应进行区分有意义吗?如果有,那么你将不得不引入多少个虚拟变量?

f. 误差组成模型是否适用于鸡蛋产量的建模?为什么是或者不是?看看你是否能用比如说 Eviews 对这样一个模型进行估计。

16.12 接着习题16.11。在决定进行混合回归之前,你需要弄清数据是否"可以混合"。为了这个目的你决定使用第 8 章中讨论过的邹至庄检验。写出所涉及的必要计算并决定混合回归是否有意义。

16.13 回到16.2 中所讨论的格伦费尔德投资函数。

a. 单独地对通用电气、通用汽车、美国钢铁以及西屋电气的投资函数进行估计。所有 80 次观测值的混合结果已经由 (16.3.1) 给出。

b. 为了确定混合回归是否合适，你决定运用第 8 章中讨论过的邹至庄检验。进行这个检验。提示：从混合回归中得到 RSS，再从每个投资函数中求得 RSS，然后应用邹至庄检验。

c. 从邹至庄检验中你得出了什么结论？如果你的结论是不能混合数据，那么对于综列数据回归方法的实用性你能说些什么呢？

16.14 表16.5 中给出了 1980—1999 年加拿大、英国以及美国的公民失业率 Y（%）和用美元表示的小时工资（指数，1992 = 100）。考虑模型：

表 16.5　美国、加拿大以及英国制造业的失业率和小时工资，1980—1999

观测值	美国 小时工资，美元/小时	美国 失业率，%	加拿大 小时工资，美元/小时	加拿大 失业率，%	英国 小时工资，美元/小时	英国 失业率，%
1980	55.6	7.1	49.0	7.2	43.7	7.0
1981	61.1	7.6	54.1	7.3	44.1	10.5
1982	67.0	9.7	59.6	10.6	42.2	11.3
1983	68.8	9.6	63.9	11.5	39.0	11.8
1984	71.2	7.5	64.3	10.9	37.2	11.7
1985	75.1	7.2	63.5	10.2	39.0	11.2
1986	78.5	7.0	63.3	9.2	47.8	11.2
1987	80.7	6.2	68.0	8.4	60.2	10.3
1988	84.0	5.5	76.0	7.3	68.3	8.6
1989	86.6	5.3	84.1	7.0	67.7	7.2
1990	90.8	5.6	91.5	7.7	81.7	6.9
1991	95.6	6.8	100.1	9.8	90.5	8.8
1992	100.0	7.5	100.0	10.6	100.0	10.1
1993	102.7	6.9	95.5	10.7	88.7	10.5
1994	105.6	6.1	91.7	9.4	92.3	9.7
1995	107.9	5.6	93.3	8.5	95.9	8.7
1996	109.3	5.4	93.1	8.7	95.6	8.2
1997	111.4	4.9	94.4	8.2	103.3	7.0
1998	117.3	4.5	90.6	7.5	109.8	6.3
1999	123.2	4.0	91.9	5.7	112.2	6.1

小时工资以美元计，指数 1992 = 100。

资料来源：*Economic Report of the President*, January 2001, Table B109, p.399.

$$Y_{it} = \beta_1 + \beta_2 X_{it} + u_{it} \tag{1}$$

a. 作为先验的,期望 Y 与 X 之间是什么关系？为什么？

b. 对每个国家的模型（1）进行估计。

c. 混合所有 60 次观测后估计模型。

d. 估计固定效应模型。

e. 估计误差组成模型。

f. 哪个模型较好，FEM 还是 ECM？说出你的理由。

【注释】

［1］一些参考书是 G.Chamberlain,"Panel Data",in *Handbook of Econometrics*, vol. Ⅱ, Z.Griliches and M.D.Intriligator, eds., North-Holland Publishers, 1984, Chap. 22; C.Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986; G.G.Judge, R.C.Hill, W.E.Griffiths, H.Lutkepohl, and T.C.Lee, *Introduction to the Theory and Practice of Econometrics*, 2d ed., John Wiley &Sons, New York, 1985, Chap.11; W.H.Greene, *Econometric Analysis*, 4th ed., Prentice-Hall, Englewood Cliffs, N.J., 2000, Chap.14; Badi H.Baltagi, *Econometric Analysis of Panel Data*, John Wiley and Sons, New York, 1995; and J.M.Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Mass., 1999.

［2］Baltagi op.cit., pp.3－6.

［3］Y.Grunfeld,"The Determinants of Corporate Investment," unpublished Ph.D. thesis, Department of Economics, University of Chicago, 1958. 这些数据在好几本书中被引用。我们的数据取自 H.D.Vinod 和 Aman Ullha, *Recent Advances in Regression Methods*, Marcel Dekker, New York, 1981, pp. 259－261。格伦费尔德研究已成为教科书作者的最爱,因为出于解释性目的,这些数据易于处理。

［4］每年我们只对回归子和回归元进行四次观测。如果考虑截距，我们将不得不估计三个参数，这样只剩下一个自由度。显然，这样的回归是没有意义的。

［5］这一讨论受到的影响来自 Judge et al., op.cit.,和 Hsiao, op. cit., pp.9－10。

［6］除了注释 1 中提到的那些书，还可参见 Terry E.Dielman, *Pooled Cross-sectional and Time Series Data Analysis*, Marcel Dekker, New York, 1989, 和 Lois W.Sayrs, *Pooled Time Series Analysis*, Sage Publications, Newbury Park, California, 1989。

［7］这就导致所谓似然无关回归(SURE)模型，它最早由 Arnold Zellner 提出。关于这个模型的讨论参见 Terry E. Dielman, op.cit.。

［8］Dielman, op.cit., Sayrs, op.cit., Jan Kmenta, *Elements of Econometrics*, 2d ed., Macmillan, New York, 1986, Chap.12.

［9］Kmenta, op. cit., p.633.

［10］有兴趣的读者可以参考 Kmenta, op.cit., pp.625－630,其中的讨论易于理解。

［11］Wooldridge, op.cit., p.450.

［12］Judge et al., op.cit., pp.489－491.

［13］Taylor 已经表明，若 $T \geqslant 3$ 以及 $N-K \geqslant 9$，这种说法成立，其中 K 表示回归元的个数。参见 W.E.Taylor, "Small Sample Considerations in Estimation from Panel

Data," *Journal of Econometrics*, vol. 13, 1980, pp. 203 – 223。

[14] J.A.Hausman, "Specification Tests in Econometrics," *Econometrica*, vol. 46, 1978, pp. 1251 – 1271.

[15] 细节参见 Baltagi, op.cit., pp. 68 – 73。

[16] Jack Johnson and John DiNardo, *Econometric Methods*, 4th ed., McGraw-Hill, 1997, p. 403.

第 17 章　动态计量经济模型：自回归与分布滞后模型

656 在涉及时间序列数据的回归分析中，如果回归模型不仅含有解释变量（诸 X）的当前值，还含有它们的滞后（过去）值，就把它称为**分布滞后模型**（distributed-lag model）。如果模型在它的解释变量中包含有因变量的一个或多个滞后值，就称它为**自回归模型**（autoregressive model）。例如：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$$

代表一个分布滞后模型，而

$$Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t$$

则代表一个自回归模型。后者由于描述了因变量相对于它的过去值的时间走径，又称**动态模型**（dynamic models）。

自回归与分布滞后模型广泛地应用于计量经济分析之中。本章对这类模型做一周密的考察以明确下述问题：

1. 滞后在经济学中的作用如何？
2. 滞后的理由是什么？
3. 在经验计量经济学中常用的滞后模型有什么理论上的依据吗？
4. 自回归与分布滞后模型之间有没有关系？如果有，又是什么关系？能从一个模型得到另一个吗？

657 5. 在估计这类模型时会遇到一些什么统计问题？

6. 变量之间的领先/滞后关系意味着因果关系吗？如果是这样，将怎样测出？

§17.1 “时间”或“滞后”在经济学中的作用

在经济学中，变量 Y（因变量）对另一（些）变量 X（解释变量）的依赖很少是瞬时的。常见的情形是，Y 对 X 的回应有一个时间的延迟，这种时间延迟就叫做滞后（lag）。为了说明滞后的性质，我们考虑几个例子。

例 17.1

消费函数。假定某人的年薪增加了 2 000 美元，并假定它是一种“永久性”增加，即这一薪金的增加将一直保持下去。那么，这种收入的增加将会对个人的年消费支出产生什么影响？

在得到收入的这种增加之后，人们通常并不急于把全部增加的收入马上花掉。比如说，受益者也许决定在收入增加后的第一年增加 800 美元的消费支出，下一年增加 600 美元，再下一年增加 400 美元，而把所余的部分用于储蓄。到第三年末，此人的年消费支出将增加 1 800 美元。于是我们可把消费函数写成：

$$Y_t = \text{常数} + 0.4X_t + 0.3X_{t-1} + 0.2X_{t-2} + u_t \tag{17.1.1}$$

其中 Y 是消费支出，而 X 是收入。

方程（17.1.1）表明，2 000 美元的收入增加的效应散布或分布在一个 3 年的期间里。因此，像（17.1.1）这样的模型，由于某一原因（收入）而产生的效果分散在若干时期里而被称为**分布滞后模型**。分布滞后模型（17.1.1）的几何意义可由图 17.1 或图 17.2 看出。

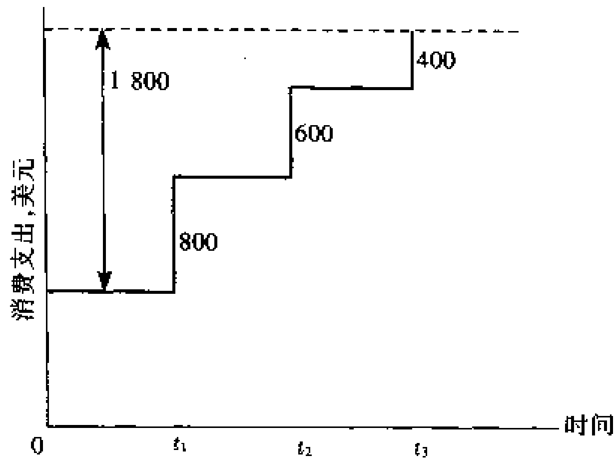


图 17.1 分布滞后的例子

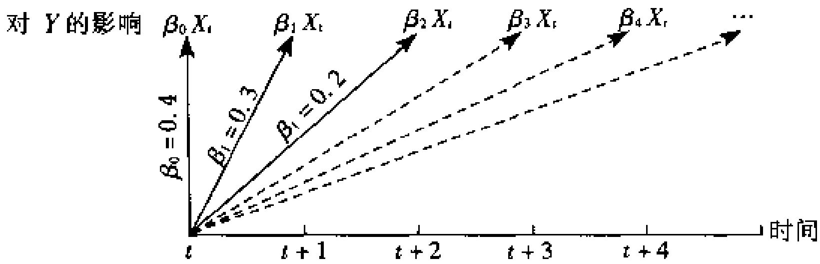


图 17.2　在时刻 t 的 X 的单位变化对时刻 t 及随后时期 Y 的影响

658 更一般地，我们可写成：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \tag{17.1.2}$$

这是带有 k 个时期的有限滞后的一个分布滞后模型。系数表示随着 X 的一单位变化，Y 均值的同期变化，故称**短期**或**即期**乘数。[1] 如果此后 X 的变化都保持在同一水平上，则（$\beta_0+\beta_1$）给出下期 Y(均值）的变化，（$\beta_0+\beta_1+\beta_2$）给出再下期 Y 的变化，如此类推。这些部分的和称**中期乘数**。最后，经过 k 期之后，我们得到：

$$\sum_{i=0}^{K} \beta_i = \beta_0 + \beta_1 + \beta_2 + \cdots + \beta_k = \beta \tag{17.1.3}$$

称**长期或总分布滞后乘数**，但这里须假定总和 β 存在（将在另处讨论）。如果我们定义：

$$\beta_i^* = \frac{\beta_i}{\sum \beta_i} = \frac{\beta_i}{\beta} \tag{17.1.4}$$

就得到"标准化" β_i。于是，标准化 β_i 的部分和将给出在某一时期所感应的冲击占长期或总冲击（即总滞后乘数）的比例。

回到消费回归（17.1.1)，我们看到短期乘数 0.4 不过是短期边际消费倾向，而长期乘数 0.4+0.3+0.2=0.9 则是长期边际消费倾向。也就是说，随着收入增加 1 美元，消费者将在收入增加的当年提高他或她的消费水平约 40 美分，在下一年里再提高 30 美分，再下一年里又再提高 20 美分，这样，收入的 1 美元增加的长期效应就是 90 美分。如果我们将每一 β_i 除以 0.9，就分别是 0.44，0.33 和 0.23，表明 X 一单位变化的总效应有 44% 被立即感应，一年以后有 77% 被感应，而到第二年末达到 100%。

例 17.2　银行货币(活期存款)的创造

假如联邦储备系统通过买进政府证券向银行系统倾注了 1 000 美元的新货币，最终将产生的银行货币或活期存款的总额是多少?

按照部分准备金体制，假定法律规定银行对它们创造的存款要持 20% 的准备金作为后盾，那么，根据熟知的乘数过程，一共要产生的活期存款总额将等于 1 000 美元[1/(1－0.8)]＝5 000 美元。当然，这 5 000 美元的活期存款不是一夜之间产生的，它有一个时间过程，如图 17.3 所示。

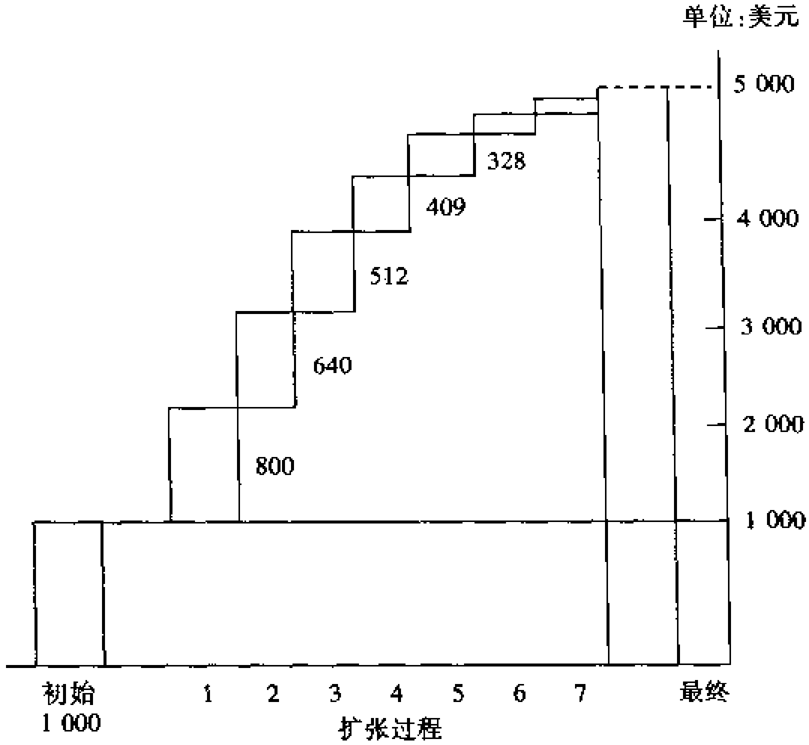


图 17.3 银行存款的累积性扩张
(初始准备金 1 000 美元及 20%的法定准备金率)

例 17.3 货币与价格之间的联结

根据货币主义学派的观点，通货膨胀实质上是一种货币现象。其意义在于一般价格的连续增长，是由于货币供给的膨胀率远远超过经济单位对货币的实际需求量所致。当然，通货膨胀和货币供给的变化之间的联带关系不是瞬时的。研究表明，两者之间的联结有一个介乎 3 到 20 个季度的滞后。表 17.1 展示这类研究的一个结果。[2] 从中看到，M1B 货币供给（= 在金融机构中的活期存款 + 支票存款）的 1% 变化被感应到长达 20 个季度里的后效。货币供给的（1%）变化对通货膨胀的长期冲击约为 $1(=\sum m_i)$，这个数字是统计上显著的。尽管中期乘数一般看来都是显著的，但约为 0.04 的短期冲击则是不显著的。顺便指出，由于 P 和 M 两者均以百分比形式表示，m_i（通常我们用符号 β_i）给出了 P 对 M 的弹性，也就是当货币供给增加 1% 时，价格在百分比变化上的响应。因此，$m_0=0.041$ 意味着对于货币供给的 1% 的增加，短期价格弹性约为 0.04%。长期弹性为 1.03%，意味着长期而论，货币供给增加 1%，反映在价格上的百分比变化也正好相当。简单说来，从长远看，货币供给增加 1%，通货膨胀率也随之增加 1%。

表 17.1　　货币—价格方程的估计:原始设定
样本时期:1955 年第 1 季度至 1969 年第 4 季度:$m_{21}=0$

$$\dot{P} = -0.146 + \sum_{i=0}^{20} m_i \dot{M}_{-i}$$
$$(0.395)$$

	系数	$\lvert t \rvert$		系数	$\lvert t \rvert$		系数	$\lvert t \rvert$
m_0	0.041	1.276	m_8	0.048	3.249	m_{16}	0.069	3.943
m_1	0.034	1.538	m_9	0.054	3.783	m_{17}	0.062	3.712
m_2	0.030	1.903	m_{10}	0.059	4.305	m_{18}	0.053	3.511
m_3	0.029	2.171	m_{11}	0.065	4.673	m_{19}	0.039	3.338
m_4	0.030	2.235	m_{12}	0.069	4.795	m_{20}	0.022	3.191
m_5	0.033	2.294	m_{13}	0.072	4.694	$\sum m_i$	1.031	7.870
m_6	0.037	2.475	m_{14}	0.073	4.468	平均滞后	10.959	5.634
m_7	0.042	2.798	m_{15}	0.072	4.202			
$\bar{R}^2$	0.525							
se	1.066							
D.W.	2.00							

符号:$\dot{P}$ = GNP 价格平缩因子的年复合变化率
$\dot{M}$ = MIB 的年复合变化率
资料来源:Keith M.Carlson,"The Lag from Money to Prices,"*Review*,Federal Reserve Bank of St.Louis,October 1980,Table 1,P.4.

661 **例 17.4　R&D 支出与生产力之间的滞后**

研究与开发(R&D)的投资决定以至最终获致生产力增进的报偿,涉及相当长期的滞后,诸如"……资金投放与发明创造开始出现的时间之间的滞后,意念上或装置上的发明创造与发展到商业上可应用的阶段之间的滞后,以及扩散过程中的滞后:在旧机器全被较好的新机器替换之前需要时间。"[3]

上述诸例只不过是在经济学中使用到滞后的一个样本。无疑,读者还能根据他或她自己的经验提供一些例子。

例 17.5　国际经济学中的 *J* 曲线

学习过国际经济学的学生应该对 J 曲线相当熟悉了,该曲线显示了贸易平衡和货币贬值的关系。假定其他条件不变,随着一个国家的货币贬值,起初贸易平衡恶化,但最终会得到改善。该曲线如图 17.4 所示。

例 17.6　投资的加速模型

在其最简单的形式中,投资理论的加速原理揭示出投资是与产出的变化

成比例的。象征性地，

$$I_t = \beta(X_t - X_{t-1}) \qquad \beta > 0 \tag{17.1.5}$$

此处 I_t 是指 t 时刻的投资，X_t 是指 t 时刻的产出，X_{t-1} 是指（$t-1$）时刻的产出。

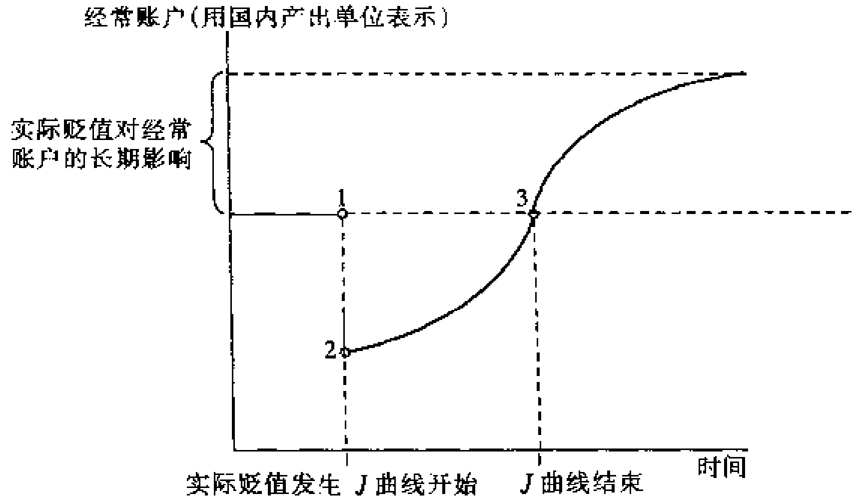


图 17.4　J 曲线

资料来源：Paul R. krugman and Maurice Obstfeld, *International Economics: Theory and Practice*, 3d ed., Harper Collins, New York, 1994, p.465.

§ 17.2　滞后的原因

虽然第 17.1 节所引用的例题指出了滞后现象的性质，但并没有完全解释为什么会出现这些滞后。滞后的主要原因有三：

1. **心理上的原因**。作为一种习惯势力（惰性）的结果，人们在价格上升或收入增加之后，并不马上改变他们的消费习惯，也许因为改变的过程会带来一些直接的负效应。例如，由于彩票中奖瞬时变成百万富翁的人们，也许不会改变他们已长期适应的生活方式，因为他们不知道怎样对这种意外之财立即做出反应。当然，给予合理的时间，他们也许会学会怎样使用他们新得的巨款。再则，人们不一定知道某种变化是“永久”性的或“临时”性的。因此，我对我的收入增加的反应要看这种增加是否永久而定：如果这是一种不再现的增加，而且在以后的时期里我的收入将回到原来的水平，那么我也许会把全部增加的收入储蓄起来。但若别人是我的话，也许会“把它花掉”。

663 2. **技术上的原因**。假使相对于劳动力而言，资本的价格下跌致使用资本代替劳动力较为经济。无疑，资本的添置需要时间（孕育时期）。此外，如果人们预期价格下跌是暂时现象，特别是在资本价格的暂时下跌之后会回升到原先水平上，厂商就不会匆忙用资本去代替劳动力。有时，不完全的知

识也说明滞后现象。目前各种性能和价格的计算机充斥着个人计算机市场，而且，自从 20 世纪 70 年代后期个人计算机被引入以来，大多数价格均急剧下跌。结果，个人计算机的未来用户均等待观望各种竞争商标的性能与价格而在购买上迟疑不决。之所以迟疑不决，也许因为他们期待着价格进一步下跌和技术创新。

3. **制度上的原因**。这种原因也造成滞后。例如，契约上的义务也许妨碍厂商从一个劳动力或原料来源转换到另一个来源。作为另一个例子，一些人已将他们的奖金存放到有固定存期比如说一年、三年或七年的长期储蓄账户之中，那么，尽管货币市场情况表明资金的另行处置会给他们带来更高的收益，可是他们已基本上被“锁定”了。同理，雇主常常让雇员在几个健康保险计划之中选择一个。但是一旦选定，雇员也就在一年之内不能从一个保险计划调换到另一个。虽然这种规定是为了行政管理上的便利而做出的，却把雇员锁定了一年之久。

由于上述原因，滞后在经济学中占有中心的地位。这点明显地反映在经济学的短期—长期方法论中。正是出于这种理由，我们说短期价格或收入弹性一般小于（从绝对值上看）相应的长期弹性，以及短期边际消费倾向一般小于长期边际消费倾向。

§ 17.3 分布滞后模型的估计

既然分布滞后模型在经济学中扮演着一个高度有用的角色，我们怎样去估计这样的模型呢？具体地说，假如我们有如下的一个解释变量的分布滞后模型[5]：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + u_t \tag{17.3.1}$$

其中我们尚未规定滞后的长度，也就是说我们没有明确要回到多远的过去。这样的一个模型叫做**无限（滞后）模型**[infinite (lag) model]，而像（17.1.2）一类的模型，则由于滞后长度 k 已设定而叫做**有限（滞后）分布滞后模型**[finite (lag) distributed-lag model]。但如我们将会看到的，（17.3.1）易于数学上的处理，我们将继续使用它。[6]

怎样估计（17.3.1）中的 α 和诸 β 呢？可以采用两种方法：(1) 现式估计法（ad hoc estimation）和（2）限定诸 β 遵从某种变化模式的先验约束法。本节将考虑现式估计法，在第 17.4 节中再考虑另一方法。

分布滞后模型的现式估计法

664 因解释变量 X_t 被假定为非随机的（或至少是与干扰项 u_t 无关的），故

X_{t-1}，X_{t-2}，等等也是非随机的。因此，原则上，普通最小二乘法可用于(17.3.1)的估计。这是阿尔特（Alt）[7]和廷伯根（Tinbergen）[8]所采取的方法。他们建议序贯地对（17.3.1）进行估计，即首先将 Y_t 对 X_t 回归，然后将 Y_t 对 X_t 和 X_{t-1} 回归，然后将 Y_t 对 X_t，X_{t-1} 和 X_{t-2} 回归，如此类推。这一序贯程序将终止于滞后变量的回归系数开始变成统计上不显著或至少有一个变量的系数改变符号（由正变负或由负变正）之时。按照这一规则，阿尔特曾求得燃烧油耗量 Y 对新订货量 X 的回归。他根据 1930—1939 年期间的季度数据，获得以下结果：

$$\hat{Y}_t = 8.37 + 0.171X_t$$

$$\hat{Y}_t = 8.27 + 0.111X_t + 0.064X_{t-1}$$

$$\hat{Y}_t = 8.27 + 0.109X_t + 0.071X_{t-1} - 0.055X_{t-2}$$

$$\hat{Y}_t = 8.32 + 0.108X_t + 0.063X_{t-1} + 0.022X_{t-2} - 0.020X_{t-3}$$

阿尔特选择第二个回归作为“最优”的一个，这是因为，在最后的两个方程中，X_{t-2} 符号表现不稳定，并且在最后的一个方程中，X_{t-3} 的符号为负而难于从经济意义上做出解释。

现式估计法虽然看起来简单明了，却有如下的一些缺点：

1. 滞后的最大长度是什么，没有任何先验性的指引。[9]

2. 在估计相继的滞后过程中，剩下来的自由度越来越少，致使统计推断多少有点不可靠。通常经济学家没有那般幸运，能获得一个足够长的数据序列，去继续不断地估计许许多多的滞后。

3. 更重要的是，在经济时间序列数据中，相继的（滞后）值一般都是高度相关的；多重共线性的阴影笼罩着整个估计问题。如第 10 章所表明的，多重共线性导致不准确的估计；就是说，标准误相对于所估系数来说有变大的倾向。结果，根据通常计算的 t 比率，我们就会倾向于（错误地）声称（诸）滞后系数是在统计上不显著的。

4. 对滞后长度的序贯寻找，将使研究者受到**数据开采**（data mining）的指控。而且，如我们在第 13.4 节中所看到的，在这种序贯寻找中，检验假设的名义和真实显著性水平将成为一个重要的争论问题。

鉴于上述问题，现式估计法没有什么值得我们推荐的。显然，如果我们要在估计的问题上取得进展，则必须对诸 β 做出某些先验性的或理论上的考虑。

§ 17.4 分布滞后模型的考伊克方法

考伊克（Koyck）曾提出一种估计分布滞后模型的巧妙方法。假使我们从无限滞后的分布滞后模型（17.3.1）开始，设想全部 β 都有相同的符号，考伊克假定它们是按如下的几何级数项衰减的。[10]

$$\beta_k = \beta_0 \lambda^k \qquad k = 0, 1, \cdots \tag{17.4.1}$$ [11]

其中 λ $(0<\lambda<1)$ 称为分布滞后的衰减率，而 $1-\lambda$ 称调节速度。

(17.4.1) 所公设的是，每个后继的 β 系数在数值上要小于先前的 β (由 $\lambda<1$ 推知)，意味着当我们追溯到越是遥远的过去，该滞后对 Y 的影响就越小。这是一个相当合理的假定。毕竟，可以预料，当前和不久前的收入要比遥远过去的收入对当前消费的影响来得大。图 17.5 描绘了考伊克模式的几何意义。

666 如图所示，滞后系数 β_k 的值，除了与公共的 β_0 有关外，还依赖于 λ 值。λ 越靠近 1，β_k 衰减的速度越慢，而 λ 越靠近零，β_k 衰减的速度越快。在前一种情形中，X 的遥远过去值对 Y 仍有可观的影响，而在后一种情形中，这种影响很快就消失。这种模式可清晰地从下面的数字说明中辨认出来：

γ	β_0	β_1	β_2	β_3	β_4	β_5	$\cdots$	β_{10}
0.75	β_0	$0.75\beta_0$	$0.56\beta_0$	$0.42\beta_0$	$0.32\beta_0$	$0.24\beta_0$	$\cdots$	$0.06\beta_0$
0.25	β_0	$0.25\beta_0$	$0.06\beta_0$	$0.02\beta_0$	$0.004\beta_0$	$0.001\beta_0$	$\cdots$	0.0

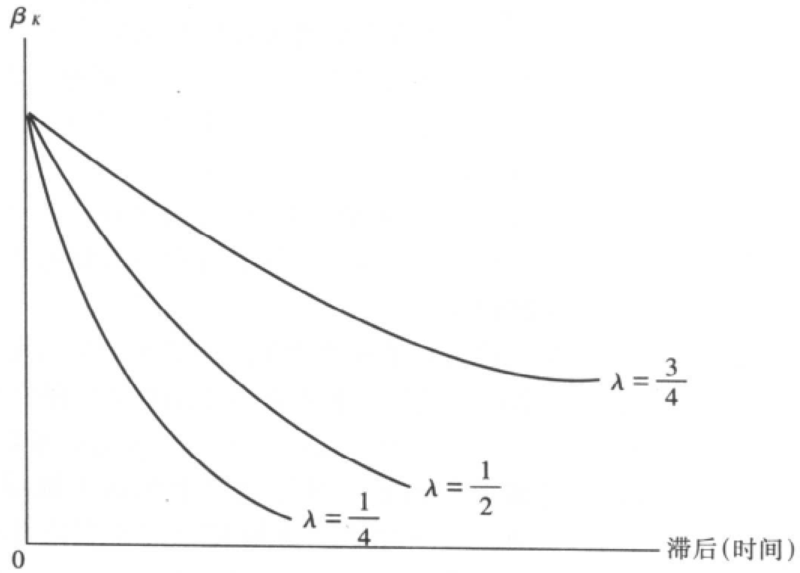


图 17.5 考伊克模式（下降的几何分布）

注意考伊克模式的以下特点：(1) 一旦假定 λ 取非负值，即排除诸 β 变号的可能性；(2) 假定了 $\lambda<1$，即对遥远的诸 β 比对近期的诸 β 给予较小的权重；以及 (3) 确保长期乘数即 β 的总和是有限的值，就是，

$$\sum_{k=0}^{\infty} \beta_k = \beta_0 \left(\frac{1}{1-\lambda} \right) \tag{17.4.2}$$ [12]

由于 (17.4.1)，无限滞后模型 (17.3.1) 可写为：

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \lambda X_{t-1} + \beta_0 \lambda^2 X_{t-2} + \cdots + u_t \tag{17.4.3}$$

从现在的情形看，由于大量的（字面上看是无限多的）参数有待估计，而且 λ 以高度非线性形式出现，故模型还不是易于估计的：严格地说，（对参数而言的）线性回归分析方法不适用于这类模型。然而考伊克提出了一个创造性的解决方法。他将（17.4.3）滞后一期而得到：

$$Y_{t-1}=\alpha+\beta_0X_{t-1}+\beta_0\lambda X_{t-2}+\beta_0\lambda^2X_{t-3}+\cdots+u_{t-1} \tag{17.4.4}$$

然后用 λ 乘（17.4.4）得：

$$\lambda Y_{t-1}=\lambda\alpha+\lambda\beta_0X_{t-1}+\beta_0\lambda^2X_{t-2}+\beta_0\lambda^3X_{t-3}+\cdots+\lambda u_{t-1} \tag{17.4.5}$$

从（17.4.3）减去（17.4.5）又得：

$$Y_t-\lambda Y_{t-1}=\alpha(1-\lambda)+\beta_0X_t+(u_t-\lambda u_{t-1}) \tag{17.4.6}$$

667 经过整理即有：

$$Y_t=\alpha\ (1-\lambda)\ +\beta_0X_t+\lambda Y_{t-1}+v_t \tag{17.4.7}$$

其中 $v_t=(u_t-\lambda u_{t-1})$，为 **$u_t$ 和 u_{t-1} 的一个移动平均。**

刚才描述的程序称为**考伊克变换**。拿（17.4.7）同（17.3.1）相比，我们看出考伊克成功地获得了一个巨大的简化结果。在这之前，我们必须估计 α 以及无限多个 β 值，而现在我们只须估计 3 个未知参数：α，β_0 和 λ。现在再没有预期着多重共线性的理由。从某种意义上来说，我们用单一变量 Y_{t-1} 代替 X_{t-1}，X_{t-2}，…，就解决了多重共线性的问题。但注意考伊克变换的以下特点：

1. 由于 Y_{t-1} 为解释变量之一，我们从一个分布滞后模型开始，却以一个自回归模型告终。这一变换表明我们怎样能把一个分布滞后模型“转变”为一个自回归模型。

2. Y_{t-1} 的出现很可能产生一些统计上的问题。Y_{t-1} 和 Y_t 一样，是随机的，意味着我们的模型包含有一个随机的解释变量。回顾一下经典最小二乘理论是建立在这样的假定基础上的：就是，解释变量或者是非随机的，或者，如果是随机的话，将是独立于随机干扰项的。因此，我们必须明确 Y_{t-1} 是否满足这一假定。（在第 17.8 节中我们将回到这个问题上。）

3. 在原始模型（17.3.1）中干扰项是 u_t，而在转换模型中它变为 $v_t=(u_t-\lambda u_{t-1})$。$v_t$ 的统计性质依赖于我们对 u_t 的统计性质的假定。以后我们将会看到，即使原始的 u_t 是序列无关的，v_t 却是序列相关的。因此，除了随机解释变量 Y_{t-1} 之外，我们还要面对序列相关的问题。我们将在第 17.8 节中处理这个问题。

4. 滞后 Y 的出现违背了使用德宾-沃森检验的基本假定之一。因此，在滞后 Y 出现时，还有必要推导出检验序列相关的另外方法。方法之一是**德宾 h 检验**（Durbin h test），将在第 17.10 节中讨论。

668 如在（17.1.4）中所看到的，标准化 β_i 的部分告诉我们在一段时间里感应到的长期或总效应的比例。然而，在实际中常用来刻划一个分布滞后模

型的滞后结构性质的却是**平均**或**中位滞后**（mean or median lag）。

中位滞后

中位滞后是指在 X 的一单位持续变化之后，Y 的前半变化，即变化达到其总变化的 50% 所需要的时间：对于考伊克模型，中位滞后有如下式（参看习题 17.6）：

$$\text{考伊克模型：中位滞后} = -\frac{\log 2}{\log \lambda} \tag{17.4.8}$$

于是，若 $\lambda = 0.2$，则中位滞后是 0.430 6，但若 $\lambda = 0.8$，则中位滞后是 3.106 7。用文字来说，对于前一情形，Y 的总变化的 50% 可在少于半个时期内完成，而对于后一情形则需要经过多于 3 个时期才能完成 50% 的变化。这一对比并不奇怪，因为我们知道 λ 值越高，调整速度越慢，而 λ 值越低，调整速度越快。

平均滞后

假使所有 β_k 都是正的，则平均滞后的定义是：

$$\text{平均滞后} = \frac{\sum_0^\infty k\beta_k}{\sum_0^\infty \beta_k} \tag{17.4.9}$$

这不外是以各个 β 系数为权的全部有关滞后的加权平均。扼要地说，它是一个时间的滞后加权平均。对于考伊克模型，平均滞后是（参看习题 17.7）：

$$\text{考伊克模型：平均滞后} = \frac{\lambda}{1-\lambda} \tag{17.4.10}$$

这样，若 $\lambda = \frac{1}{2}$，则平均滞后是 1。

从以上讨论可清楚地看出，中位和平均滞后都是 Y 对 X 回应的速度的一个总度量。对表 17.1 所给的例子来说，平均滞后约为 11 个季度，表明平均而言要经过相当长的时间，货币供给变化对价格变化的影响才会被感应出来。

669 **例 17.7　人均消费**

这个例子调查了 1970—1999 年间美国人均消费支出（PPCE）和人均可个人支配收入（PPDI）的关系，所有的数据均以 1996 年美元计算。为了说

明考伊克模型，考虑表 17.2 中的数据。PPCE 对 PPDI 和滞后 PPCE 的回归如下所示：

$$
\begin{aligned}
\text{PPCE}_t &= -1\,242.169 + 0.603\,3\text{PPDI}_t + 0.410\,6\text{PPCE}_{t-1} \\
\text{se} &= \quad (402.578\,4)\quad (0.150\,2) \qquad (0.154\,6) \\
t &= \quad (-3.085\,5)\quad (4.015\,5) \qquad (2.656\,1) \\
& \qquad R^2 = 0.992\,6 \quad d = 1.005\,6 \quad \text{Durbin } h = 5.119
\end{aligned}
$$

注：Durbin h（德宾 h 统计量）的计算将在 17.10 节讨论。

如果我们假定此模型是对一个考伊克类型变换的估计结果，则 λ 是 0.410 6。中位滞后是 $-\dfrac{\log(2)}{\log\lambda} = -\dfrac{\log(2)}{\log(0.410\,6)} = 0.778\,6$，而平均滞后是 $\dfrac{\lambda}{1-\lambda} = \dfrac{0.410\,6}{0.589\,4} = 0.696\,6$。换句话说，PPCE 对 PDPI 的调整能在相当短时间内完成。

表 17.2　　PPCE 和 PPDI，1970—1999

年份	PPCE	PPDI	年份	PPCE	PPDI
1970	11 300	12 823	1985	16 020	18 229
1971	11 581	13 218	1986	16 541	18 641
1972	12 149	13 692	1987	16 398	18 870
1973	12 626	14 496	1988	17 463	19 522
1974	12 407	14 268	1989	17 760	19 833
1975	12 551	14 393	1990	17 899	20 058
1976	13 155	14 873	1991	17 677	19 919
1977	13 583	15 256	1992	17 989	20 318
1978	14 035	15 845	1993	18 399	20 384
1979	14 230	16 120	1994	18 910	20 709
1980	14 021	16 063	1995	19 294	21 055
1981	14 069	16 265	1996	19 727	21 385
1982	14 105	16 328	1997	20 232	21 838
1983	14 741	16 673	1998	20 989	22 672
1984	15 401	17 799	1999	21 901	23 191

注：PPCE＝人均消费支出，以 1996 年美元计算

PPDI＝人均可个人支配收入，以 1996 年美元计算

资料来源：*Economic Report of the President*, 2001, Table B-31, p.311.

§ 17.5 考伊克模型的合理化：适应性预期模型

670 考伊克模型（17.4.7）虽然十分精美，但由于它纯是一种代数演算的结果，不免有些特别；它缺少任何理论上的耕耘。然而，如果我们从一个不同的视角开始思考，这一空隙就可以得到填补。假如我们公设如下的一个模型：

$$Y_t = \beta_0 + \beta_1 X_t^* + u_t \tag{17.5.1}$$

其中 Y = 对货币（实际现金余额）的需求

X^* = 均衡、最优、预期的长期或正常利率

u = 误差项

方程（17.5.1）设想，货币需求是预期（预测意义的）利率的函数。

由于预期变量 X^* 不可直接观测，我们对预期的形成做如下的设想：

$$X_t^* - X_{t-1}^* = \gamma(X_t - X_{t-1}^*) \tag{17.5.2}$$ [13]

其中 γ 为 $0 < \gamma \leqslant 1$，称**期望系数**（coefficient of expectation）。假设（17.5.2）称**适应性预期**（adaptive expectation）或**累进式期望**（progressive expectation）或**错误中学习**（error learning）假设，曾被卡根（Cagan）[14]和弗里德曼[15]推广而得以普及。

（17.5.2）的含义是，“经济行为者（们）将根据过去的经验修改他（们）的期望，特别是要从错误中学习”[16]。更具体地说，（17.5.2）表明他们每期都按变量的现期值与前期期望值之间的差距的一个分数 γ 去修改期望值。拿我们的模型来说，这就意味着每期都按现期观测的利率与它的前期预测值之差的一个分数 γ 去修改利率的期望值。表达这种预期的另一方法是把（17.5.2）写成：

$$X_t^* = \gamma X_t + (1-\gamma) X_{t-1}^* \tag{17.5.3}$$

671 从而说明时间 t 的利率期望值是时间 t 的利率真实值与它的前期期望值各以 γ 和 $1-\gamma$ 为权的加权平均。如果 $\gamma = 1$，则 $X_t^* = X_t$ 意谓期望是立即（亦即在同一时期里）全部实现的。另一方面，如果 $\gamma = 0$，则 $X_t^* = X_{t-1}^*$，意谓期望是静止的，即“今天出现的情况将在今后的一切时期里继续维持下去。预期的未来值将与现在值相重合。”[17]

将（17.5.3）代入（17.5.1），我们得到：

$$\begin{aligned} Y_t &= \beta_0 + \beta_1[\gamma X_t + (1-\gamma) X_{t-1}^*] + u_t \qquad (17.5.4) \\ &= \beta_0 + \beta_1 \gamma X_t + \beta_1 (1-\gamma) X_{t-1}^* + u_t \end{aligned}$$

现将（17.5.1）滞后一期并乘以 $1-\gamma$，然后将此乘积从（17.5.4）中减去，

经过简单的代数运算便得：

$$
\begin{aligned}
Y_t &= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + u_t - (1-\gamma)u_{t-1} \\
&= \gamma\beta_0 + \gamma\beta_1 X_t + (1-\gamma)Y_{t-1} + v_t
\end{aligned}
\tag{17.5.5}
$$

其中 $v_t = u_t - (1-\gamma)u_{t-1}$。

在继续讨论之前，让我们先来看看（17.5.1）和（17.5.5）的差别。前者中 β_1 度量着 Y 对 X^* 一单位变化的响应，而在（17.5.5）中，$\gamma\beta_1$ 度量 Y 对实际或观测的 X 值一单位变化的平均响应。当然，除非 $\gamma=1$，也就是现期和长期的 X 值相同，否则这两种响应是不相等的。在实践中，我们先估计（17.5.5），一旦从滞后 Y 的系数得到 γ 的一个估计值，便容易将 X_t 的系数（$=\gamma\beta_1$）除以 γ 而得出 β_1。

虽然适应性预期模型（17.5.5）和考伊克模型（17.4.7）的系数有不同的含义，但两个模型的相似性是很显然的。我们看到，和考伊克模型一样，适应性预期模型也是自回归的，并且它的误差项类似于考伊克的误差项。在第 17.8 节中，我们将讨论适应性预期模型的估计问题，并在第 17.12 节中给出一些例题。现在我们既已对适应性预期（AE）模型做了描述，但它究竟有多少真实性呢？无疑它要比纯粹代数的考伊克方法更有吸引力。但 AE 假设的合理性如何？赞成 AE 假说的人可以这样说：

672 它除了对经济代理商的作用公设了一种富有意义的行为模式外，同时也为经济理论中的预期提供了一个相当简单的建模手段。人们要从经验中学习的信念，同象征着静止预期论调的全然忘记回忆的隐含假定相比，显然是一个更有意义的出发点。此外，断言更遥远的经验要比更新近的经验发挥更小的作用，这也符合常识，并且看来已被简单的观测有力地证实。[18]

直至**理性预期**（rational expectation，RE）**假说**降临之前，AE 假说在经验经济学中一直是很流行的。RE 假说最先由 J. 穆特（Muth）提出，尔后由 R. 卢卡斯（Lucas）和 T. 萨金特（Sargent）推广。RE 的支持者声称，AE 假设是不适宜的，因为它在期望的形成中只依靠一个变量的过去值[19]，而 RE 则假定："各个经济行为者在建立他们的期望时，利用了当前所能获得的有关信息，并不纯粹依赖于过去的经验。"[20]简言之，RE 假设称辩说，"预期之所以是合理的，是因为这些预期在其形成时就已有效地容纳了所有能够得到的全部信息。"[21]而不仅仅是过去的信息。

虽然理性预期假说本身也有许多批评者，但 RE 支持者针对 AE 假设的批判是言之有理的。[22]我们将不陷入这些颇带激情的辩论言辞中，也许我们可以妥协于 S. 麦克尼斯（Stephen McNees）的话："充其量，适应性预期假设只能当作一种'参考假设（working hypothesis）'，代表着一个更为复杂也许不断在变化的预期形成机制而加以维护。"[23]

例 17.8　再议例 17.7

如果我们把模型（17.4.11）看作是由适应性预期机制（即 PPCE 是预

期 PPDI 的一个函数）产生的，那么预期系数 γ 就可由（17.5.5）算出为 $1-0.4106=0.5894$。这样，按照前面关于 AE 模型的讨论，就可以说，约 70% 的实际与预期 PDPI 的差异在一年之内消失，这是一个相当快的调整。

§17.6 考伊克模型理性化的另一形式：存量调整或部分调整模型

673 适应性预期模型是考伊克模型理性化的一种方式，而 M. 尼洛夫的所谓**存量调整**（stock adjustment）**或部分调整模型**（partial adjustment model）又提供了模型理性化的另一方式。[24]为了说明此模型，考虑经济理论中的**柔性加速（器）模型**（flexible accelerator model）。该模型假定在给定的技术状态、利率等条件下，存在有给定的产出所需资本存量的均衡、最优、理想或长期额度。为简单起见，假定此理想资本水平 Y_t^* 是产出 X 的如下线性函数：

$$Y_t^* = \beta_0 + \beta_1 X_t + u_t \tag{17.6.1}$$

由于理想资本水平是不可直接观测的，尼洛夫提出以下所谓**部分调整**或**存量调整假设**：

$$Y_t - Y_{t-1} = \delta(Y_t^* - Y_{t-1}) \tag{17.6.2}$$[25]

其中 δ 为 $0<\delta\leqslant 1$，称**调整系数**，$Y_t - Y_{t-1}=$ 实际变化，而 $(Y_t^* - Y_{t-1})=$ 理想变化。

因为两个时期之间的资本存量的变化就是投资，故（17.6.2）又可写为：

$$I_t = \delta(Y_t^* - Y_{t-1}) \tag{17.6.3}$$

其中 $I_t=$ 时期 t 的投资。

方程（17.6.2）假定任意给定时期 t 的资本存量的实际变化（投资）是该时期的理想变化的某个分数 δ。如果 $\delta=1$，则意味着实际资本存量等于理想存量；即实际存量瞬时（在同一时期内）调整到理想的水平。但若 $\delta=0$，则意味着时期 t 的实际存量无异于前一时期所观测的存量，所以什么都没有改变。由于行为上的僵性、惰性和契约上的义务等等，资本存量的调整往往是不完全理想的。因此，典型地说，δ 预期会落在这两个极端值之间，从而得各部分调整模型。注意，调整机制（17.6.2）又有另一表述方式：

$$Y_t = \delta Y_t^* + (1-\delta) Y_{t-1} \tag{17.6.4}$$

这说明在时期 t 观测到的资本存量是该时期的理想资本存量与前一时期实有资本存量分别以 δ 和 $1-\delta$ 为权的一个加权平均。现将（17.6.1）代入（17.6.4）得：

$$
\begin{aligned}
Y_t &= \delta(\beta_0 + \beta_1 X_t + u_t) + (1-\delta) Y_{t-1} \\
&= \delta\beta_0 + \delta\beta_1 X_t + (1-\delta) Y_{t-1} + \delta u_t
\end{aligned}
\tag{17.6.5}
$$

此模型被称为**部分调整模型**（PAM）。

因（17.6.1）是对资本存量的长期或均衡需求，故（17.6.5）又可称为对资本存量的短期需求函数。短期中现有的资本存量不一定等于它的长期水平。一旦我们估计出短期函数（17.6.5）并且又（从 Y_{t-1}的系数）估计得调整系数 δ，只要用 δ 去除 $\delta\beta_0$ 和 $\delta\beta_1$ 并略去滞后 Y 项，就可容易地导出长期函数（17.6.1）。

675 从几何上看，部分调整模型可用图 17.6 来说明。[26]图中 Y^* 是理想资本存量，而 Y_1 是当前实际资本存量。为了说明，假定 $\delta=0.5$，这意味着厂商计划每期填补实际和理想的资本存量之间差距的一半。这样，在第 1 期里，它把资本存量移至 Y_2，使投资等于（$Y_2 - Y_1$），也就是（$Y^* - Y_1$）的一半。在往后的每一时期里，都再填补该期期初的资本存量与理想存量 Y^*之间的差距的一半。

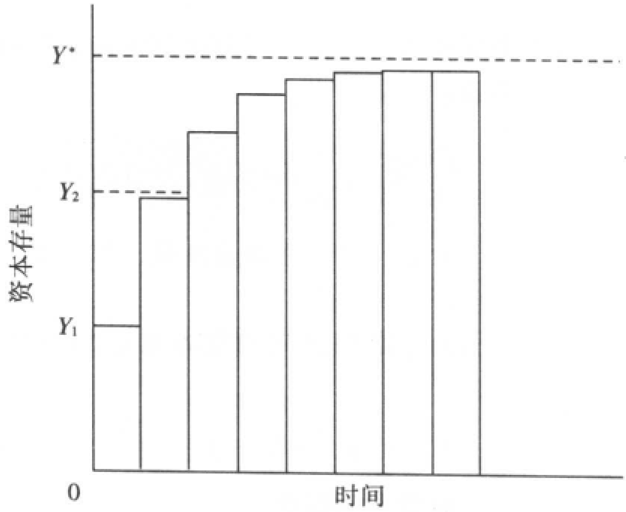


图 17.6　资本存量的逐渐调整

部分调整模型类似于考伊克模型和适应性预期模型，也是自回归的。但它有一个简单得多的干扰项：原始的干扰项乘以常数 δ。但应记住，适应性预期和部分调整模型虽然外表相似，在概念上却很不相同。前者以（关于未来的价格，利率等等）不确定性为依据，而后者则出于对技术或制度上的僵性、惰性，变化的代价等等的考虑。然而，这两种模型在理论上都要比考伊克模型健全得多。

由于从表面上看适应性预期模型和部分调整模型难以区分，如果假定后者可用于收入—支出一例（即认为理想的或期望的个人消费支出是当前的每人可支配收入的线性函数），则适应性预期模型的 γ 系数0.589 4也可解释为

存量调整模型的 δ 系数。

由于考伊克、适应性预期和存量调整三个模型——且不问误差项在形式上的差异——都产生同样的最终估计模型，我们必须极其仔细地告知读者，研究者用的是哪一模型，并且为什么。记住这一点是重要的。因而，研究者必需明确道出他们所用模型的理论基础。

§17.7 适应性预期与部分调整模型的组合

考虑如下模型：

$$Y_t^* = \beta_0 + \beta_1 X_t^* + u_t \tag{17.7.1}$$

其中 Y_t^* = 理想的资本存量，而 X_t^* = 预期的产出水平。

由于 Y_t^* 和 X_t^* 都是不可直接观测的，故可对 Y_t 启用部分调节机制，并对 X_t^* 启用适应性预期模型以得到下述估计方程（参看习题 17.2)：

$$\begin{aligned} Y_t &= \beta_0\delta\gamma + \beta_1\delta\gamma X_t + [(1-\gamma)+(1-\delta)]Y_{t-1} \\ &\quad -(1-\delta)(1-\gamma)Y_{t-2} + [\delta u_t - \delta(1-\gamma)u_{t-1}] \\ &= \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + v_t \end{aligned} \tag{17.7.2}$$

676 其中 $v_t = \delta[u_t - (1-\gamma)u_{t-1}]$，此模型也是自回归性质的。它同纯适应性预期模型的惟一差别在于 Y_{t-2} 和 Y_{t-1} 作为自变量同时出现。和考伊克以及 AE 模型一样，(17.7.2) 的误差项遵循一个移动平均过程。此模型的另一特点是，虽然该模型对诸 α 来说是线性的，但它对原来的参数而言则是非线性的。

(17.7.1) 的一个著名的应用要算弗里德曼的永久收入假说，意谓“永久”或长期收入是“永久”或长期收入的一个函数。[27]

因 (17.7.2) 和考伊克或适应性预期模型一样，都是带有同样误差结构的自回归模型，所以有着同样的估计问题。此外，(17.7.2) 还涉及一些非线性估计问题，对此我们在习题 17.1 中做了简要的考虑，但在本书中，我们不予深究。

§17.8 自回归模型的估计

讨论至此，我们有了以下的三个模型：

考伊克：

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (u_t - \lambda u_{t-1}) \tag{17.4.7}$$

适应性预期：

$$Y_t=\gamma\beta_0+\gamma\beta_1X_t+(1-\gamma)Y_{t-1}+[(u_t-(1-\gamma)u_{t-1}] \qquad (17.5.5)$$

部分调整：

$$Y_t=\delta\beta_0+\delta\beta_1X_t+(1-\delta)Y_{t-1}+\delta u_t \qquad (17.6.5)$$

所有这些模型都有如下的共同形式：

$$Y_t=\alpha_0+\alpha_1X_t+\alpha_2Y_{t-1}+v_t \qquad (17.8.1)$$

即它们都属于自回归性质。因此，经典最小二乘法未必对它们直接适用，而有必要考虑这类模型的估计问题。**理由有二：随机解释变量的出现以及序列相关的可能性。**

677 如前所见，现在要启用经典最小二乘理论，还必须表明随机解释变量 Y_{t-1}的分布与干扰项 v_t 无关。为了明确这点是否属实，重要的是要知道 v_t 的性质。即使我们假定原始的干扰项 u_t 满足全部经典假设，诸如$E(u_t)=0$，$\text{var}(u_t)=\sigma^2$（同方差性假定）和$\text{cov}(u_t,u_{t+s})=0$ 当 $s\neq0$（无自相关假定），v_t 也未必继承所有这些性质。例如，且看考伊克模型中的误差项 $v_t=u_t-\lambda u_{t-1}$。给定对 u_t 的假定容易证明：

$$E(v_tv_{t-1})=-\lambda\sigma^2 \qquad (17.8.2)^{[28]}$$

这是个非零值（除非 λ 碰巧是零），所以 v_t 是序列相关的。再由于 Y_{t-1}作为解释变量出现在考伊克模型中，它就必然和 v_t 相关（通过其中的 u_{t-1}而出现）。事实上，可以证明：

$$\text{cov}[Y_{t-1},(u_t-\lambda u_{t-1})]=-\lambda\sigma^2 \qquad (17.8.3)$$

这与（17.8.2）相同。读者可以证实，这些对适应性预期模型来说也是对的。

我们发现在考伊克模型和适应性预期模型中随机变量 Y_{t-1}与误差项 v_t 相关，这意味着什么呢？如前所指出的，**如果在一个回归模型中，解释变量与随机干扰项相关，则 OLS 估计量不仅是偏误的，甚至是非一致性的；就是说，即使样本含量无限地增大，估计量仍不迫近其真实总体值。**[29]**因此，用平常的 OLS 法去估计考伊克和适应性预期模型可能产生严重误导的结果。**

然而，部分调整模型却不一样。在此模型中 $v_t=\delta u_t$，其中 $0<\delta\leqslant1$。因此，如果 u_t 满足以前所给的关于经典线性回归模型的假定，则 δu_t 也能满足。从而部分调整模型的 OLS 估计将给出一致性估计，尽管这些估计（在有限或小样本中）有偏误的倾向。[30]直觉地看，一致性的理由是：虽然 Y_{t-1}依赖于 u_{t-1}和全部先前的干扰项，但它却与当前的误差项 u_t 无关。因此，只要 u_t 是序列上独立的，Y_{t-1}也就独立于 u_t 或至少与它不相关，从而满足 OLS 的一个重要的假定，即解释变量与随机干扰项无相关关系。

678 虽然由于存量或部分调整模型有简单的误差项结构使得它的 OLS 估计是一致性的，却不要认为，不妨把考伊克或适应性预期模型修改成适用于

OLS的（部分调整）情形。[31]读者切勿做这种削足适履的事情。模型的选择应建立在强有力的理论思考的基础上，而不要仅仅因为它易于统计上的估计。每一模型都应根据其本身的优点来考虑，适当注意其中出现的随机干扰。如果遇到像考伊克或适应性预期模型那样的情形，OLS不能直接应用，就需要设计出解决估计问题的方法。现在有好几种方法可供利用，下节中我们将考虑这类方法中的一个。

§17.9 工具变量法

OLS之所以不能适用于考伊克或适应性预期模型，是因为解释变量 Y_{t-1} 势必与误差项 v_t 相关。如果这种相关性能多少被消除掉，则如前所述，OLS就可用来获取一致性估计。（注：这里将涉及一些小的偏误。）怎样能做到这点？利维亚坦（Liviatan）曾提出以下解答。[32]

假使我们找到一个与 Y_{t-1} 高度相关但与 v_t 不相关的变量作为 Y_{t-1} 的替代，其中 v_t 是出现在考伊克或适应性预期模型中的误差项，这样的替代变量叫做**工具变量**（IV）。[33]利维亚坦建议用 X_{t-1} 作为 Y_{t-1} 的工具变量，并且还建议回归（17.8.1）的参数可由下述正规方程解得：

$$\begin{aligned}\sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1\sum X_t + \hat{\alpha}_2\sum Y_{t-1}\\ \sum Y_tX_t &= \hat{\alpha}_0\sum X_t + \hat{\alpha}_1\sum X_t^2 + \hat{\alpha}_2\sum Y_{t-1}X_t\\ \sum Y_tX_{t-1} &= \hat{\alpha}_0\sum X_{t-1} + \hat{\alpha}_1\sum X_tX_{t-1} + \hat{\alpha}_2\sum Y_{t-1}X_{t-1}\end{aligned} \tag{17.9.1}$$

679 注意，如果我们直接对（17.8.1）应用OLS，平常的OLS正规方程将是（参看第7.4节）：

$$\begin{aligned}\sum Y_t &= n\hat{\alpha}_0 + \hat{\alpha}_1\sum X_t + \hat{\alpha}_2\sum Y_{t-1}\\ \sum Y_tX_t &= \hat{\alpha}_0\sum X_t + \hat{\alpha}_1\sum X_t^2 + \hat{\alpha}_2\sum Y_{t-1}X_t\\ \sum Y_tY_{t-1} &= \hat{\alpha}_0\sum Y_{t-1} + \hat{\alpha}_1\sum X_tY_{t-1} + \hat{\alpha}_2\sum Y_{t-1}^2\end{aligned} \tag{17.9.2}$$

两组正规方程的差异应是很明显的。利维亚坦曾证明，从（17.9.1）估计出来的诸 α 是一致性的，而从（17.9.2）估计出来的则不是。这是因为 Y_{t-1} 和 v_t $[=u_t-\lambda u_{t-1}$ 或 $u_t-(1-\gamma)u_{t-1}]$ 可能相关，而 X_t 和 X_{t-1} 与 v_t 都不相关。（为什么？）

虽说利维亚坦技术在一旦找到适当的替代变量之后是容易应用的，但因进入正规方程（17.9.1）的 X_t 和 X_{t-1} 很可能是高度相关的（如在第12章所提到的，典型地说，大多数经济时间序列都在其相继值之间表现出高度的相关性），它不免受到多重共线性问题的困扰。这意味着，虽然利维亚坦方法给出一致性估计，但估计量很可能是低效的。[34]

在我们继续往下讲之前，一个明显的问题是：怎样去寻找 Y_{t-1} 的“好”

替代变量，使得它和 Y_{t-1} 虽然高度相关却与 v_t 不相关？文献中有一些建议，我们将通过习题的形式（见习题 17.5）予以讨论。但必须声明，要找到好的替代变量，并不总是容易的，这时 IV 方法就没有多少实用价值。而有必要借助于最大似然技术，后者超出了本书的范围。[35]

那么存在一种能用来查明你所选取的工具是否有效的检验吗？丹尼斯·萨根因此而提出了一种检验，即 **SARG 检验**。该检验在附录 17A 的 17A.1 节有所描述。

§17.10 在自回归模型中侦察自相关：德宾 h 检验

诚如所见，误差项中可能的序列相关会使自回归模型的估计问题变得相当复杂。如果原始模型中的误差项为序列无关，则存量调整模型的误差项 v_t 就不会是序列相关的。然而对于考伊克和适应性预期模型，即使 u_t 序列无关，v_t 仍可能序列相关。于是问题是，怎样知道自相关模型中的误差项是否序列相关呢？

680 如第 12 章所提到的，德宾-沃森 d 统计量不宜用于侦察自回归模型中的（一阶）自相关。这是因为在这类模型中，所估算的 d 通常都有偏向 2 的偏误，而 2 是纯随机序列的 d 的期望值。这就是说，如果我们照例对这类模型计算 d 统计量，就会有一种妨碍我们发现（一阶）序列相关的内在偏误。尽管如此，许多研究者由于没有更好的方法仍然计算着 d。然而，最近，德宾本人提出了自回归模型一阶序列相关的一个大样本（large-sample）检验[36]，称之为 ***h* 统计量**。

在习题 12.36 中我们已经讨论过德宾 h 检验。为了方便，我们再次写出 h 统计量（符号上有一点小变化）：

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\operatorname{var}(\hat{\alpha}_2)]}} \tag{17.10.1}$$

其中 n = 样本含量，$\operatorname{var}(\hat{\alpha}_2)$ = (17.8.1) 中滞后 $Y_t(=Y_{t-1})$ 的系数的方差，$\hat{\rho}$ = 最初在第 12 章讨论的一阶序列相关 ρ 的估计值。

正如在习题 12.36 中所提到的，对于大的样本含量，德宾曾证明在 $\rho=0$ 的假设条件下，(17.10.1) 中的 h 统计量遵循标准正态分布。即：

$$h_{\text{asy}} \sim N(0,1) \tag{17.10.2}$$

其中 asy 的意思是渐近地。

实际上，正如在第 12 章所提到的，ρ 的估计为：

$$\hat{\rho} \approx 1 - \frac{d}{2} \tag{17.10.3}$$

我们可以观察到一个有趣的现象：尽管我们不能用德宾 d 来检验自回归模型的自相关性，但是在计算 h 统计量时可以将其作为输入量。

我们用例 17.7 来说明 h 统计量的使用。在这个例子中，$n=30$，$\hat{\rho}\approx(1-d/2)=0.497\,2$（注：$d=1.005\,6$），并且 $\text{var}(\hat{\alpha}_2)=\text{var}(\text{PPCE}_{t-1})=(0.154\,6)^2=0.023\,9$。将这些值代入(17.10.1)。我们得到：

$$h=0.497\,2\sqrt{\frac{30}{1-30(0.023\,9)}}=5.119\,1 \qquad (17.10.4)$$

681 因为在虚拟假设条件下 h 值服从标准正态分布，所以得到如此高的 h 值的概率是很小的。回忆一下一个标准正态变量超过这个值的概率是非常小的。从而在目前这个例子中，我们的推论是存在（正）自相关性。当然，要记住 h 渐近地服从标准正态分布。我们选取的 30 个观察值的样本不一定足够大。

注意 h 统计量的如下特性：

1. 不管回归模型中含有多少个 X 变量或多少个 Y 的滞后值，都可应用。计算 h 时只需考虑滞后 Y_{t-1} 的系数的方差。

2. 如果$[n\ \text{var}(\hat{\alpha}_2)]$超过 1，检验便不适用。（为什么？）不过，在实践中，这种情形不常发生。

3. 由于该检验是一种大样本检验，它在小样本中的应用严格地说，如因德尔（Inder）[37]和基维埃（Kiviet）[38]所表明的，是不能赞同的。曾有人提出，第 12 章中所讲的 BG 检验，又称拉格朗日乘数检验，不仅在大样本中，而且在有限或小样本中，都有统计上更强的功效，因而优于 h 检验。[39]

§17.11 一个数值例子：加拿大的货币需求，1979 年第 1 季度至 1988 年第 4 季度

为了说明我们已经深入讨论过的该模型的使用，考虑前面提到的一个经验应用，即货币需求（或者对实际现金余额的需求）。特别地，考虑下面这个模型。[40]

$$M_t^* = \beta_0 R_t^{\beta_1} Y_t^{\beta_2} e^{u_t} \qquad (17.11.1)$$

其中 M_t^* = 理想或长期货币（实际现金余额）需求

R_t = 长期利率，%

Y_t = 实际总国民收入

为了做统计估计，可将（17.11.1）方便地表达成对数的形式：

$$\ln M_t^* = \ln \beta_0 + \beta_1 \ln R_t + \beta_2 \ln Y_t + u_t \qquad (17.11.2)$$

682 由于理想需求变量不可直接观测，且假定存量调整假说，即：

$$\frac{M_t}{M_{t-1}}=\left(\frac{M_t^*}{M_{t-1}}\right)^{\delta} \qquad 0<\delta\leqslant 1 \qquad (17.11.3)$$

方程（17.11.3）表示在每一时期（年）内，都有真实的与理想的实际现金

余额之间的差距的一定百分比（为什么?）被消去。方程（17.11.3）又可用对数形式表达为：

$$\ln M_t - \ln M_{t-1} = \delta(\ln M_t^* - \ln M_{t-1}) \tag{17.11.4}$$

将（17.11.2）的 $\ln M_t^*$ 代入方程（17.11.4）并加以整理，得：

$$\ln M_t = \delta \ln \beta_0 + \beta_1 \delta \ln R_t + \beta_2 \delta \ln Y_t + (1-\delta)\ln M_{t-1} + \delta u_t \qquad (17.11.5)^{[41]}$$

此式可称为短期货币需求函数。(为什么?)

为了说明实际现金余额的短期和长期需求，考虑表 17.3 给出的数据。这些按季度计算的数据是属于加拿大 1979—1988 年间的一些数据。变量定义为：M[由 M1 货币供给所定义，加拿大元(C\$)，百万]，P(绝对通货紧缩，1981 = 100)，GDP 以不变的 1981 年价格计算(C\$，百万)和 R(90 天主要公司利率，%)。[42]M1 被 P 缩小以得到实际现金余额的数据。经过推理可认为，实际现金余额与 GDP 正相关(正的收入效应)，而与 R 负相关(利率越高，持有货币的机会成本越高，因为即使 M1 货币需要支付利息，所支付的利息也会很低)。

回归结果如下所示[43]：

$$\begin{aligned}
\widehat{\ln M_t} = {} & 0.8561 - 0.0634 \ln R_t - 0.0237 \ln \mathrm{GDP}_t + 0.9607 \ln M_{t-1} \\
\mathrm{se} = {} & (0.5101) \quad (0.0131) \quad (0.0366) \quad (0.0414) \\
t = {} & (1.6782) \quad (-4.8134) \quad (-0.6466) \quad (23.1972) \\
& R^2 = 0.9482 \quad d = 2.4582 \quad F = 213.7234
\end{aligned} \qquad (17.11.6)^{[43]}$$

表 17.3　货币、利率、价格指数和 GDP，加拿大

观察值	M1	R	P	GDP
1979 - 1	22 175.00	11.133 33	0.779 47	334 800
1979 - 2	22 841.00	11.166 67	0.808 61	336 708
1979 - 3	23 461.00	11.800 00	0.826 49	340 096
1979 - 4	23 427.00	14.183 33	0.848 63	341 844
1980 - 1	23 811.00	14.383 33	0.866 93	342 776
1980 - 2	23 612.33	12.983 33	0.889 50	342 264
1980 - 3	24 543.00	10.716 67	0.915 53	340 716
1980 - 4	25 638.66	14.533 33	0.937 43	347 780
1981 - 1	25 316.00	17.133 33	0.965 23	354 836
1981 - 2	25 501.33	18.566 67	0.987 74	359 352
1981 - 3	25 382.33	21.016 66	1.013 14	356 152
1981 - 4	24 753.00	16.616 65	1.034 10	353 636

1982 - 1	25 094.33	15.350 00	1.057 43	349 568
1982 - 2	25 253.66	16.049 99	1.077 48	345 284
1982 - 3	24 936.66	14.316 67	1.096 66	343 028
1982 - 4	25 553.00	10.883 33	1.116 41	340 292
1983 - 1	26 755.33	9.616 670	1.123 03	346 072
1983 - 2	27 412.00	9.316 670	1.133 95	353 860
1983 - 3	28 403.33	9.333 330	1.147 21	359 544
1983 - 4	28 402.33	9.550 000	1.160 59	362 304
1984 - 1	28 715.66	10.083 33	1.171 17	368 280
1984 - 2	28 996.33	11.450 00	1.174 06	376 768
1984 - 3	28 479.33	12.450 00	1.177 95	381 016
1984 - 4	28 669.00	10.766 67	1.184 38	385 396
1985 - 1	29 018.66	10.516 67	1.189 90	390 240
1985 - 2	29 398.66	9.666 670	1.206 25	391 580
1985 - 3	30 203.66	9.033 330	1.214 92	396 384
1985 - 4	31 059.33	9.016 670	1.218 05	405 308
1986 - 1	30 745.33	11.033 33	1.224 08	405 680
1986 - 2	30 477.66	8.733 330	1.228 56	408 116
1986 - 3	31 563.66	8.466 670	1.239 16	409 160
1986 - 4	32 800.66	8.400 000	1.253 68	409 616
1987 - 1	33 958.33	7.250 000	1.271 17	416 484
1987 - 2	35 795.66	8.300 000	1.284 29	422 916
1987 - 3	35 878.66	9.300 000	1.295 99	429 980
1987 - 4	36 336.00	8.700 000	1.310 01	436 264
1988 - 1	36 480.33	8.616 670	1.323 25	440 592
1988 - 2	37 108.66	9.133 330	1.332 19	446 680
1988 - 3	38 423.00	10.050 00	1.350 65	450 328
1988 - 4	38 480.66	10.833 33	1.366 48	453 516

注：M1 = C＄，百万

P = 绝对通货紧缩（1981 = 100）

R = 90 天主要公司利率，%

GDP = C＄，百万（1981 年的价格）

资料来源：Rao，前引文献，第 210 - 213 页。

684 所估计的短期需求函数表明，短期利率弹性有正确的符号，并且在统计上是显著的，因为它的 p 值几乎为零。让人惊奇的是短期收入弹性为负，尽管在统计上它无异于零。调节系数为 $\delta = (1 - 0.960\ 7) = 0.039\ 3$，意味着在理想的和真实的现金余额的差异中约有 4% 在一个季度里被消去，这是个

十分缓慢的调节过程。

要回到长期需求函数（17.11.2），只需将短期需求函数通除以 δ（为什么?）并丢掉 $\ln M_{t-1}$ 一项。［因为长期而言，（17.11.6）中的 $\ln M_{t-1}$ 无异于 $\ln M_t$，将二者合并即得一系数。可作为长期的 $\ln M_t^*$ 的系数。——译者注］结果是：

$$\ln \overline{M}_t^* = 21.788\,8 - 1.613\,2\ln R_t - 0.603\,0\ln \mathrm{GDP} \qquad (17.11.7)^{[44]}$$

可见，货币需求的长期利率弹性在绝对值上远远大于相应的短期弹性，对收入弹性来说同样如此，尽管在目前这个例子中它的经济和统计显著性是不确定的。

注意所估计的德宾-沃森 d 是 2.458 2，接近于 2。从而证实我们先前的评语：在自回归模型中计算的 d 通常接近于 2，因此不能靠计算 d 来判断我们的数据是否有序列相关。本例中所选取的样本容量有 40 个观察值，这对于应用 h 检验来说可能是相当大的。在本例中，读者可以证明估计的 h 值为 $-1.500\,8$，这在 5% 水平上是不显著的，也许意味着误差项没有一阶自相关性。

§ 17.12　说明性例子

在本节中举几个例子以表明研究者如何在他们的经验研究工作中应用分布滞后模型。

例 17.9　联邦储备与实际利率

为了评估 M1（现钞加支票存款）的增长对 Aaa 债券的实际利率变动的影响，G.J. 桑通尼（Santoni）和 C.C. 斯通（Stone）[45]利用月数据对美国估计了如下的分布滞后模型：

$$r_t = 常数 + \sum_{i=0}^{11} a_i \dot{M}_{t-i} + u_t \qquad (17.12.1)$$

其中 r_t = 穆迪 Aaa 债券收益指数减去经季节调整的前 36 个月消费者价格指数的年平均变化率，以此作为实际利率的度量，而 $\dot{M}_t$ = 每月的 M1 增长。

685 根据“货币中性理论”实际经济变量诸如产出、就业、经济增长、实际利率等等均不致于永久地受货币增长的影响，从而与货币政策基本上无关。……按照这种论调，联邦储备理应对实际利率无永久性的影响。[46]

如果这种理论正确，则可以预期诸分布滞后系数以及它们的总和都在统计上无异于零。为了判明情况是否如此，作者们对两个不同的时期估计了（17.12.1）。一个从 1951 年 2 月到 1979 年 9 月，而另一个从 1979 年 10 月到 1982 年 11 月。后一时期考虑到联邦储备的货币政策的改变。自从 1979 年

10月以来，联邦储备银行改变以前主要注意利率变化的政策，更多地注意了货币供给的增长率。表17.4列出其回归结果。这些结果似乎支持了“货币中性理论”。因为，从1951年2月至1979年9月，现期的和滞后的货币增长对实际利率的变化都无统计上显著的影响。而且，对于后一时期，由于$\sum a_i$不显著地异于零；只有系数a_i是显著的，但它的符号却是错误的。(为什么?) 货币中性理论也似乎是成立的。

表17.4　每月M1增长对Aaa债券实际利率变化的影响：1951年2月—1982年11月

$$r = \text{常数} + \sum_{i=0}^{11} a_i \dot{M}_{1t-i}$$

	1951年2月—1979年9月		1979年10月—1982年11月	
	系数	$\|t\|$*	系数	$\|t\|$
常数	1.488 5†	2.068	1.036 0	0.801
a_0	−0.000 88	0.388	0.008 40	1.014
a_1	0.001 71	0.510	0.039 60†	3.419
a_2	0.001 70	0.423	0.031 12	2.003
a_3	0.002 33	0.542	0.027 19	1.502
a_4	−0.002 49	0.553	0.009 01	0.423
a_5	−0.001 60	0.348	0.019 40	0.863
a_6	0.002 92	0.631	0.024 11	1.056
a_7	0.002 53	0.556	0.014 46	0.666
a_8	0.000 00	0.001	−0.000 36	0.019
a_9	0.000 74	0.181	−0.004 99	0.301
a_{10}	0.000 16	0.045	−0.011 26	0.888
a_{11}	0.000 25	0.107	−0.001 78	0.211
$\sum a_i$	0.007 37	0.221	0.154 9	0.926
$\bar{R}^2$	0.982 6		0.866 2	
$D-W$	2.07		2.04	
RH01	1.27†	24.536	1.40†	9.838
RH02	−0.28†	5.410	−0.48†	3.373
NOB	344.		38.	
SER(=RSS)	0.154 8		0.389 9	

* $|t|=t$的绝对值。

† 在0.05水平上显著地异于零。

资料来源：G.J.Santoni and Courtenay C.Stone, "The Fed and the Real Rate of Interest", *Review*, Federal Reserve Bank of St. Louis, December 1982, p.16.

例17.10　1967—1993年斯里兰卡短期与长期总消费

假定消费C和永久收入X^*有线性关系：

$$C_t = \beta_1 + \beta_2 X_t^* + u_t \tag{17.12.2}$$

由于 X_t^* 不可直接观测，有必要明确产生永久收入的机制。假如我们采用（17.5.2）中的适应性预期假说，利用（17.5.2）并作简化，就得到如下的估计方程（比较 17.5.5）：

$$C_t = \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \tag{17.12.3}$$

其中 $\alpha_1 = \gamma\beta_1$

$\alpha_2 = \gamma\beta_2$

$\alpha_3 = (1-\gamma)$

$v_t = [u_t - (1-\gamma)u_{t-1}]$

我们知道，β_2 给出消费对每增加一美元（比如说）永久收入的平均响应，而给出消费对每增加一美元现期收入的平均响应。

根据 1967—1993 年期间的斯里兰卡年度数据，我们可以得到下面的回归结果[47]：

$$\begin{aligned}\hat{C} &= 1\,038.403 + \quad 0.404\,3X_t + 0.500\,9C_{t-1}\\ \text{se} &= (2\,501.455) \quad (0.091\,9) \quad (0.121\,3)\\ t &= \quad (0.415\,1) \; (4.397\,9) \quad (4.129\,3)\\ &\qquad R^2 = 0.991\,2 \quad d = 1.416\,2 \quad F = 1\,298.466\end{aligned} \tag{17.12.4}$$

其中 C = 个人消费支出，X = GDP，二者均以固定不变的价格计算。我们也可以把实际利率引入该模型，但它在统计上不是显著的。

此回归表明，短期边际消费倾向为 0.404 3，这意味着当前的或观测的实际收入（用实际 GDP 来计算）每增加 1 卢比，消费平均增加约 0.40 卢比。但若收入的这一增加一直持续下去，则这一永久收入的 MPC 最终将是 $\beta_2 = \gamma\beta_2/\gamma = 0.404\,3/0.499\,1 = 0.810\,0$ 或约为 0.81 卢比。换句话说，当消费者有足够的时间调整到与 1 卢比的收入变化相适应时，他们的消费将增加约 0.81 卢比。

表 17.5　　私人消费支出（PCON）和 GDP，斯里兰卡

年份	PCON	GDP	年份	PCON	GDP
1967	61 284	78 221	1981	120 477	152 846
1968	68 814	83 326	1982	133 868	164 318
1969	76 766	90 490	1983	148 004	172 414
1970	73 576	92 692	1984	149 735	178 433
1971	73 256	94 814	1985	155 200	185 753
1972	67 502	92 590	1986	154 165	192 059
1973	78 832	101 419	1987	155 445	191 288
1974	80 240	105 267	1988	157 199	196 055
1975	84 477	112 149	1989	158 576	202 477
1976	86 038	116 078	1990	169 238	223 225
1977	96 275	122 040	1991	179 001	233 231

1978	101 292	128 578	1992	183 687	242 762
1979	105 448	136 851	1993	198 273	259 555
1980	114 570	144 734			

注：PCON＝私人消费支出

GDP＝国内生产总值

资料来源：见章末注 47。

687 现假定我们的消费函数是：

$$C_t^* = \beta_1 + \beta_2 X_t + u_t \tag{17.12.5}$$

在这一构架中，永久或长期消费 C_t 是当前的或观测的收入的一个线性函数。由于 C_t^* 不可直接观测，不妨借助于部分调整模型（17.6.2）。利用此模型，经过一些代数运算，我们得到：

$$\begin{aligned} C_t &= \delta\beta_1 + \delta\beta_2 X_t + (1-\delta) C_{t-1} + \delta u_t \\ &= \alpha_1 + \alpha_2 X_t + \alpha_3 C_{t-1} + v_t \end{aligned} \tag{17.12.6}$$

从表面看，此模型和适应性预期模型（17.12.3）没有区别，因此（17.12.4）所给的回归结果在这里同样适用。然而，怎样解释这两个模型却大有区别，且不谈模型（17.12.3）为自回归且可能为序列相关所带来的估计问题。模型（17.12.5）是长期或均衡消费函数，而（17.12.6）是短期消费函数。β_2 衡量着长期 MPC，而 α_2（$=\delta\beta_2$）则给出短期 MPC；前者可以从后者除以调节系数 δ 得出。

回到（17.12.4），我们现在可以解释 0.404 3 这个短期 MPC。由于 $\delta=$ 0.499 1，长期 MPC 是 0.81。注意，约为 0.50 的调节系数说明，消费者在任一时期里，都只把他们的消费水平在走向其理想的或长期的水平的路程上调整约 1/2。

本例提出一个关键性问题，即从外表看，适应性预期和部分调整模型，甚或考伊克模型如此相似，单从诸如（17.12.4）这样的回归估计看，将无从得知哪个是正确地设定的模型。这正说明为什么在选择模型做经验分析之前要先明确模型的理论基础，然后再适当地做下去是根本重要的。如果习惯或惯性象征着消费行为，则部分调整模型是适宜的。另一方面，如果消费行为是前瞻性的，即是以预期的未来收入作为依据，则适应性预期模型是适宜的。如果属于后一情形，则还必须密切注意估计方法，以求得一致性估计量。而对于前一种情形，OLS 将提供一致性估计量，如果通常的 OLS 假定得到满足的话。

§ 17.13 分布滞后模型的阿尔蒙方法：阿尔蒙或多项式分布滞后[48]

虽然考伊克分布滞后模型在实际中广为应用，但它是建立在这样的假定上，即随着滞后的延长，β 系数按几何方式下降（见图 17.5）。这一假定对某些情况来说未免过于苛求。例如，考虑图 17.7。

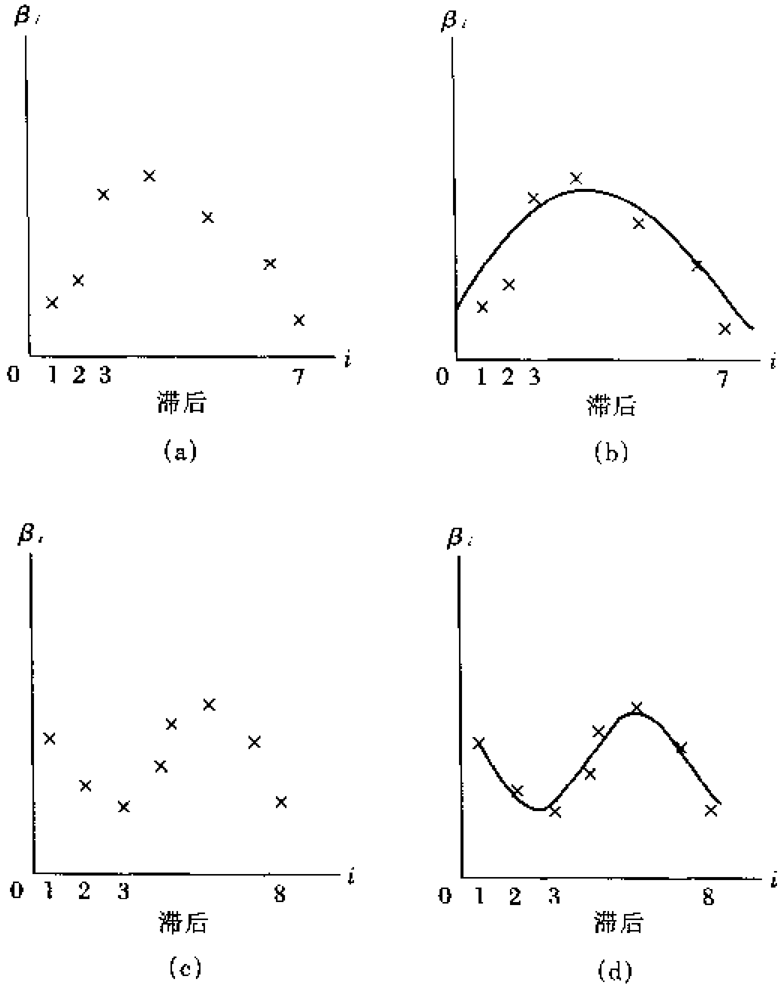


图 17.7 阿尔蒙多项式滞后模式

图 17.7a 假定 β 先增后减，而图 17.7c 假定这些 β 有一循环模式。显然，对于这些情形，分布滞后模型的考伊克模式将不适用。然而，在细察图 17.7a 和 c 之后，似可把 β_i 表达为滞后（时间）长度 i 的函数，并拟合适当的曲线以反映两者的函数关系，如图 17.7b 和 d 所示。这就是 S. 阿尔蒙(Shirley Almon) 提出的方法。为了说明她的方法，让我们回到前面考虑过的有限分布滞后模型，即：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t \tag{17.1.2}$$

688 这又可简洁地写为：

$$Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t \tag{17.13.1}$$

689 根据数学中的一个定理，称**韦亚斯特拉斯定理**（Weierstrass' theorem），阿尔蒙假定 β_i 可用滞后长度 i 的一个适当高次的多项式来逼近。[49]例如，如果 17.7a 所展示的滞后模式合适的话，就可写为：

$$\beta_i = a_0 + a_1 i + a_2 i^2 \tag{17.13.2}$$

这是 i 的一个二次多项式（参看图 17.7 b）。然而，如果 β_i 遵循图 17.7c 的模式，则可写：

$$\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3 \tag{17.13.3}$$

这是 i 的一个三次多项式（参看图 17.7d）。更一般地，我们可写：

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m \tag{17.13.4}$$

这是 i 的一个 m 次多项式。这里假定 m（多数式的次数）小于 h（滞后的最大长度）。

为说明怎样用阿尔蒙模式，且假定 β_i 符合图 17.7a 所展现的样式，从而用二次多项式逼近是合适的。将（17.13.2）代入（17.13.1），我们得到：

$$\begin{aligned} Y_t &= \alpha + \sum_{i=0}^{k} (a_0 + a_1 i + a_2 i^2) X_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^{k} X_{t-i} + a_1 \sum_{i=0}^{k} i X_{t-i} + a_2 \sum_{i=0}^{k} i^2 X_{t-i} + u_t \end{aligned} \tag{17.13.5}$$

定义：

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^{k} X_{t-i} \\ Z_{1t} &= \sum_{i=0}^{k} i X_{t-i} \\ Z_{2t} &= \sum_{i=0}^{k} i^2 X_{t-i} \end{aligned} \tag{17.13.6}$$

就可把（17.13.5）写为：

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + u_t \tag{17.13.7}$$

在阿尔蒙模式中，Y 要对构造变量 Z 而不是对原始变量 X 做回归。应看到（17.13.7）可用平常的 OLS 方法去估计。如果随机干扰项 u 满足经典线性回归模型的假定的话，则这样得到的 α 和 a_i 的估计值将具有全部的优良统计性质。在这一方面，阿尔蒙技术和考伊克方法相比有明显的优点。因为，我们已经看到，由于随机解释变量 Y_{t-1} 的出现，并且很可能与干扰项相关而涉及严重的估计问题。

690 一旦从（17.13.7）估计出诸 α，即可从（17.13.2）［或更一般地还可

从（17.13.4）］估计原始的诸 β 系数如下：

$$
\begin{aligned}
\hat{\beta}_0 &= \hat{a}_0 \\
\hat{\beta}_1 &= \hat{a}_0 + \hat{a}_1 + \hat{a}_2 \\
\hat{\beta}_2 &= \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 \\
\hat{\beta}_3 &= \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 \\
&\cdots\cdots\cdots\cdots \\
\hat{\beta}_k &= \hat{a}_0 + k\hat{a}_1 + k^2\hat{a}_2
\end{aligned}
\tag{17.13.8}
$$

在应用阿尔蒙技术之前，还必须解决以下实际问题。

1. 必须事先规定滞后 k 的最大长度。这里也许可以采纳戴维森和麦金农的意见：

> 解决滞后长度问题的最好方法，也许是先从一个很大的 q 值［作为滞后长度］开始，而不对分布滞后的形状施加任何约束，然后看模型的拟合是否会随 q 的减小而显著恶化。[50]

这种意见是和第 14 章所讨论的亨德里的一般到具体（Hendry's top-down）方法论的精神相一致的。记住，如在第 13 章中所见，如果存在有某个“真实”的滞后长度，则选择过小的长度将导致“漏掉有关变量”的偏误；这时系数仍可由 OLS 做一致性估计，尽管系数的方差不那么有效。

我们可以用第 13 章讨论的赤池或施瓦兹信息准则选取合适的滞后长度。这些准则也可以用来讨论多项式的合适的次数以及进行第二点（即下面一点）的讨论。

2. 定出 k 后，还必须定出多项式的次数 m。一般地说，多项式的次数应至少比联系着 i 和 β_i 的曲线的转向点个数大 1。例如，在图 17.7a 中只有一个转向点，从而一个二次多项式就将是一个良好的逼近。在图 17.7c 中有两个转向点，因而一个三次多项式将会提供一个良好的逼近。我们也许不能先验地知道转向点的个数，因而对 m 的选择，大体上是主观的。对某些情况而言，理论也许能提示我们一个具体的形状。在实践中，我们希望一个相当低次（比如 $m=2$ 或 3）的多项式能给出良好的结果。在选定 m 的一个具体值后，如果我们想看看一个更高次的多项式是否会给出更好的结果，则可进行如下：

691 假使我们必须在二和三次多项式之间做出选择。对于二次多项式，估计方程由（17.13.7）给出，对于三次多项式，相应的方程是：

$$
Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + a_3 Z_{3t} + u_t \tag{17.13.9}
$$

其中 $Z_{3t} = \sum_{i=0}^{k} i^3 X_{t-i}$。在做回归（17.13.9）后，如果我们发现 a_2 是统计上显著的，而 a_3 不显著，就可认为二次多项式给出了良好的近似。

或者，如戴维森和麦金农所建议的，“在决定 q［滞后的长度］后，就可试图决定 d［多项式的次数］，再次从一个大数开始，然后把它减下来。”然而我们必须当心多重共线性的问题。因为，如（17.13.6）［还有（17.13.10）］所示，Z 变量是从 X 变量构造出来的，故多重共线性的问题容

易发生。如第 10 章所表明的，当多重共线性的情形严重时，a_3 在统计上不显著并不是因为真的 a_3 为零，只是因为手头中的样本不允许我们估计出 Z_3 对 Y 的分别影响。因此，在上述说明中，在我们接受三次多项式不是正确的选择这一结论之前，我们必须先肯定多重共线性问题还不致严重到不能用第 10 章所讨论的技术去把它处理好。

3. 一旦确定了 m 和 k，就容易构造出诸 Z。例如，若 $m=2$ 和 $k=5$，则诸 Z 是：

$$Z_{0t}=\sum_{i=0}^{5}X_{t-i}=(X_t+X_{t-1}+X_{t-2}+X_{t-3}+X_{t-4}+X_{t-5})$$

$$Z_{1t}=\sum_{i=0}^{5}iX_{t-i}=(X_{t-1}+2X_{t-2}+3X_{t-3}+4X_{t-4}+5X_{t-5})$$

$$(17.13.10)$$

$$Z_{2t}=\sum_{i=0}^{5}i^2X_{t-i}=(X_{t-1}+4X_{t-2}+9X_{t-3}+16X_{t-4}+25X_{t-5})$$

注意，诸 Z 是原始的诸 X 的线性组合，并且注意诸 Z 为什么会表现多重共线性。

在讲到一个数值例子之前，先看看阿尔蒙方法有哪些优点。第一，它给出了一个涵盖形形色色的滞后结构的灵活方法（参看习题 17.17）。而考伊克技术则拘泥于假定诸 β 系数是几何递减的。其次，在用阿尔蒙方法时，不像考伊克技术那样，我们不必担心滞后因变量作为解释变量出现在模型中从而产生估计问题。最后，如果可以拟合一个足够低次的多项式，则待估系数（指诸 a）的个数要比原先系数（指诸 β）的个数少得多。

692 但让我们再次强调使用阿尔蒙技术的问题。第一，多项式次数以及滞后的最大期数基本上是一种主观臆断；其次，由于前述理由，诸 Z 变量很可能带有大的标准误（相对于那些系数的估计值而言），从而使一或多个系数在通常 t 检验的基础上成为统计上不显著的。但这并不一定意味着一或多个原始系数将是统计上不显著的。（关于这一陈述的证明有些复杂，但在习题 17.18 中所有提示。）这样一来，多重共线性的问题也许没有人们所想像的那样严重。此外，我们知道，在多重共线性的情形中，即使无法准确估计某一个别的系数，而这些系数的某一线性组合［**可估函数**（estimable function）］却是可准确估计的。

例 17.11　阿尔蒙分布滞后模型的说明

为了阐明阿尔蒙技术，表 17.6 给出 1954—1999 年美国制造业的库存 Y 和销售 X。为了说明，假定库存依赖于当年和前 3 年的销售量如下：

$$Y_t=\alpha+\beta_0X_t+\beta_1X_{t-1}+\beta_2X_{t-2}+\beta_3X_{t-3}+u_t \qquad (17.13.11)$$

此外，再假定 β_i 可由一个二次多项式逼近，如（17.13.2）所示。于是，按照（17.13.5）可写出：

$$Y_t=\alpha+a_0Z_{0t}+a_1Z_{1t}+a_2Z_{2t}+u_t \qquad (17.13.12)$$

其中：

$$Z_{0t}=\sum_{i=0}^{3}X_{t-i}=(X_t+X_{t-1}+X_{t-2}+X_{t-3})$$
$$Z_{1t}=\sum_{i=0}^{3}iX_{t-i}=(X_{t-1}+2X_{t-2}+3X_{t-3}) \tag{17.13.13}$$
$$Z_{2t}=\sum_{i=0}^{3}i^2X_{t-i}=(X_{t-1}+4X_{t-2}+9X_{t-3})$$

如此构造出的诸 Z 变量，见于表 17.6。利用 Y 和诸 Z 的数据，我们求出以下回归：

$$\begin{aligned}\hat{Y}_t &= 25\,845.06 + 1.114\,9Z_{0t} - 0.371\,3Z_{1t} - 0.060\,0Z_{2t}\\ \text{se} &= (6\,596.998)\quad(0.538\,1)\quad(1.374\,3)\quad(0.454\,9)\\ t &= \quad(3.917\,7)(2.071\,8)\ (-0.270\,2)\ (-0.131\,9)\\ R^2 &= 0.975\,5 \qquad d=0.164\,3 \qquad F=517.765\,6\end{aligned} \tag{17.13.14}$$

（注：因为我们假定一个 3 年滞后，故观测总个数从 46 减到 43。）

表 17.6　　美国制造业库存 Y、销售 X 以及构造的 Z

年份	库存	销售	Z_0	Z_1	Z_2
1954	41 612	23 355	NA	NA	NA
1955	45 069	26 480	NA	NA	NA
1956	50 642	27 740	NA	NA	NA
1957	51 871	28 736	106 311	150 765	343 855
1958	50 203	27 248	110 204	163 656	378 016
1959	52 913	30 286	114 010	167 940	391 852
1960	53 786	30 878	117 148	170 990	397 902
1961	54 871	30 922	119 334	173 194	397 254
1962	58 172	33 358	125 444	183 536	427 008
1963	60 029	35 058	130 216	187 836	434 948
1964	63 410	37 331	136 669	194 540	446 788
1965	68 207	40 995	146 742	207 521	477 785
1966	77 986	44 870	158 254	220 831	505 841
1967	84 646	46 486	169 682	238 853	544 829
1968	90 560	50 229	182 580	259 211	594 921
1969	98 145	53 501	195 086	277 811	640 003
1970	101 599	52 805	203 021	293 417	672 791
1971	102 567	55 906	212 441	310 494	718 870
1972	108 121	63 027	225 239	322 019	748 635
1973	124 499	72 931	244 669	333 254	761 896
1974	157 625	84 790	276 654	366 703	828 193

1975	159 708	86 589	307 337	419 733	943 757
1976	174 636	98 797	343 107	474 962	1 082 128
1977	188 378	113 201	383 377	526 345	1 208 263
1978	211 691	126 905	425 492	570 562	1 287 690
1979	242 157	143 936	482 839	649 698	1 468 882
1980	265 215	154 391	538 433	737 349	1 670 365
1981	283 413	168 129	593 361	822 978	1 872 280
1982	311 852	163 351	629 807	908 719	2 081 117
1983	312 379	172 547	658 418	962 782	2 225 386
1984	339 516	190 682	694 709	1 003 636	2 339 112
1985	334 749	194 538	721 118	1 025 829	2 351 029
1986	322 654	194 657	752 424	1 093 543	2 510 189
1987	338 109	206 326	786 203	1 155 779	2 688 947
1988	369 374	224 619	820 140	1 179 254	2 735 796
1989	391 212	236 698	862 300	1 221 242	2 801 836
1990	405 073	242 686	910 329	1 304 914	2 992 108
1991	390 905	239 847	943 850	1 389 939	3 211 049
1992	382 510	250 394	969 625	1 435 313	3 340 873
1993	384 039	260 635	993 562	1 458 146	3 393 956
1994	404 877	279 002	1 029 878	1 480 964	3 420 834
1995	430 985	299 555	1 089 586	1 551 454	3 575 088
1996	436 729	309 622	1 148 814	1 639 464	3 761 278
1997	456 133	327 452	1 215 631	1 745 738	4 018 860
1998	466 798	337 687	1 274 316	1 845 361	4 261 935
1999	470 377	354 961	1 329 722	1 921 457	4 434 093

注：Y 和 X 以百万美元计，经季节调整。

资料来源：*Economic Report of the President*, 2001, Table B-57, p.340. 诸 Z 如(17.13.13)所示。

694 对上述结果的一个简短的评论：在三个 Z 变量中，只有 Z_0 在 5% 水平上是独立统计显著的，而其他两个则不是，而且 F 值如此之高以至于我们可以拒绝总体上诸 Z 对 Y 没有影响的虚拟假设。正如同你可能会怀疑的一样，这也许应归功于多重共线性。同时，请注意计算的 d 值非常低。这并不一定意味着残值是由于自相关性造成的。更可能的是，残值意味着我们所用的模型可能是误设的（mis-specified）。

根据（17.13.3）所给出的 α 系数的估计值，我们可以轻易地估计出初始 β 值，如（17.13.8）所示。在当前这个例子中，结果如下所示：

$$
\begin{aligned}
\hat{\beta}_0 &= \hat{a}_0 = 1.114\,9 \\
\hat{\beta}_1 &= (\hat{a}_0 + \hat{a}_1 + \hat{a}_2) = 0.683\,6 \\
\hat{\beta}_2 &= (\hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2) = 0.132\,1
\end{aligned}
\tag{17.13.15}
$$

$$\hat{\beta}_3=(\hat{a}_0+3\hat{a}_1+9\hat{a}_2)=-0.539\,4$$

于是，所估计的、对应于（17.13.11）的分布滞后模型为：

$$\begin{aligned}
\hat{Y}_t &= 25\,845.0 + 1.115\,0X_0+0.683\,6X_{t-1}+0.132\,1X_{t-2}-0.539\,4X_{t-3}\\
\text{se} &= (6\,596.99)\quad (0.538\,1)\quad (0.467\,2)\quad (0.465\,6)\quad (0.565\,6)\\
t &= (3.917\,7)\quad (2.071\,8)\quad (1.463\,0)\quad (0.283\,7)\quad (-0.953\,7)
\end{aligned}\tag{17.13.16}$$

估计的 β_i 值的几何图形如图 17.8 所示。

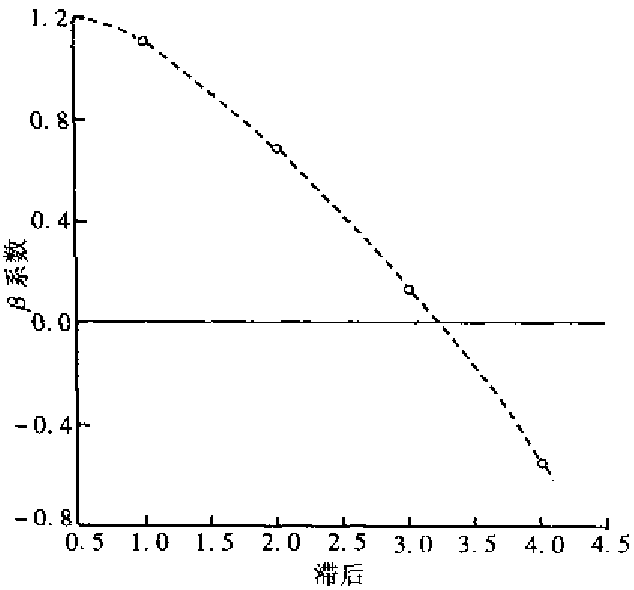


图 17.8　说明性例子的滞后结构

可用上述说明性例子来表明阿尔蒙程序的另外几个特点：

1. a 系数的标准误可从 OLS 回归（17.13.14）直接求得，但我们主要关心的系数的某些标准误却不能如此获得。但这些标准误可通过统计学中的一个熟知的公式，从所估 a 系数的标准误容易地算出。习题 17.18 给出这个
695 公式。当然，没必要用手工来做这个工作，因为大多数统计软件包能例行公事般地完成它。(17.13.15) 所给出的标准误可以用 Eviews 4 来得到。

2.（17.13.16）所求的诸 β 系数，因未受任何先验性约束，故称**无约束估计值**。然而，在某些情况中，人们也许想对诸 β 施加所谓**端点约束**（end-point restrictions），而令 β_0 和 β_k（即现期和第 k 个滞后系数）为零。由于心理、制度或技术上的原因，解释变量在现期中也许还不会对因变量的现值有什么影响，从而说明 β_0 取值零。同理，超过一定时期 k，自变量也许不再对因变量起作用，故可假定 β_k 为零。在存货这个例子中，X_{t-3} 的系数为负值，这可能没有经济意义。因此，我们可以限定该系数为零。[51] 当然，你不必限制两端；你可以只限制第一个系数（称为近端限制），或者只限制最后

一个系数（称为远端限制）。关于存货这个例子，习题 17.28 进行了说明。有时我们在诸 β 的总和为零的限制条件下来估计 β 值。但我们不能不经过仔细思索就加上这些限制条件，因为这些限制也会影响到其他（未受限制的）滞后系数的值。

3. 因为滞后系数的数值和多项式的次数的选择取决于建模者的判断力，所以试错法是不可避免的，尽管建模者可以控制数据的采集。这里就可以用到第 13 章讨论的**赤池和施瓦茨信息准则**。

4. 因为我们用三期滞后和二次多项式来估计（17.13.16），它是一个受限制的最小二乘模型。假定我们决定用三期滞后，但不使用阿尔蒙多项式法。也就是说，我们用 OLS 来估计（17.13.11）。那么会出现什么情况呢？我们首先来看如下结果：

$$
\begin{aligned}
\hat{Y}_t &= 26\,008.60 \quad + \; 0.977\,1X_t + 1.013\,9X_{t-1} - 0.202\,2X_{t-2} - 0.393\,5X_{t-3} \\
\text{se} &= (6\,691.12) \quad (0.682\,0) \quad (1.092\,0) \quad (1.102\,1) \quad (0.718\,6) \\
t &= \quad (3.887\,0)(1.432\,7) \quad (0.928\,4) \quad (-0.183\,5) \quad (-0.547\,6) \\
& \qquad R^2 = 0.975\,5 \qquad d = 0.157\,1 \qquad F = 379.51 \qquad (17.13.17)
\end{aligned}
$$

如果将这些结果同（17.13.16）所给出的结果进行比较，你将看出总的 R^2 实际上差不多是一样的，尽管（17.13.17）中的滞后模式表明它比（17.13.16）所展示的形状更加凸起。

696 正如该例所说明的那样，在使用阿尔蒙分布滞后技巧时必须小心，因为结果可能对于多项式的次数和/或滞后系数个数的选择比较敏感。

§17.14 经济学中的因果关系：格兰杰检验[52]

在第 1.4 节中我们说过，虽然回归分析考虑一个变量依赖于另一个变量，但这不一定意味着因果关系。换言之，变量间某种关系的存在不能够证明是因果关系或者影响的方向。但在涉及时间序列数据的回归中，情况可能有一点不同，因为正如一个作者所说的：

……时间不会倒退。即如果事件 A 在事件 B 之前发生，那么可能是 A 引起了 B。但是，不可能是 B 引起了 A。换言之，过去的事件能够引起今天发生的事件，将来的事件却不能引起今天的事件。[53]（着重强调。）

在所谓的格兰杰因果检验（Granger causality test）后面有一个粗略的想法。[54]但是值得注意的是，因果问题相当复杂，而且有着各种各样的争论。一个极端是相信“万物引起万物”的人们，而另一个极端是否认任何因果关系存在的人们。[55]计量经济学家爱德华·里默喜欢先后关系这个术语甚于因果关系。弗朗西斯·狄波德更喜欢用预计因果关系这个术语。他写道：

……“y_i 是 y_j 的原因”这种表达只是为了更精确而采用的一种简洁明了的表达方式，而另外一种冗长的表达方式是“除了这个系统中其他变量的

历史之外，y_i 包含着可预见 y_j 的有用的信息（从线性最小二乘的意义上讲）。”为了节省空间，我们简单的说 y_i 是 y_j 的原因。[56]

格兰杰检验

697 为了解释格兰杰检验，我们考虑在宏观经济学中经常会问到的问题：是GDP“引起”货币供给 M（GDP→M）还是货币供给 M“引起”GDP（M→GDP），这里箭头指出了因果关系的方向。格兰杰因果关系检验假定了有关GDP或 M 变量的预测信息全部包含在这两个变量的时间序列之中。检验要求估计以下的回归：

$$\mathrm{GDP}_t = \sum_{i=1}^{n} \alpha_i M_{t-i} + \sum_{j=1}^{n} \beta_j \mathrm{GDP}_{t-j} + u_{1t} \tag{17.14.1}$$

$$M_t = \sum_{i=1}^{n} \lambda_i M_{t-i} + \sum_{j=1}^{n} \delta_j \mathrm{GDP}_{t-j} + u_{2t} \tag{17.14.2}$$

其中干扰项 u_{1t} 和 u_{2t} 假定为不相关的。顺便提一下，请注意因为有两个变量，所以我们正在处理双向因果关系。在关于时间序列的章节中，我们将通过**向量自回归**（VAR）的技巧来扩展到多元因果关系。

方程（17.14.1）假拟当前GDP与GDP自身以及 M 的过去值有关，而（17.14.2）对 M_t 也假拟了类似的行为。注意，这些回归还可被铸造成增长的形式 $\dot{\mathrm{GDP}}$ 和 $\dot{M}$，其中变量上方的圆点表示增长率。现在分四种情形讨论：

1. 如果对（17.14.1）中的滞后 M 所估计的系数作为一个群体是统计上异于零的（即 $\sum \alpha_i \neq 0$），并且对（17.14.2）中的滞后GDP所估计的系数的集合不是统计上异于零的（即 $\sum \delta_j \neq 0$），则表明有从 M 到GDP的单向因果关系（unidirectional causality）。

2. 反之，如果（17.14.1）中的滞后 M 的系数集不是统计上异于零的（即 $\sum \alpha_i = 0$），而（17.14.2）中的滞后GDP的系数集却是统计上异于零的（即 $\sum \delta_j \neq 0$），则存在从GDP到 M 的单向因果关系。

3. 如果 M 和GDP的系数集在两个回归中都是统计上异于零的，则表示有反馈（feedback）或双向（bilateral）因果关系。

4. 最后，如果 M 和GDP的系数集在两个回归中都不是统计上显著的，就表示两者之间各自的独立性（independence）。

此外，一般地，由于将来不能预测过去，如果变量 X 是变量 Y 的（格兰杰）原因，则 X 的变化应先于 Y 的变化。因此，在做 Y 对其他变量（包括自身的过去值）的回归时，如果把 X 的过去或滞后值包括进来能显著地改进对 Y 的预测，我们就可以说 X 是 Y 的（格兰杰）原因。类似地定义 Y 是 X 的（格兰杰）原因。

做格兰杰因果关系检验的步骤如下，现通过方程（17.14.1）所给的

GDP—货币一例加以说明：

698 1. 将当前的 GDP 对所有的滞后 GDP 项以及别的什么变量（如果有的话）做回归，但在这一回归中没有把滞后 M 变量包括进来。根据第 8 章，这是一个受约束的回归。然后从它得到受约束的残差平方和 RSS_R。

2. 现在做含有滞后 M 项的回归，用第 8 章的语言，这是一个无约束的回归，由此回归得到无约束的残差平方和 RSS_{UR}。

3. 虚拟假设是 H_0: $\sum \alpha_i = 0$，即滞后 M 项不属于此回归。

4. 为了检验此假设，我们利用（8.7.9）所给的 F 检验，即：

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} \tag{8.7.9}$$

它遵循自由度为 m 和 $(n-k)$ 的 F 分布。在本例中 m 等于滞后 M 项的个数，而 k 是无约束回归中待估参数的个数。

5. 如果在选定的显著性水平上计算的 F 值超过临界 F 值，则拒绝虚拟假设，这样滞后 M 项就属于此回归。这是 M 导致 GDP 的另一种说法。

6. 为了检验模型（17.14.2），即检验 GDP 是否是导致 M 的原因，可重复步骤 1 至 5。

在我们说明格兰杰因果关系检验之前，有几件事必须引起注意：

1. 假定这两个变量（即 GDP 和 M）是平稳的。我们前面已经用直观的词语讨论过平稳性的概念，并将在第 21 章更正式的讨论。如果变量在水平的形式上已经不平稳，有时我们对变量进行一阶微分来使其平稳。

2. 在因果关系检验中引入的滞后期的个数是一个重要的应用性的问题。因为在分布滞后模型的这个例子中，我们可能不得不使用赤池或者施瓦茨信息准则来做出选择。但补充说明一点，即因果关系的方向可能严格依赖于所包含的滞后期的个数。

3. 我们假定进入因果关系检验的误差项是无关联的。如果情况不是这样，那么就不得不采取恰当的变换（正如第 12 章所讨论的）。[57]

4. 因为兴趣在于检验因果关系，所以我们不必明显地给出模型（17.14.1）和（17.14.2）的估计系数（为了节省篇幅）；只须给出（8.7.9）中的 F 检验结果便足够了。

699 **例 17.12 货币和收入间的因果关系**

R.W. 哈弗（Hafer）利用格兰杰检验以判明 1960 年第 1 季度至 1980 年第 4 季度期间美国的 GNP（而不是 GDP）与 M 之间究竟是怎样一种因果关系。他用这些变量的增长率（$\dot{GNP}$ 和 $\dot{M}$）来代替它们的总值，并且在前面给出的两个回归中对每个变量用了四期滞后。结果如下所示。[58] 每个例子中的虚拟假设是：所考虑的变量不是另一变量的“格兰杰原因”。

因果关系的方向	F 值	决定
$\dot{M} \rightarrow \dot{\text{GNP}}$	2.68	拒绝
$\dot{\text{GNP}} \rightarrow \dot{M}$	0.56	不拒绝

这些结果意味着因果关系的方向是从货币增长到 GNP 增长，因为估计的 F 在 5%水平是显著的；临界的 F 值是 2.50（对于 4 和 71 的自由度而言）。另一方面，不存在"反向"因果关系，即从 GNP 的增长到 M 的增长，因为 F 值是统计上不显著的。

例 17.13 加拿大货币和利率间的因果关系

参照表 17.3 给出的加拿大的数据。假定我们想查明加拿大 1979—1988 年间每季度货币供给和利率之间是否存在任何因果关系。为了表明格兰杰因果检验十分依赖于模型中引入的滞后期的个数，我们提供了使用几个（季度）滞后的 F 检验的结果。在每种情况下，虚拟假设为利率不是货币供给的（格兰杰）原因，反之亦然。

因果关系的方向	滞后期的数量	F 值	决定
$R \rightarrow M$	2	12.92	拒绝
$M \rightarrow R$	2	3.22	拒绝
$R \rightarrow M$	4	5.59	拒绝
$M \rightarrow R$	4	2.45	拒绝（在 7%水平）
$R \rightarrow M$	6	3.516 3	拒绝
$M \rightarrow R$	6	2.71	拒绝
$R \rightarrow M$	8	1.40	不拒绝
$M \rightarrow R$	8	1.62	不拒绝

注意上述 F 检验结果的特性：一直到第 6 期滞后，货币供给与利率之间才存在双向的因果关系。但是，在第 8 期滞后，这两个变量之间不存在统计上可辨别的关系。这更加能说明我们前面得出的观点：格兰杰检验的结果对于模型中引入的滞后期的个数是敏感的。

700 **例17.14 九个东亚国家（地区）的 GDP 增长率和总储蓄率之间的因果关系**

GDP 增长率（g）和总储蓄率（s）之间的双向因果关系的一项研究表明如表 17.7 所示的结果。[59] 为了进行对比，该表也提供了美国的结果。表 17.7 所提供的结果大体上表明：对于大多数东亚国家（地区），因果关系是从 GDP 增长率到总储蓄率。与此对照，美国 1950—1988 年间直到第 3 期滞

后，因果关系仍是双向的，但到了第4期和第5期滞后，因果关系是从GDP到储蓄率，而不是相反的方向。

为了对格兰杰检验的讨论做出推断，请记住我们检验的问题是：当两个变量之间暂时存在先导和滞后关系时，人们是否能从统计上查明其因果关系。如果因果关系已经建立了，则意味着人们可以用一个变量来更好地预测另一变量，这种预测比简单地根据另一变量过去的历史所做出的预测更准确。在东亚经济的例子中，似乎我们能通过考虑GDP增长率的滞后值来更好地预测总储蓄率，这种预测比仅仅考虑总储蓄率的滞后值所做出的预测更准确。

表17.7 实际人均GDP增长率和总储蓄率之间的双变量格兰杰因果关系检验

国家（地区），年份	滞后年数	等式右侧滞后变量储蓄	增长	国家（地区），年份	滞后年数	等式右侧滞后变量储蓄	增长
中国香港，1960—1988	1	Sig	Sig	菲律宾，1950—1988	1	NS	Sig
	2	Sig	Sig		2	NS	Sig
	3	Sig	Sig		3	NS	Sig
	4	Sig	Sig		4	NS	Sig
	5	Sig	Sig		5	NS	Sig
印度尼西亚，1965	1	Sig	Sig	新加坡，1960—1988	1	NS	NS
	2	NS	Sig		2	NS	NS
	3	NS	Sig		3	NS	NS
	4	NS	Sig		4	Sig	NS
	5	NS	Sig		5	Sig	NS
日本，1950—1988	1	NS	Sig	中国台湾，1950—1988	1	Sig	Sig
	2	NS	Sig		2	NS	Sig
	3	NS	Sig		3	NS	Sig
	4	NS	Sig		4	NS	Sig
	5	NS	Sig		5	NS	Sig
朝鲜，1955—1988	1	Sig	Sig	泰国，1950—1988	1	NS	Sig
	2	NS	Sig		2	NS	Sig
	3	NS	Sig		3	NS	Sig
	4	NS	Sig		4	NS	Sig
	5	NS	Sig		5	NS	Sig
马来西亚，1955—1988	1	Sig	Sig	美国，1950—1988	1	Sig	Sig
	2	Sig	Sig		2	Sig	Sig

3	NS	NS	3	Sig	Sig
4	NS	NS	4	NS	Sig
5	NS	Sig	5	NS	Sig

Sig：显著；NS：不显著

注：增长是以 1985 年的国际价格计算的实际人均 GDP 增长。

资料来源：World Bank，*The East Asian Miracle*：*Economic Growth and Public Policy*，Oxford University Press，New York，1993，p.244，（Table A5－2）. The original source is Robert Summers and Alan Heston，"The Penn World Tables（Mark 5）：An Expanded Set of International Comparisons，1950－1988，" *Quarterly Journal of Economics*，vol.105，no.2，1991.

※关于因果关系和外生性的一个注解

701 当我们在本书的第 4 部分研究联立方程模型时，经济变量常常被划分成两大类，即**内生变量**和**外生变量**。宽松地讲，内生变量是单一方程回归模型中因变量的等价物，而外生变量是此类模型中 X 变量或者回归元的等价物，只要 X 变量和该方程的误差项不相关。[60]

现在会出现一个有趣的问题：假定在一个格兰杰因果关系检验中，我们发现 X 变量是 Y 变量的（格兰杰）原因，而后者不是前者的原因（即无双向因果关系）。那么能否把 X 变量当成外生变量呢？换句话说，我们能否用格兰杰因果关系（或非因果关系）来建立外生性呢？

为了回答这个问题，我们需要区分三种类型的外生性：（1）弱的，（2）强的，和（3）超级的。为了使说明简单，假定我们只考虑两个变量 Y_t 和 X_t，并且进一步假定把 Y_t 对 X_t 进行回归。如果 Y_t 又不能解释 X_t，则称 X_t 是弱外生的。这时，以 X_t 的值为条件，就可以做回归模型的估计和检验。实际上，回到第 2 章，你将会意识到我们的回归模型的建立是以给定 X_t 变量为条件的。如果现期和滞后的 Y 值都不解释 X_t（也就是没有反馈关系），则称 X_t 是强外生的。如果即使是 X 值发生变化，在 Y 和 X 回归中的参数仍然不变，则称 X_t 是超级外生的；也就是说，参数值是不随 X 值的变化而变化的。如果实际情况是这样的话，那么著名的"卢卡斯批判"站不住脚。[61]

区分三类外生性的原因在于："一般而言，弱外生性是估计和检验所必需的，强外生性对于预测是必要的，而超级外生性对于政策分析是必要的。"[62]

702 回到格兰杰因果关系，如果变量 Y 不是另一变量 X 的原因，那么我们能否假设后者是外生性的呢？不幸的是，答案不是直截了当的。如果我们正在谈论弱外生性，那么它将表明格兰杰因果关系对于建立外生性既不是必要的也不是充分的。另一方面，格兰杰因果关系对于强外生性来说是必要的（但不是充分的）。这些陈述的证明超出了本书的范围。[63]对于我们而言，有

必要把格兰杰因果关系和外生性的概念区别开来，并且将前者当作分析时间序列的一个有益的描述性工具。

§17.15 要点与结论

1. 心理上、技术上和制度上的理由，一个回归子对一（多）个回归元的回应会带有一定的时滞。考虑时间滞后的回归模型称为**动态**或**滞后回归模型**。

2. 有两类滞后回归模型：**分布滞后**和**自回归**。在前一种情形中，由诸回归元的当前和滞后值作为解释变量而出现；而在后一种情形中，有回归子的滞后值作为解释变量而出现。

3. 一个纯粹的分布滞后模型可用 OLS 去估计，但这时由于一个回归元的相继滞后值的相关趋向而会有多重共线性的问题。

4. 为此，人们设计了简洁的方法，包括考伊克、适应性预期和部分调整等机制，前者纯粹是一个代数方法，而后两者则有经济原理作为其依据。

5. 但**考伊克**、**适应性预期**和**部分调整诸模型**有一个独特的性质，就是它们全是自回归模型；回归子的滞后值都作为解释变量出现。

6. 自回归性质在估计问题上提出了挑战；如果滞后回归子和误差项相关，则这些模型的 OLS 估计量不仅是有偏误的，而且是非一致性的。考伊克和适应性预期模型属于有偏误和非一致性情形；部分调整模型则有所不同，尽管滞后回归元出现，仍可用 OLS 获得一致性估计。

7. 为了得到考伊克和适应性预期模型的一致性估计，最广为应用的方法是**工具变量法**。工具变量是滞后回归子的一个替代变量，但它具有与误差项无关的性质。

8. 取代方才讨论的滞后回归模型的另一方法，是**阿尔蒙多项式分布滞后模型**。它避免了自回归模型带来的估计问题。然而，阿尔蒙方法的主要问题是，使用者必须事先规定滞后长度和多项式次数。解决滞后长度和多项式次数的选择问题，有正式的和非正式的两种方法。

9. 分布滞后和自回归模型由于明确地考虑了时间的作用，而把（不然的话将是）静态的经济理论变成了动态的，故在经验经济学中显得极为有用。尽管它们有估计上的困难，却是可以克服的。这些模型，帮助我们区分相对于解释变量的一个单位变化的因变量的短期和长期响应。例如，为了估计短期和长期价格、收入的替代以及其他弹性，这些模型是非常有用的。[64]

10. 分布滞后和（或）自回归模型由于涉及滞后而引发了经济变量中的因果关系的讨论。在应用研究中，格兰杰因果模型受到了广泛的注意。但因格兰杰方法对模型中所取的滞后长短异常敏感，故在应用中要保持高度警觉。

11. 即使变量 X 是另一个变量 Y 的“格兰杰原因”，这并不意味着 X 是

外生性的。我们区分了三种类型的外生性——弱的，强的和超级的，并且指出了这种区分的重要性。

习题

问答题

17.1 用简单的理由说明以下的陈述是正确的、错误的或不确定的：

a. 所有计量经济模型本质上都是动态的。

b. 如果有某些分布滞后系数是正的，而另一些是负的，考伊克模型就没有多大意义。

c. 如果用 OLS 估计考伊克和适应性预期模型，则估计量将是有偏误，但却是一致性的。

d. 在部分调整模型中，OLS 估计量在有限样本中是有偏误的。

e. 当一（多）个随机回归元和一个自相关误差项同时出现时，工具变量法将产生无偏且一致性估计。

f. 当一个滞后回归子作为一个回归元出现时，用德宾-沃森 d 统计量去侦破自相关性实际上是无效的。

g. 德宾 h 检验对大和小样本都是有效的。

h. 格兰杰检验与其说是因果关系检验，不如说是领先滞后检验。

17.2 建立方程 (17.7.2)。

17.3 证明方程 (17.8.3)。

17.4 假定价格是按照如下的适应性预期假设形成的：

$$P_t^* = \gamma P_{t-1} + (1-\gamma)P_{t-1}^*$$

其中 P^* 是预期价格而 P 是真实价格。

假定 $\gamma = 0.5$[1]，试填补下表中的空格。

时期	P^*	P
$t-3$	100	110
$t-2$		125
$t-1$		155
t		185
$t+1$		—

17.5 考虑模型：$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 Y_{t-1} + v_t$，假定 Y_{t-1}和 v_t 相关。为了消除相关，假定我们采取以下的工具变量方法：先求 Y_t 对 X_{1t}和 X_{2t}的回归，并从此回归得到估计值 $\hat{Y}_t$。然后做回归：

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 \hat{Y}_{t-1} + v_t$$

其中 $\hat{Y}_{t-1}$是从第一步回归估计出来的。

704 a. 这一方法是怎样把原模型中的 Y_{t-1}和 v_t 之间的相关消除的？

b. 和利维亚坦(Liviatan)方法比较，以上建议的方法有些什么优点？

***17.6** a. 建立（17.4.8）。

b. 对 $\lambda=0.2$，0.4，0.6，0.8 估计中位滞后。

c. λ 值与中位滞后值之间有何规律性的关系？

17.7 a. 证明对于考伊克模型平均滞后有如（17.4.10）所示。

b. 如果 λ 比较大，这将意味着什么？

17.8 利用（17.4.9）所给的平均滞后公式，证实表 17.1 中展示的平均滞后 10.959 个季节无误。

17.9 假令：$M_t = \alpha + \beta_1 Y_t^* + \beta_2 R_t^* + u_t$

其中 M = 对实际现金余额的需求，Y^* = 期望实际收入，以及 R^* = 预期利率。假定期望值的形成方式如下：

$$Y_t^* = \gamma_1 Y_t + (1-\gamma_1) Y_{t-1}^*$$
$$R_t^* = \gamma_2 R_t + (1-\gamma_2) R_{t-1}^*$$

其中 γ_1 和 γ_2 是期望系数，均落在 0 与 1 之间。

705 a. 你怎样用可观测的数量来表达 M_t？

b. 你预见到什么估计问题？

***17.10** 如果你用 OLS 估计（17.7.2），你能不能导出原始参数的估计值呢？你预见到什么问题？（关于细节，参看 Roger N. Waud。[2]）

17.11 **序列相关模型**。考虑如下模型：

$$Y_t = \alpha + \beta X_t + u_t$$

假定 u_t 遵循第 12 章中所给的马尔可夫一阶自回归模式，即：

$$u_t = \rho u_{t-1} + \varepsilon_t$$

其中 ρ 是（一阶）自相关系数，而 ε_t 满足全部经典 OLS 假定。于是，如第 12 章所表明的，模型：

$$Y_t = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \rho Y_{t-1} + \varepsilon_t$$

将有一序列无关的误差项，使得 OLS 估计成为可能。但这个被称为序列相关模型（serial correlation model）的模型非常像考伊克、适应性预期和部分调整模型。那么，你怎样会知道在任一情况中，上述诸模型中的哪一个适用。[3]

17.12 考虑（17.4.7）所给的考伊克（或者适应性预期也行）模型，即：

$$Y_t = \alpha(1-\lambda) + \beta_0 X_t + \lambda Y_{t-1} + (u_t - \lambda u_{t-1})$$

假定在原始模型中 u_t 遵循一阶自回归模式 $u_t - \rho u_{t-1} = \epsilon_t$，其中 ρ 是自相关系数，而 ϵ_t 满足全部经典 OLS 假定。

a. 如果 $\rho = \lambda$，那么能不能用 OLS 估计考伊克模型？

b. 这样得到的估计值将是无偏的？一致性的？为什么或为什么不？

c. 假定 $\rho = \lambda$ 的合理性如何？

17.13 **三角形或算术分布滞后模型。**[4]此模型假定刺激变量（解释变量）在当前时期发挥它的最大影响，然而随着它奔向遥远的过去，影响按等差级数下降到零。从几何上看如图 17.9 所示。

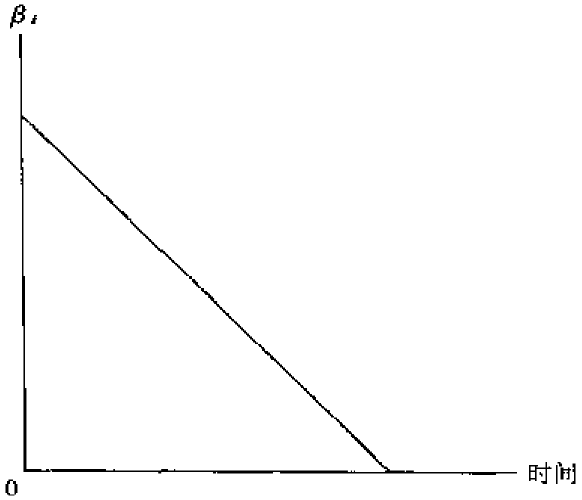


图 17.9　（费希尔的）三角形或算术滞后模式

706 假使按照这种分布，我们做如下的一连串回归：

$$Y_t = \alpha + \beta\left(\frac{2X_t + X_{t-1}}{3}\right)$$

$$Y_t = \alpha + \beta\left(\frac{3X_t + 2X_{t-1} + X_{t-2}}{6}\right)$$

$$Y_t = \alpha + \beta\left(\frac{4X_t + 3X_{t-1} + 2X_{t-2} + X_{t-3}}{10}\right)$$

等等，并选择有最高 R^2 的回归作为“最优”回归，对此策略试加评论。

17.14 根据 1950—1960 年间的季度数据，F.P.R. 布列奇凌（Brechling）得到如下英国经济的劳动需求函数（括号中的数字是标准误）[5]：

$$\dot{E}_t = 14.22 + 0.172Q_t - 0.028t - 0.000\,7t^2 - 0.297E_{t-1}$$
$$(2.61)(0.014)\quad(0.015)\quad(0.000\,2)\quad(0.033)$$
$$R^2 = 0.76 \qquad d = 1.37$$

其中 $\dot{E}_t = (E_t - E_{t-1})$

Q = 产出

t = 时间

上述方程依据了一个假定，即就业的理想水平是产出、时间和时间平方的函数，和一个假设，即 $E_t - E_{t-1} = \delta(E_t^* - E_{t-1})$，其中调整系数 δ 落在 0 与 1 之间。

a. 解释上述回归。

b. δ 值为何？

c. 从所估的短期需求函数导出长期需求函数。

d. 你怎样检验上述模型中的序列相关？

707 **17.15** 格里利谢斯曾用以下模型研究农场对拖拉机的需求[6]：

$$T_t^* = \alpha X_{1,t-1}^{\beta_1} X_{2,t-1}^{\beta_2}$$

其中 T^* = 拖拉机的理想存量

X_1 = 拖拉机的相对价格

X_2 = 利率

他利用存量调整模型，对 1921—1957 年得到如下结果：

$$\widehat{\log T_t} = \text{常数} - 0.218\log X_{1,t-1} - 0.855\log X_{2,t-1} + 0.864\log T_{t-1}$$
$$(0.051) \qquad (0.170) \qquad (0.035)$$
$$R^2 = 0.987$$

这里，括号中的数字是估计的标准误。

a. 估计的调整系数为何？

b. 短期的、长期的价格弹性各为多少？

c. 相应的利息弹性为多少？

d. 在本模型中出现高或低的调整速度的理由是什么？

17.16 每当滞后因变量作为一个解释变量出现时，R^2 通常都要比它不出现时要高许多。观察到这种现象的理由是什么？

17.17 考虑图 17.10 中的滞后模式，你会用多高次的多项式去拟合滞后的这些结构，为什么？

17.18 考虑方程（17.13.4）：

$$\beta_i = a_0 + a_1 i + a_2 i^2 + \cdots + a_m i^m$$

为了从 $\hat{\alpha}_i$ 的方差得到 $\hat{\beta}_i$ 的方差，我们利用如下公式：

$$\begin{aligned}\operatorname{var}(\hat{\beta}_i) &= \operatorname{var}(\hat{\alpha}_0 + \hat{\alpha}_1 i + \hat{\alpha}_2 i^2 + \cdots + \hat{\alpha}_m i^m) \\ &= \sum_{j=0}^{m} i^{2j}\operatorname{var}(\hat{\alpha}_j) + 2\sum_{j<p} i^{(j+p)}\operatorname{cov}(\hat{\alpha}_j\hat{\alpha}_p)\end{aligned}$$

a. 利用上述公式求如下表达的 $\hat{\beta}_i$ 的方差：

$$\hat{\beta}_i = \hat{a}_0 + \hat{a}_1 i + \hat{a}_2 i^2$$
$$\hat{\beta}_i = \hat{a}_0 + \hat{a}_1 i + \hat{a}_2 i^2 + \hat{a}_3 i^3$$

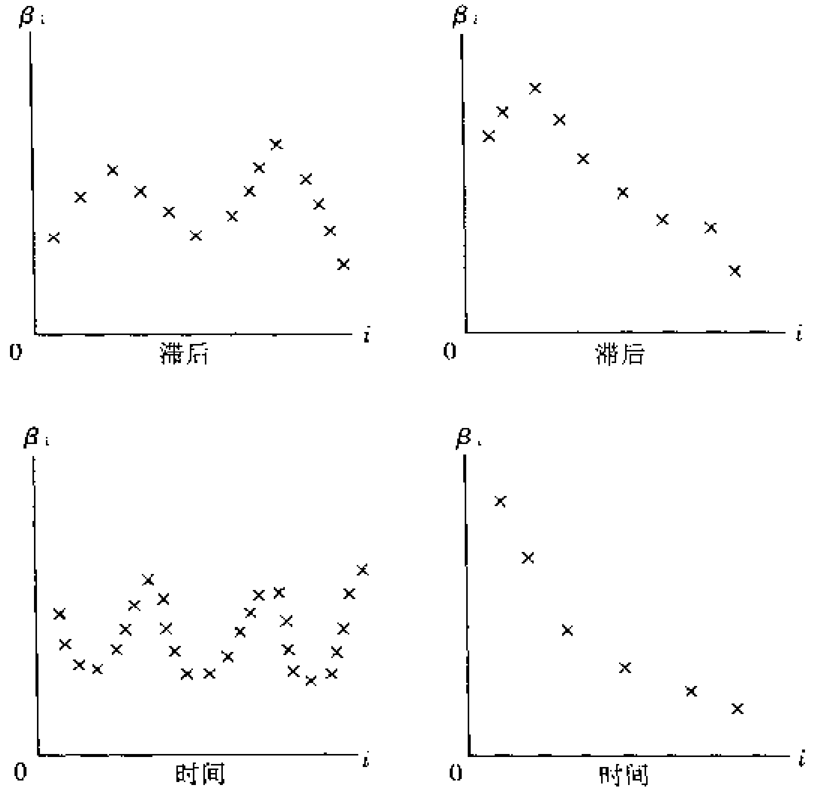


图 17.10 假想滞后结构

b. 如果诸 a_i 的方差相对于它们本身较大，则 $\hat{\beta}_i$ 的方差也将较大？为什么或为什么不？

708 **17.19** 考虑如下的分布滞后模型：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

假定 β_i 可适当地用二次多项式表达如下：

$$\beta_i = a_0 + a_1 i + a_2 i^2$$

如果你想施加约束 $\beta_0 = \beta_4 = 0$，你将怎样估计这些 β_i？

709 **17.20** **倒 V 形分布滞后模型**。考虑 k 期有限分布滞后模型：

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + u_t$$

F. 狄利乌（DeLeeuw）曾提出像图 17.11 那样的 β 结构，诸 β_i 按照一个倒转的 V 形而变化。为简单起见，假定 k（滞后的最大长度）是偶数并假定 β_0 和 β_k 是零。狄利乌建议对诸 β_i 采用如下模式[7]：

$$\begin{aligned} \beta_i &= i\beta & 0 \leqslant i \leqslant \frac{k}{2} \\ &= (k-i)\beta & \frac{k}{2} \leqslant i < k \end{aligned}$$

你怎样用狄利乌的模式去估计上述 k 期分布滞后模型呢？

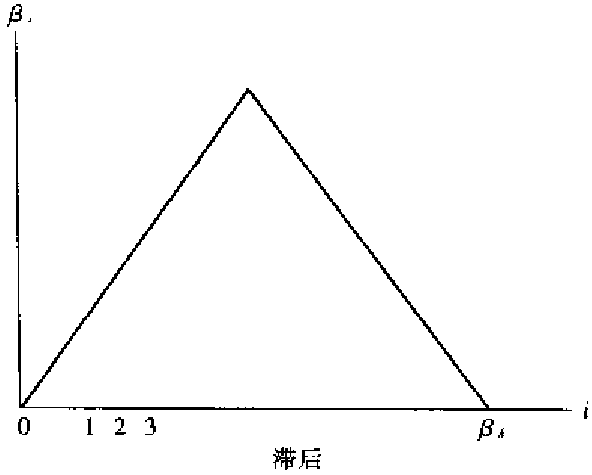


图 17.11　倒 V 形分布滞后模型

17.21　参照习题 12.15，既然那里显示的 d 值对（一阶）自相关的发现没有什么用处（为什么），对于这种情形，你会怎样检验自相关？

解答题

17.22　考虑如下模型：

$$Y_t^* = \alpha + \beta_0 X_t + u_t$$

其中 Y_t^* = 理想的或长期的新建厂房设备企业开支，X_t = 销售量，及 t = 时间，利用存量调整模型，从表 17.8 中的数据估计出对新厂房设备开支的长期和短期需求函数中的参数。

你怎样去发现数据中是否有序列相关？

710

表 17.8　　1970—1991 年间美国制造业固定厂房设备投资 Y 和销量 X_2，以 10 亿美元计，并经季节调整

年份	厂房开支 Y	销量 X_2	年份	厂房开支 Y	销量 X_2
1970	36.99	52.805	1981	128.68	168.129
1971	33.60	55.906	1982	123.97	163.351
1972	35.42	63.027	1983	117.35	172.547
1973	42.35	72.931	1984	139.61	190.682
1974	52.48	84.790	1985	152.88	194.538
1975	53.66	86.589	1986	137.95	194.657
1976	58.53	98.797	1987	141.06	206.326
1977	67.48	113.201	1988	163.45	223.541

1978	78.13	126.905	1989	183.80	232.724
1979	95.13	143.936	1990	192.61	239.459
1980	112.60	154.391	1991	182.81	235.142

资料来源：*Economic Report of the President*, 1993. Y 数据取自 Table B－52，第 407 页；X_2 数据取自 Table B－53，第 408 页。

17.23 利用习题 17.22 的数据，考虑如下模型：

$$Y_t^* = \beta_0 X_t^{\beta_1} e^{u_t}$$

用存量调整模型（为什么?）估计新厂房设备开支的销售量短期和长期弹性。将你的结果同习题 17.22 的结果相比较，你会选择哪个模型，为什么？数据中是否有序列相关？你怎样知道的？

17.24 利用习题 17.22 的数据，但假定：

$$Y_t = \alpha + \beta X_t^* + u_t$$

其中 X_t^* 为理想销售量。估计此模型的参数，并将结果同习题 17.22 所得到的相比较。你怎样决定哪个模型是适当的模型？根据 h 统计量，你会得出数据中有序列相关的结论吗？

17.25 假如有人使你相信企业的新厂房和设备开支与销售量有如下的关系：

$$Y_t^* = \alpha + \beta X_t^* + u_t$$

其中 Y^* 是理想的开支而 X^* 是预想的或期望的销售量。用习题 17.22 所给的数据去估计此模型，并评论你的结果。

17.26 利用习题 17.22 所给的数据，判定厂房开支是销售量的格兰杰原因，抑或销售量是厂房开支的格兰杰原因呢？使用直至 6 期为止的滞后并评述你的结果。你从这个练习中引出什么重要的结论？

711 **17.27** 假定习题17.22中的销售量对厂房设备支出有一个分布滞后效应，试用一个阿尔蒙滞后模型拟合该数据。

17.28 将方程（17.13.16）加上（1）近端限制，(2) 远端限制，和（3）两端限制等限制条件重新进行估计，并将所得结果与方程（17.13.16）进行比较，你能得出什么结论？

17.29 表 17.9 给出了如下数据：私人在信息处理和设备上的固定投资（Y，10 亿美元），总制造业和贸易销售额（X_2，百万美元），利率（X_3，穆迪 Aaa 公司债券利率，%）；Y 和 X_2 数据经季节调整

a. 检验 Y 和 X_2 之间的双向因果关系，注意滞后长度

b. 检验 Y 和 X_3 之间的双向因果关系，注意滞后长度

c. 考虑到销售对投资的分布滞后效应，假定你决定用阿尔蒙滞后法，列出估计模型，适当注意滞后长度和多项式的次数。

表 17.9　　投资、销售和利率，美国，1960—1999

年份	投资	销售	利率	年份	投资	销售	利率
1960	4.9	60 827	4.41	1980	69.6	327 233	11.94
1961	5.2	61 159	4.35	1981	82.4	355 822	14.17
1962	5.7	65 662	4.33	1982	88.9	347 625	13.79
1963	6.5	68 995	4.26	1983	100.8	369 286	12.04
1964	7.3	73 682	4.40	1984	121.7	410 124	12.71
1965	8.5	80 283	4.49	1985	130.8	422 583	11.37
1966	10.6	87 187	5.13	1986	137.6	430 419	9.02
1967	11.2	90 820	5.51	1987	141.9	457 735	9.38
1968	11.9	96 685	6.18	1988	155.9	497 157	9.71
1969	14.6	105 690	7.03	1989	173.0	527 039	9.26
1970	16.7	108 221	8.04	1990	176.1	545 909	9.32
1971	17.3	116 895	7.39	1991	181.4	542 815	8.77
1972	19.3	131 081	7.21	1992	197.5	567 176	8.14
1973	23.0	153 677	7.44	1993	215.0	595 628	7.22
1974	26.8	177 912	8.57	1994	233.7	639 163	7.96
1975	28.2	182 198	8.83	1995	262.0	684 982	7.59
1976	32.4	204 150	8.43	1996	287.3	718 113	7.37
1977	38.6	229 513	8.02	1997	325.2	753 445	7.26
1978	48.3	260 320	8.73	1998	367.4	779 413	6.53
1979	58.6	297 701	9.63	1999	433.0	833 079	7.04

注：投资＝私人在信息处理设备和软件上的固定投资，10 亿美元，经季节调整

销售＝总制造业和贸易销售额，百万美元，经季节调整

利率＝穆迪 Aaa 公司债券利率，%

资料来源：*Economic Report of the President*，2001，Table B－18，B－57，and B－73.

17.30　表 17.10 给出了 1960—1999 年间美国经济的商业部门每小时实际补偿指数和每小时产量指数的数据，这两个指数均以 1992＝100 为基数，民间失业率（X_3）也是指这一时期。

712

表 17.10 补偿（COMP）、生产力和失业率（UNRate），美国，1960—1999

年份	COMP	产量	UNRate	年份	COMP	产量	UNRate
1960	60.0	48.8	5.5	1980	89.5	80.4	7.1
1961	61.8	50.6	6.7	1981	89.5	82.0	7.6
1962	63.9	52.9	5.5	1982	90.9	81.7	9.7
1963	65.4	55.0	5.7	1983	91.0	84.6	9.6
1964	67.9	57.5	5.2	1984	91.3	87.0	7.5
1965	69.4	59.6	4.5	1985	92.7	88.7	7.2
1966	71.9	62.0	3.8	1986	95.8	91.4	7.0
1967	73.8	63.4	3.8	1987	96.3	91.9	6.2
1968	76.3	65.4	3.6	1988	97.3	93.0	5.5
1969	77.4	65.7	3.5	1989	95.9	93.9	5.3
1970	78.9	67.0	4.9	1990	96.5	95.2	5.6
1971	80.4	69.9	5.9	1991	97.5	96.3	6.8
1972	82.7	72.2	5.6	1992	100.0	100.0	7.5
1973	84.5	74.5	4.9	1993	99.9	100.5	6.9
1974	83.5	73.2	5.6	1994	99.7	101.9	6.1
1975	84.4	75.8	8.5	1995	99.3	102.6	5.6
1976	86.8	78.5	7.7	1996	99.7	105.4	5.4
1977	87.9	79.8	7.1	1997	100.4	107.6	4.9
1978	89.5	80.7	6.1	1998	104.3	110.5	4.5
1979	89.7	80.7	5.8	1999	107.3	114.0	4.2

注：COMP＝每小时实际补偿指数（1992＝100）

产量＝每小时产量指数（1992－100）

UNRate＝民间失业率，%

资料来源：*Economic Report of the President*，2001，Table B－49，p.332.

a. 你如何判断是工资补偿决定劳动生产力还是恰好相反？

b. 提出一个合适的模型来检验你在（a）中的推测，并提供有用的统计量。

c. 你认为失业率对工资补偿有影响吗？如果有，你如何将之考虑进去？列出必要的统计分析。

17.31 **西姆斯的因果关系检验**（Sims′test of Causality）。[8]在格兰杰因果关系的纠缠中，西姆斯另辟蹊径，发现将来不能是现在的原因。

713 假如我们想弄清楚 X 是不是 Y 的原因。现在考虑如下模型：

$$Y_t = \alpha + \beta_k X_{t-k} + \beta_{k-1} X_{t-k-1} + \cdots + \beta_1 X_{t-1} + B_0 X_t + \lambda_1 X_{t+1} + \lambda_2 X_{t+2} + \cdots + \lambda_m X_{t+m} + u_t$$

此回归包含回归元 X 的滞后、现在和将来或**先导值**；像 X_{t+1} 和 X_{t+2}这样的项叫做**先导项**（lead terms）。在上述回归中有 k 个滞后和 m 个先导项。如果 X 是 Y 的格兰杰原因，则诸 X 先导项的系数的总和必须在统计上等于零。[9]

用西姆斯的因果关系检验判定销售量是否是决定投资支出的（格兰杰）原因。自行决定回归元的适当先导和滞后长度。

[习题注释]

[1] 由 G.K.Shaw，前引文献改编而来，第 26 页。

[2] "Misspecification in the 'Partial Adjustment' and 'Adaptive Expectations' Models," *International Economic Review*, vol. 9, no. 2, June 1968, pp.204 – 217.

[3] 关于系列相关模型的一个讨论，参看 Zvi Griliches, "Distributed Lags: A Survey," *Econometrica*, vol. 35, no. 1, January 1967, p.34。

[4] 该模型提出自 Irving Fisher in "Note on a Short-Cut Method for Calculating Distributed Lags," *International Statistical Bulletin*, 1937, pp.323 – 328。

[5] F.P.R.Brechling, "The Relationship between Output and Employment in British Manufacturing Industries," *Review of Economic Studies*, vol. 32, July 1965.

[6] Zvi Griliches, "The Demand for a Durable Input: Farm Tractors in the United States 1921 – 1957," in Arnold C. Harberger, ed., *The Demand for Durable Goods*, University of Chicago Press, Chicago, 1960.

[7] 参看他的论文，"The Demand for Capital Goods by Manufacturers: A Study of Quarterly Time Series," *Econometrica*, vol. 30, no. 3, July 1962, pp.407 – 423。

[8] C.A.Sims, "Money, Income, and Causality," *Amercian Economic Review*, vol. 62, 1972, pp.540 – 552.

[9] Granger 和 Sims 因果关系检验之间的选择问题还不清楚。关于这些模型的进一步的讨论，见于 G.Chamberlain, "The General Equivalence of Granger and Sims Causality," *Econometrica*, vol. 50, 1982, pp.569 – 582。

附录 17A 工具有效性的萨根检验

假定我们用一个工具变量来代替与误差项相关的自变量。那么工具变量又会多有效呢？也就是说我们如何知道所选的工具变量与误差项是独立的呢？萨根提出了一个统计量（即 SARG）来检验工具变量法中所使用的工具的有效性。[1]SARG 中所涉及的步骤如下所示[2]：

1. 将回归方程中所包括的变量分成两组，一组是独立于误差项的变量（称为 $X_1, X_2, \cdots, X_p$），另一组是不独立于误差项的变量（称为 $Z_1, Z_2, \cdots, Z_q$）。

2. 选取 $W_1, W_2, \cdots, W_s$ 为 1 中的 Z 变量的工具，其中 $s \geq q$。

3. 用诸 W 代替诸 Z 并估计初始回归，也就是说，通过 IV 来估计初始

回归并得出残值 $\hat{u}$。

4. 将 $\hat{u}$ 对一个常数进行回归，即对所有的 X 变量和 W 变量（但不包括 Z 变量）进行回归，从回归中得出 R^2。

5. 现在计算 SARG 统计量，定义：

$$\text{SARG} = (n-k)\ R^2 \tag{17A.1}$$

其中 n = 观测值的个数，k = 初始回归方程中系数的个数。萨根指出（17A.1）服从自由度为 r 的 χ^2 分布，其中 $r = s - q$。

6. 虚拟假设是所有的（w）工具变量都是有效的。如果计算出来的 χ^2 检验超过了 χ^2 检验的临界值，则拒绝虚拟假设，这意味着至少有一个工具是与误差项相关的，因而建立在所选取工具的基础上的 IV 估计是无效的。

【附录注释】

[1] Sargan, J.D., "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P.E.Hart, G. Mills, and J.K.Whitaker (eds), *Econometric Analysis for National Economic Planning*, Butterworths, London, 1964.

[2] 下面的讨论依赖于：H.R.Seddighi, K.A.Lawler and A.V.Katos, *Econometrics: A Practical Approach*, Routledge, New York, 2000, pp.155 156。

【注释】

[1] 技术上，β_0 是 Y 对 X_t 的偏导数，β_1 是 Y_1 对 X_{t-1} 的偏导数，β_2 是 Y_2 对 X_{t-2} 的偏导数，如此类推。符号上，$\partial Y_t / \partial X_{t-k} = \beta_k$。

[2] Keith M. Carlson, "The Lag from Money to Prices," *Review*, Federal Reserve Bank of St. Louis, October, 1980, Table 1, p.4.

[3] Zvi Griliches, "Distributed Lags: A Survey," *Econometrica*, vol.36, no.1, January 1967, p.16－49.

[4] 本节大量取材于 Marc Nerlove, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*, Agricultural Handbook No.141, U.S.Department of Agriculture, June 1958。

[5] 如果模型中有多于一个解释变量，那么每一个变量都可以对 Y 有一滞后影响，我们假定一个解释变量只是为了简单而已。

[6] 然而，实际上可以预料遥远的 X 值的系数对 Y 的影响是微弱的。

[7] F. F. Alt, "Distributed Lags", *Econometrica*, vol. 10, 1942, pp.113－128.

[8] J. Tinbergen, "Long-Term Foreign Trade Elasticities," *Metroeconomica*, vol. 1, 1949, pp.174－185.

[9] 如果滞后长度 k 的设定不正确，则还要对付第 13 章中所讨论的设定偏误问题。另外，要警惕数据开采的指控。

[10] L. M. Koyck, *Distributed Lags and Investment Analysis*, North Holland Publishing Company, Amsterdam, 1954.

[11] 由于[12]所给的理由，有时又把它写成：$\beta_k = \beta_0(1-\lambda)\lambda^k \quad k = 0, 1, \cdots$。

[12] 这是因为：$\sum \beta_k = \beta_0(1+\lambda+\lambda^2+\lambda^3+\cdots) = \beta_0\left(\frac{1}{1-\lambda}\right)$。这又因为 $0<\lambda<1$ 时右边括号内的表达式是一个无穷几何级数，它的和为 $1/(1-\lambda)$。顺便指出，如果 β_k 是像 [11] 中那样定义的，则 $\beta_k = \beta_0(1-\lambda)/(1-\lambda) = \beta_0$，从而保证全部权重 $(1-\lambda)\lambda^k$ 之和为 1。

[13] 有时又表达为：$X_t^* - X_{t-1}^* = \gamma\ (X_{t-1} - X_{t-1}^*)$。

[14] P. Cagan, "The Monetary Dynamics of Hyperinflations," in M. Friedman (ed.), *Studies in The Quantity Theory of Money*, University of Chicago Press, Chicago, 1956.

[15] Milton Friedman, *A Theory of the Consumption Function*, National Bureau of Economic Research, Princeton University Press, Princeton, N.J., 1957.

[16] G.K.Shaw, *Rational Expectation: An Elementary Exposition*, St.Martin's Press, New York, 1984.p.25.

[17] 同上，第 19－20 页。

[18] 同上，第 27 页。

[19] 可以证明，像考伊克模型那样，在 AE 假设下，一个变量的期望值是该变量过去值的一个指数加权平均。

[20] G. K. Shaw，前引文献，第 47 页。关于 RE 假设的更多的细节，参看 Steven M. Sheffrin, *Rational Expectations*, Cambridge University Press, New York, 1983。

[21] Stephen K. McNees, "The Phillips Curve: Forward-or Backward-Looking?" *New England Economic Review*, July-August 1979, p.50.

[22] 关于 RE 假设的一个新近的严格评价，参看 Michael C. Lovell, "Test of the Rational Expectations Hypothesis," *American Economic Review*, March 1966, pp.110－124。

[23] Stephen K. McNees，前引文献，第 50 页。

[24] Marc Nerlove, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*，前引文献。

[25] 一些作者不把干扰项 u_t 加在关系式（17.6.1）上，而把它加在此式上。他们认为如果前者真的是一个均衡关系式，就没有误差项的地位，而调整机制是不完善的，所以需要有误差项。顺便指出，（17.6.2）有时又写成 $Y_t - Y_{t-1} = \delta(Y_{t-1}^* - Y_{t-1})$。

[26] 此图取材于 Rudiger Dornbusch and Stanley Fischer, *Macroeconomics*, 3d ed., McGraw-Hill, New York, 1984，第 216 页中的图 7.4。

[27] Milton Friedman, *A Theory of Consumption Function*, Princeton University Press, Princeton, N.J., 1957.

[28] $E(v_t v_{t-1}) = E(u_t - \lambda u_{t-1})(u_{t-1} - \lambda u_{t-2}) = -\lambda E(u_{t-1})^2 = -\lambda\sigma^2$，因为由假定，$u_t$ 的协方差为零。

[29] 其证明超出了本书的范围，可参阅 Griliches，前引文献，第 36～38 页。但是，第 18 章在另一讨论中给出了证明的一个提要。

[30] 证明见 J. Johnson, *Econometric Methods*, 3d ed., McGraw-Hill, NewYork, 1984, pp. 360－362；还见 H. E. Doran and J. W. B. Guise, *Single Equation Methods in Econometrics: Applied Regression Analysis University of New Eng-*

land Teaching Monograph Series 3, Armidale, NSW, Australia, 1984, pp.236–244。

[31] 而且，如 J. Johnston 所指出（前引文献，第 350 页），"调整的模式（如部分调整模型所建议的）……有时未必合理。"

[32] N. Liviatan, "Consistent Estimation of Distributed Lags," *International Economic Review*, vol. 4, January 1963, pp.44–52.

[33] 在联立方程模型中常用到这类工具变量（见第 20 章）。

[34] 怎样能改进这些估计量的效率，可参阅 Lawrence. R. Klien, *A Textbook of Econometrics*, 2d ed., Prentice-Hall, Englewood Cliffs, N. J., 1974, p.99。还参看 William H. Greene, *Econometric Analysis*, Macmillan, 2d ed., New York, 1993, pp.535–538。

[35] 关于 ML 法的一个简明的讨论，见 J. Johnston, 前引文献，第 366–371 页以及本书附录 4A 和附录 15A。

[36] J.Durbin, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Econometrica*, Vol. 38, 1970, pp.410–421.

[37] B. Inder, "An Approximation to the Null Distribution of the Durbin-Watson Statistic in Models Containing Lagged Dependent Variables," *Econometric Theory*, vol. 2, no. 3, 1986, pp.413–428.

[38] J. F. Kiviet, "On the Vigour of Some Misspecification Tests for Modelling Dynamic Relationships," *Review of Economic Studies*, vol. 53, no. 173, 1986, pp.241–262.

[39] Gabor Korosi, Laszlo Matyas, and Istvan P. Szekely, *Practical Econometrics*, Ashgate Publishing Company, Brookfield, Vermont, 1992, p.92.

[40] 一个类似的模型，见于 Gregory C. Chow, "On the Long-Run and Short-Run Demand for Money," *Journal of Political Economy*, vol. 74, no. 2, 1966, pp.111–131。注意，乘积性函数的一个优点是，变量的指数给出弹性的直接估计（参看第 6 章）。

[41] 顺便指出，此模型本质上是对参数非线性的。因此，虽然 OLS 给出比方说 $\beta_1\delta$（合起来看）的一个无偏估计，却不见得给出 β_1 和 δ（个别地看）的无偏估计，特别是小样本的情形。

[42] 这些数据来自于 Bhaskar Rao, ed., *Cointegration for the Applied Economist*, St.Martin's Press, New York, 1994, pp.210–213。原始数据是从 1956 年第 1 季度到 1988 年第 4 季度，但出于说明的目的，我们从 1979 年第 1 季度开始分析。

[43] 注意估计的标准误的这个特色：比如说，$\ln R_t$ 的系数的标准误指的是 $\beta_1\delta$ 的一个估计量$\overline{\beta_1\delta}$的标准误。要从$\overline{\beta_1\delta}$的标准误得到$\hat{\beta}_1$ 和 $\hat{\delta}$ 的个别的标准误，并无简单的方法，尤其是当样本较小时。但是，对于大样本，$\hat{\beta}_1$ 和 $\hat{\delta}$ 的个别的标准误可以近似地得到，但计算是复杂的，可参看 Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, p.444。

[44] 注意，由于注释 [43] 讨论过的原因，我们未给出所估系数的标准误。

[45] "The Fed and the Real Rate of Interest," *Review*, Federal Reserve Bank of St. Louis, December 1982, pp.8–18.

[46] 同上，第 15 页。

[47] 数据来自于 data disk in Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, New York, 1998。原始数据来自于 World Bank's World Tables。

[48] Shirley Almon, "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, vol.33, January 1965, pp.178 - 196.

[49] 粗略地说，该定理认为，在一有限闭区间里，任意连续函数都可由一适当高次的多项式来一致地逼近。

[50] Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, pp.675 - 676.

[51] 一个具体的应用，见于 D. B. Batten and Daniel Thornton, "Polynomial Distributed Lags and the Estimation of the St. Louis Equation," *Review*, Federal Reserve Bank of St. Louis, April 1983, pp.13 - 25。

[52] 在另外一种有时会使用的因果关系检验，即所谓的西蒙斯因果关系检验。我们用习题的方式来讨论。

[53] Gary Koop, *Analysis of Economic Data*, John Wiley & Sons, New York, 2000, p.175.

[54] C. W. J. Granger, "Investigating Causal Relations by Econometric Models and CrossSpectral Methods," *Econometrica*, July 1969, pp.424 - 438. 虽然通称 Granger 因果检验，但因维纳（Wiener）更早提出此法，故宜称 **Wiener-Granger 因果检验**。参看 N. Wiener, "The Theory of Prediction," 载 E. F. Beckenback, *Modern Mathematics for Engineers*, McGraw-Hill, New York, 1956, pp.165 - 190。

[55] 关于这个专题的一个极好的讨论，参看 Arnold Zellner, "Causality and Econometrics," *Carnegie-Rochester Conference Series*, 10, K. Brunner and A. H. Meltzer, eds., North Holland Publishing Company, Amsterdam, 1979, pp.9 - 50。注意，计量经济学家 Edward Leamer 宁愿用先后关系（precedence）一词以代替因果关系（Causality）。

[56] Francis X. Diebold, *Elements of Forecasting*, South Western Publishing, 2ded., 2001, p.254.

[57] 进一步的细节，见于 Wojciech W. Charemza and Derek F. Deadman, *New Directions in Economitric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, 3d ed., Edward Elgar Publisher, 1997, Chap.6。

[58] R. W. Hafer, "The Role of Fiscal Policy in the St. Louis Equation," *Review*, Federal Reserve Bank of St. Louis, January, 1982, pp.17 - 22. 关于程序细节参阅注释 [12]。

[59] 这些结果来自于 *The East Asian Miracle: Economic Growth and Public Policy*, pubished for the World Bank by Oxford University Press, 1993, p.244。

[60] 当然，如果解释变量包括一个或一个以上内生变量的滞后期，则可能没有满足该需要。

[61] 诺贝尔奖获得者罗伯特·卢卡斯提出：当政策变化时，从模型中估计的参数对于预测来说没有什么价值，从而经济变量间存在的关系可能会发生变化。关于这一点，参阅 Oliver Blanchard, *Macroeconomics*, Prentice Hall, 1997, p.371 - 372。

[62] Keith Cuthbertson, Stephen G. Hall, and Mark P. Taylor, *Applied Econo-*

metric Techniques, University of Michigan Press, 1992, p.100.

[63] 关于一个相当简单的讨论，见于 G.S.Maddala, *Introduction to Econometrics*, 2d ed., Macmillan, New York, 1992, pp.394－395, and also David F.Hendry, *Dynamic Ecomometrics*, Oxford University Press, New York, 1995, Chap.5。

[64] 关于这些模型的应用，参看 Arnold C. Harberger, *The Demand for Durable Goods*, University of Chicago Press, Chicago, 1960。

第4篇 联立方程模型

715 随便翻阅一下已经出版的关企业和经济学方面的经验性作品，都将发现许多的经济关系式是属于单一方程类型的。这正说明为什么我们把本书的前三篇全用来讨论单一方程类型。在这类模型中，一个变量（因变量 Y）被表达成一或多个其他变量（解释变量 X）的一个线性函数。这类模型有一个隐含的假定，就是 Y 和诸 X 之间如果有因果关系的话，则这种关系是单向的：解释变量是原因，而因变量是结果。

然而，有许多情形经济变量之间的影响是双向的；即一个经济变量影响另一（或多）个经济变量，而反过来又受另一个（或多）个变量的影响。例如，在货币 M 对利率 r 的回归中，单一方程方法论隐含地假定利率是（比方说由联邦储备系统）制定的，并试图求出货币需求对于利率水平变化的反应。但如果利率依赖于货币需求，又会出现什么情况呢？因为这时 M 依赖于 r 而 r 又依赖于 M，所以本书至今为止讲的条件回归分析（conditional regression analysis）就未必合适。这样一来，我们需要两个方程，一个把 M 联系到 r 和另一个把 r 联系到 M，从而导致对联立方程模型的考虑。在其中有多于一个以上的回归方程，每个互相依赖的变量占有一个。

在第 4 篇中，我们对**联立方程模型**（simultaneous-equation models）这个复杂的问题做一个非常粗浅的而且常常是直觉的介绍。详细的论述请看参考文献。

716 在第 18 章中，我们讲几个联立方模型的例子，并说明为什么前面考虑的普通最小二乘法，一般地说不适用于估计模型中的各个方程中的参数。

在第 19 章中，我们考虑所谓**认识问题**（identification problem）。如果在一个含有两或多个方程的联立方程组中，由于一些方程是观测上无区别的（observationally indistinguishable），或者是看上去过于相似，而无法有区别地获得每一方程中的每一参数的估值，我们就有了识别的问题。例如，在数量 Q 对价格 P 的回归中，由于 Q 和 P 均进入需求和供给两函数中，那么我们得到的回归方程是需求函数呢，还是供给函数呢？因此，如果我们除了 Q 和 P 数据外而别无其他信息，要说它是需求或供给函数，即使不是不可能的，也将是非常困难的。所以，在我们进行估计之前，必须先解决识别问题，否则我们不知道我们估计的是什么，估计本身就是无意的。

在第 20 章中，我们考虑专门为估计联立方程模型而设计的一些估计方法及其优异性和局限性。

第 18 章　联立方程模型

717 在本章和下两章中，我们讨论联立方程模型。具体地说，我们讨论它们的特点，对它们的估计，以及与它们有关的某些统计问题。

§ 18.1　联立方程模型的性质

在本教材的第 1～3 篇里，我们仅考虑单一方程模型，也就是有单一因变量 Y 和一或多个解释变量 X 的模型。在这些模型中，重点放在估计和（或）预测以 X 变量的固定值为条件的 Y 均值。因此在这样的模型中因果关系是从诸 X 到 Y。

但有许多情形这种单向因果关系是没有意义的。如果 Y 决定于诸 X 而某些 X 又反过来决定于 Y，就会出现这种情形。简言之，Y 和（某些）X 之间有一双向或联立关系，致使因变量和解释变量之间的划分成为可疑的情况。较好的方法是把一组变量合在一起，它们是能由另一组变量联合地决定的——这正是联立方程模型所要做的。在这类模型中有多于一个方程，每个
718 相互或共同依赖的变量，或称**内生变量**（endogenous variables），占有一个方程。[1]不像单一方程模型那样，在联立方程模型中，我们在估计一个方程的

参数时不可以不考虑方程组中其他方程所提供的信息。

如果，比方说，用OLS法去估计每一方程的参数而不考虑方程组中的其他方程，那么会出现什么情况呢？记得OLS法的关键假定之一是，解释变量 X 是非随机的，或者，虽是随机的，却独立于随机误差项而分布。如果两种情形都不满足，则如后面所示，最小二乘估计量不但是有偏误的，而且是非一致性的；就是说，即使样本无限地增大，估计量仍不收敛于它们的真（总体）值。例如，在下列假设的方程组中[2]：

$$Y_{1i} = \beta_{10} + \beta_{12} Y_{2i} + \gamma_{11} X_{1i} + u_{1i} \tag{18.1.1}$$

$$Y_{2i} = \beta_{20} + \beta_{21} Y_{1i} + \gamma_{21} X_{1i} + u_{2i} \tag{18.1.2}$$

其中 Y_1 和 Y_2 是相依或内生变量，而 X_1 是外生变量。u_1 和 u_2 是随机干扰项。变量 Y_1 和 Y_2 都是随机的。因此，除非能表明（18.1.1）中的随机解释变量 Y_2 的分布独立于 u_1，并且（18.1.2）中的随机解释变量 Y_1 的分布独立于 u_2，对这些方程应用经典的OLS将导致非一致性估计。

在本章的其余部分，我给出联立方程模型的几个例子，并说明对这类模型直接应用最小二乘法所涉及的偏误。在第19章讨论所谓的识别问题之后，我们在第20章讨论一些用以处理联立方程模型的特殊方法。

§18.2 联立方程模型举例

例18.1 需求与供给模型

众所周知，一个商品的价格 P 和它的出售量 Q 是由该商品需求和供给曲线的交点来决定的。比如，为简单起见，假定供求曲线是线性的，那么加上随机干扰项 u_1 和 u_2，我们就可写出经验需求与供给函数为：

需求函数：$Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t} \qquad \alpha_1 < 0 \qquad (18.2.1)$

供给函数：$Q_t^s = \beta_0 + \beta_1 P_t + u_{2t} \qquad \beta_1 > 0 \qquad (18.2.2)$

均衡条件：$Q_t^d = Q_t^s$

719 其中 Q^d = 需求量

Q^s = 供给量

t = 时间

而诸 α 和诸 β 是参数。先验地，预期 α_1 为负（向下倾斜的需求曲线），而 β_1 为正（向上倾斜的供给曲线）。

现在不难看出 P 和 Q 是联合因变量。例如，由于影响 Q_t^d 的其他变量（诸如收入，财富和嗜好）的改变，（18.2.1）中的 u_{1t} 将改变。如果 u_{1t} 是正的，需求曲线将向上方迁移；如果 u_{1t} 是负的，需求曲线将向下方迁移。图18.1表明了这些迁移。

如图所示，需求曲线的迁移同时改变了 P 和 Q。同理，u_{2t} 的改变（由

于罢工、气候、进出口限制，等等）将使供给曲线迁移，又会影响 P 和 Q 两者。由于 Q 和 P 之间的这种同时相依性，（18.2.1）中的 u_{1t}和 P_t 以及（18.2.2）中的 u_{2t}和 P_t 不可能是独立的。因此，像（18.2.1）那样的 Q 对 P 的回归将破坏经典线性回归模型的一个重要假定，即解释变量与干扰项之间不相关的假定。

720

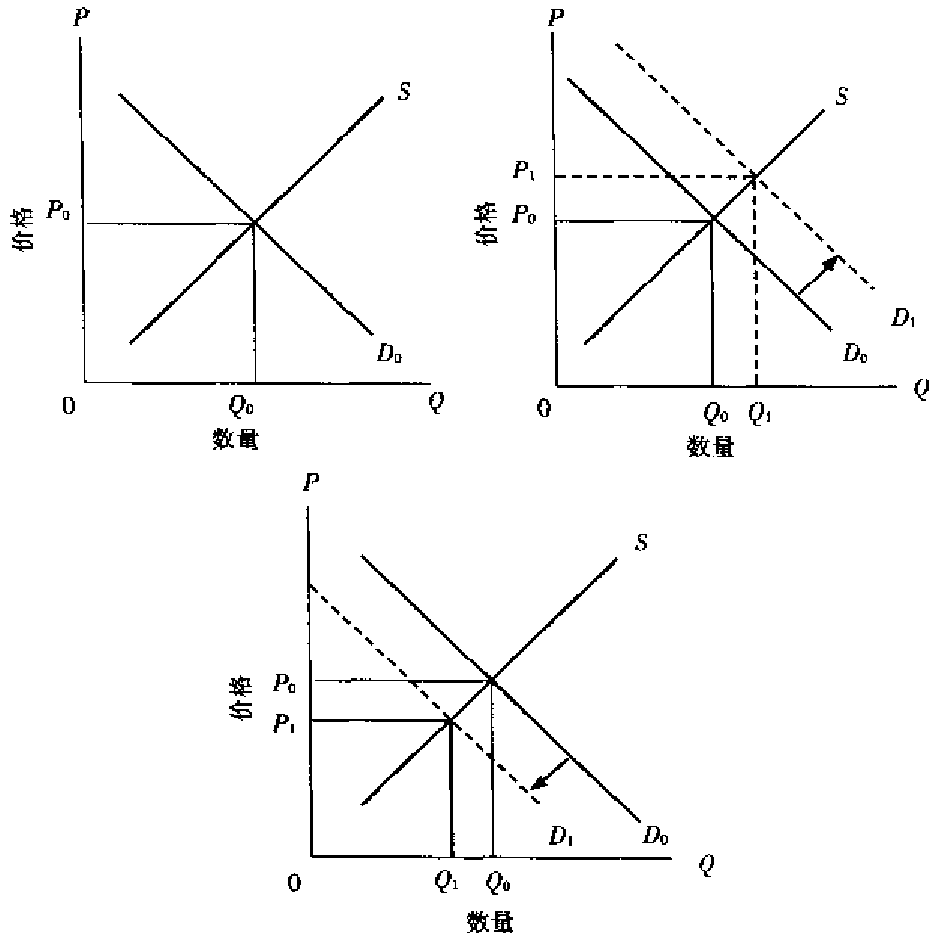


图 18.1　价格与数量的相依性

例18.2　凯恩斯收入决定模型

考虑简单的凯恩斯收入决定模型：

$$\text{消费函数：}\ C_t=\beta_0+\beta_1 Y_t+u_t \qquad 0<\beta_1<1 \tag{18.2.3}$$

$$\text{收入恒等式：}\ Y_t=C_t+I_t\ (=S_t) \tag{18.2.4}$$

其中 C = 消费支出

Y = 收入

I = 投资（假定为外生）

S = 储蓄

t = 时间

u = 随机干扰项

β_0 和 β_1 = 参数

参数 β_1 称边际消费倾向（增加 1 美元收入带来的消费支出的增加）。由经济理论预料 β_1 位于 0 与 1 之间。方程（18.2.3）是（随机）消费函数；而（18.2.4）是个收入恒等式，意味着总收入等于总消费加总投资，这又意味着总投资等于总储蓄。用图形表示，我们得到图 18.2。

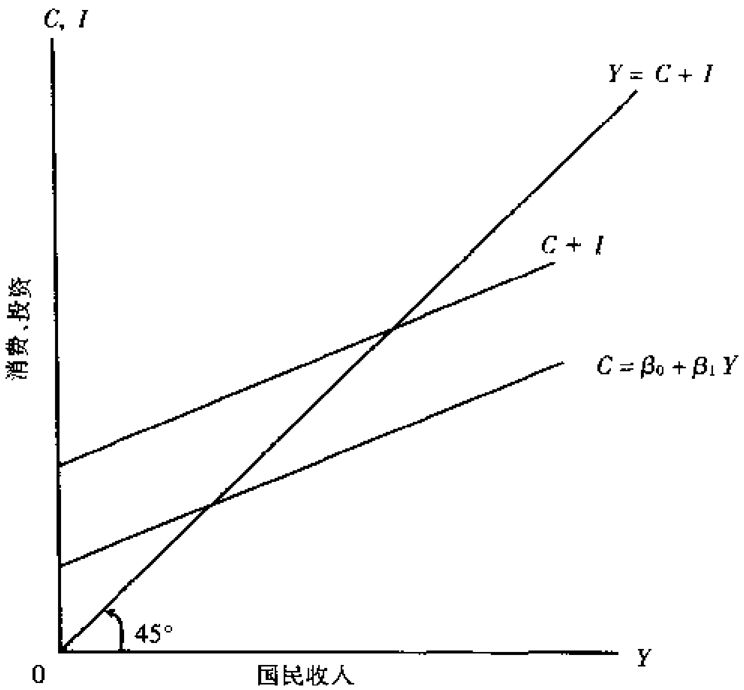


图 18.2 收入决定的凯恩斯模型

从这个公设的消费函数和图 18.2 明显看出，C 和 Y 是互相依赖的，并且不能指望（18.2.3）中的 Y_t 会独立于干扰项。因为，当 u_t 移动时（由于误差项包含着种种因素），消费函数也随之移动，而消费的变动又反过来影响 Y_t。由此，再次表明经典最小二乘法对（18.2.3）不适用。如果要用，所得的估计量将如后所示是非一致性的。

721 **例18.3 工资—价格模型**

考虑如下货币工资与价格决定的菲利普斯类模型：

$$\dot{W}_t = \alpha_0 + \alpha_1 \mathrm{UN}_t + \alpha_2 \dot{P}_t + u_{1t} \qquad (18.2.5)$$

$$\dot{P}_t = \beta_0 + \beta_1 \dot{W}_t + \beta_2 \dot{R}_t + \beta_3 \dot{M}_t + u_{2t} \qquad (18.2.6)$$

其中 $\dot{W}$ = 货币工资变化率

UN = 失业率，%

$\dot{P}$ = 价格变化率

$\dot{R}$ = 资本成本变化率

$\dot{M}$ = 进口原材料变化率

t = 时间

u_1，u_2 = 随机干扰项

由于价格变量 $\dot{P}$ 进入工资方程，并且工资变量 $\dot{W}$ 进入价格方程，故两变量是相互依赖的。因此预料随机解释变量与有关的随机干扰项是相关的，再次使经典 OLS 法不适宜于用来对两方程一个个地进行参数估计。

例18.4 宏观经济学中的 IS 模型

宏观经济学中著名的 IS 或商品货物市场均衡模型[3]的非随机形式可表达为：

消费函数：	$C_t = \beta_0 + \beta_1 Y_{dt} \quad 0 < \beta_1 < 1$	(18.2.7)
税收函数：	$T_t = \alpha_0 + \alpha_1 Y_t \quad 0 < \alpha_1 < 1$	(18.2.8)
投资函数：	$I_t = \gamma_0 + \gamma_1 r_t$	(18.2.9)
定义：	$Y_{dt} = Y_t - T_t$	(18.2.10)
政府支出：	$G_t = \bar{G}$	(18.2.11)
国民收入恒等式：	$Y_t = C_t + I_t + G_t$	(18.2.12)

其中 Y = 国民收入

C = 消费开支

I = 计划的或希望的净投资

$\bar{G}$ = 给定的政府支出水平

T = 税收

Y_d = 可支配收入

r = 利率

722 如果你将 (18.2.10) 和 (18.2.8) 代入 (18.2.7)，并将所得的 C 方程以及方程 (18.2.9) 和 (18.2.11) 代入 (18.2.12)，你应得到 IS 方程：

$$Y_t = \pi_0 + \pi_1 r_t \tag{18.2.13}$$

其中：

$$\pi_0 = \frac{\beta_0 - \alpha_0\beta_1 + \gamma_0 + \bar{G}}{1 - \beta_1(1 - \alpha_1)} \tag{18.2.14}$$

$$\pi_1 = \frac{1}{1 - \beta_1(1 - \alpha_1)}$$

方程 (18.2.13) 就是 IS 或商品市场均衡方程。就是说，它给出了商品市场出清或均衡的那些利率与收入水平的组合。图 18.3 展示了 IS 曲线的几何意义。

如果，比如说，我们孤立地估计消费函数，会出现什么情况呢？我们能得到 β_0 和 β_1 的无偏和（或）一致性估计吗？这种结果是不大可能的。因为，消费依赖于可支配收入，可支配收入依赖于国民收入，而后者又依赖于 r 和 $\bar{G}$ 以及进入 π_0 的其他参数。因此，除非我们把所有这些影响都考虑进来，一个简单的 C 对 Y_d 的回归注定要给出 β_0 和 β_1 的有偏误且非一致性的估计。

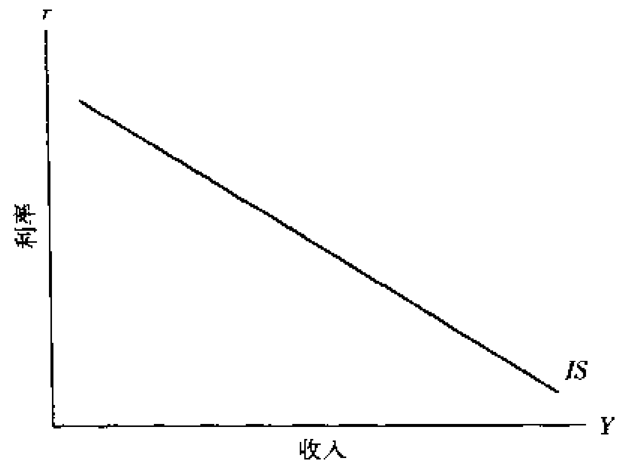


图 18.3　IS 曲线

例 18.5　LM 模型

著名的 IS－LM 范式的另一半是 LM 或货币市场均衡关系式，它给出货币市场出清即货币供需相等的那些利率与收入水平的组合，代数上，模型的非随机形式可表达为：

货币需求函数：　$M_t^d = a + bY_t - cr_t$　(18.2.15)

货币供给函数：　$M_t^s = \bar{M}$　(18.2.16)

均衡条件：　$M_t^d = M_t^s$　(18.2.17)

723 其中 Y＝收入，r＝利率及 $\bar{M}$＝给定的货币量，比方说，由联邦银行来决定的货币量。

令货币需求函数等于供给函数并加以简化，我们得到：

LM 方程：　$Y_t = \lambda_0 + \lambda_1 \bar{M} + \lambda_2 r_t$　(18.2.18)

其中

$$\lambda_0 = -a/b$$

$$\lambda_1 = 1/b$$

$$\lambda_2 = c/b \qquad (18.2.19)$$

给定 $M = \bar{M}$，LM 曲线所代表的关系式如图 18.4 所示。

IS 和 LM 曲线分别展示与商品市场均衡和货币市场均衡相适应的整个利率系列。当然，只有一个利率和一个收入水平能同时适应于两个均衡。为了得到这个利率和这个收入水平，只须令（18.2.13）等于（18.2.18）。习题 18.4 要求你表述同时与货物和货币市场均衡相适应的利率和收入水平。

例 18.6　计量经济模型

一些计量经济学家在他们构造的计量经济模型中曾广泛地使用联立方程模型。这一领域的一位早期的先驱者是宾夕法尼亚州立大学沃顿商学院（Wharton School of the University of Pennsylvania）的 L. 克莱因教授。他的开创性模型，名为**克莱因模型 I**，有如下述：

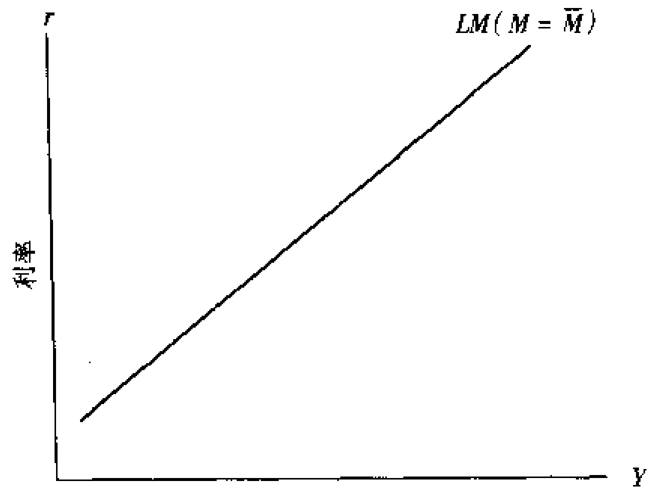


图 18.4　LM 曲线

消费函数：　$C_t = \beta_0 + \beta_1 P_t + \beta_2 (W + W')_t + \beta_3 P_{t-1} + u_{1t}$

724 投资函数：　$I_t = \beta_4 + \beta_5 P_t + \beta_6 P_{t-1} + \beta_7 K_{t-1} + u_{2t}$

劳动需求：　$W_t = \beta_8 + \beta_9 (Y + T - W')_t + \beta_{10} (Y + T - W')_{t-1} + \beta_{11} t + u_{3t}$

恒等式：　$Y_t + T_t = C_t + I_t + G_t$　　(18.2.20)

恒等式：　$Y_t = W'_t + W_t + P_t$

恒等式：　$K_t = K_{t-1} + I_t$

其中　C = 消费支出

I = 投资支出

G = 政府支出

P = 利润

W = 民间工资账单

W' = 政府工资账单

K = 资本存量

T = 税收

Y = 税后收入

t = 时间

u_1, u_2, u_3 = 随机干扰项[4]

在上述模型中，变量 C、I、W、Y、P 和 K 被看作联合因变量或内生变量，而变量 P_{t-1}，K_{t-1}和 Y_{t-1}被看作前定的（predetermined）。[5]总共有 6 个方程（包括 3 个恒等式）用以研究 6 个内生变量的相互依赖性。

在第 20 章中，我们将会看到怎样去估计这类计量经济模型。目前，只须注意由于内生变量之间的相互依赖性。一般地说，它们是不独立于随机干

扰项的，因此，不适宜于用 OLS 法逐个地估计方程组中的方程。如第 18.3 节将要表明的，这样得来的估计量将是非一致性的；即使样本很大，它们也不收敛到它们的真总体值。

§18.3 联立方程偏误：OLS 估计量的非一致性

如前所说，最小二乘法不适宜于用来估计包含在一个联立方程组中的单一方程，因为，如果在该方程中有一或多个解释变量与干扰项相关，这样得
725 到的估计量就是非一致性的。为说明这点，让我们回到 18.2 所给的简单的凯恩斯收人决定模型。假使我们要估计消费函数（18.2.3）的参数。假定 $E(u_t)=0, E(u_t^2)=\sigma^2, E(u_tu_{t+j})=0$（当 $j\neq0$），和 $\operatorname{cov}(I_t,u_t)=0$，也就是经典线性回归模型中的那些假定，我们首先表明（18.2.3）中的 Y_t 和 u_t 相关，然后证明 $\hat{\beta}_1$ 是 β_1 的一个非一致性估计量。

为了证明 Y_t 和 u_t 相关，我们如下进行。将（18.2.3）代入（18.2.4）得：

$$Y_t=\beta_0+\beta_1Y_t+u_t+I_t$$

即，

$$Y_t=\frac{\beta_0}{1-\beta_1}+\frac{1}{1-\beta_1}I_t+\frac{1}{1-\beta_1}u_t \tag{18.3.1}$$

现在：$E(Y_t)=\dfrac{\beta_0}{1-\beta_1}+\dfrac{1}{1-\beta_1}I_t$ （18.3.2）

其中用到了 $E(u_t)=0$ 这一假定，而 I_t 既然是外生的或前定的（因已事先给定），其期望值就是它本身。

因此，从（18.3.1）减去（18.3.2），结果是：

$$Y_t-E(Y_t)=\frac{u_t}{1-\beta_1} \tag{18.3.3}$$

另外，

$$u_t-E(u_t)=u_t \qquad \text{（为什么？）} \tag{18.3.4}$$

由此得：

$$\begin{aligned}\operatorname{cov}(Y_t,u_t)&=E[Y_t-E(Y_t)][u_t-E(u_t)]\\&=\frac{E(u_t^2)}{1-\beta_1} \qquad \text{利用(18.3.3)和(18.3.4)}\\&=\frac{\sigma^2}{1-\beta_1}\end{aligned} \tag{18.3.5}$$

因为按假定 σ^2 是正的（为什么？），故（18.3.5）中 Y 和 u 的协方差必不为零。[6] 从而可以预料（18.2.3）中的 Y_t 和 u_t 是相关的，这就违反了经典线性回归模型的假定：干扰项与解释变量相独立或至少不相关。如前面看到的，在

这种情形中，OLS 估计量是非一致性的。

726 为了表明由于 Y_t 和 u_t 之间的相关，OLS 估计量 $\hat{\beta}_1$ 是 β_1 的一个非一致性估计量，我们进行如下：

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum(C_t-\bar{C})(Y_t-\bar{Y})}{\sum(Y_t-\bar{Y})^2}\\ &= \frac{\sum c_t y_t}{\sum y_t^2}\\ &= \frac{\sum C_t y_t}{\sum y_t^2}\end{aligned} \tag{18.3.6}$$

其中，和平常一样，小写字母表示对（样本）均值的离差。将（18.2.3）代入 C_t 得：

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum(\beta_0+\beta_1 Y_t+u_t)y_t}{\sum y_t^2}\\ &= \beta_1+\frac{\sum y_t u_t}{\sum y_t^2}\end{aligned} \tag{18.3.7}$$

其中在最后一步里用到了 $\sum y_t=0$ 和 $(\sum Y_t y_t/\sum y_t^2)=1$（为什么？）这两个关系式。

如果对（18.3.7）两边取期望值，则得：

$$E(\hat{\beta}_1)=\beta_1+E\left[\frac{\sum y_t u_t}{\sum y_t^2}\right] \tag{18.3.8}$$

可惜，由于期望运算子是一个线性运算子［注：$E(A/B)\neq E(A)/E(B)$］，我们未能估算 $E(\sum y_t u_t/\sum y_t^2)$。但直观上应明显看出，除非 $(\sum y_t u_t/\sum y_t^2)$ 这项是零，$\hat{\beta}_1$ 将是 β_1 的一个有偏误的估计量。但在（18.3.5）中，我们并未证明 Y 和 u 之间的协方差是非零的，因而 $\hat{\beta}_1$ 会不会是无偏的呢？答案是不尽然。因为 cov（Y_t，u_t）是一个总体概念，而 $\sum y_t u_t$ 是一个样本测度，虽然随着样本大小无限地加大时后者趋向于前者，两者究竟不完全是一回事。但是，如果样本无限地增大，我们可以求助于一致性估计量的概念，去判明当样本大小 n 趋于无穷大时出现的情况。总之，当我们不能明确估算如同在（18.3.8）中那样的估计量的期望值时，我们可以把注意力转到它的大样本性态上来。

我们说，如果一个估计量的**概率极限**（probability limit 或简记 plim）[7]等于它的真（总体）值，它就是一致性的。因此，为了表明（18.3.7）中的 $\hat{\beta}_1$ 是非一致性的，我们必须证明它的 plim 不等于真 β_1。对（18.3.7）应用

727 概率极限法则，我们得到[8]：

$$\begin{aligned}\text{plim}(\hat{\beta}_1) &= \text{plim}(\beta_1)+\text{plim}\left(\frac{\sum y_t u_t}{\sum y_t^2}\right)\\ &= \text{plim}(\beta_1)+\text{plim}\left(\frac{\sum y_t u_t/n}{\sum y_t^2/n}\right)\\ &= \beta_1+\frac{\text{plim}(\sum y_t u_t/n)}{\text{plim}(\sum y_t^2/n)}\end{aligned}\tag{18.3.9}$$

其中的第二步我们用了样本中的总观测个数 n 分别地去除 $\sum y_t u_t$ 和 $\sum y_t^2$，以使括号中的量现在变为 Y 和 u 的样本协方差和 Y 的方差。

用语言表达，(18.3.9) 是说，$\hat{\beta}_1$ 的概率极限等于真 β_1 加上 Y 和 u 的样本协方差的概率极限与 Y 的样本方差的概率极限之比。现在，可以预期，随着样本大小 n 无限地增大，Y 和 u 的样本协方差将迫近其真总体协方差 $E[Y_t-E(Y_t)][u_t-E(u_t)]$，而后者由 (18.3.5) 等于 $[\sigma^2/(1-\beta_1)]$。类似地，随着 n 趋于无穷大，Y 的样本方差将迫近其总体方差，且记为 σ_Y^2。因此方程 (18.3.9) 可写为：

$$\begin{aligned}\text{plim}(\hat{\beta}_1) &= \beta_1+\frac{\sigma^2/(1-\beta_1)}{\sigma_Y^2}\\ &= \beta_1+\frac{1}{1-\beta_1}\left(\frac{\sigma^2}{\sigma_Y^2}\right)\end{aligned}\tag{18.3.10}$$

对于 $0<\beta_1<1$ 且 σ^2 和 σ_Y^2 都是正的，显然由 (18.3.10) 知 $\text{plim}(\hat{\beta}_1)$ 总比 β_1 大；就是说，$\hat{\beta}_1$ 将过高地估计了真 β_1。[9] 换句话说，$\hat{\beta}_1$ 是一个有偏误的估计量，并且这个偏误将不会消失，不管样本多么大。

§18.4 联立方程偏误：一个数值例子

为了阐明上节中的某些问题，让我们回到例 18.2 所给的简单凯恩斯收

728 入决定模型，并完成以下的蒙特卡罗研究。[10] 假定投资 I 的取值由表 18.1 的第 (3) 列给出。再假定：

$$E(u_t)=0$$
$$E(u_t u_{t+j})=0 \quad (j\neq 0)$$
$$\text{var}(u_t)=\sigma^2=0.04$$
$$\text{cov}(u_t, I_t)=0$$

根据这些假定产生的 u_t 见于第 (4) 列。

假定消费函数 (18.2.3) 中的真参数值已知为 $\beta_0=2$ 和 $\beta_1=0.8$。

由 β_0 和 β_1 的假定值以及已产生的 u_t 值，我们可从 (18.3.1) 产生收入 Y_t 的数值，见表 18.1 的第 (1) 列。一旦 Y_t 已知，并且知道了 β_0，β_1

和 u_t，就容易从（18.2.3）产生消费 C_t 的数值。把如此产生的 C 值列在表 18.1 的第（2）列。

由于真 β_0 和 β_1 为已知，再由于我们的样本误差正好是“真实”误差（因来自我们设计蒙特卡罗研究的方法），如果我们用表 18.1 的数据做 C_t 对 Y_t 的回归，我们应该得到 $\beta_0=2$ 和 $\beta_1=0.8$，假使 OLS 是无偏的话。但从（18.3.7）我们知道，如果回归元 Y_t 与干扰项 u_t 相关，情形就不会是这样。现在从我们的数据不难验证，Y_t 与 u_t 的（样本）协方差是 $\sum y_t u_t=3.8$，
729 而 $\sum y_t^2=184$。于是，如（18.3.7）所表明的，我们应得到：

$$\begin{aligned}\hat{\beta}_1 &= \beta_1+\frac{\sum_t y_t u_t}{\sum y_t^2}\\ &= 0.8+\frac{3.8}{184}\\ &=0.820\,65\end{aligned}\tag{18.4.1}$$

就是说 $\hat{\beta}_1$ 有 0.020 65 的过高偏误。

现在，让我们用表 18.1 中的数据做 C_t 对 Y_t 的回归（OLS 回归。——译者注）。回归的结果是：

$$\begin{aligned}\hat{C}_t &= \quad 1.494\,0+\quad 0.820\,65\,Y_t\\ \text{se} &= (0.354\,13)\ (0.014\,34)\\ t &= (4.218\,8)\ (57.209)\qquad\qquad R^2=0.994\,5\end{aligned}\tag{18.4.2}$$

表 18.1

Y_t (1)	C_t (2)	I_t (3)	u_t (4)
18.156 97	16.156 97	2.0	−0.368 605 5
19.599 80	17.599 80	2.0	−0.800 408 4E−01
21.934 68	19.734 68	2.2	0.186 935 7
21.551 45	19.351 45	2.2	0.110 290 6
21.884 27	19.484 27	2.4	−0.231 453 5E−01
22.426 48	20.026 48	2.4	0.852 954 4E−01
25.409 40	22.809 40	2.6	0.481 880 7
22.695 23	20.095 23	2.6	−0.609 548 1E−01
24.364 65	21.564 65	2.8	0.729 298 3E−01
24.393 34	21.593 34	2.8	0.786 681 9E−01
24.092 15	21.092 15	3.0	−0.181 570 3

24.874 50	21.874 50	3.0	−0.250 990 0E−01
25.315 80	22.115 80	3.2	−0.136 839 8
26.304 65	23.104 65	3.2	0.609 294 6E−01
25.782 35	22.382 35	3.4	−0.243 529 8
26.080 18	22.680 18	3.4	−0.183 963 8
27.244 40	23.644 40	3.6	−0.151 120 0
28.009 63	24.409 63	3.6	0.192 673 9E−02
30.893 01	27.093 01	3.8	0.378 601 5
28.987 06	25.187 06	3.8	−0.258 885 2E−02

资料来源：与章末注［10］同，第 132 页。

如所预料，所估的 β_1 恰恰是（18.4.1）所预测的 β_1 值。顺便指出，所估的 β_0 也是有偏误的。

一般地说，$\hat{\beta}_1$ 中偏误的多少依赖于 β_1，σ^2 和var(Y)，而且特别依赖于 Y 与 u 之间的协方差程度。[11]如 K. 怀特等人所说的："关于联立方程偏误不外就是这些。和单一方程模型对比，我们不再能假定方程右边的变量与误差项是不相关的了。"[12]要记住，即使在大样本中，这个偏误仍存在。

鉴于在联立方程中应用 OLS 的潜在严重后果，是否有某种联立性检验能告诉我们，在某一给定的事例中有没有联立性的问题？为此，可利用**豪斯曼设定检验**（Hausman specification test）的一个形式，我们将在第 20 章里讨论它。

§18.5 要点与结论

1. 与单一方程模型对比，联立方程模型涉及多于一个因变量或**内生**变量，从而有多少个内生变量就需要有多少个方程。

2. 联立方程模型的一个特有的性质是，一个方程中的内生变量（即回归子）作为一个解释变量而出现在方程组中的另一方程中。

3. 结果是，这样的**内生解释变量**变成了随机的，而且常常和它作为解释变量的所在方程中的误差项有相关关系。

4. 在这种情况下，经典 OLS 法未必适用，因为这样得到的估计量是非一致性的。就是说，不管样本多大，这些估计量都不会收敛于真总体值。

730 5. 课文中介绍的蒙特卡罗例题，表明了当一个回归方程中的回归元与干扰项相关（这正是联立方程模型的典型情况）时用 OLS 去估计其参数所涉及的偏误的性质。

6. 因为联立方程模型是常常用得着的，特别是在计量经济模型中，所以不同作者曾研究出一些其他估计方法，这些方法将在第 20 章中讨论。但在这之前，先在第 19 章中考虑识别问题，这是逻辑上先于估计问题的一个专题。

习 题

问答题

18.1 建造一个美国对牙科医生的供需联立方程模型。明确模型中的内生和外生变量。

18.2 建造一个简单的美国货币需求和供给模型，并将你的模型同 K. 布伦纳（Brunner）和 A.H 梅尔策（Meltzer）[1] 以及 R. 蒂金（Tiegen）[2]所研制的相比较。

18.3 a. 对例 18.1 的供求模型求 $\hat{\alpha}_1$ 的概率极限表达式。

b. 在什么条件下这个概率极限会等于真 α_1?

18.4 对于课文中讨论的 IS－LM 模型，求出同时与商品以及货币两市场均衡相协调的利率与收入水平。

18.5 为了研究通货膨胀与普通股收益的关系，B. 奥迭特（Oudet）[3] 使用以下的模型：

$$R_{bt} = \alpha_1 + \alpha_2 R_{st} + \alpha_3 R_{bt-1} + \alpha_4 L_t + \alpha_5 Y_t + \alpha_6 \mathrm{NIS}_t + \alpha_7 I_t + u_{1t}$$
$$R_{st} = \beta_1 + \beta_2 R_{bt} + \beta_3 R_{bt-1} + \beta_4 L_t + \beta_5 Y_t + \beta_6 \mathrm{NIS}_t + \beta_7 E_t + u_{2t}$$

其中 L = 实际人均基础货币

Y = 实际人均收入

I = 预期通货膨胀率

NIS = 一种新发行股票变量

E = 预期期末股票回报，由滞后股价比率作为代理变量

R_{bt} = 债务收益

R_{st} = 普通股回报

731 a. 试从理论上说出此模型的道理。看看你的理解是否和奥迭特所理解的一致。

b. 模型中哪些是内生变量？哪些是外生变量？

c. 你怎样看待滞后 R_{bt}——内生还是外生？

18.6 J.U. 法利（Farley）和 H.J. 莱维特（Levitt）在他们的论文《有注册商标的个人用品在牙买加的流通模型》（A Model of the Distribution of Branded Personal Products in Jamaica）[4]中建立了如下模型（所考虑的产品是剃须霜、面霜、卫生纸和牙膏）：

$$
\begin{aligned}
Y_{1i} &= \alpha_1 + \beta_1 Y_{2i} + \beta_2 Y_{3i} + \beta_3 Y_{4i} + u_{1i} \\
Y_{2i} &= \alpha_2 + \beta_4 Y_{1i} + \beta_5 Y_{5i} + \gamma_1 X_{1i} + \gamma_2 X_{2i} + u_{2i} \\
Y_{3i} &= \alpha_3 + \beta_6 Y_{2i} + \gamma_3 X_{3i} + u_{3i} \\
Y_{4i} &= \alpha_4 + \beta_7 Y_{2i} + \gamma_4 X_{4i} + u_{4i} \\
Y_{5i} &= \alpha_5 + \beta_8 Y_{2i} + \beta_9 Y_{3i} + \beta_{10} Y_{4i} + u_{5i}
\end{aligned}
$$

其中 Y_1 = 储存此产品的商店所占百分比

Y_2 = 每月销售单位数

Y_3 = 与产品进口商和制造商直接洽谈指数

Y_4 = 地区的批发活动指数

Y_5 = 产品储存的商标深度指数（指经营该类产品的商店存有产品的商标种类平均个数）

X_1 = 产品的目标人口总体

X_2 = 地区所在行政区的人均收入

X_3 = 从人口中心到首都金斯敦的距离

X_4 = 从人口中心到最近一个批发城的距离

a. 你能识别上述模型中的内生和外生变量吗？

b. 能用最小二乘法去估计模型中的一或多个方程吗？为什么或为什么不？

18.7 为了研究广告费与香烟销售量的关系，F. 巴斯（Bass）使用如下模型[5]：

$$
\begin{aligned}
Y_{1t} &= \alpha_1 + \beta_1 Y_{3t} + \beta_2 Y_{4t} + \gamma_1 X_{1t} + \gamma_2 X_{2t} + u_{1t} \\
Y_{2t} &= \alpha_2 + \beta_3 Y_{3t} + \beta_4 Y_{4t} + \gamma_3 X_{1t} + \gamma_4 X_{2t} + u_{2t} \\
Y_{3t} &= \alpha_3 + \beta_5 Y_{1t} + \beta_6 Y_{2t} + u_{3t} \\
Y_{4t} &= \alpha_4 + \beta_7 Y_{1t} + \beta_8 Y_{2t} + u_{4t}
\end{aligned}
$$

732 其中 Y_1 = 滤嘴香烟销售量（支数）的对数除以年龄 20 岁以上的人口

Y_2 = 无滤嘴香烟销售量（支数）的对数除以年龄 20 岁以上的人口

Y_3 = 滤嘴香烟美元广告费的对数除以 20 岁以上的人口再除以广告价格指数

Y_4 = 无滤嘴香烟美元广告费的对数除以 20 岁以上的人口再除以广告价格指数

X_1 = 可支配个人收入的对数除以 20 岁以上的人口再除以消费者价格指数

X_2 = 无滤嘴香烟每包价格的对数除以消费者价格指数

a. 在上述模型中诸 Y 是内生的，而诸 X 是外生的。为什么作者认为 X_2 是外生的呢？

b. 如果把 X_2 看作一个内生变量，你会怎样修改上述模型？

18.8 G. 门杰斯（Menges）对西德经济构造了如下的计量经济模型[6]：

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 I_t + u_{1t} \\
I_t &= \beta_3 + \beta_4 Y_t + \beta_5 Q_t + u_{2t} \\
C_t &= \beta_6 + \beta_7 Y_t + \beta_8 C_{t-1} + \beta_9 P_t + u_{3t} \\
Q_t &= \beta_{10} + \beta_{11} Q_{t-1} + \beta_{12} R_t + u_{4t}
\end{aligned}
$$

其中 Y = 国民收入

I = 净资本形成

C = 个人消费

Q = 利润

P = 生活费用指数

R = 工业生产力

t = 时间

u = 随机干扰项

a. 你认为哪些变量是内生的和哪些是外生的？

b. 方程组中有没有可以用单一方程最小二乘法去估计的方程？

c. 把变量 P 包含在消费函数中依据的是什么理由？

733 18.9 L.E. 加拉韦（Gallaway）和 P.E. 史密斯（Smith）对美国经济构造了一个简单的模型如下[7]：

$$
\begin{aligned}
Y_t &= C_t + I_t + G_t \\
C_t &= \beta_1 + \beta_2 \mathrm{YD}_{t-1} + \beta_3 M_t + u_{1t} \\
I_t &= \beta_4 + \beta_5 (Y_{t-1} - Y_{t-2}) + \beta_6 Z_{t-1} + u_{2t} \\
G_t &= \beta_7 + \beta_8 G_{t-1} + u_{3t}
\end{aligned}
$$

其中 Y = 国民总产值

C = 个人消费支出

I = 私人国内总投资

G = 政府支出加对外净投资

YD = 可支配或税后收入

M = 季初货币供给

Z = 税前财产收入

t = 时间

u_1，u_2，u_3 = 随机干扰项

所有变量均以一阶差分形式度量。

根据 1948—1957 年的季度数据，作者们对每一方程逐个地应用最小二乘法并得到如下结果：

$$
\begin{aligned}
C_t &= 0.09 + 0.43\mathrm{YD}_{t-1} + 0.23 M_t & R^2 &= 0.23 \\
I_t &= 0.08 + 0.43(Y_{t-1} - Y_{t-2}) + 0.48 Z_t & R^2 &= 0.40 \\
G_t &= 0.13 + 0.67 G_{t-1} & R^2 &= 0.42
\end{aligned}
$$

a. 你怎样为在本例中使用单一方程最小二乘法做辩护？

b. 为什么这些 R^2 值都相当低？

解答题

18.10 表 18.2 给定以下 1970—1999 年期间的 Y(国内生产总值)，C(个人消费支出)和 I(国内总私人投资)数据，均以 10 亿美元计。假定像例 18.2 的简单凯恩斯收入决定模型那样，C 和 Y 的关系是线性的。求消费函数中的参数的 OLS 估计值。保留你的计算结果，以便和第 20 章讲的用另一观点处理同样数据做比较。

734

表 18.2 个人消费支出、私人国内总投资、国内生产总值，美国，1970—1999(以 1996 年 10 美元计)

年份	C	I	Y
1970	2 317.5	436.2	3 578.0
1971	2 405.2	485.8	3 697.7
1972	2 550.5	543.0	3 998.4
1973	2 675.9	606.5	4 123.4
1974	2 653.7	561.7	4 099.0
1975	2 710.9	462.2	4 084.4
1976	2 868.9	555.5	4 311.7
1977	2 992.1	639.4	4 511.8
1978	3 124.7	713.0	4 760.6
1979	3 203.2	735.4	4 912.1
1980	3 193.0	655.3	4 900.9
1981	3 236.0	715.6	5 021.0
1982	3 275.5	615.2	4 913.3
1983	3 454.3	673.7	5 132.3
1984	3 640.6	871.5	5 505.2
1985	3 820.9	863.4	5 717.1
1986	3 981.2	857.7	5 912.4
1987	4 113.4	879.3	6 113.3
1988	4 279.5	902.8	6 368.4
1989	4 393.7	936.5	6 591.9
1990	4 474.5	907.3	6 707.9
1991	4 466.6	829.5	6 676.4
1992	4 594.5	899.8	6 880.0

1993	4 748.9	977.9	7 062.6
1994	4 928.1	1 107.0	7 347.7
1995	5 075.6	1 140.6	7 543.8
1996	5 237.5	1 242.7	7 813.2
1997	5 423.9	1 393.3	8 159.5
1998	5 678.7	1 566.8	8 515.7
1999	5 978.8	1 669.7	8 875.8

注：C = 个人消费支出

I = 私人国内总投资

Y = 国内总产值

资料来源：*Economic Report of the President*, 2001, Table B－2, p.276.

18.11 利用习题 18.10 中的数据，求国内总投资 I 对国内生产总值（GDP）的回归。保留计算结果做后面一章中进一步分析之用。

18.12 考虑宏观经济学恒等式：

$$C + I = Y \qquad (= \text{GDP})$$

如前，假定：

$$C_t = \beta_0 + \beta_1 Y_t + u_t$$

并且按照宏观经济学的**加速数模型**（accelerator model）令：

$$I_t = \alpha_0 + \alpha_1 (Y_t - Y_{t-1}) + v_t$$

其中 u 和 v 是误差项。用习题 18.10 中的数据去估计加速数模型，并把计算结果留做进一步研究之用。

【习题注释】

[1] "Some Further Evidence on Supply and Demand Functions for Money", *Journal of Finance*, vol.19, May 1964, pp.240－283.

[2] "Demand and Supply Functions for Money in the United States," *Econometrica*, vol.32, no.4, October 1964, pp.476－509.

[3] Bruno A. Oudet, "The Variation of the Return on Stocks in Periods of Inflation," *Journal of Financial and Quantitative Analysis*, vol. 8, no. 2, March 1973, pp.247－258.

[4] *Journal of Marketing Research*, November 1968, pp.362－368.

[5] "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research*, vol.6, August 1969, pp.291－300.

[6] G. Menges, "Ein Ökonometriches Modell der Bundesrepublik Deutschland (Vier Strukturgleichungen)," I.F.O.Studien, vol.5, 1959, pp.1－22.

[7] "A Quarterly Econometric Model of the United States," *Journal of American Statistical Association*, vol.56, 1961, pp.379－383.

【注释】

[1] 在联立方程模型的用语中，共同依赖的变量叫做内生变量。那些真正非随机或可看作非随机的变量叫做外生（exogenous）或前定（predetermined）变量。（更多的细节见第19章。）

[2] 这些节约的、但不说自明的符号将在第19章中被推广到多于两个方程上面。

[3]"商品市场均衡曲线或IS曲线给出计划开支等于收入的那些利率与产出水平的组合。"参见 Rudiger Dornbusch and Stanley Fischer, *Macroeconomics*, 3d ed., McGraw-Hill, New York, 1984, p.102。注意，为简单起见，我们省去了国外贸易部门。

[4] L. R. Klein, *Economic Fluctuations in the United States, 1921－1941*, John Wiley & Sons, New York, 1950.

[5] 模型构造人必须指明一个模型中的哪些变量是内生的，哪些是前定的。K_{t-1}和Y_{t-1}之所以是前定的，是因为在时刻t时它们的值是已知的。（对此第19章中有更多的讨论。）

[6] 只要MPC β_1落在0与1之间，它一定大于0。但若β_1大于1，它将是负的。当然一个大于1的MPC值不会有多少经济意义，因此在现实中可以预期Y_t和u_t之间的协方差是正的。

[7] 关于概率极限的定义，参看附录A。

[8] 如在附录A中所述，常数（例如β_1）的plim就是该常数，以及(A/B)的$\text{plim} = \text{plim}(A)/\text{plim}(B)$。但注意，$E(A/B) \neq E(A)/E(B)$。

[9] 然而，一般地说，偏误的方向将依赖于模型的具体结构和回归系数的真值。

[10] 本研究取自 Kenneth J. White, Nancy G. Horsman, and Justin B. Wyatt, *SHAZAM: Computer Handbook for Econometrics for Use with Basic Econometrics*, McGraw-Hill, New York, 1985, pp.131－134。

[11] 参看方程（18.3.5）。

[12] Kenneth J. White等，前引文献，第133－134页。

第 19 章　识别问题

735 在本章我们考虑识别问题的性质和意义。识别问题的症结在于：回顾第 18.2 节引进的供求模型。假使我们仅有 Q 和 P 的时间序列数据而没有更多的（诸如消费者收入，前期流行价格及气候）信息，那么识别问题就是要寻求下述问题的答案：仅仅给出 P 和 Q 的数据，我们怎样知道我们是在估计需求函数呢还是供给函数呢？或者问，如果我们想着我们是在拟合一个需求函数，我们又怎样保证我们所估计的确实是需求函数而不是别的什么呢？

稍加思索便知，在我们估计这个需求函数之前，必须先回答上述问题。本章中，我们将说明怎样解决识别的问题。我们先引进一些符号和定义，再用几个例子阐明识别问题。然后，给出一些规则，用以判断联立方程模型中的一个方程是否可以识别。就是说，它确实是我们要估计的那个关系式，不管它是需求或供给函数，或什么别的。

§19.1　符号与定义

为了便于讨论，我们引进如下符号和定义：

736 一般的 M 个内生或联合因变量的 M 个方程模型可写如方程组

(19.1.1)：

$$
\begin{aligned}
Y_{1t} &= \qquad\quad \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \cdots + \beta_{1M}Y_{Mt} \\
&\qquad\qquad + \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + \cdots + \gamma_{1K}X_{Kt} + u_{1t} \\
Y_{2t} &= \beta_{21}Y_{1t} \qquad + \beta_{23}Y_{3t} + \cdots + \beta_{2M}Y_{Mt} \\
&\qquad\qquad + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + \cdots + \gamma_{2K}X_{Kt} + u_{2t} \\
Y_{3t} &= \beta_{31}Y_{1t} + \beta_{32}Y_{2t} \qquad + \cdots + \beta_{3M}Y_{Mt} \\
&\qquad\qquad + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + \cdots + \gamma_{3K}X_{Kt} + u_{3t} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_{MT} &= \beta_{M1}Y_{1t} + \beta_{M2}Y_{2t} + \cdots + \beta_{M,M-1}Y_{M-1,t} \\
&\qquad\qquad + \gamma_{M1}X_{1t} + \gamma_{M2}X_{2t} + \cdots + \gamma_{MK}X_{Kt} + u_{Mt}
\end{aligned}
\tag{19.1.1}
$$

其中　Y_1，Y_2，…，$Y_M = M$ 个内生或联合因变量

X_1，X_2，…，$X_K = K$ 个前定变量（这些 X 变量之一可取值 1，以使每个方程有一截距项）

u_1，u_2，…，$u_M = M$ 个随机干扰项

$t = 1$，2，…，$T =$ 总观测个数

诸 $\beta =$ 内生变量系数

诸 $\gamma =$ 前定变量系数

顺便指出，并不需要每个变量都出现在每一方程之中。事实上，我们将在第 19.2 节中看到，如果有一个方程可以识别的话，就一定不能出现这种情形。

如方程（19.1.1）所表明的，进入联立方程模型的变量可分为两类：**内生的**（endogenous），指其值要从模型内部决定；和**前定的**（predetermined），指其值由模型外部决定。内生变量被视为随机的，而前定变量则被视为非随机的。

前定变量又分为两类：**外生的**（exogenous），包括当前的或滞后的；以及**滞后内生的**（lagged endogenous），例如，X_{1t}是当前（现时）外生变量，而 $X_{1(t-1)}$是滞后一期的滞后外生变量。$Y_{(t-1)}$是滞后一期的内生变量，但因在当前时期里 $Y_{1(t-1)}$值已知，故把它看作非随机的，由此得名前定变量。[1]总之，当前外生、滞后外生和滞后内生变量都被认为是前定的；在当前时期里，它们的值不是由模型决定的。

737 哪些变量是内生的，哪些是前定的，由模型构造者来裁定。虽然一些（非经济）变量如气温、雨量等等明显是外生或前定的，模型构造者在划分经济变量为内生或前定时必须倾注大量的注意力。他或她必须能在先验或理论的基础上为这个分类做出辩护。尽管如此，在本章的后一部分，我们将为外生性提供一种统计检验。

因为出现在（19.1.1）中的方程也许描述一个经济社会的（一个经济模型的）结构，或者描述一个经济人（如消费者或生产者）的行为，所以把这些方程称为**结构**（structural）或**行为**（behavioral）方程。诸 β 和诸 γ 则称**结构参数**（structural parameters）或**系数**。

从结构方程组可以解出 M 个内生变量并导出诱导型（reduced-form）方程和相应的诱导型系数。**所谓诱导型方程，是指单纯由前定变量和随机干扰项来表达一个内生变量的方程。**为了说明，考虑第 18 章讲的凯恩斯收入决定模型：

$$\text{消费函数：} C_t = \beta_0 + \beta_1 Y_t + u_t \qquad 0 < \beta_1 < 1 \tag{18.2.3}$$

$$\text{收入恒等式：} Y_t = C_t + I_t \tag{18.2.4}$$

在此模型中，C（消费）和 Y（收入）是内生变量，而 I（投资支出）被视为外生变量。这两个方程都是结构方程。（18.2.4）则是一个恒等关系。如同平常，假定 MPC β_1 落在 0 与 1 之间。

如果将（18.2.3）代入（18.2.4），经过简单的代数运算，就得到：

$$Y_t = \Pi_0 + \Pi_1 I_t + w_t \tag{19.1.2}$$

其中

$$\begin{aligned} \Pi_0 &= \frac{\beta_0}{1-\beta_1} \\ \Pi_1 &= \frac{1}{1-\beta_1} \\ w_t &= \frac{u_t}{1-\beta_1} \end{aligned} \tag{19.1.3}$$

方程（19.1.2）是一个**诱导型方程**，它把内生变量 Y 表达为仅仅是外生（或前定）变量 I 和随机干扰项 u 的函数。Π_0 和 Π_1 是相应的**诱导型系数**。注意，这些诱导型系数是结构系数的非线性组合。

738 将（19.1.2）中的 Y 值代入（18.2.3）的 C，就得到另一诱导型方程：

$$C_t = \Pi_2 + \Pi_3 I_t + w_t \tag{19.1.4}$$

其中

$$\begin{aligned} \Pi_2 &= \frac{\beta_0}{1-\beta_1} \qquad \Pi_3 = \frac{\beta_1}{1-\beta_1} \\ w_t &= \frac{u_t}{1-\beta_1} \end{aligned} \tag{19.1.5}$$

诱导型系数 Π_1 和 Π_3 度量着外生变量取值的单位变化时内生变量的即期影响[2]，所以也称为**即时或冲击**（impact）**乘数或短期乘数**。比如说，在上述凯恩斯模型中，投资支出增加 1 美元，并假定 MPC 是 0.8，那么，由（19.1.3）我们得到 $\Pi_1 = 5$。这个结果是说，增加 1 美元的投资，将立即（在现期里）导致 5 美元的收入增加，也就是 5 倍的收入增加。同理，在假定的条件下，（19.1.5）中的 $\Pi_3 = 4$ 表示，投资支出增加 1 美元，将立即导致消费支出增加 4 美元。

在计量经济模型的言辞中，像（18.2.4）这样的方程或 $Q_t^d = Q_t^s$（需求量等于供给量），均称均衡条件（equilibrium conditions）。恒等式（18.2.4）是说，总收入 Y 必定等于总支出（即消费支出加投资支出）。均衡一旦实现，诸内生变量便取得它们的均衡值。[3]

注意诱导型方程的一个令人感兴趣的特点，由于前定变量和随机干扰均出现在这些方程的右边，并且由于假定了前定变量与干扰项不相关，因此可用 OLS 法估计诱导型方程的系数（诸 Π_i）。以后将表明，我们也许能从所估计的诱导型系数计算出结构系数（诸 β_i）。这种方法名为**间接最小二乘**（indirect least squares，ILS）法，而所估计的结构系数称 ILS 估计。

739 在第 20 章中，我们将详细研究 ILS 法。现在我们看到，既然诱导型系数可由 OLS 法来估计，并且这些系数又是结构系数的组合，这就存在着从诱导型系数“恢复”结构系数本来面目的可能性，而正是结构系数的估计才是我们的最终兴趣所在。怎样从诱导系数复原到结构性系数呢？答案见于第 19.2 节。这个答案表明了识别问题的症结所在。

§ 19.2 识别问题

所谓**识别问题**，是指能否从所估计的诱导型系数求出一个结构方程的参数的数值估计值。如果能够，就说该方程是可以识别的（identified）。如果不能，就说所考虑的方程是不可识别的（unidentified）或不足识别的（underidentified）。

一个可识别方程或者是被恰好（充分或刚好）识别的，或者是被过分识别的。恰好识别是指能够得到结构参数的一个独一无二的值，过分识别是指可获得结构参数的值不止一个。在随后的讨论中将分别展示以上这两种情况。

由于对于同样一组数据，不同组的结构参数可能同样适用，因此产生了识别问题。对于不同的情况，一个简化方程可能对于不同的结构方程和不同的假设都是适用的，但它可能很难告诉我们在研究的是哪一个具体的假设。在这一部分的剩余环节我们来考虑几个例子以展示识别问题的本质。

不足识别

再次考虑供求模型（18.2.1）和（18.2.2）以及供求相等的市场出清或均衡条件。由均衡条件我们得到：

$$\alpha_0 + \alpha_1 P_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.2.1}$$

解（19.2.1），我们得到均衡价格：

$$P_t = \Pi_0 + v_t \tag{19.2.2}$$

其中

$$\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \tag{19.2.3}$$

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \tag{19.2.4}$$

740 将（19.2.2）的 P_t 代入（18.2.1）或（18.2.2），我们得到下面的均衡数量：

$$Q_t = \Pi_1 + w_t \tag{19.2.5}$$

其中

$$\Pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \tag{19.2.6}$$

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \tag{19.2.7}$$

顺便指出，误差项 v_t 和 w_t 是原始误差项 u_1 和 u_2 的线性组合。

方程（19.2.2）和（19.2.5）为诱导型方程。现在我们的供求模型含有 4 个结构系数 α_0、α_1、β_0 和 β_1，但我们没有估计它们的惟一方法。为什么？答案见于（19.2.3）和（19.2.6）所给的两个诱导型系数。这些诱导型系数含有全部 4 个结构系数，但没有方法能从仅仅两个诱导型系数估计出 4 个结构性未知数来。回忆中学代数便知，要估计 4 个未知数，我们必须有 4 个（独立）方程。并且，一般地说，要估计 k 个未知数，必须有 k 个（独立）方程。顺便看到，如果我们做诱导型回归（19.2.2）和（19.2.5），将仅有常数项而没有任何解释变量，并且这些常数项仅给出 P 和 Q 的均值（为什么?）。

所有这些都意味着，给定 P（价格）和 Q（数量）的时间序列数据而无任何其他信息，研究者将无法保证他所估计的是需求函数还是供给函数。就是说，一双给定的 P 和 Q，由于供求相等的均衡条件，仅代表适当的需求和供给曲线的交点。为了清楚地看到这个问题，且考虑图 19.1 所示意的散点图。

图 19.1a 给出几个联系着 Q 和 P 的散点，每一散点代表一条需求曲线和一条供给曲线的交点，如图 19.1b 所示。现在拿一个点来考虑，比如图 19.1c 中的那一点。我们无法肯定这个点是由图中整个（供求）曲线族中的哪一供求线产生的。为此，显然需要有关于供求曲线的性质的一些其他信息，比方说，如果由于收入、嗜好等等的变化，需求曲线随时间而迁移，而供给曲线则保持相对稳定，如图 19.1d 那样，则散点将描画出一条供给曲线。这对我们来说，供给曲线是可识别的。同理，如果由于气候条件的变化（在考虑农产品时）或其他外部因素的变化，供给曲线随时间而迁移，但需求曲线保持相对稳定，如图 19.1a 那样，则散点将描画出一条需求曲线，对我们来说需求曲线是可识别的。

741 还有另一种也许是更有启发性的看待识别问题的方法。假使我们用 $\lambda(0\leqslant\lambda\leqslant1)$去乘（18.2.1），同时用（$1-\lambda$）去乘（18.2.2），以得到下列方程（注：我们省掉了 Q 的上标）：

$$\lambda Q_t = \lambda\alpha_0 + \lambda\alpha_1 P_t + \lambda u_{1t} \tag{19.2.8}$$

$$(1-\lambda)Q_t = (1-\lambda)\beta_0 + (1-\lambda)\beta_1 P_t + (1-\lambda)u_{2t} \tag{19.2.9}$$

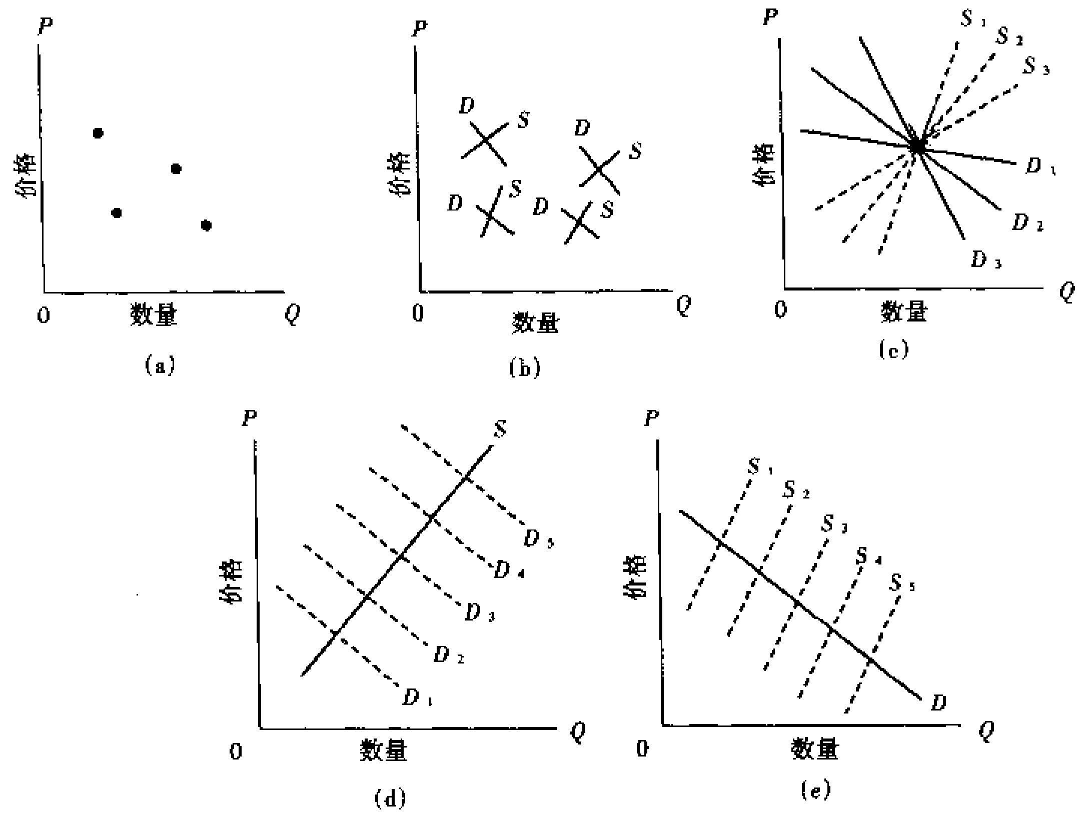


图 19.1 假想供求函数与识别问题

将两方程相加得到原始供求方程的如下**线性组合**：

$$Q_t = \gamma_0 + \gamma_1 P_t + w_t \tag{19.2.10}$$

其中

$$\begin{aligned} \gamma_0 &= \lambda\alpha_0 + (1-\lambda)\beta_0 \\ \gamma_1 &= \lambda\alpha_1 + (1-\lambda)\beta_1 \\ w_t &= \lambda u_{1t} + (1-\lambda)u_{2t} \end{aligned} \tag{19.2.11}$$

741 这个“伪造的”或“混杂的”方程（19.2.10）和（18.2.1）或（18.2.2）都是观测上无区别的（observationally indistinguishable）；它们都是 Q 和 P 的回归。因此，如果我们只有 P 和 Q 的时间序列数据，则（18.2.1），（18.2.2）或（19.2.10）中的任一个都会和同样的数据相协调。换句话说，同样的数据可以适合于“假设”（18.2.1），（18.2.2）或（19.2.10），我们无法告知我们是在检验哪一个假设。

要使一个方程成为可识别的，也就是使它的参数能被估计，必须证明给定的数据集不会产生表面看来像是我们所感兴趣的方程那样的一个结构方程。如果我们计划去估计需求函数，我们必须表明所给数据不适合于供给函数或某些混杂方程。

恰好或恰可识别

上述需求函数或供给函数不能识别的理由是在两个函数中出现同样的变量 P 和 Q，而且再没有其他诸如图 19.1d 或 e 所表示的那种信息。然而，假使我们考虑下述需求与供给模型：

$$\text{需求函数：}\ Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \alpha_1 < 0,\ \alpha_2 > 0 \tag{19.2.12}$$

$$\text{供给函数：}\ Q_t = \beta_0 + \beta_1 P_t + u_{2t} \qquad \beta_1 > 0 \tag{19.2.13}$$

其中 I = 消费者的收入，为一外生变量，而其他变量定义如前。

注意，上述模型和我们原来的供求模型之间的惟一差别，是在需求函数中多了一个变量收入。从需求的经济理论得知，收入常常是对大多数商品和服务需求的一个重要决定因素。因此，把它包含进需求函数中来，将给我们提供关于消费者行为的某些其他信息。对大多数商品来说，可以预料收入对消费有正的影响（$\alpha_2 > 0$）。

利用市场结清机制：需求量 = 供给量，我们有：

$$\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.2.14}$$

由此解出 P_t 的均衡值如下：

$$P_t = \Pi_0 + \Pi_1 I_t + v_t \tag{19.2.15}$$

748 其中的诱导型系数是：

$$\Pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\Pi_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1} \tag{19.2.16}$$

和

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

将 P_t 的均衡值代入上述需求或供给函数，我们得到如下的均衡数量：

$$Q_t = \Pi_2 + \Pi_3 I_t + w_t \tag{19.2.17}$$

其中：

$$\Pi_2 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}$$

$$\Pi_3 = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} \tag{19.2.18}$$

和

$$w_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

因为（19.2.15）和（19.2.17）都是诱导型方程，故可用 OLS 法估计它们的参数。现在供求模型（19.2.12）和（19.2.13）含有 5 个结构系数，即 α_0、α_1、α_2、β_1 和 β_2。但只有 4 个方程，即由（19.2.16）和（19.2.18）给出的 4 个诱导型系数 Π_0、Π_1、Π_2 和 Π_3（方程）可用来估计它们。因此，要得到全部结构系数的惟一解是不可能的。但容易看出，供给函数的参数是可识别的（可被估计的）。这是因为：

$$\beta_0 = \Pi_2 - \beta_1 \Pi_0 \tag{19.2.19}$$
$$\beta_1 = \frac{\Pi_3}{\Pi_1}$$

然而没有估计需求函数的参数的惟一方法；因此需求函数仍不可识别。顺便指出，结构系数 β_1 是诱导型系数的一个非线性函数，如我们将在第 20 章中看到的，这会给 β_1 估计值的标准误的估计带来一些问题。

744 为了验证需求函数（19.2.12）不可识别（不能估计），让我们用 $\lambda(0 \leqslant \lambda \leqslant 1)$ 去乘它，再用 $(1-\lambda)$ 去乘（19.2.13），然后把它们加起来得到以下的“混杂”方程：

$$Q_t = \gamma_0 + \gamma_1 P_t + \gamma_2 I_t + w_t \tag{19.2.20}$$

其中

$$\begin{aligned} \gamma_0 &= \lambda\alpha_0 + (1-\lambda)\beta_0 \\ \gamma_1 &= \lambda\alpha_1 + (1-\lambda)\beta_1 \\ \gamma_2 &= \lambda\alpha_2 \end{aligned} \tag{19.2.21}$$

和

$$w_t = \lambda u_{1t} + (1-\lambda) u_{2t}$$

方程（19.2.20）虽然有别于不含解释变量 I 的供给函数（19.2.13），却与需求函数（19.2.12）在观测上不可区别。因此需求函数仍是不可识别的。

注意一个有趣的事实：正是在需求函数中添加了一个变量，使得我们能识别供给函数！为什么？在需求方程中放进收入变量将对函数的变异提供一些额外信息，如图 19.1d 所示。该图表明，稳定的供给曲线与迁移中的需求曲线的交点是怎样能使我们去跟踪（识别）供给曲线的。如我们即将看到的，一个方程的可识别性常常依赖于它是否排除了包含在模型里其他方程中的一或多个变量。

但若我们考虑如下供求模型：

$$\text{需求函数：} Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1t} \qquad \alpha_1 < 0,\ \alpha_2 > 0 \tag{19.2.12}$$
$$\text{供给函数：} Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \quad \beta_1 > 0,\ \beta_2 > 0 \tag{19.2.22}$$

其中需求函数和前面的一样，但供给函数包含有另一解释变量即滞后一期的价格。该供给函数设想，一个商品的供给量依赖于它的当时和前期价格，这是常用来解释许多农产品商品供给的一个模型。注意，P_{t-1} 在时间 t 是已知的，所以是一个前定变量。

利用市场出清机制，我们有

$$\alpha_0+\alpha_1 P_t+\alpha_2 I_t+u_{1t}=\beta_0+\beta_1 P_t+\beta_2 P_{t-1}+u_{2t} \tag{19.2.23}$$

745 解此方程得到如下的均衡价格：

$$P_t=\Pi_0+\Pi_1 I_t+\Pi_2 P_{t-1}+v_t \tag{19.2.24}$$

其中

$$\begin{aligned}\Pi_0&=\frac{\beta_0-\alpha_0}{\alpha_1-\beta_1}\\ \Pi_1&=-\frac{\alpha_2}{\alpha_1-\beta_1}\\ \Pi_2&=\frac{\beta_2}{\alpha_1-\beta_1}\\ v_t&=\frac{u_{2t}-u_{1t}}{\alpha_1-\beta_1}\end{aligned} \tag{19.2.25}$$

将均衡价格代人需求或供给方程，便得到对应的均衡数量：

$$Q_t=\Pi_3+\Pi_4 I_t+\Pi_5 P_{t-1}+w_t \tag{19.2.26}$$

其中诱导系数：

$$\begin{aligned}\Pi_3&=\frac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1}\\ \Pi_4&=-\frac{\alpha_2\beta_1}{\alpha_1-\beta_1}\\ \Pi_5&=\frac{\alpha_1\beta_2}{\alpha_1-\beta_1}\end{aligned} \tag{19.2.27}$$

和

$$w_t=\frac{\alpha_1 u_{2t}-\beta_1 u_{1t}}{\alpha_1-\beta_1}$$

方程（19.2.12）和（19.2.22）所给的供求模型共含 6 个结构系数——α_0、α_1、α_2、β_0、β_1 和 β_2——和用以估计它们的 6 个诱导系数——Π_0、Π_1、Π_2、Π_3、Π_4 和 Π_5。这样，我们就有含 6 个未知数的 6 个方程。在正常情况下，我们应能得到惟一的估计值。因此，需求和供给两方程的参数都是可识别的，从而整个模型是可识别的。（习题 19.2 要求读者用前面所给的 6 个诱导型系数把 6 个结构系数表达出来，以表明模型的惟一估计是可能的。）

746 为了核实上述供求函数是可识别的，仍可沿用“混合”的办法，即用 λ（$0\leqslant\lambda\leqslant1$）乘需求方程（19.2.12），用 $1-\lambda$ 乘供给方程（19.2.22），将它们相加，得到一个混杂方程。这个混杂方程将含有 I_t 和 P_{t-1} 两个前定变量，从而它在观测上有别于需求和供给方程两者。因为，前者不含 P_{t-1}，而后者不含 I_t。

过度识别

对某些商品和服务来说，消费者的收入和财富都同样是需求的重要决定因素。因此我们把需求函数（19.2.12）修改如下，但保持供给函数如前：

$$\text{需求函数：}\quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \tag{19.2.28}$$

$$\text{供给函数：}\quad Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t} \tag{19.2.22}$$

其中除了已定义的变量外，R 代表财富；对大多数商品和服务来说，财富和收入一样，预料会对消费产生正的影响。

令需求等于供给，便得到以下的均衡价格和数量：

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + \Pi_3 P_{t-1} + v_t \tag{19.2.29}$$

$$Q_t = \Pi_4 + \Pi_5 I_t + \Pi_6 R_t + \Pi_7 P_{t-1} + w_t \tag{19.2.30}$$

其中

$$\begin{aligned}
\Pi_0 &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} & \Pi_1 &= -\frac{\alpha_2}{\alpha_1 - \beta_1} \\
\Pi_2 &= -\frac{\alpha_3}{\alpha_1 - \beta_1} & \Pi_3 &= \frac{\beta_2}{\alpha_1 - \beta_1} \\
\Pi_4 &= \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} & \Pi_5 &= -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1} \\
\Pi_6 &= -\frac{\alpha_3\beta_1}{\alpha_1 - \beta_1} & \Pi_7 &= \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} \\
w_t &= \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} & v_t &= \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}
\end{aligned} \tag{19.2.31}$$

上述供求模型含有 7 个结构系数，但用以估计它们的有 8 个方程——（19.2.31）所给的 8 个诱导型系数；就是说方程个数大于未知数个数。其结
747 果是，要对我们这个模型的全部参数求惟一估计值是不可能的。这点是容易说明的。由上述诱导型系数，我们能得到：

$$\beta_1 = \frac{\Pi_6}{\Pi_2} \tag{19.2.32}$$

或：

$$\beta_1 = \frac{\Pi_5}{\Pi_1} \tag{19.2.33}$$

就是说，对供给方程中的价格函数有两个估计值，但不保证这两个估计值将是相同的。[4]此外，由于 β_1 出现在所有诱导型系数的分母中，在 β_1 估计中的含糊性还会传递给其他的估计值。

为什么在方程组（19.2.12）和（19.2.22）中供给函数是可识别的，而在方程组（19.2.28）和（19.2.22）中尽管供给函数仍是一样的却不能（恰好）识别呢？答案是为了识别供给曲线，我们有了“太多”的或“过于充分”的信息。这种情形和太少信息的不足识别情形恰好相反。信息的过于充

分是由如下事实造成的：在由（19.2.12）和（19.2.22）构成的模型中，把收入变量从供给函数中排除出去便足以识别供给函数。但在由（19.2.28）和（19.2.22）构成的模型中，供给函数不仅排除收入变量，还排除了财富变量。换言之，在后一模型中，我们对供给函数施加了“过多”的约束，要求它排除多于识别它所必需的变量个数。然而，这种情况并不意味着过度识别一定是坏事。因为，在第 20 章中，我们将看到，我们是怎样处理好过多信息或过多约束的问题的。

至此，我们已列举了所有的情形，如以上讨论所表明的，联立方程模型中的一个方程可以是不足识别的或可识别的（过度或恰好）。如果模型中的每一方程都是可识别的，就说整个模型是可识别的。为了判明识别问题，我们求助于诱导型方程。但在第 19.3 节中，我们将考虑判明联立方程模型中的一个方程是否可区别于另一个。这也许是较省时的方法。

§ 19.3 识别规则

如第 19.2 节诸例所表明的，原则上可借助于诱导型方程来决定联立方
748 程组中的某一方程的可识别性。但这些例子也表明了这个识别过程多么地耗时与费力。幸而这种程序并不是非用它不可的。所谓的**识别的阶和秩条件**（order and rank conditions of identification），由于提供了一种系统性的例行程序而减轻了这一任务。

为了了解阶和秩条件，我们引进以下符号：

M = 模型中内生变量的个数

m = 给定方程中内生变量的个数

K = 模型中前定变量的个数

k = 给定方程中前定变量的个数

可识别性的阶条件[5]

可识别性的一个必要（但非充分）条件，称**阶条件**（order condition），可用两种不同但等价的方式叙述如下（稍后即将介绍识别的必要条件和充分条件）：

> **定义 19.1** 在一个含有 M 个联立方程的模型中，为了使一个方程能被识别，它必须排除**至少** $M-1$ 个在模型中出现的变量（内生或前定）。如果它恰好排除 $M-1$ 个变量，则该方程是恰好识别的，如果它排除多于 $M-1$ 个变量，则它是过度识别的。

定义 19.2 在一个含有 M 个联立方程的模型中，为了使一个方程能被识别，该方程所排除的前定变量的个数必须不少于它所含有的内生变量的个数减 1，即：

$$K-k \geqslant m-1 \tag{19.3.1}$$

如果 $K-k=m-1$，则方程是恰好识别的；但如果 $K-k>m-1$，则它是过度识别的。

习题 19.1 要求读者证明上述两个可识别性定义是等价的。

为了说明阶条件，让我们回到前面的例子。

749 **例19.1**

需求函数： $$Q_t=\alpha_0+\alpha_1 P_t+u_{1t} \tag{18.2.1}$$

供给函数： $$Q_t=\beta_0+\beta_1 P_t+u_{2t} \tag{18.2.2}$$

此模型有两个内生变量 P 和 Q 而无前定变量。为了能被识别，每个方程至少要排除 $M-1=1$ 个变量。但情形并非如此，故没有哪个方程是可识别的。

例 19.2

需求函数： $$Q_t=\alpha_0+\alpha_1 P_t+\alpha_2 I_t+u_{1t} \tag{19.2.12}$$

供给函数： $$Q_t=\beta_0+\beta_1 P_t+u_{2t} \tag{19.2.13}$$

在此模型中，Q 和 P 是内生的，而 I 是外生的。应用（19.3.1）所给的阶条件，我们看到需求函数是不可识别的。另一方面，供给函数恰好排除 $M-1=1$ 个变量 I_t，所以是恰好识别的。

例 19.3

需求函数： $$Q_t=\alpha_0+\alpha_1 P_t+\alpha_2 I_t+u_{1t} \tag{19.2.12}$$

供给函数： $$Q_t=\beta_0+\beta_1 P_t+\beta_2 P_{t-1}+u_{2t} \tag{19.2.22}$$

给定 P_t 和 Q_t 为内生变量以及 I 和 P_{t-1} 为前定变量，方程（19.2.12）恰好排除 1 个变量 P_{t-1}，并且方程（19.2.22）也恰好排除 1 个变量 I_t，因此根据阶条件每个方程都是可识别的，从而整个模型是可识别的。

例 19.4

需求函数： $$Q_t=\alpha_0+\alpha_1 P_t+\alpha_2 I_t+\alpha_3 R_t+u_{1t} \tag{19.2.28}$$

供给函数： $$Q_t=\beta_0+\beta_1 P_t+\beta_2 P_{t-1}+u_{2t} \tag{19.2.22}$$

在此模型中 P_t 和 Q_t 为内生，而 I_t，R_t 和 R_{t-1} 为前定。需求函数恰好排除 1 个变量 P_{t-1}，所以按照阶条件它是恰好识别的。但供给函数排除了两个变量 I_t 和 R_t，因而是过度识别的。如前所见，似此情形，将存在有两种估计价格变量系数的方法。

注意这里的一个略为复杂的情况。由阶条件，需求函数是可识别的。但如果我们试图从（19.2.31）所给的诱导型系数去估计此方程的参数的话，

则由于在计算中 β_1 的介入、估计值将不是惟一的。β_1 有两个值，我们还要决定哪一个值是适当的。不过，这一复杂性是可以避免的。如在第 20 章中将要说明的，在遇到过度识别的情形时，间接最小二乘法是不适宜的，而应代之以其他方法，其中的一个方法就是我们将要在第 20 章里详细讨论的**二阶最小二乘**（two-stage least squares）法。

750 如上面诸例所表明的，如果在一个联立方程模型中，某个方程不含有包含在模型别处的一或多个变量，就有可能去识别这个方程。这种识别方法称为（变量的）排除（排斥）准则或零约束准则（方程中不出现的变量，可视其系数为零）。这个准则是获至或确定一个方程的识别性的最常用方法。但注意，零约束准则是以某些变量不出现于某一方程的先验性或理论性预期为依据的。它任由研究者去明辨为什么他/她预料某些变量出现于一些方程而不出现于另一些方程。

可识别性的秩条件[6]

前面讨论的阶条件是识别的必要条件而非充分条件；就是说，即使它得到满足，方程也会出现不能识别的情形。例如在例 19.2 中，供给方程排除了出现于需求函数中的收入变量 I_t，按照阶条件来说是可识别的。但是，识别的实现还只有当需求函数中 I_t 的系数 α_2 不为零时，也就是，收入变量不仅仅有可能进入而且确实进入了需求函数时，才能识别。

说得更一般，即使一个方程满足了阶条件 $K-k \geqslant m-1$，它仍会是不可识别的。因为该方程所排除的、出现于模型中的那些变量也许不是独立的，以致结构系数（诸 β_i）与诱导系数（诸 Π_i）之间没有一一对应关系。就是说，如同我们即将表明的，我们也许不能从诱导型系数估计出结构系数。因此，我们需要一个既必要而又充分的识别条件。可识别性的**秩条件**（rank condition of identification）提供了这一条件，现陈述如下：

> **可识别性的秩条件** 在一个含 M 个内生变量的 M 个方程的模型中，一个方程是可识别的，当且仅当我们能从模型（其他方程）所含而该方程所不含的诸变量（内生或前定）的系数矩阵中构造出至少一个$(M-1)\times(M-1)$阶的非零行列式来。

751 作为可识别性的秩条件的一个说明，考虑以下假想的联立方程组，其中 Y 变量为内生而 X 变量为前定。[7]

$$Y_{1t}-\beta_{10}\quad-\beta_{12}Y_{2t}-\beta_{13}Y_{3t}-\gamma_{11}X_{1t}=u_{1t}\qquad(19.3.2)$$

$$Y_{2t}-\beta_{20}\quad-\beta_{23}Y_{3t}-\gamma_{21}X_{1t}-\gamma_{22}X_{2t}=u_{2t}\qquad(19.3.3)$$

$$Y_{3t}-\beta_{30}-\beta_{31}Y_{1t}\quad-\gamma_{31}X_{1t}-\gamma_{32}X_{2t}=u_{3t}\qquad(19.3.4)$$

$$Y_{4t}-\beta_{40}-\beta_{41}Y_{1t}-\beta_{42}Y_{2t}\qquad\qquad\qquad -\gamma_{43}X_{3t}=u_{4t}\qquad(19.3.5)$$

为了便于识别性判断，我们将上面的方程组写成表 19.1。

表 19.1

	变量的系数							
方程编号	1	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
(19.3.2)	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	γ_{11}	0	0
(19.3.3)	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
(19.3.4)	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
(19.3.5)	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	γ_{43}

首先应用可识别性的阶条件，如表 19.2 所示。按阶条件，每个方程都是可识别的。再按秩条件复查。先看第一个方程；它排除了变量 Y_4、X_2 和 X_3（这由表 19.1 第 1 行中的那些零表示出来）。若此方程能被识别，我们必须从其他方程所含而此方程所不含的诸变量的系数（矩阵）中找到至少一个非零的 3×3 阶行列式。为了得到此行列式，我们先找出包含在其他方程中的变量 Y_4、X_2 和 X_3 的有关系数矩阵。在本例中，这样的矩阵（方阵）只有一个，
752 且称它为 A，其定义如下：

$$\mathbf{A}=\begin{bmatrix}0 & -\gamma_{22} & 0\\ 0 & -\gamma_{32} & 0\\ 1 & 0 & -\gamma_{43}\end{bmatrix}\qquad(19.3.6)$$

不难看出，此矩阵的行列式为零：

$$\det\mathbf{A}=\begin{vmatrix}0 & \gamma_{22} & 0\\ 0 & -\gamma_{32} & 0\\ 1 & 0 & -\gamma_{43}\end{vmatrix}\qquad(19.3.7)$$

表 19.2

方程编号	被排除的前定变量个数 $(K-k)$	被包含的内生变量个数减 1 $(m-1)$	识别情况
(19.3.2)	2	2	恰好
(19.3.3)	1	1	恰好
(19.3.4)	1	1	恰好
(19.3.5)	2	2	恰好

既然行列式为零，矩阵（19.3.6）的秩，记为$\rho(\mathbf{A})$，就小于3。因此，方程（19.3.2）不满足秩条件，从而它是不可识别的。

我们说过，秩条件是可识别性的一个必要且充分条件。因此，虽然方程（19.3.2）满足识别的阶条件，但不满足识别的秩条件，表明它是不可识别的。显然，（19.3.6）所给矩阵**A**的列或行不是（线性）独立的，这意味着变量Y_4、X_2与X_3之间有某种关系，致使我们未能得到足够的信息用以估计方程（19.3.2）中的参数；上述模型的诱导型方程将表明不可能从诱导型系数求出该方程的结构系数。读者应能验证，按照秩条件，方程（19.3.3）和（19.3.4）也是不可识别的，但方程（19.3.5）可以识别。

如以上讨论所表明的，秩条件告诉我们所考虑的方程是否可识别，而阶条件告诉我们它是恰好可识别还是过度可识别的。

为了应用秩条件，我们可按以下步骤进行：

1. 像表19.1那样，把方程组写成表格形式。
2. 划掉被考虑的方程所在的一行的系数。
3. 再划掉与步骤2中的非零系数对应的列。
4. 表中余下的系数将构成方程组所含而被考虑的方程所不含的变量的系数矩阵，从这一系数矩阵形成所有可能的像**A**那样的$M-1$阶方阵，并求出相应的行列式。如果能找到至少一个非消失的或非零的行列式，则所讨论的方程是（恰好或过度）可识别的。这时矩阵**A**的秩恰好是$M-1$。如果所有可能的$(M-1)\times(M-1)$行列式皆是零，则矩阵**A**的秩小于$M-1$，从而所研究的方程是不可识别的。

753 我们对可识别性的阶和秩条件的讨论，导致了关于在M个联立方程组中的一个结构方程的可识别性的一般原则如下：

1. 如果$K-k>m-1$且**A**矩阵的秩是$M-1$，则方程是过度可识别的。
2. 如果$K-k=m-1$且**A**矩阵的秩是$M-1$，则方程是恰好可识别的。
3. 如果$K-k\geqslant m-1$而矩阵**A**的秩小于$M-1$，则方程是不可识别的。
4. 如果$K-k<m-1$，则结构方程是不可识别的。这时**A**矩阵的秩必定小于$M-1$。（为什么？）

从此以后，凡是谈到识别情况，都指恰好识别或过度识别。考虑不可识别或不足识别的方程是没有意义的。因为这时无论数据有多广泛，结构方程都是不可估计的。而另一方面，如在第20章中所表明的，过度识别和恰好识别的方程中的参数都是能估计的。

在实践中我们应该使用哪一条件：阶或秩？对大型联立方程模型来说，秩条件的应用是一件令人生畏的任务。为此，哈维（Harvey）指出，

> 幸亏，阶条件通常已足以保证可识别性，虽然当心秩条件是重要的，但不去验证它，一般不会造成什么危害。[8]

※§19.4 联立性检验[9]

如果没有联立方程或**联立性问题**（simultaneity problem），OLS 估计量将产生一致且有效估计。而另一方面，如果存在有联立性，则 OLS 估计量甚至不是一致性的。如在第 20 章中将要表明的，当联立性出现时，**二阶最小二乘**（2SLS）和**工具变量**将给出一致性且有效估计量。说来有点奇怪，如果我们在没有联立性情况下应用这些不同于 OLS 的方法，仍可得到一致性的、但非有效的估计量。所有这些讨论表明，在我们摒弃 OLS 而倾向于用其他方法之前，应检验联立性的问题。

754 如前些时候已表明的，联立性问题之所以出现，是因为一些回归元是内生的，并因而很可能与干扰或误差项相关。因此，联立性检验在本质上是检验（一个内生）回归元是否与误差项相关。如果是，就有联立性的问题，这时需要找出不同于 OLS 的估计方法；如果不是，就可使用 OLS。为在某一事例中判定是或不是，可采取豪斯曼的设定误差检验。

豪斯曼设定检验

豪斯曼设定误差检验的一种形式可用于检验联立性问题。现解释如下[10]：

为了建立概念，考虑如下的二方程模型：

$$\text{需求函数：}\quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t} \tag{19.4.1}$$

$$\text{供给函数：}\quad Q_t = \beta_0 + \beta_1 P_t + u_{2t} \tag{19.4.2}$$

其中 P = 价格

Q = 数量

I = 收入

R = 财富

诸 u = 误差项

假定 I 和 R 为外生。当然，P 和 Q 为内生。

现考虑供给函数（19.4.2）。如果没有联立性问题（即 P 与 Q 互相独立），则 P_t 与 u_{2t} 应是不相关的。（为什么？）另一方面，如果有联立性，则 P_t 与 u_{2t} 将是相关的。要判明是哪一种情形，豪斯曼检验的程序如下：

首先，从（19.4.1）和（19.4.2）得到以下的诱导型方程：

$$P_t = \Pi_0 + \Pi_1 I_t + \Pi_2 R_t + v_t \tag{19.4.3}$$

$$Q_t = \Pi_3 + \Pi_4 I_t + \Pi_3 R_t + w_t \tag{19.4.4}$$

其中 v 和 w 为诱导型误差项，用 OLS 估计（19.4.5）得：

$$\hat{P}_t = \hat{\Pi}_0 + \hat{\Pi}_1 I_t + \hat{\Pi}_2 R_t \tag{19.4.5}$$

因此，

$$P_t = \hat{P}_t + \hat{v}_t \tag{19.4.6}$$

755 其中 $\hat{P}_t$ 代表估计的 P_t，而 $\hat{v}_t$ 为估计的残差。将（19.4.6）代入（19.4.2）得：

$$Q_t = \beta_0 + \beta_1 \hat{P}_t + \beta_1 \hat{v}_t + u_{2t} \tag{19.4.7}$$

注：$\hat{P}_t$ 和 v_t 有相同的系数。

现在，在无联立性的虚拟假设下，$\hat{v}_t$ 与 u_{2t} 之间的相关应在渐近意义下等于零。因此，如果我们做回归（19.4.7），并发现（19.4.7）中 v_t 的系数在统计上为零，就可得到不存在有联立性问题的结论。当然，如果我们发现这个系数是统计上显著的，就把结论反过来。

于是豪斯曼检验基本上涉及两个步骤：

步骤 1. 求 P_t 对 I_t 和 R_t 的回归以得到 $\hat{v}_t$。

步骤 2. 求 Q_t 对 $\hat{P}_t$ 和 $\hat{v}_t$ 的回归并对 $\hat{v}_t$ 的系数做 t 检验。如果它是显著的，就不拒绝联立性假设；否则拒绝之。[11]然而，为了更有效的估计，平狄克（Pindyck）和鲁宾费尔德（Rubinfeld）建议做 Q_t 对 P_t 和 $\hat{v}_t$ 的回归。[12]

例 19.5：

平狄克-鲁宾费尔德公共支出模型[13]

为了研究美国州和地方政府支出行为，两位作者提出以下的联立方程模型：

$$\text{EXP} = \beta_1 + \beta_2\text{AID} + \beta_3\text{INC} + \beta_4\text{POP} + u_i \tag{19.4.8}$$

$$\text{AID} = \delta_1 + \delta_2\text{EXP} + \delta_3\text{PS} + v_i \tag{19.4.9}$$

其中 EXP＝州与地方政府公共支出

AID＝联邦拨款水平

INC＝州收入

POP＝州人口

PS＝初级与次级学校在校儿童人口

u 和 v＝误差项

在此模型中，INC、POP 和 PS 被视为外生的。

由于 EXP 和 AID 之间有联立性的可能，作者先求 AID 对 INC、POP 和 PS 的回归（即诱导型回归）。令此回归的误差项为 w_i。由此回归求得残差

756 $\hat{w}_i$。然后，作者求 EXP 对 AID、INC、POP 和 $\hat{w}_i$ 的回归而得到以下结果：

$$\widehat{EXP} = -89.41 + 4.50AID + 0.000\,13INC - 0.518POP - 1.39\hat{w}_i$$
$$t = (-1.04)\quad(5.89)\quad(3.06)\quad(-4.63)\quad(-1.73)$$
$$(19.4.10)^{[10]}$$
$$R^2 = 0.99$$

在5%显著水平上 $\hat{w}_i$ 的系数不是统计上显著的，并因此在此水平上没有联立性问题。然而，在10%显著水平上它却是统计上显著的，故存在联立性问题的可能性似仍可考虑。

不妨看看，(19.4.8) 的 OLS 估计如下：

$$\widehat{EXP} = -46.81 + 3.24AID + 0.000\,19INC - 0.597POP$$
$$t = (-0.56)\quad(13.64)\quad(8.12)\quad(-5.71)\qquad(19.4.11)$$
$$R^2 = 0.993$$

注意，在 (19.4.10) 和 (19.4.11) 所给的结果中有一个有趣的现象。当我们明确地考虑联立性问题时，AID 变量的应用尽管在数值上变大了一些，但它的显著性反而减小了。

* §19.5 外生性检验

我们前面曾说过，明确哪些变量是内生的，哪些是外生的，责任在研究者。这将与研究者考虑的问题和他拥有的先验信息有关。但能不能推出一种像格兰杰因果关系检验那样的外生性检验?

第 19.4 节讨论的豪斯曼检验可用来回答此问题。假使我们有一个三内生变量的三方程模型，并假定有三个外生变量 X_1、X_2 和 X_3。再进一步假定模型的第一个方程是：

$$Y_{1i} = \beta_0 + \beta_2 Y_{2i} + \beta_3 Y_{3i} + \alpha_1 X_{1i} + u_{1i} \qquad (19.5.1)$$

如果 Y_2 和 Y_3 是真的内生变量，我们就不能用 OLS 去估计 (19.5.1)。(为什么?) 但我们如何知道? 可进行如下检验：我们求 Y_2 和 Y_3 的诱导型方程 (注：诱导型方程的右边将仅有前定变量)。由这些诱导型方程我们分别得到 Y_{2i} 和 Y_{3i} 的预测值 $\hat{Y}_{2i}$ 和 $\hat{Y}_{3i}$。然后根据先前讨论的豪斯曼检验的思想，我们用 OLS 估计下述方程：

757
$$Y_{1i} = \beta_0 + \beta_2 Y_{2i} + \beta_3 Y_{3i} + \alpha_1 X_{1i} + \lambda_2 \hat{Y}_{2i} + \lambda_3 \hat{Y}_{3i} + u_{1i} \qquad (19.5.2)$$

可通过 F 检验来检验假设：$\lambda_2 = \lambda_3 = 0$。如果此假设被拒绝，则可认为 Y_2 和 Y_3 是内生的，但如果它不被拒绝，就可视同外生。一个具体的例子见于习题 19.16。

---

* 供选读。

§ 19.6 要点与结论

1. 识别问题的考虑应先于估计问题。

2. 识别问题是问我们能否从估计得的诱导型系数求出结构系数的惟一数值估计值。

3. 如果能做到，就说联立方程组中的某个方程是可识别的。如果做不到，该方程就是不可或不足识别的。

4. 一个可识别的方程可以是恰好可识别或过度可识别的。在前一种情形中，可以得到结构系数的惟一值；而在后一种情形中，也许对一或多个结构参数有多于一个估计值。

5. 识别问题之所以出现，是因为同样的数据集适合于不同的结构系数集，也就是适合于不同的模型。例如，在一个价格只对数量的回归中，很难说人们是在估计供给函数还是需求函数，因为价格和数量同样进入两个方程。

6. 要判断一个结构方程的可识别性，我们可以应用**诱导型方程**的技术，把一个内生变量表达为纯粹是前定变量的一个函数。

7. 然而，这种耗时的程序由于**阶条件**或**秩条件**的利用而得以避免。虽然阶条件易于应用，但它仅是可识别性的一个必要条件。另一方面，秩条件则是识别的既必要又充分的一个条件。如果秩条件被满足，阶条件也一定被满足。但反过来未必真。尽管如此，在实践中，阶条件一般地说能较好地保证可识别性。

8. 当联立性出现时，如第 18 章中所表明的，OLS 一般而言是不适用的。但如果我们仍想用它，则必须明白地做出联立性检验，为此，可利用**豪斯曼设定检验**。

9. 虽然在实践中一个变量是内生或外生的，是凭判断而做决定的，但我们可以用豪斯曼设定检验判定一个或一组变量是内生还是外生的。

675 10. 因果关系和外生性虽属于同一类问题，但它们的概念却是不同的。其中一个概念并不蕴涵另一个概念。在实践中，仍然是把这两个概念区分开来为好（见 17.14 节）。

习 题

问答题

19.1 说明可识别性的阶条件的两个定义是等价的。

19.2 从（19.2.25）和（19.2.27）所给的诱导型系数导出结构系数。

19.3 求出以下模型的诱导型，从而判定每一种情形的结构方程是不可识别的，恰好识别的或过度识别的。

a. 第 18 章例 18.2。

b. 第 18 章例 18.3。

c. 第 18 章例 18.6。

19.4 兼用阶和秩条件检查习题 19.3 的模型的可识别性。

19.5 在课文的模型（19.2.22）和（19.2.28）中，我们曾表明供给方程是过度识别的，能否对结构方程的参数作些约束使得此方程变为恰好识别的？说明这种约束的理由。

19.6 由模型：

$$Y_{1t}=\beta_{10}+\beta_{12}Y_{2t}+\gamma_{11}X_{1t}+u_{1t}$$
$$Y_{2t}=\beta_{20}+\beta_{21}Y_{1t}+\gamma_{22}X_{2t}+u_{2t}$$

得到如下的诱导型方程：

$$Y_{1t}=\Pi_{10}+\Pi_{11}X_{1t}+\Pi_{12}X_{2t}+w_t$$
$$Y_{2t}=\Pi_{20}+\Pi_{21}X_{1t}+\Pi_{22}X_{2t}+v_t$$

a. 这些结构方程是可识别的吗？

b. 如果先验地知道 $\gamma_{11}=0$，识别情况会有什么变化？

19.7 参照习题 19.6. 估计的诱导型方程如下：

$$Y_{1t}=4+3X_{1t}+8X_{2t}$$
$$Y_{2t}=2+6X_{1t}+10X_{2t}$$

a. 求结构参数的值。

b. 你会怎样检验虚拟假设 $\gamma_{11}=0$?

19.8 模型：

$$Y_{1t}=\beta_{10}+\beta_{12}Y_{2t}+\gamma_{11}X_{1t}+u_{1t}$$
$$Y_{2t}=\beta_{20}+\beta_{21}Y_{1t}+u_{2t}$$

759 产生以下诱导型方程：

$$Y_{1t}=4+8X_{1t}$$
$$Y_{2t}=2+12X_{1t}$$

a. 有没有哪些结构系数是能够从诱导型系数估计出来的？说明你的见解。

b. 如果预先知道（1）$\beta_{12}=0$ 和（2）$\beta_{10}=0$，你怎样改变对（a）的答案？

19.9 习题 18.8 所给模型的结构方程是可识别的吗？

19.10 参照习题 18.7 并找出哪些结构方程是可识别的。

19.11 下面是一个含 5 个内生变量 Y 和 4 个外生变量 X 的 5 方程模型：

方程编号	变量的系数								
	Y_1	Y_2	Y_3	Y_4	Y_5	X_1	X_2	X_3	X_4
1	1	β_{12}	0	β_{14}	0	γ_{11}	0	0	γ_{14}
2	0	1	β_{23}	β_{24}	0	0	γ_{22}	γ_{23}	0
3	β_{31}	0	1	β_{34}	β_{35}	0	0	γ_{33}	γ_{34}
4	0	β_{42}	0	1	0	γ_{41}	0	γ_{43}	0
5	β_{51}	0	0	β_{54}	1	0	γ_{52}	γ_{53}	0

借助于可识别性的阶和秩条件，判定每一方程的可识别性。

19.12 考虑以下扩展的凯恩斯收入决定模型：

消费函数： $C_t = \beta_1 + \beta_2 Y_t - \beta_3 T_t + u_{1t}$

投资函数： $I_t = \alpha_0 + \alpha_1 Y_{t-1} + u_{2t}$

税收函数： $T_t = \gamma_0 + \gamma_1 Y_t + u_{3t}$

收入恒等式： $Y_t = C_t + I_t + G_t$

其中 C = 消费支出

Y = 收入

I = 投资

T = 税收

诸 u = 干扰项

模型中的内生变量是 C、I、T 和 Y，而前定变量是 G（政府支出）和 Y_{t-1}。

用阶条件检查方程组中每一方程和整个方程组的可识别性，
760 假定有一个作为外生变量的利率 γ_t 出现在投资函数的右侧，将会出现什么情况？

19.13 参照第 18 章表 18.1 所给的数据。利用这些数据估计诱导型回归 (19.1.2) 和 (19.1.4)，你能估计 β_0 和 β_1 吗？说明你的计算。模型是可识别的吗？为什么或为什么不？

19.14 假使我们提出可识别性的阶条件的另一个定义：

$$K \geqslant m + k - 1$$

就是说方程组中的前定变量的个数不可少于待识别的方程中所含未知系数的个数。说明此定义和课文中所给阶条件的另两个定义是等价的。

19.15 休茨（Suits）的西瓜市场模型的一个简化形式如下[1]：

需求方程： $P_t = \alpha_0 + \alpha_1 (Q_t/N_t) + \alpha_2 (Y_t/N_t) + \alpha_3 F_t + u_{1t}$

作物供给函数： $Q_t = \beta_0 + \beta_1 (P_t/W_t) + \beta_2 P_{t-1} + \beta_3 C_{t-1} + \beta_4 T_{t-1} + u_{2t}$

其中　P = 价格

(Q/N) = 人均需求量

(Y/N) = 人均收入

F_t = 运费

(P/W) = 相对于农场工资的价格

C = 棉花价格

T = 其他蔬菜价格

N = 人口

P 和 Q 为内生变量。

a. 求出诱导型。

b. 判定需求函数，供给函数，或两者是否均可识别？

19.16 考虑以下的货币供求模型：

货币需求：　$M_t^d = \beta_0 + \beta_1 Y_t + \beta_2 R_t + \beta_3 P_t + u_{1t}$

货币供给：　$M_t^s = \alpha_0 + \alpha_1 Y_t + u_{2t}$

其中　M = 货币

Y = 收入

R = 利率

P = 价格

诸 u = 误差项

761 假定 R 和 P 是外生的，而 M 和 Y 是内生的。下表给出 1970—1991 年美国的 M（M_2 定义），Y（国内生产总值），R（3 个月国债利率）和 P（消费者价格指数）数据。

a. 需求函数可识别吗？

b. 供给函数可识别吗？

c. 求出 M 和 Y 的诱导型方程的表达式。

d. 对供给函数做联立性检验。

e. 我们怎样知道货币供给函数中的 Y 确实是内生的？

表 19.4　货币、国内总产值、利率，以及消费者价格指数，美国，1970—1999

年份	M2	GDP	TBRATE	CPI
1970	626.400 0	3 578.000	6.458 000	38.800 00
1971	710.100 0	3 697.700	4.348 000	40.500 00
1972	802.100 0	3 998.400	4.071 000	41.800 00
1973	855.200 0	4 123.400	7.041 000	44.400 00
1974	901.900 0	4 099.000	7.886 000	49.300 00
1975	1 015.900	4 084.400	5.838 000	53.800 00
1976	1 151.700	4 311.700	4.989 000	56.900 00

1977	1 269.900	4 511.800	5.265 000	60.600 00
1978	1 365.500	4 760.600	7.221 000	65.200 00
1979	1 473.100	4 912.100	10.041 00	72.600 00
1980	1 599.100	4 900.900	11.506 00	82.400 00
1981	1 754.600	5 021.000	14.029 00	90.900 00
1982	1 909.500	4 913.300	10.686 00	96.500 00
1983	2 126.000	5 132.300	8.630 000	99.600 00
1984	2 309.700	5 505.200	9.580 000	103.900 0
1985	2 495.400	5 717.100	7.480 000	107.600 0
1986	2 732.100	5 912.400	5.980 000	109.600 0
1987	2 831.100	6 113.300	5.820 000	113.600 0
1988	2 994.300	6 368.400	6.690 000	118.300 0
1989	3 158.400	6 591.900	8.100 000	124.000 0
1990	3 277.600	6 707.900	7.510 000	130.700 0
1991	3 376.800	6 676.400	5.420 000	136.200 0
1992	3 430.700	6 880.000	3.450 000	140.300 0
1993	3 484.400	7 062.600	3.020 000	144.500 0
1994	3 499.000	7 347.700	4.290 000	148.200 0
1995	3 641.900	7 543.800	5.510 000	142.400 0
1996	3 813.300	7 813.200	5.020 000	156.900 0
1997	4 028.900	8 159.500	5.070 000	160.500 0
1998	4 380.600	8 515.700	4.810 000	163.000 0
1999	4 643.700	8 875.800	4.660 000	166.600 0

注：M2＝M2 货币供给（10 亿美元计，经季节性调整）
GDP＝国内生产总值（10 亿美元计，经季节性调整）
TBRATE＝3 个月国债利率，%
CPI＝消费者价格指数（1982—1984 年＝100）
资料来源：*Economic Report of the President*, 2001、Tables B-2、B-60、B-73、B-69.

【习题注释】

[1] D.B. Suits, "An Econometric Model of the Watermelon Market," *Journal of Farm Economics*, vol. 37, 1955, pp.237-251.

【注释】

[1] 这里隐含地假定了随机干扰项 u 没有序列相关性。如果有的话，Y_{t-1}将与现期干扰项 u_t 相关，从而我们不能把它看作前定的了。

[2] 在计量经济学中，外生变量扮演着重要的角色。这类变量常常被置于政府

直接控制之下，例如个人与公司的税率、津贴、失业救济等等。

[3] 有关细节，参看 Jan Kmenta, *Elements of Econometrics*, 2d ed., Macmillan, New York, 1986, pp.723 – 731。

[4] 注意不足与过度识别之间的差异。对于前一情形，要得到结构参数的估计值是不可能的，而对于后一情形，则可能有一或多个结构系数的几个估计值。

[5] 名词“阶”指一个矩阵的阶，即出现在矩阵中的行和列的个数。参看附录B。

[6] 名词“秩”指一个矩阵的秩，它由（该矩阵所含）行列式不为零的最高阶方阵给出。或者，换一种说法，一个矩阵的秩是指该矩阵的线性独立的最大行或列数。

[7]（19.1.1）中所给的联立方程组可用下面的另一种形式表现出来，以便于矩阵运算。

[8] Andrew Harvey, *The Econometric Analysis of Time Series*, 2d ed., The MIT Press, Cambridge, Mass., 1990, p.328.

[9] 以下讨论取自 Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 3d ed., McGraw-Hill, New York, 1991, pp.303 – 305。

[10] J. A. Hausman, "Specification Tests in Econometrics," *Econometrica*, vol. 46, November 1976, pp.1251 – 1271. See also A. Nakamura and M. Nakamura, "On the Relationship among Several Specification Error Tests Presented by Durbin, Wu, and Hausman," *Econometrica*, vol. 49, November 1981, pp.1583 – 1588.

[11] 如果涉及多于 1 个内生回归元，我们将使用 F 检验。

[12] Pindyck and Rubinfeld，前引文献，第 304 页。注意回归元是 P_t，不是 $\hat{P}_t$。

[13] Pindyck and Rubinfeld，前引文献，第 176 – 177 页。符号略有改动。

[14] 根据注 [12]，作者们用AID而不用 $\widehat{AID}$ 作为回归元。

[15] G.S. Maddala, *Introduction to Econometrics*, Macmillan, 2d ed., New York, 1992, pp.389 – 395.

第 20 章　联立方程方法

762 已经在前两章中讨论过联立方程模型的性质，本章我们转向这类模型的参数估计问题。首先让我们指出，由于存在着有不同统计性质的许多估计技术，估计问题是相当复杂的。像本书这种导引性质的教材，我们将只考虑少数的估计技术。我们的讨论将是简单的，并且常常是直觉指引的。一些细致的问题则留给阅读参考文献解决。

§20.1　估计的方法

考虑（19.1.1）所给的一般性 M 个内生变量的 M 个方程模型。为了估计结构方程，可采取两种方法，即单一方程法，又称**有限信息法**（limited information methods）和方程组法（或系统法），又称**完全信息法**（full information methods）。在即将考虑的单一方程法中，我们个别地估计（联立）方程组中的每一方程，仅考虑对该方程的约束（如对某些变量的排除）而不考
763 虑对其他方程的约束[1]，由此得名有限信息法。而另一方面，在方程组法中，我们同时地估计模型中的全部方程，适当考虑了因某些变量被排除而对方程组造成的全部约束（回想一下，这些约束对识别来说是关键性的），由

此得到完全信息法。

作为一个例子，考虑如下的 4 方程模型：

$$
\begin{aligned}
Y_{1t} &= \beta_{10} + \beta_{12}Y_{2t} + \beta_{13}Y_{3t} + \gamma_{11}X_{1t} + u_{1t} \\
Y_{2t} &= \beta_{20} + \beta_{23}Y_{3t} + \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + u_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}Y_{1t} + \beta_{34}Y_{4t} + \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + u_{3t} \\
Y_{4t} &= \beta_{40} + \beta_{42}Y_{2t} + \gamma_{43}X_{3t} + u_{4t}
\end{aligned}
\tag{20.1.1}
$$

其中诸 Y 为内生变量，而诸 X 为外生变量。如果，比如说，我们旨在估计第 3 个方程，则单一方程法将仅考虑此方程，即仅注意变量 Y_2 和 X_3 被排除在此方程之外。而另一方面，在方程组法中，我们要同时地估计全部 4 个方程，把对方程组中多个方程的全部约束都考虑进来。

为了保持联立方程模型的品质，最理想的应是使用方程组法，诸如**完全信息最大似然**（full information maximum likelihood，FIML）法。[2] 然而实际上，由于多种原因，这类方法并不常被使用。第一，计算上的负担太大。例如，比较小的（20 个方程）1955 年克莱因-戈德伯格构造的美国经济模型就有 151 个非零系数。作者们用时间序列数据仅估计其中的 51 个系数。布鲁金斯社会科学研究院（Brookings-Social Science Research Council，SSRC）于 1965 年出版的美国计量经济模型则开始含有 150 个方程。[3] 尽管这些精心制作的模型能对各个经济部门做出详细的描述，但计算工作量之大，即使在高速计算机时代的今天也是惊人的，更不必说成本上的耗费。第二，像 FIML 这样的系统方法常常导致参数的高度非线性解，以致难于确定。第三，如果方程组中的一或多个方程有设定误差（比如说，一个错误的函数形式或漏掉有关的变量），则误差将传递至其余的方程。其结果是，方程组法变成对设定误差非常敏感。

764 因此，在实践中常常使用单一方程法。如克莱因所说的，

> 在联立方程组的构架中，单一方程法在下述意义上也许对设定误差不那么敏感，即方程组的正确设定部分受另一部分的设定误差的影响也许不很大。[4]

在本章的其余部分，我们将仅讨论单一方程法。具体地说，我们将讨论如下的单一方程方法：

1. 普通最小二乘法（ordinary least squares，OLS）
2. 间接最小二乘法（indirect least squares，ILS）
3. 二阶最小二乘法（two-stage least squares，2SLS）

§20.2 递归模型与普通最小二乘法

我们在第 18 章中看到，因为随机干扰项和（诸）内生解释变量之间的相互依赖性，所以 OLS 法不适宜用来估计联立方程组中的方程。如果错误

地应用，那么，如我们在第 18.3 节所看到的，估计量不但是偏误的（在小样本中），而且是非一致性的，即不管样本含量多大，偏误也不会消失。然而，有一种情形，即使在联立方程的构架中，OLS 也是适用的。这就是**递归**（recursive），**三角形**（triangular）或**因果性**（causal）模型的情形。为了看清楚这种模型的性质，且考虑以下的 3 方程组：

$$
\begin{aligned}
Y_{1t} &= \beta_{10} & &+ \gamma_{11}X_{1t} + \gamma_{12}X_{2t} + u_{1t} \\
Y_{2t} &= \beta_{20} + \beta_{21}Y_{1t} & &+ \gamma_{21}X_{1t} + \gamma_{22}X_{2t} + u_{2t} \\
Y_{3t} &= \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} & &+ \gamma_{31}X_{1t} + \gamma_{32}X_{2t} + u_{3t}
\end{aligned}
\tag{20.2.1}
$$

如常，其中诸 Y 和诸 X 分别是内生和外生变量。干扰项有如下性质：

$$\operatorname{cov}(u_{1t}, u_{2t}) = \operatorname{cov}(u_{1t}, u_{3t}) = \operatorname{cov}(u_{2t}, u_{3t}) = 0$$

就是说，不同方程中的同期干扰项是不相关的［用专门术语说，这是一种**零同期相关**（zero contemporaneous correlation）假定］。

现考虑（20.2.1）中的第一个方程。因为它的右边仅含有外生变量，又
765 因为假定外生变量与干扰项 u_{1t} 不相关，所以此方程满足经典 OLS 解释变量与干扰项不相关的基本假定。因而 OLS 可直接应用于此方程的估计。再考虑（20.2.1）中的第二个方程。它不但含有非随机的诸 X，还含有 Y_1 作为解释变量。那么，如果 Y_{1t} 和 u_{2t} 是不相关的话，OLS 就可应用于此方程。但是不是这种情形呢？因为影响 Y_1 和 u_1 按假定是和 u_2 不相关的，所以答案是肯定的。因此，为了一切实际目的，在考虑 Y_2 的形成中就可把 Y_1 看作前定的，从而我们可以用 OLS 估计第二个方程。把这种推理再推进一步，由于 Y_1 和 Y_2 都与 u_3 不相关，我们又可对（19.2.1）中的第三个方程应用 OLS。

于是，在递归系统中，OLS 可分别地应用于每一个方程。其实在这种情况下，我们并没有联立方程的问题。从这种系统的结构看，显然不存在内生变量之间的相互依赖性（interdependence）。比方说，Y_1 影响 Y_2 但 Y_2 不影响 Y_1。类似地，Y_1 和 Y_2 影响 Y_3，而反过来并不受 Y_3 的影响。换言之，每个方程都展现一种单向的因果依赖性，由此得名因果性模型。[5]对此，图 20.1 给出一个图解。

作为递归系统的一个例子，不妨设想如下的一个工资与价格决定模型：

$$
\begin{aligned}
&\text{价格方程：}\quad \dot{P}_t = \beta_{10} + \beta_{11}\dot{W}_{t-1} + \beta_{12}\dot{R}_t + \beta_{13}\dot{M}_t + \beta_{14}\dot{L}_t + u_{1t} \\
&\text{工资方程：}\quad \dot{W}_t = \beta_{20} + \beta_{21}\mathrm{UN}_t + \beta_{32}\dot{P}_t + u_{2t}
\end{aligned}
\tag{20.2.2}
$$

766 其中 $\dot{P}$ = 单位产品的价格变化率

$\dot{W}$ = 每职工的工资变化率

$\dot{R}$ = 资本的价格变化率

$\dot{M}$ = 进口价格变化率

$\dot{L}$ = 劳动生产力变化率

UN = 失业率，%[6]

价格方程假定了当前价格变化率是资本和原料价格变化率，劳动生产变

化率以及前期工资变化率的函数。工资方程则表示当前工资变化率决定于当前价格变化率和失业率。显然，因果链的奔向是从 $\dot{W}_{t-1} \to \dot{P}_t \to \dot{W}_t$，因此，OLS 可分别地用于估计这两个方程的参数。

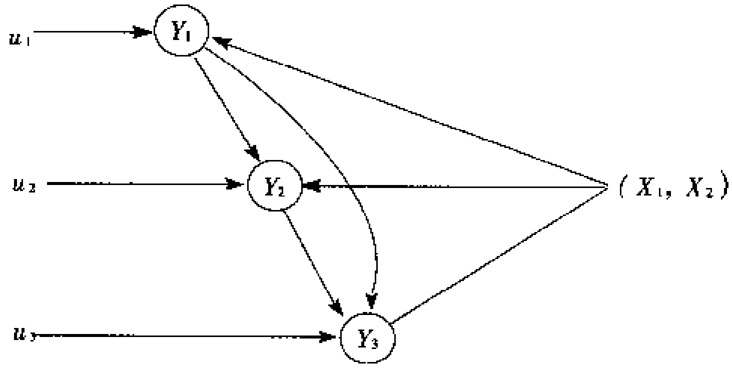


图 20.1　递归模型

虽然递归模型是有用场的，但大多数联立方程模型并不展现这种单向因果关系。因此，一般地说，在联立方程模式中，用 OLS 去估计其中的单一个方程是不适宜的。[7]

有些人称辩，虽然 OLS 一般地说不适用于联立方程模型，但如果只是为了提供一种比较的标准或规范，则还是可以使用的，就是说，我们可以用 OLS 去估计一个结构方程，尽管这种估计带有偏误、非一致性等性质。然后，用专门为对付联立性问题而设计的其他方法去估计同样的方程，再比较（至少，定性地比较）这两种方法所得的结果。如我们后面将要看到的，在许多应用中，用不适当的 OLS 法估计的结果和用更复杂的方法得到的结果相差可能不大。原则上，只要同时给出为联立方程模型而设计的其他方法的估计结果。我们就不应过多地反对列出基于 OLS 的估计结果，事实上，采取这种策略能使我们意识到，当 OLS 不适用而被使用时，它的表现会有多坏。[8]

§20.3　恰可识别方程的估计：间接最小二乘法

对一个恰好可识别的结构方程，从诱导型系数的 OLS 估计值获得结构系数估计值的方法叫做**间接最小二乘法**，而如此得到的估计值称间接最小二乘估计值。ILS 涉及以下三个步骤：

步骤 1.　先求诱导型方程。如第 19 章所表明的，从结构方程组解出诸诱导型方程，使得在每一方程中因变量成为惟一的内生变量，并且仅仅是前定（外生或滞后内生）变量和随机误差项的函数。

步骤 2. 对诱导型方程逐个地应用 OLS。因为这些方程中的解释变量是前定的并因而与随机干扰项不相关，所以这一运算是合适的，由此得到的估计值是一致性的。[9]

步骤 3. 从步骤 2 得到的诱导型系数的估计值求原始结构系数的估计值。如第 19 章所表明的，若方程恰可识别，则结构与诱导型系数之间有一一对应关系；就是说可以从后者导出前者的惟一估计值。

上述三步法表明，结构系数（对大多数情形来说是研究问题的主要目的）是从诱导型系数的估计值间接求出的，因此取名 ILS。

一个说明性例子

考虑第 19.2 节中引进的供求模型，为方便起见再次给出如下，符号稍有变化：

$$\text{需求函数：}\quad Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_{1t} \tag{20.3.1}$$

$$\text{供给函数：}\quad Q_t = \beta_0 + \beta_1 P_t + u_{2t} \tag{20.3.2}$$

其中 Q = 数量

P = 价格

X = 收入或支出

假定 X 是外生的。如前所示，供给函数是恰可识别的，而需求函数是不可识别的。

768 与上述结构方程组相对应的诱导型方程是

$$P_t = \Pi_0 + \Pi_1 X_t + w_t \tag{20.3.3}$$

$$Q_t = \Pi_2 + \Pi_3 X_t + v_t \tag{20.3.4}$$

其中的诸 Π 是诱导型系数，而且，如方程（19.2.16）和（19.2.18）所示，是结构系数的（非线性）组合，但是，其中的 w 和 v 则结构性干扰项 u_1 和 u_2 的线性组合。

注意到每一诱导型方程仅含一个内生变量，而这就是方程的因变量，并且仅仅是外生变量 X（收入）和随机干扰项的函数，由此知上述诱导型方程可由 OLS 来估计。这些估计值是：

$$\hat{\Pi}_1 = \frac{\sum p_t x_t}{\sum x_t^2} \tag{20.3.5}$$

$$\hat{\Pi}_0 = \bar{P} - \hat{\Pi}_1 \bar{X} \tag{20.3.6}$$

$$\hat{\Pi}_3 = \frac{\sum q_t x_t}{\sum x_t^2} \tag{20.3.7}$$

$$\hat{\Pi}_2 = \bar{Q} - \hat{\Pi}_3 \bar{X} \tag{20.3.8}$$

其中，和平常一样，小写字母表示对样本均值的离差，而 $\bar{Q}$ 和 $\bar{P}$ 是 Q 和 P 的样本均值。如前所说，诸 $\hat{\Pi}_i$ 是一致性估计量并在适当假定下还是最小方差、无偏或渐近有效的（参看章末注 9）。

由于我们的主要目的是定出结构系数，让我们看看能否从诱导型系数把它们估计出来。现在如第 19.2 节所表明的，供给函数是恰好可识别的。因此，它的参数从诱导型系数惟一地估计如下：

$$\beta_0 = \Pi_2 - \beta_1 \Pi_0 \qquad \text{和} \qquad \beta_1 = \frac{\Pi_3}{\Pi_1}$$

由此知这些参数的估计值可从诱导型系数的估计值计算如下：

$$\hat{\beta}_0 = \hat{\Pi}_2 - \hat{\beta}_1 \hat{\Pi}_0 \tag{20.3.9}$$

$$\hat{\beta}_1 = \frac{\hat{\Pi}_3}{\hat{\Pi}_1} \tag{20.3.10}$$

这些就是 ILS 估计量。注意，需求函数的参数不能这样得到（参看习题 20.13）。

769 为了给出一些数值结果，我们取得表 20.1 的数据。首先我们将价格和数量分别对人均实际消费支出求回归，结果如下：

$$\begin{aligned} \hat{P}_t &= 72.309\,1 + 0.004\,3X_t \\ \text{se} &= (9.200\,2) \quad (0.000\,9) \\ t &= (7.859\,5) \quad (4.410\,4) \qquad R^2 = 0.493\,0 \end{aligned} \tag{20.3.11}$$

$$\begin{aligned} \hat{Q}_t &= 84.070\,2 + 0.002\,0X_t \\ \text{se} &= (4.896\,0) \quad (0.000\,5) \\ t &= (17.171\,1) \quad (3.783\,9) \qquad R^2 = 0.417\,2 \end{aligned} \tag{20.3.12}$$

利用（20.3.9）和（20.3.10）我们得到这些 ILS 估计值：

$$\hat{\beta}_0 = 51.056\,2 \tag{20.3.13}$$

$$\hat{\beta}_1 = 0.456\,6 \tag{20.3.14}$$

表 20.1　1970—1991 年美国作物产量，作物价格与人均个人消费支出；以 1982 年美元计

年份	作物生产指数（1977 年 = 100）Q	农民收到的作物价格指数（1977 年 = 100）P	人均实际个人消费支出 X
1970	77	52	3 152
1971	86	56	3 372
1972	87	60	3 658

1973	92	91	4 002
1974	84	117	4 337
1975	93	105	4 745
1976	92	102	5 241
1977	100	100	5 772
1978	102	105	6 384
1979	113	116	7 035
1980	101	125	7 677
1981	117	134	8 375
1982	117	121	8 868
1983	88	128	9 634
1984	111	138	10 408
1985	118	120	11 184
1986	109	107	11 843
1987	108	106	12 568
1988	92	126	13 448
1989	107	134	14 241
1990	114	127	14 996
1991	111	130	15 384

资料来源：*Economic Report of the President*，1993. Q 数据见表 B-94，P 数据见表 B-96，X 数据见表 B-5。

770 因此，所估计的 ILS 回归是[10]：

$$\hat{Q}_t = 51.056\,2 + 0.456\,6P_t \tag{20.3.15}$$

为了比较，我们给出 Q 对 P 的（不适当地应用的）OLS 回归结果：

$$\begin{aligned} \hat{Q}_t &= 65.171\,9 + 0.327\,2P_t \\ \text{se} &= (9.329\,4)\quad (0.083\,5) \\ t &= (6.985\,6)\quad (3.920\,3) \qquad R^2 = 0.434\,5 \end{aligned} \tag{20.3.16}$$

这些结果表明了当 OLS 被不适当地应用时，“真”象被歪曲的情况。

ILS 估计量的性质

我们曾看到，诱导型系数的估计量是一致性的，并且在适当的假定下还是最优无偏的或渐近有效的（见章末 9）。这些性质会传递到 ILS 估计量上吗？可以证明，ILS 估计量继承了诱导型估计量的全部渐近性质，诸如一致性和渐近有效性。但像无偏性这样的（小样本）性质一般地说不再成立。在附录 20A，第 20A.1 节中，我们证明了上述供给函数的 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是有偏误

的，但这一偏误将随样本含量无限地增大而消失（就是说，估计量是一致性的）。[11]

§20.4 过度识别方程的估计：二阶最小二乘法

考虑如下模型：

$$\text{收入函数：}\quad Y_{1t}=\beta_{10}+\beta_{11}Y_{2t}+\gamma_{11}X_{1t}+\gamma_{12}X_{2t}+u_{1t}\tag{20.4.1}$$

$$\text{货币供给函数：}\quad Y_{2t}=\beta_{20}+\beta_{21}Y_{1t}+u_{2t}\tag{20.4.2}$$

771 其中 Y_1 = 收入

Y_2 = 货币存量

X_1 = 投资支出

X_2 = 政府对商品和服务的支出

变量 X_1 和 X_2 是外生的。

收入方程，收入决定的数量理论与凯恩斯方法的一个混合物，说的是，收入由货币供给、投资支出和政府支出来决定。**货币供给函数**则设想货币的存量是在收入水平的基础上决定的。显然，我们有了一个联立方程的问题，这可由第 19 章中讨论的联立性检验加以查核。

应用可识别性的阶条件，可以看出收入方程是不可识别的，而货币供给方程则是过度可识别的。关于收入方程，如果不改变模型的设定，便无计可施。而过度识别的货币供给函数由于存在 β_{21} 的两个 ILS 估计值（读者应能通过诱导型系数加以证实），也不能用 ILS 去估计它。

从实际考虑，人们也许想用 OLS 去估计货币供给方程，但这样得到的估计值，由于随机解释变量 Y_1 和随机干扰项 u_2 之间可能的相关关系，将是非一致性的。然而，假令我们能找到随机解释变量 Y_1 的这样一个"代理变量"，它和 Y_1 相像（意思是它和 Y_1 高度相关），而又与 u_2 不相关。这样的代理变量又称**工具变量**（见第 17 章）。如果能找到这样的一个代理变量，OLS 就可直接用于估计货币供给函数。但怎样能得到这样的一个工具变量呢？由 H. 泰尔[12]和 R. 巴斯曼（Robert Basmann）[13]各自独立发现的**二阶最小二乘**（2SLS）法给出了一个答案。此法顾名思义，涉及 OLS 的两次相继的应用。过程如下：

阶段 1. 为摆脱 Y_1 和 u_1 之间的可能的相关性。先求 Y_1 对整个方程组中（不仅仅是所考虑的方程中）的全部前定变量的回归。在本例中，这意味着求 Y_1 对 X_1 和 X_2 的回归如下：

$$Y_{1t}=\hat{\Pi}_0+\hat{\Pi}_1X_{1t}+\hat{\Pi}_2X_{2t}+\hat{u}_t\tag{20.4.3}$$

其中 $\hat{u}_t$ 是平常的 OLS 残差。由方程（20.4.3）我们得到：

$$\hat{Y}_{1t}=\hat{\Pi}_0+\hat{\Pi}_1 X_{1t}+\hat{\Pi}_2 X_{2t} \tag{20.4.4}$$

772 其中 $\hat{Y}_{1t}$是以固定 X 值为条件的 Y 均值的一个估计值。注意，(20.4.3) 的右边仅出现外生或前定变量，所以它不外是一个诱导型回归而已。

现在，方程 (20.4.3) 可表达为：

$$Y_{1t}=\hat{Y}_{1t}+\hat{u}_t \tag{20.4.5}$$

表明随机的 Y_1 由两部分构成：作为非随机 X 的一个线性组合的 $\hat{Y}_{1t}$和随机成分 $\hat{u}_t$。按照 OLS 理论，$\hat{Y}_{1t}$和 $\hat{u}_t$ 是不相关的。(为什么?)

阶段 2. 现在，过度识别的货币供给方程可写为：

$$\begin{aligned} Y_{2t} &= \beta_{20}+\beta_{21}(\hat{Y}_{1t}+\hat{u}_t)+u_{2t} \\ &= \beta_{20}+\beta_{21}\hat{Y}_{1t}+(u_{2t}+\beta_{21}\hat{u}_t) \\ &= \beta_{20}+\beta_{21}\hat{Y}_{1t}+u_t^* \end{aligned} \tag{20.4.6}$$

其中 $u_t^*=u_{2t}+\beta_{21}\hat{u}_t$。

比较 (20.4.6) 和 (20.4.2)，我们看到它们外表上非常相似，惟一的差别是 Y_1 被 $\hat{Y}_1$ 所代替。(20.4.6) 有什么好处? 可以证明，虽然原始货币供给方程中的 Y_1 和干扰项 u_2 是或很可能是相关的 (从而使 OLS 不适用)，但 (20.4.6) 中的 $\hat{Y}_1$ 则渐近地即在大样本中 (或说得更准确，随着样本含量无限地增大) 与 u_t^* 不相关。这样一来，OLS 可应用于 (20.4.6)，以得出货币供给函数的参数的一致性估计。[14]

这个二阶程序表明，2SLS 的基本思想是，从随机解释变量 Y_1 中把随机干扰 u_2 的影响"清除"掉，即通过求 Y_1 对方程组中全部前定变量的诱导型回归 (阶段 1)，得到估计值 $\hat{Y}_{1t}$，再用 $\hat{Y}_{1t}$代替原方程中的 Y_{1t}，然后对这样变换的方程应用 OLS (阶段 2)，就达到了这一目的。这样得到的估计量是一致性的，即随着样本的无限增大，这些估计量收敛于其真值。

773 为了进一步说明 2SLS，现将收入—货币供给模型修改如下：

$$Y_{1t}=\beta_{10}+\beta_{12}Y_{2t}+\gamma_{11}X_{1t}+\gamma_{12}X_{2t}+u_{1t} \tag{20.4.7}$$

$$Y_{2t}=\beta_{20}+\beta_{21}Y_{1t}+\gamma_{23}X_{3t}+\gamma_{24}X_{4t}+u_{2t} \tag{20.4.8}$$

其中，除已定义的变量外，X_3 = 前一时期的收入，而 X_4 = 前一时期的货币供给。X_3 和 X_4 都是前定的。

容易验证，方程 (20.4.7) 和 (20.4.8) 都是过度可识别的。为了应用 2SLS，我们进行如下：在阶段 1 中，我们求内生变量对方程组中**全部**前定变量的回归，即：

$$Y_{1t}=\hat{\Pi}_{10}+\hat{\Pi}_{11}X_{1t}+\hat{\Pi}_{12}X_{2t}+\hat{\Pi}_{13}X_{3t}+\hat{\Pi}_{14}X_{4t}+\hat{u}_{1t} \tag{20.4.9}$$

$$Y_{2t}=\hat{\Pi}_{20}+\hat{\Pi}_{21}X_{1t}+\hat{\Pi}_{22}X_{2t}+\hat{\Pi}_{23}X_{3t}+\hat{\Pi}_{24}X_{4t}+\hat{u}_{2t} \tag{20.4.10}$$

在阶段 2 中，将原始 (结构) 方程中的 Y_1 和 Y_2 代以它们从上述二回归中得到的估计值，然后做如下的 OLS 回归：

$$Y_{1t}=\beta_{10}+\beta_{12}\hat{Y}_{2t}+\gamma_{11}X_{1t}+\gamma_{12}X_{2t}+u_{1t}^* \tag{20.4.11}$$

$$Y_{2t}=\beta_{20}+\beta_{21}\hat{Y}_{1t}+\gamma_{23}X_{3t}+\gamma_{24}X_{4t}+u_{2t}^* \tag{20.4.12}$$

其中 $u_{1t}^* = u_{1t} + \beta_{12}\hat{u}_{2t}$ 和 $u_{2t}^* = u_{2t} + \beta_{21}\hat{u}_{1t}$。这样得到的估计将是一致性的。

注意 2SLS 的如下特点：

1. 它可以应用于方程组中的某一个别的方程而无须考虑方程组中的任何其余方程。因此，在求解涉及大量方程的计量经济模型时，2SLS 提供了一个节约的方法。由于这一原因，此法在实际中被广泛应用。

2. 与 ILS 相比，ILS 为过度识别的方程提供参数的多个估计值，而 2SLS 对每个参数只提供一个。

3. 它只需知道方程组中一共有多少个外生或前定变量而无须知道方程组中的任何其他变量，故易于应用。

4. 此法虽然专为过度识别的方程而设计，但同样适用于恰好识别的方程。但这时 ILS 和 2SLS 将给出相同的估计。(为什么?)

5. 如果诱导型回归（即阶段 1 的回归）的 R^2 值很高，比如说高于 0.8，
774 则经典 OLS 估计和 2SLS 估计将相差无几。但这不应有什么可奇怪的，因为如果第一阶段的 R^2 值很高，就意味着内生变量的估计值和它们的真实值非常接近，从而知道后者和原始方程中的随机干扰项有较小的相关。(为什么?)[15]然而，如果 R^2 在第一阶段回归中很低，则表明 2SLS 估计实际上是无意义的，因为我们将要在第二阶段回归中用第一阶段估计得的 $\hat{Y}$ 去代替原始的 Y，而 $\hat{Y}$ 在很大程度上代表第一阶段回归中的干扰项。换句话说，这时 $\hat{Y}$ 是原始 Y 的很糟糕的代理变量。

6. 注意，在报告（20.3.15）中的 ILS 回归时，我们没有给出所估系数的标准误（理由已在章末注 10 中陈述）。但我们能对 2SLS 估计值给出这个标准误。这是因为这时结构系数是直接从第二阶段（OLS）回归估计出来的。但要当心一个问题，如我们从方程（20.4.6）看到的，误差项 u_t^* 事实上是原始误差项 u_{2t} 加上 $\beta_{21}\hat{u}_t$，因此第二阶段回归中所估计的标准误还需要修改。即 u_t^* 的方差还不正好等于原始 u_{2t} 的方差。然而，所需要的修改是容易通过附录 20A，第 20A.2 节所给的公式加以实现的。

7. 在使用 2SLS 时，应注意 H. 泰尔的以下评语：

> 2SLS 的统计合理性是属于大样本一类的。在没有滞后内生变量情况下……如果外生变量在重复样本中保持不变，并且［出现在各个行为或结构方程中的］（诸）干扰项是有零均值和有限方差的独立且同分布变量，则 2SLS 系数估计量是一致性的……如果上述两条件得到满足，则 2SLS 系数估计量的抽样分布对大样本来说将是近似于正态的，……
>
> 当方程组含有滞后内生变量时，2SLS 系数估计量的一致性和大样本正态性还需要有另一条件……随着样本的增大，每一滞后内生变量取值的均方差在概率上收敛到一个正的极限……
>
> 如果出现在各个结构方程中的干扰项不是独立地分布的，则诸滞后内生变量就不独立于方程组的当前运作……这意味着，这些变量并非真正前定。如果在 2SLS 的实施过程中仍把这些变量看作前定的，结果得到的估计量就不是一致性的了。[16]

§20.5 2SLS：一个数值例子

775 为了说明 2SLS 法，考虑前面由方程（20.4.1）和（20.4.2）给出的收入—货币供给模型。如已表明的，货币供给方程是过度可识别的。为了估计该方程的参数，我们求助于二阶最小二乘法。分析中所需要的数据见于表 20.2。

表中还给出为了回答习题中的一些问题所必需的数据。

阶段 1 回归。我们首先求代表 GDP 的随机解释变量：收入 Y_1 对前定变量：私人投资 X_1 和政府支出 X_2 的回归，得到如下结果：

776
$$\begin{aligned}\hat{Y}_{1t} &= 2\,587.351 + 1.670\,7X_{1t} + 1.969\,3X_{2t} \\ \text{se} &= \quad (72.001\,1) \quad (0.164\,6) \quad (0.098\,3) \qquad (20.5.1) \\ t &= \quad (35.934\,9) \quad (10.148\,9) \quad (20.020\,0) \qquad R^2 = 0.994\,7\end{aligned}$$

表 20.2 1970—1991 年美国 GDP，M2 货币供给，私人国内总投资，6 月期国库券利率数据

年份	GDP (Y_1)	M2 (Y_2)	GPDI (X_1)	FEDEXP (X_2)	TB6 (X_3)
1970	3 578.000	626.400 0	436.200 0	198.600 0	6.562 000
1971	3 697.700	710.100 0	485.800 0	216.600 0	4.511 000
1972	3 998.400	802.100 0	543.000 0	240.000 0	4.466 000
1973	4 123.400	855.200 0	606.500 0	259.700 0	7.178 000
1974	4 099.000	901.900 0	561.700 0	291.200 0	7.926 000
1975	4 084.400	1 015.900	462.200 0	345.400 0	6.122 000
1976	4 311.700	1 151.700	555.500 0	371.900 0	5.266 000
1977	4 511.800	1 269.900	639.400 0	405.000 0	5.510 000
1978	4 760.600	1 365.500	713.000 0	444.200 0	7.572 000
1979	4 912.100	1 473.100	735.400 0	489.600 0	10.017 00
1980	4 900.900	1 599.100	655.300 0	576.600 0	11.374 00
1981	5 021.000	1 754.600	715.600 0	659.300 0	13.776 00
1982	4 913.300	1 909.500	615.200 0	732.100 0	11.084 00
1983	5 132.300	2 126.000	673.700 0	797.800 0	8.750 000
1984	5 505.200	2 309.700	871.500 0	856.100 0	9.800 000
1985	5 717.100	2 495.400	863.400 0	924.600 0	7.660 000
1986	5 912.400	2 732.100	857.700 0	978.500 0	6.030 000
1987	6 113.300	2 831.100	879.300 0	1 018.400	6.050 000
1988	6 368.400	2 994.300	902.800 0	1 066.200	6.920 000

1989	6 591.900	3 158.400	936.500 0	1 140.300	8.040 000
1990	6 707.900	3 277.600	907.300 0	1 228.700	7.470 000
1991	6 676.400	3 376.800	829.500 0	1 287.600	5.490 000
1992	6 880.000	3 430.700	899.800 0	1 418.900	3.570 000
1993	7 062.600	3 484.400	977.900 0	1 471.500	3.140 000
1994	7 347.700	3 499.000	1 107.000	1 506.000	4.660 000
1995	7 543.800	3 641.900	1 140.600	1 575.700	5.590 000
1996	7 813.300	3 813.300	1 242.700	1 635.900	5.090 000
1997	8 159.500	4 028.900	1 393.300	1 678.800	5.180 000
1998	8 515.700	4 380.600	1 566.800	1 705.000	4.850 000
1999	8 875.800	4 643.700	1 699.700	1 750.200	4.760 000

注：Y_1 = GDP = 国内生产总值（10 亿美元，经季节性调整）
Y_2 = M2 = M2 货币供给（10 亿美元，经季节性调整）
X_1 = GPDI = 私人国内总投资（10 亿美元，经季节性调整）
X_2 = FEDEXP = 联邦国内总投资（10 亿美元，经季节性调整）
X_3 = TB6 = 6 个月国库券利率（%）
资料来源：*Economic Report of the President*, 2001, Tables B-2, B-69, B-73, B-84.

阶段 2 回归。我们现在估计货币供给函数（20.4.2），用得自（20.5.1）的 Y_1 的估计值（$=\hat{Y}_1$）代替内生的 Y_1。结果如下：

$$\begin{aligned}\hat{Y}_{2t} &= -2\,198.297 \quad + \quad 0.791\,6\hat{Y}_{1t} \\ \text{se} &= \quad (139.098\,6) \quad (0.023\,2) \\ t &= \quad (-15.803\,8) \quad (34.050\,2) \qquad R^2 = 0.976\,4\end{aligned} \tag{20.5.2}$$

我们曾指出，所估计的（20.5.2）中的标准误需要按照附录 20A，第 20A.2 节所建议的方法加以校正。经过这一校正（大多数标准计量经济程序包都按例做到这点），我们得到如下结果：

$$\begin{aligned}\hat{Y}_{2t} &= -2\,198.297 \quad + \quad 0.791\,5\hat{Y}_{1t} \\ \text{se} &= \quad (126.959\,8) \quad (0.021\,2) \\ t &= \quad (-17.314\,9) \quad (37.305\,7) \qquad R^2 = 0.980\,3\end{aligned} \tag{20.5.3}$$

如在附录 20A 第 20A.2 中所指出的，由于阶段 1 回归中的 R^2 很高，（20.5.3）所给的标准误和（20.5.2）所给的相差无几。

OLS 回归。为了比较，我们给出由（20.4.2）表示的货币存量对收入的回归，而不去"清除掉"随机变量 Y_{1t} 中受干扰项影响的成分：

$$\begin{aligned}\hat{Y}_{2t} &= -2\,195.468 \quad + \quad 0.791\,1Y_{1t} \\ \text{se} &= \quad (126.646\,0) \quad (0.021\,1) \\ t &= \quad (-17.335\,4) \quad (37.381\,2) \qquad R^2 = 0.980\,3\end{aligned} \tag{20.5.4}$$

777 将此"不适当的"OLS 结果同阶段 2 回归比较，我们看到这两个回归基本上是一样的。这是否说 2SLS 程序是不值得做的呢？完全不这样。在本例中，如前所说，由于在第一阶段中的 R^2 值很高，致使估计值 $\hat{Y}_{1t}$ 和实测值

Y_{1t}基本相同，因而两个结果也基本一致，这是没有什么可奇怪的。因此，在本例中，OLS 和第二阶段回归将是多少有些相似的。但这并不保证在每一应用中情况皆如此。由此引出一个含义：在过度可识别的方程中，我们不可以不经过第二阶段回归的核对就接受经典 OLS 程序。

GDP 和货币供给的联立性。让我们看看 GDP（Y_1）和货币供给（Y_2）是否相互依赖。为此，我们应用第 19 章中所讨论的豪斯曼联立性检验。

首先，我们求 GDP 对方程组中的外生变量 X_1（投资支出）和 X_2（政府支出）的回归（也就是，我们估计诱导型回归）。由此回归我们得到估计的 GDP 和残差 $\hat{v}_t$，如方程（19.4.7）所示。然后求货币供给对所估 GDP 和 $\hat{v}_t$ 的回归，结果如下：

$$
\begin{aligned}
\hat{Y}_{2t} &= -2\,198.297 \quad + \quad 0.791\,5\hat{Y}_{1t} + 0.698\,4\hat{v}_t \\
\text{se} &= \quad (129.054\,8) + (0.021\,5) \quad (0.297\,0) \qquad\qquad (20.5.5)\\
t &= \quad (-17.033\,8)\ (36.700\,16) \quad (2.351\,1)
\end{aligned}
$$

由于 $\hat{v}_t$ 的 t 值是统计上显著的（p 值是 0.026 3），我们不能拒绝货币供给和 GDP 之间的联立性假设，这是没有什么可奇怪的。（注：严格地说。这一结论仅在大样本中，或用专门术语说，随着样本无限地增大时，才是有效的。）

假设检验。假如我们要检验收入对货币需求无影响这一假设。能不能用通常从所估回归（20.5.2）得到的 t 比率来检验此假设呢？是的，如果样本是大样本并且我们对（20.5.3）中的标准误做了如同在（20.5.3）中那样的校正，我们就能用 t 检验去检验每一个别系数的显著性，和用公式（8.5.7）中的 F 检验去检验两个或多个系数的联合显著性。[17]

如果一个结构方程中的误差项是自相关的，或者它同方程组中的另一结构方程中的误差项有相关关系，又会出现什么情况？要对此问题做出全面答复，将超出本书的讨论范围，还是留给读者去阅读参考文献为好。（参看章
778 末注 7 中的参考文献。）不管怎样，用以处理这种复杂性的估计技术［如泽尔纳（Zellner）的 SURE 技术］是存在的。

§20.6 说明性例子

本节中我们讨论联立方程方法的一些应用。

例 20.1：广告费、集中度与价差

为了研究广告费、集中度（由集中比率来衡量）和价格—成本差距之间的相互关系，A.D. 斯特里克兰（Strickland）和 L.W. 韦斯（Weiss）构造了如下的三方程模型。[18]

广告深度函数：

$$\text{Ad}/S = a_0 + a_1 M + a_2(\text{CD}/S) + a_3 C + a_4 C^2 + a_5 \text{Gr} + a_6 \text{Dur} \qquad (20.6.1)$$

集中度函数：

$$C = b_0 + b_1(\mathrm{Ad}/S) + b_2(\mathrm{MES}/S) \tag{20.6.2}$$

价格—成本差距函数：

$$M = c_0 + c_1(K/S) + c_2\mathrm{Gr} + c_3 C + c_4\mathrm{GD} + c_5(\mathrm{Ad}/S) + c_6(\mathrm{MES}/S) \tag{20.6.3}$$

其中 Ad＝广告费用

S＝发货价值（value of shipments）

C＝四厂集中比率

CD＝消费者需求

MES＝最小有效规模

M＝价格/成本差距

Gr＝工业生产的年增长率

Dur＝耐用品工业虚拟变量

K＝资本存量

GD＝产品的地区分散性度量

由识别的阶条件知，方程（20.6.2）是过度可识别的，而（20.6.1）和（20.6.3）是恰好可识别的。

用于分析的数据，基本上来自1963年制造业者普查资料，包括了417个4位数字制造工业中的408个。作者们先用OLS估计这三个方程，得到了表20.3中的结果。为校正联立方程偏误起见，再用2SLS重新估计模型，随之又得到表20.4中的结果。我们把它留给读者加以比较。

779

表20.3　三方程的OLS估计（括号中为 t 比率）

	因变量		
	Ad/S Eq.(20.6.1)	C Eq.(20.6.2)	M Eq.(20.6.3)
常数	−0.031 4(−7.45)	0.263 8(25.93)	0.168 2(17.15)
C	0.055 4(3.56)	—	0.062 9(2.89)
C^2	−0.056 8(−3.38)	—	—
M	0.112 3(9.84)	—	—
CD/S	0.025 7(8.94)	—	—
Gr	0.038 7(1.64)	—	0.225 5(2.61)
Dur	−0.002 1(−1.11)	—	—
Ad/S	—	1.161 3(3.3)	1.653 6(11.00)
MES/S	—	4.185 2(18.99)	0.068 6(0.54)
K/S	—	—	0.112 3(8.03)
GD	—	—	−0.000 3(−2.90)
R^2	0.374	0.485	0.402
df	401	405	401

表 20.4　　三方程的二阶最小二乘估计(括号中为 t 比率)

	因变量		
	Ad/S	C	M
	Eq.(20.6.1)	Eq.(20.6.2)	Eq.(20.6.3)
常数	−0.024 5(−3.86)	0.259 1(21.30)	0.173 6(14.66)
C	0.073 7(2.84)	—	0.037 7(0.93)
C^2	−0.064 3(−2.64)	—	—
M	0.054 4(2.01)	—	—
CD/S	0.026 9(8.96)	—	—
Gr	0.053 9(2.09)	—	0.233 6(2.61)
Dur	−0.001 8(−0.93)	—	—
Ad/S	—	1.534 7(2.42)	1.625 6(5.52)
MES/S	—	4.169(18.84)	0.172 0(0.92)
K/S	—		0.116 5(7.30)
GD	—	—	−0.000 3(−2.79)

例 20.2：克莱因的模型Ⅰ

在例 18.6 中，我们曾扼要地讨论过克莱因的开创性模型。最初，此模型曾对 1920—1941 年时期进行估计。所依据的数据见于表 20.5；其 OLS，诱导型及 2SLS 诸估计均列于表 20.6 中。我们留给读者去解释这些结果。

780

表 20.5　　克莱因模型Ⅰ所依据的数据

年份	C^*	P	W	I	K_{-1}	X	W'	G	T
1920	39.8	12.7	28.8	2.7	180.1	44.9	2.2	2.4	3.4
1921	41.9	12.4	25.5	−0.2	182.8	45.6	2.7	3.9	7.7
1922	45.0	16.9	29.3	1.9	182.6	50.1	2.9	3.2	3.9
1923	49.2	18.4	34.1	5.2	184.5	57.2	2.9	2.8	4.7
1924	50.6	19.4	33.9	3.0	189.7	57.1	3.1	3.5	3.8
1925	52.6	20.1	35.4	5.1	192.7	61.0	3.2	3.3	5.5
1926	55.1	19.6	37.4	5.6	197.8	64.0	3.3	3.3	7.0
1927	56.2	19.8	37.9	4.2	203.4	64.4	3.6	4.0	6.7
1928	57.3	21.1	39.2	3.0	207.6	64.5	3.7	4.2	4.2
1929	57.8	21.7	41.3	5.1	210.6	67.0	4.0	4.1	4.0
1930	55.0	15.6	37.9	1.0	215.7	61.2	4.2	5.2	7.7

1931	50.9	11.4	34.5	-3.4	216.7	53.4	4.8	5.9	7.5
1932	45.6	7.0	29.0	-6.2	213.3	44.3	5.3	4.9	8.3
1933	46.5	11.2	28.5	-5.1	207.1	45.1	5.6	3.7	5.4
1934	48.7	12.3	30.6	-3.0	202.0	49.7	6.0	4.0	6.8
1935	51.3	14.0	33.2	-1.3	199.0	54.4	6.1	4.4	7.2
1936	57.7	17.6	36.8	2.1	197.7	62.7	7.4	2.9	8.3
1937	58.7	17.3	41.0	2.0	199.8	65.0	6.7	4.3	6.7
1938	57.5	15.3	38.2	-1.9	201.8	60.9	7.7	5.3	7.4
1939	61.6	19.0	41.6	1.3	199.9	69.5	7.8	6.6	8.9
1940	65.0	21.1	45.0	3.3	201.2	75.7	8.0	7.4	9.6
1941	69.7	23.5	53.3	4.9	204.5	88.4	8.5	13.8	11.6

* 各列标题的含义已在例 18.6 中列出。

资料来源：这些数据取自 G.S. Maddala, *Econometrics*, McGraw-Hill, New York, 1977, p.238。

表 20.6*　　克莱因模型 I 的 OLS 诱导型和 2SLS 估计

OLS:

$$C = 16.237 + 0.193P + 0.796(W + W') + 0.089P_{-1} \quad \bar{R}^2 = 0.978 \quad \mathrm{DW} = 1.367$$
$$(1.203)\quad(0.091)\quad(0.040)\qquad(0.090)$$

$$I = 10.125 + 0.479P + 0.333P_{-1} - 0.112K_{-1} \quad \bar{R}^2 = 0.919 \quad \mathrm{DW} = 1.810$$
$$(5.465)\quad(0.097)\quad(0.100)\quad(0.026)$$

$$W = 0.064 + 0.439X + 0.146X_{-1} + 0.130t \quad \bar{R}^2 = 0.985 \quad \mathrm{DW} = 1.958$$
$$(1.151)(0.032)\quad(0.037)\quad(0.031)$$

诱导型：

$$P = 46.383 + 0.813P_{-1} - 0.213K_{-1} + 0.015X_{-1} + 0.297t - 0.926T + 0.443G$$
$$(10.870)(0.444)\quad(0.067)\quad(0.252)\quad(0.154)\quad(0.385)\quad(0.373)$$
$$\bar{R}^2 = 0.753 \quad \mathrm{DW} = 1.854$$

$$W + W' = 40.278 + 0.823P_{-1} - 0.144K_{-1} + 0.115X_{-1} + 0.881t - 0.567T + 0.859G$$
$$(8.787)\quad(0.359)\quad(0.054)\quad(0.204)\quad(0.124)\quad(0.311)\quad(0.302)$$
$$\bar{R}^2 = 0.949 \quad \mathrm{DW} = 2.395$$

$$X = 78.281 + 1.724P_{-1} - 0.319K_{-1} + 0.094X_{-1} + 0.878t - 0.565T + 1.317G$$
$$(18.860)(0.771)\quad(0.110)\quad(0.438)\quad(0.267)\quad(0.669)\quad(0.648)$$
$$\bar{R}^2 = 0.882 \quad \mathrm{DW} = 2.049$$

2SLS:

$$C = 16.543 + 0.019P + 0.810(W + W') + 0.214P_{-1} \quad \bar{R}^2 = 0.972\,6$$
$$(1.464)\quad(0.130)\quad(0.044)\qquad(0.118)$$

$$I = 20.284 + 0.149P + 0.616P_{-1} - 0.157K_{-1} \qquad \bar{R}^2 = 0.864\,3$$
$$(8.361)\ (0.191)\ (0.180)\quad (0.040)$$
$$W = 0.065 + 0.438X + 0.146X_{-1} + 0.130t \qquad R^2 = 0.985\,2$$
$$(1.894)(0.065)\ (0.070)\quad (0.053)$$

* 变量的含义已在习题18.6中列出（括号内为标准误）。

资料来源：G. S. Maddala, *Econometrics*, McGraw-Hill, New York, 1977, p. 242.

例 20.3：　表达为递归方程组的资本资产定价模型

C.F. 李（Lee）和 W.P. 劳埃德（Lloyd）[19]在一项颇不平凡的递归联立方程建模的应用中，估计了如下的石油工业模型：

$$\begin{aligned}
R_{1t} &= \alpha_1 && + \gamma_1 M_t + u_{1t}\\
R_{2t} &= \alpha_2 + \beta_{21}R_{1t} && + \gamma_2 M_t + u_{2t}\\
R_{3t} &= \alpha_3 + \beta_{31}R_{1t} + \beta_{32}R_{2t} && + \gamma_3 M_t + u_{3t}\\
R_{4t} &= \alpha_4 + \beta_{41}R_{1t} + \beta_{42}R_{2t} + \beta_{43}R_{3t} && + \gamma_4 M_t + u_{4t}\\
R_{5t} &= \alpha_5 + \beta_{51}R_{1t} + \beta_{52}R_{2t} + \beta_{53}R_{3t} + \beta_{54}R_{4t} && + \gamma_5 M_t + u_{5t}\\
R_{6t} &= \alpha_6 + \beta_{61}R_{1t} + \beta_{62}R_{2t} + \beta_{63}R_{3t} + \beta_{64}R_{4t} + \beta_{65}R_{5t} && + \gamma_6 M_t + u_{6t}\\
R_{7t} &= \alpha_7 + \beta_{71}R_{1t} + \beta_{72}R_{2t} + \beta_{73}R_{3t} + \beta_{74}R_{4t} + \beta_{75}R_{5t} + \beta_{76}R_{6t} && + \gamma_7 M_t + u_{7t}
\end{aligned}$$

其中　R_1 = 证券1（=“帝国”石油）的回报率

R_2 = 证券2（=“太阳”石油）的回报率

781 ⋮

R_7 = 证券7（=“印第安纳标准”）

M_t = 市场回报率指数

u_{it} = 干扰项（$i = 1, 2, \cdots, 7$）

在我们介绍这个结果之前，一个显然的问题是：我们怎样选择哪一种证券作为证券1，哪一种作为证券2，等等？李和劳埃德纯粹地凭经验回答了这个问题。他们求证券 i 的回报率对其余6种证券的回报率的回归并观察所得的 R^2。于是有7个这样的回归。然后将所估计的 R^2 从低到高排队。把有最低 R^2 的证券算做证券1，而有最高 R^2 的证券算做证券7。这样做的思想背景，明显地简单。比方说，如果“帝国”石油的回报率对其余6种证券来说，R^2 是最低的，这就告知我们，这一证券受其他证券回报率变动的影响最小。因此，如果有因果顺序的话，它就是从这一证券奔向其他证券的，而没有从其他证券反馈回来的作用。

虽然我们可以对这种因果顺序的纯经验方法提出异议，仍不妨把他们的经验结果列出来，如表20.7所示。

782

表 20.7　　石油产业的递归方程组估计

	线性型因变量						
	印第安纳标准	壳牌石油	菲利普斯石油	联合石油	俄亥俄标准	太阳石油	帝国石油
印第安纳标准							
壳牌石油	0.210 0*						
	(2.859)						
菲利普斯石油	0.229 3*	0.079 1					
	(2.176)	(1.065)					
联合石油	0.175 4*	0.217 1*	0.222 5*				
	(2.472)	(3.177)	(2.337)				
俄亥俄标准	−0.079 4	0.014 7	0.424 8*	0.146 8*			
	(−1.294)	(0.235)	(5.501)	(1.735)			
太阳石油	0.124 9	0.171 0*	0.047 2	0.133 9	0.049 9		
	(1.343)	(1.843)	(0.355)	(0.908)	(0.271)		
帝国石油	−0.107 7	0.052 6	0.035 4	0.158 0	−0.254 1*	0.082 8	
	(−1.412)	(0.680 4)	(0.319)	(1.290)	(−1.691)	(0.971)	
常数	0.086 8	−0.038 4	−0.012 7	−0.203 4	0.300 9	0.201 3	0.371 0*
	(0.681)	(1.296)	(−0.068)	(0.986)	1.204	(1.399)	(2.161)
市场指数	0.368 1*	0.499 7*	0.288 4	0.760 9*	0.908 9*	0.716 1*	0.643 2*
	(2.165)	(3.039)	(1.232)	(3.069)	(3.094)	(4.783)	(3.774)
$\bar{R}^2$	0.502 0	0.465 8	0.410 6	0.253 2	0.098 5	0.240 4	0.124 7
德宾-沃森	2.108 3	2.471 4	2.230 6	2.346 8	2.218 1	2.310 9	1.959 2

* 指在双侧检验中的 0.10 或更好的水平上显著。

注：系数下方括号内为 t 值。

资料来源：Cheng F. Lee and W. P. Lloyd，前引文献，Table 3b。

在习题 5.5 中，我们曾介绍现代投资理论的特征线，它不外是证券 i 的回报率对市场回报率的回归。以 β 系数为名的斜率系数度量着该证券回报的波动性。李-劳埃德回归结果所表明的是，除了由市场组合证券代表的共同的市场影响外，还有证券与证券之间的回报率的显著产业内（intra-industry）关系需要考虑。例如“印第安纳标准”的回报率不仅依赖于市场回报率，还依赖于“壳牌石油”、“菲利普斯石油”和“联合石油”的回报率。换句话说，如果除了市场回报率外，我们还考虑“壳牌石油”、“菲利普斯石油”和“联合石油”所经历的回报率的话，那么“印第安纳标准”的回报率的变动就会得到更好的解释。

例 20.4：圣路易斯模型的修订版本[20]

著名的、常有争议的圣路易斯模型原先于20世纪60年代后期问世，其后曾经多次修改。其中的一个修订版本有如表20.8所示。根据该修订版得到的经验结果见于表20.9。（注：变量上方的圆点表示该变量的增长率。）
788 模型基本上由表20.8中的方程（1）、（2）、（4）和（5）构成，其他方程代表各种定义。方程（1）曾由OLS估计，方程（1）、（2）和（4）均用带（端点）约束的阿尔蒙分布滞后方法估计。在认为合适时，方程还对一阶（ρ_1）和/或二阶（ρ_2）序列相关做了校正。

从结果的分析看，我们注意到主要决定（名义）GNP增长率的因素是货币供给的增长率，而不是为高就业而支出的增长率。M 系数的总和是1.06，表示随着货币供给的1%的（持续）增加平均而言导致名义GNP的约1.06%的增加。另一方面，E 系数的总和约为0.05，表明政府高就业支出的变化对名义GNP没有什么影响。表20.9中报道的其他回归结果，留给读者去解释。

表 20.8　　圣路易斯模型

(1) $$\dot{Y}_t = C1 + \sum_{i=0}^{4} CM_i(\dot{M}_{t-i}) + \sum_{i=0}^{4} CE(\dot{E}_{t-i}) + \epsilon 1_t$$

(2) $$\dot{P}_t = C2 + \sum_{i=1}^{4} \mathrm{CPE}_i(\dot{\mathrm{PE}}_{t-i}) + \sum_{i=0}^{5} CD_i(\dot{X}_{t-i} - \dot{XF}^*_{t-i1}) + \mathrm{CPA}(\dot{\mathrm{PA}}_t) + \mathrm{CDUM1(DUM1)} + \mathrm{CDUM2(DUM2)} + \epsilon 2_t$$

(3) $$\dot{\mathrm{PA}}_t = \sum_{i=1}^{21} C\mathrm{PRL}_i(\dot{P}_{t-i})$$

(4) $$\mathrm{RL}_t = C3 + \sum_{i=0}^{20} \mathrm{CPRL}_i(\dot{P}_{t-i}) + \epsilon 3_t$$

(5) $$U_t - \mathrm{UF}_t = CG(\mathrm{GAP}_t) + CG1(\mathrm{GAP}_{t-1}) + \epsilon 4_t$$

(6) $$Y_t = (P_t/100)(X_t)$$

(7) $$\dot{Y}_t = [(Y_t/Y_{t-i})^4 - 1]100$$

(8) $$\dot{X}_t = [(X_t/X_{t-i})^4 - 1]100$$

(9) $$\dot{P}_t = [(P_t/P_{t-i})^4 - 1]100$$

(10) $$\mathrm{GAP}_t = [(\mathrm{XF}_t/X_t)/\mathrm{XF}_t]100$$

(11) $$\dot{\mathrm{XF}}^*_t = [(\mathrm{XF}_t/X_{t-1})^4 - 1]100$$

Y = 名义GNP

M = 货币存量（M1）

E = 高就业支出

P = GNP平缩因子（1972年 = 100）

PE = 能源相对价格

X = 1972年美元产出

XF = 潜在产出（由Rasche/Tatom定义）

RL = 公司债券利率

U = 失业率

UF = 充分就业时的失业率

DUM1 = 虚拟控制期变量(1971 年第 3 季度至 1973 年第 1 季度 = 1;其他时期为零)

DUM2 = 虚拟控制期后变量(1973 年第 2 季度至 1975 年第 1 季度 = 1;其他时期为零)

资料来源:Federal Reserve Bank of St. Loius, *Review*, May 1982, p. 14.

784

样本期内估计:1960 年第 1 季度至 1980 年第 4 季度

表 20.9 **(括号内为 t 统计量的绝对值)**

(1)
$$\dot{Y}_t = \underset{(2.15)}{2.44} + \underset{(3.38)}{0.40\dot{M}_t} + \underset{(5.06)}{0.39\dot{M}_{t-1}} + \underset{(2.18)}{0.22\dot{M}_{t-2}} + \underset{(0.82)}{0.06\dot{M}_{t-3}} - \underset{(0.11)}{0.01\dot{M}_{t-4}}$$
$$+ \underset{(1.46)}{0.06\dot{E}_t} + \underset{(0.63)}{0.02\dot{E}_{t-1}} - \underset{(0.57)}{0.02\dot{E}_{t-2}} - \underset{(0.52)}{0.02\dot{E}_{t-3}} + \underset{(0.34)}{0.01\dot{E}_{t-4}}$$
$$R^2 = 0.39 \quad se = 3.50 \quad DW = 2.02$$

(2)
$$\dot{P}_t = \underset{(2.53)}{0.96} + \underset{(0.75)}{0.01\dot{PE}_{t-1}} + \underset{(1.96)}{0.04\dot{PE}_{t-2}} - \underset{(0.73)}{0.01\dot{PE}_{t-3}} + \underset{(1.38)}{0.02\dot{PE}_{t-4}}$$
$$- \underset{(0.18)}{0.00(\dot{X}_t - \dot{XF}_t^*)} + \underset{(1.43)}{0.01(\dot{X}_{t-1} - \dot{XF}_{t-1}^*)} + \underset{(4.63)}{0.02(\dot{X}_{t-2} - \dot{XF}_{t-2}^*)}$$
$$+ \underset{(3.00)}{0.02(\dot{X}_{t-3} - \dot{XF}_{t-3}^*)} + \underset{(2.42)}{0.02(\dot{X}_{t-4} - \dot{XF}_{t-4}^*)} + \underset{(2.16)}{0.01(\dot{X}_{t-5} - \dot{XF}_{t-5})}$$
$$+ \underset{(10.49)}{1.03(\dot{PA}_t)} - \underset{(1.02)}{0.61(\text{DUM1}_t)} + \underset{(2.71)}{1.65(\text{DUM2}_t)}$$
$$R^2 = 0.80 \quad se = 1.28 \quad DW = 1.97 \quad \hat{\rho} = 0.12$$

(4)
$$\text{RL}_t = \underset{(3.12)}{2.97} + \underset{(5.22)}{0.96} \sum_{i=0}^{20} \dot{P}_{t-i}$$
$$R^2 = 0.32 \quad se = 0.33 \quad DW = 1.76 \quad \hat{\rho} = 0.94$$

(5)
$$U_t - \text{UF}_t = \underset{(11.89)}{0.28(\text{GAP}_t)} + \underset{(6.31)}{0.14(\text{GAP}_{t-1})}$$
$$R^2 = 0.63 \quad se = 0.17 \quad DW = 1.95 \quad \hat{\rho}_1 = 1.43 \quad \hat{\rho}_2 = 0.52$$

资料来源:Federal Reserve Bank of St. Louis, *Review*, May 1982, p. 14.

§20.7 要点与结论

1. 假定在联立方程模型中的一个方程是可识别的(恰好或过度)，我们有几种估计它的方法。

2. 这些方法分为两大类：单一方程方法和方程组(或系统)方法。

3. 为了节省并考虑到设定误差等原因，单一方程法的使用远较普遍。这些方法的一个独特性质是，我们可以估计在一个多方程模型中的单一个方程，而不必过多地顾虑方程组中的其他方程。(注：为了可识别性，系统中的其他方程是有关系的。)

4. 三个常用的单一方程方法是 **OLS**、**ILS** 和 **2SLS**。

5. 虽然，一般地说，在联立方程模型的构架中 OLS 是不适宜的，但它可用于内生变量之间有确定的、单向因果关系的所谓**递归模型**。

785 6. ILS 法适用于恰可识别的方程。在用这个方法时，先将 OLS 应用于诱导型方程，然后从诱导型系数估计原始结构系数。

7. 2SLS 法虽然也可用于恰好可识别方程，却是专门为过度可识别方程而设计的。当方程是恰好可识别时，2SLS 和 ILS 有相同结果。2SLS 的基本思想是将(随机)内生解释变量代之以模型中的前定变量的一个线性组合，并用该组合代替原始的内生变量。因此，2SLS 法类似于**工具变量**的估计方法，它用了前定变量的线性组合作为内生回归元的工具或代理变量。

8. ILS 和 2SLS 都有一个值得注意的特点，就是所得到的估计值是一致性的，就是说，随着样本无限地增大，估计值收敛于其真的总体值。这些估计值未必满足诸如无偏性和最小方差性等小样本性质。因此，对在小样本中使用这些方法而得到的结果做(统计)推断时，要谨慎地加以解释。

习 题

问答题

20.1 判断以下的每一陈述是正确的或错误的：

a. OLS 法不适用于估计联立方程模型中的结构方程。

b. 如果一个方程不可识别，2SLS 是不适用的。

c. 在一个递归联立方程模型中不会有联立性问题。

d. 联立性问题和外生性问题是同一回事。

e. 估计结构方程的 2SLS 和其他方法只在大样本中才有优良的统计性质。

f. 并不存在有一个对整个联立方程模型而言的 R^2。

*g. 如果联立方程组中的方程误差是自相关的，并且(或)在不同方程之间是相关的，则 2SLS 和其他方法将不适宜于用来估计结构方程。

h. 如果一个方程是恰好可识别的，则 ILS 和 2SLS 将给出相同的结果。

20.2 为什么没有必要用二阶最小二乘法去估计恰好可识别方程？

786 **20.3** 考虑以下修改的凯恩斯收入决定模型：

$$C_t = \beta_{10} + \beta_{11} Y_t + u_{1t}$$
$$I_t = \beta_{20} + \beta_{21} Y_t + \beta_{22} Y_{t-1} + u_{2t}$$
$$Y_t = C_t + I_t + G_t$$

其中 C = 消费支出

I = 投资支出

Y = 收入

G = 政府支出

G_t 和 Y_{t-1} 假定是前定的。

a. 求诱导型方程并判定上述方程中哪些是可识别的(恰好或过度)。

b. 你将用什么方法估计过度可识别方程和恰好可识别方程中的参数？说出你答案中的理由。

20.4 考虑如下结果[1]：

$$\text{OLS}: \dot{W}_t = 0.276 + 0.258\dot{P}_t + 0.046\dot{P}_{t-1} + 4.959 V_t \quad R^2 = 0.924$$
$$\text{OLS}: \dot{P}_t = 2.693 + 0.232\dot{W}_t - 0.544\dot{X}_t + 0.247\dot{M}_t + 0.064\dot{M}_{t-1} \quad R^2 = 0.982$$
$$\text{2SLS}: \dot{W}_t = 0.272 + 0.257\dot{P}_t + 0.046\dot{P}_{t-1} + 4.966 V_t \quad R^2 = 0.920$$
$$\text{2SLS}: \dot{P}_t = 2.686 + 0.233\dot{W}_t - 0.544\dot{X}_t + 0.246\dot{M}_t + 0.046\dot{M}_{t-1} \quad R^2 = 0.981$$

其中 $\dot{W}_t$、$\dot{P}_t$、$\dot{M}_t$ 和 $\dot{X}_t$ 分别是收益，价格，进口价格以及劳动生产率的百分率变化(所有百分率变化，均相对于上一年而言)，而 V_t 代表未填补的职位空缺率(相对于职工总人数的百分率)。

“由于 OLS 和 2SLS 结果基本相同，故 2SLS 是无意义的。”试加评论。

***20.5** 假定生产可由柯布-道格拉斯生产函数来刻画：

$$Q_i = AK_i^{\alpha}L_i^{\beta}$$

其中 Q = 产出

K = 资本投入

* 选做题。

L - 劳动投入

A,α 和 β = 参数

i = 第 i 个厂家

787 给定最终产品价格 P，劳动价格 W 和资本价格 R，并假定利润最大化，我们得出以下经验生产模型：

生产函数：
$$\ln Q_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln u_{1i} \tag{1}$$

劳动的边际生产函数：
$$\ln Q_i = -\ln\beta + \ln L_i + \ln\frac{W}{P} + \ln u_{2i} \tag{2}$$

资本的边际生产函数：
$$\ln Q_i = -\ln\alpha + \ln K_i + \ln\frac{R}{P} + \ln u_{3i} \tag{3}$$

其中 u_1, u_2 和 u_3 是随机干扰项。

在上述模型中有三个内生变量 Q, L 和 K 的三个方程。P, R 和 W 是外生的。

a. 如果 $\alpha + \beta = 1$ 即有恒定的规模报酬，你在估计模型时会遇到什么问题？

b. 即使 $\alpha + \beta \neq 1$，你能估计这些方程吗？通过考虑方程组的可识别性做出回答。

c. 如果方程组不可识别，怎样能使它可以识别？

注：方程(2)和(3)的推导如下：将 Q 对劳动和资本分别求导并令它们等于 W/P 和 R/P，再把所得的表达式转换成对数并加上误差项(的对数)。

20.6 考虑如下的货币供求模型：

货币需求：$M_t^d = \beta_0 + \beta_1 Y_1 + \beta_2 R_t + \beta_3 P_t + u_{1t}$

货币供给：$M_t^s = \alpha_0 + \alpha_1 Y_t + u_{2t}$

其中 M = 货币

Y = 收入

R = 利率

P = 价格

假定 R 和 P 是前定的。

a. 需求函数可识别吗？

b. 供给函数可识别吗？

c. 你会用什么方法去估计可识别的方程中的参数？为什么？

d. 假使我们把供给函数加以修改，多加进两个解释变量 Y_{t-1} 和 M_{t-1}，会出现些什么识别的问题？你还会用你在(c)中用的方法吗？为什么或为什么不？

788 **20.7** 参照习题 18.10。求出那里的 2 方程组的诱导型方程并估计其参数。估计消费对收入的间接最小二乘回归，再将你的结果同 OLS 回归做比较。

解答题

20.8 考虑以下模型：

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$$

其中 M_t(货币供给)是外生的，R_t 是利率，而 Y_t 是 GDP。

a. 此模型的合理性何在？

b. 这些方程可识别吗？

c. 利用表 20.2 的数据，估计可识别的方程，说明你所用方法的理由。

20.9 假使我们把习题 20.8 中的模型改变如下：

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + \beta_3 Y_{t-1} + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + u_{2t}$$

a. 判明此方程组是否可识别。

b. 利用表 20.2 中的数据，估计可识别的方程的参数。

20.10 考虑以下模型：

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$$

其中变量定义同习题 20.8。把 I（国内投资）和 M 看作外生的，判定此方程组的可识别性。用表 20.2 的数据估计可识别的方程。

20.11 假使我们将习题 20.10 的模型改变如下：

$$R_t = \beta_0 + \beta_1 M_t + \beta_2 Y_t + u_{1t}$$
$$Y_t = \alpha_0 + \alpha_1 R_t + \alpha_2 I_t + u_{2t}$$
$$I_t = \gamma_0 + \gamma_1 R_t + u_{3t}$$

假定 M 是由外部决定的。

a. 判别哪个方程是可识别的。

b. 用表 20.2 所给数据估计可识别方程的参数。说明你所用方法的理由。

20.12 证实在（20.5.3）中报道的标准误。

20.13 回到方程(20.3.1)和(20.3.2)所给的供求模型。假使供给函数改为：

$$Q_t = \beta_0 + \beta_1 P_{t-1} + u_{2t}$$

789 其中 P_{t-1} 是前一时期流行的价格。

a. 如果 X（支出）和 P_{t-1} 是前定的，是否就没有联立性问题。

b. 如果是的话，那么供求函数都是可识别的吗？如果确实如此，就求出它们的诱导型方程并用表 20.1 中的数据去估计

它们。

c. 你能从诱导型系数导出结构系数吗？说明必要的计算。

20.14 课堂练习：考虑 1960—1999 年期间美国经济得到如下简单宏观经济模型[2]：

私人消费函数：

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t} \qquad \alpha_1 > 0,\ 0 < \alpha_2 < 1$$

私人总投资函数：

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 R_t + \beta_3 I_{t-1} + u_{2t} \quad \beta_1 > 0,\ \beta_2 < 0,\ 0 < \beta_3 < 1$$

一个货币需求函数：

$$R_t = \lambda_0 + \lambda_1 Y_1 + \lambda_2 M_{t-1} + \lambda_3 P_t + \lambda_4 R_{t-1} + u_{3t}$$
$$\lambda_1 > 0,\ \lambda_2 < 0,\ \lambda_3 > 0,\ 0 < \lambda_4 < 1$$

收入恒等式：

$$Y_t = C_1 + I_t + G_t$$

其中 C = 真实私人消费；I = 真实私人总投资；G = 真实政府支出，Y = 真实 GDP，M = 在当期价格水平上 M2 的货币供给，R = 长期利率（%），P = 消费者价格指数。内生变量为 C，I，R 和 Y。前定变量为 C_{t-1}，I_{t-1}，M_{t-1}，P_t，R_{t-1}。G_t 为截距项。诸 u 均表示误差项。

a. 用识别的阶条件判定这四个方程中哪些可以识别，是恰好识别还是过度识别。

b. 你用什么方法估计可识别的方程？

c. 从政府或私人部门获得适当的数据来估计这个模型，并对你的结果进行评论。

附录 20A

20A.1 间接最小二乘估计量的偏误

ILS 估计量虽然是一致性的，却是有偏误的。我们利用方程（20.3.1）和（20.3.2）所给的供求模型来说明这点。由（20.3.10）我们得到：

$$\hat{\beta}_1 = \frac{\hat{\Pi}_3}{\hat{\Pi}_1}$$

790 现在：

$$\hat{\Pi}_3 = \frac{\sum q_t x_t}{\sum x_t^2} \qquad \text{利用(20.3.7)}$$

和

$$\hat{\Pi}_1 = \frac{\sum p_t x_t}{\sum x_t^2} \qquad \text{利用(20.3.5)}$$

经代入，我们得到：

$$\hat{\beta}_1 = \frac{\sum q_t x_t}{\sum p_t x_t} \tag{1}$$

利用（20.3.3）和（20.3.4），我们得到：

$$p_t = \Pi_1 x_t + (w_t - \bar{w}) \tag{2}$$

$$q_t = \Pi_3 x_t + (v_t - \bar{v}) \tag{3}$$

其中 $\bar{w}$ 和 $\bar{v}$ 分别是 w_t 和 v_t 的均值。

将（2）和（3）代入（1），我们得到：

$$\begin{aligned}\hat{\beta}_1 &= \frac{\Pi_3 \sum x_t^2 + \sum (v_t - \bar{v}) x_t}{\Pi_1 \sum x_t^2 + \sum (w_t - \bar{w}) x_t} \\ &= \frac{\Pi_3 + \sum (v_t - \bar{v}) x_t / \sum x_t^2}{\Pi_1 + \sum (w_t - \bar{w}) x_t / \sum x_t^2}\end{aligned} \tag{4}$$

因为期望值运算子 E 是一个线性运算子，我们不能取（4）的期望值，虽然一般地说 $\hat{\beta}_1 \neq (\Pi_3/\Pi_1)$（为什么?），这是显然的。

但随着样本含量趋于无穷大，可以得到：

$$\operatorname{plim}(\hat{\beta}_1) = \frac{\operatorname{plim}\Pi_3 + \operatorname{plim}\sum (v_t - \bar{v}) x_t / \sum x_t^2}{\operatorname{plim}\Pi_1 + \operatorname{plim}\sum (w_t - \bar{w}) x_t / \sum x_t^2} \tag{5}$$

这里利用了 plim 的性质，即：

$$\operatorname{plim}(A + B) = \operatorname{plim}A + \operatorname{plim}B \quad \text{和} \quad \operatorname{plim}\left(\frac{A}{B}\right) = \frac{\operatorname{plim}A}{\operatorname{plim}B}$$

现在，随着样本无限地变大，（5）的分子和分母中的第二项均趋于零（为什么?），从而给出：

$$\operatorname{plim}(\hat{\beta}_1) = \frac{\Pi_3}{\Pi_1} \tag{6}$$

这表明，$\hat{\beta}_1$ 虽有偏误，却是 β_1 的一个一致性估计量。

20A.2 2SLS 估计量的标准误的估计

本附录的目的是要表明，用适合于 OLS 估计的公式去估计 2SLS 程序中

第二阶段回归的估计值的标准误，将不是对“真”标准误的“恰当”估计。我们利用（20.4.1）和（20.4.2）所给的收入—货币供给模型来说明这个问题。且考虑从第二阶段回归：

$$Y_{2t} = \beta_{20} + \beta_{21}\hat{Y}_{1t} + u_t^* \tag{20.4.6}$$

其中

$$u_t^* = u_{2t} + \beta_{21}\hat{u}_t \tag{7}$$

估计过度可识别的供给函数中的参数 $\hat{\beta}_{21}$。当我们做（20.4.6）的 OLS 回归时，$\hat{\beta}_{21}$的标准误是用以下的表达式计算的：

$$\operatorname{var}(\hat{\beta}_{21}) = \frac{\hat{\sigma}_{u^*}^2}{\sum \hat{y}_{1t}^2} \tag{8}$$

其中

$$\hat{\sigma}_{u^*}^2 = \frac{\sum(\hat{u}_t^*)^2}{n-2} = \frac{\sum(Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}\hat{Y}_{1t})^2}{n-2} \tag{9}$$

但 σ_{u*}^2 不等于 $\hat{\sigma}_{u2}^2$，后者是 u_2 的真方差的一个无偏估计。这一差别容易从（7）看到。为了求出（如前所定义的）真 $\sigma_{u_2}^2$，我们进行如下：

$$\hat{u}_{2t} = Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}Y_{1t}$$

其中 $\hat{\beta}_{20}$和 $\hat{\beta}_{21}$是得自第二阶段回归的估计值。因此，

$$\hat{\sigma}_{u_2}^2 = \frac{\sum(Y_{2t} - \hat{\beta}_{20} - \hat{\beta}_{21}Y_{1t})^2}{n-2} \tag{10}$$

注意（9）和（10）之间的差异：在（10）中我们用实测 Y 而不用从第一阶段回归得到的估计值 $\hat{Y}_1$。

一旦估计出（10），校正第二阶段回归所估计的系数的标准误，最容易的方法是用 $\hat{\sigma}_{u_2}/\hat{\sigma}_{u^*}$乘每一个系数的标准误。注意，如果 Y_{1t}和 $\hat{Y}_{1t}$非常相近，即第一阶段回归中的 R^2 非常之高，则校正因子 $\hat{\sigma}_{u_2}/\hat{\sigma}_{u^*}$将非常接近于1。在这种情况中，就不妨把第二阶段回归估计的标准误当作真估计值。在其他情况中，我们将有必要使用上述校正因子。

【习题注释】

[1] 资料来源：*Prices and Earnings in 1951—1969: An Econometric Assessment*, Department of Employment, United Kingdom, Her Majesty's Stationery Office, London, 1971, p.30.

[2] 节选自 H. R. Seddighi, K. A. Lawler, and A. V. Katos, *Econometrics: A Practical Approach*, Routledge, New York, 2000, p.204。

【注释】

[1] 然而，为了识别的目的，有必要考虑其他方程所提供的信息。如第 19 章所说的，只有对（恰好或过度）可识别的情形，方程的估计才是可能的。在本章中，

我们假定识别问题已通过第 19 章的技术得到了解决。

[2] 对这种方法的一个简单的论述，参看 Carl F. Christ, *Econometric Models and Methods*, John Wiley & Sons, New York, 1966, pp.395－401。

[3] James S. Duesenberry, Gary Fromm, Lawrence R. Klein, and Edwin Kuh, eds., *A Quarterly Model of the United States Economy*, Rand McNally, Chicago, 1965.

[4] Lawrence R. Klein, *A Textbook of Econometrics*, 2d ed., Prentice-Hall, Englewood Cliffs, NJ., 1974, p. 150.

[5] 另名"三角形模型"之称来源于以下事实：若将 (20.2.1) 中内生变量的系数排成一个阵列，我们就会得到以下的三角形矩阵：

$$
\begin{array}{cc}
 & \begin{array}{ccc} Y_1 & \quad Y_2 & \quad Y_3 \end{array} \\
\begin{array}{c} \text{方程 1} \\ \text{方程 2} \\ \text{方程 3} \end{array} &
\begin{bmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ \beta_{31} & \beta_{32} & 1 \end{bmatrix}
\end{array}
$$

注意，主对角线上方的元素都是零。(为什么?)

[6] 注：打圆点的符号表示"对时间求导数"，例如，$\dot{P}=dP/dt$。对于离散时间序列，有时用 $\Delta P/\Delta t$ 来迫近 dP/dt，其中 Δ 是我们最初在第 12 章中介绍过的一阶差分运算子。

[7] 重要的是，须记住，我们是在假定不同方程之间的干扰项无同期相关。不然的话，我们也许有必要求助于泽尔纳 (Zellner) 的似无关回归 (seemingly unrelated regressions) 估计技术，以估计递归系统中的参数。参看 A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, vol.57, 1962, pp.348－368。

[8] 还应注意，在小样本中，其他估计量和 OLS 估计量一样，也是有偏误的，但 OLS 估计量有这样的"优点"，就是，它和其他估计量相比，有最小方差。然而，这仅对小样本而言是对的。

[9] 除一致性外，估计值"还会是最优无偏的和（或）渐近有效的，这将分别视：(*i*) 诸 z [＝诸 X] 是否不仅是前定的而且是外生的 [即不含内生变量的滞后值] 和（或） (*ii*) 干扰是否正态分布的"而定。 (W.C, Hood and Tjalling C. Koopmans, *Studies in Econometric Method*, John Wiley & Sons, New York, 1953, p.133.)

[10] 我们没有给出所估结构系数的标准误。原因是，如前所说，这些系数一般地说都是诱导型系数的非线性函数，因而没有从诱导型系数的标准误推出结构系数的标准误的简单估计方法。然而，对于大样本，结构系数的标准误是可以近似地得到的。关于细节，参看 Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, p.444。

[11] 这个问题可从直观上分析如下：若 $E(\hat{\Pi}_3/\hat{\Pi}_1)=(\Pi_3/\Pi_1)$，则 $E(\hat{\beta}_1)=\beta_1$。但现在即使 $E(\hat{\Pi}_3)=\Pi_3$ 和 $E(\hat{\Pi}_1)=\Pi_1$，可以证明 $E(\hat{\Pi}_3/\hat{\Pi}_1)\neq E(\hat{\Pi}_3)/E(\hat{\Pi}_1)$；就是说，两个变量的比率的期望值不等于这两个变量的期望值的比率。然而，如附录 20A.1 中所证明的，由于 $\hat{\Pi}_3$ 和 $\hat{\Pi}_1$ 是一致性估计量，$\text{plim}(\hat{\Pi}_3/\hat{\Pi}_1)=\text{plim}(\hat{\Pi}_3)/\text{plim}(\hat{\Pi}_1)=\Pi_3/\Pi_1$。

[12] Henri Theil, "Repeated Least-Squares Applied to Complete Equation Systems," The Hague: The Central Planning Bureau, The Netherlands, 1953, (mimeographed).

[13] Robert L. Basmann, "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, vol.25, 1957, pp.77~83.

[14] 但要注意，在小样本中 $\hat{Y}_{1t}$很可能与 u_t^* 相关。理由如下：由（20.4.4）我们看到 $\hat{Y}_{1t}$是以诸 $\hat{\Pi}_i$ 为权的诸 X_i 的加权线性组合。那么，即使前定变量确实是非随机的，而作为计量的 $\hat{\Pi}_i$ 却是随机的。因此 $\hat{Y}_{1t}$仍是随机的。于是，根据我们对诱导型方程和间接最小二乘估计的讨论，显然诱导系数 $\hat{\Pi}_i$ 是随机干扰项 u_2 等的函数。既然 $\hat{Y}_{1t}$依赖于诸 $\hat{\Pi}_i$，所以 $\hat{Y}_{1t}$很可能与u_2 相关。而 u_2 是 u_t^* 的一个成分，因此预料 $\hat{Y}_{1t}$与u_t^* 相关。但如前所说，随着样本无限增大，这一相关将消失。所有这些话的要点在于，在小样本中，2SLS 程序可能导致偏误估计。

[15] 至于极端的情形，如果在第 1 阶段的回归中，$R^2=1$，则原始（过度识别的）方程中的内生变量将是实际上非随机的了。(为什么?)

[16] Henri Theil, *Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N. J., 1978, pp.341-342.

[17] 但要当心一个问题：分子中的受限制和无限制 RSS 必须用预测的 Y 值去计算（像在 2SLS 的阶段 2 中那样）。而分母中的 RSS 则用回归元的实测值而不是预测值去计算。关于这个问题的一个易懂的讨论，参看 T. Dudley Wallace and J. Lew Silver, *Econometrics: An Introduction*, Addison-Wesley, Reading, Mass., 1988, Sec. 8.5。

[18] 参看他们的 "Advertising, Concentration and Price-Cost Margins," *Journal of Political Economy*, vol.84, no.5, 1976, pp.1109-1121。

[19] "The Capital Asset Pricing Model Expressed as a Recursive System: An Empirical Investigation," *Journal of Financial and Quantitative Analysis*, June 1976, pp.237-249.

[20] *Review*, Federal Reserve Bank of St.Louis, May 1982, p.14.

第 21 章　时间序列计量经济学：一些基本概念

792　在第 1 章中我们便已注意到，经济分析中所用的两类重要数据之一便是时间序列数据。因为这类数据向计量经济学家和同行业者提出了若干挑战，所以我们要在本章和下一章中对它们做进一步的审视。

第一，以时间序列数据为依据的经验工作都假定有关的时间序列是平稳的。虽然在第 1 章中我们介绍了平稳性的直觉理念，本章仍将对它做更充分的讨论。说得更具体些，我们力图明确平稳性究竟有什么重要意义，为什么要担心一个时间序列会是不平稳的。

第二，我们在有关自相关的第 12 章中讨论了导致自相关的几个原因。有时候，原本的时间序列的非平稳性也能导致自相关。

第三，在用一个时间序列对另一个时间序列做回归时，虽然两者之间并无任何有意义的关系，但常会得到一个很高的 R^2 值（超过 0.9），而有时候我们预计两个变量之间没有关系，但一个变量对另一个变量的回归通常表现出一种显著关系。这种情况表征了**谬误或无调回归**（spurious or nonsense regression）的问题，其性质稍后解释。因此，判明经济变量之间的关系是真实的还是谬误的，就非常重要。本章中，我们将会看到，如果时间序列不是平稳的话，谬误回归会怎样产生。

793　第四，诸如股票价格之类的某些金融时间序列表现出所谓的**随机步游现象**（random walk phenomenon）。这就意味着，对一只股票（比如 IBM）明天

价格的最佳预测，就等于今天的价格加上一个纯粹随机的冲击（或误差项）。若果真如此，预测资产价格将是一件徒劳无益的事情。

第五，涉及时间序列数据的回归模型常常被用来做预测。鉴于以上的讨论，我们会想知道，如果所依据的时间序列不是平稳的，这种预测是否仍然有效。

最后，我们在第 17 章讨论的格兰杰和西姆斯因果性检验，都假定分析中所涉及的时间序列是平稳的。因此，平稳性检验应先于因果性检验。

作为开始，首先要做一个声明。由于时间序列分析专题如此广泛而又深入，而且时间序列各种方法背后的数学又如此复杂，所以在我们这样一本初级教材中，充其量只能让读者粗略地了解时间序列分析中的一些基本概念。对那些欲进一步深入钻研这一专题的读者，我们提供一些参考文献。[1]

§21.1 选看美国经济的一些时间序列

为便于以后的分析，并让读者对本章中要讨论的时间序列分析中的一些难以理解的概念有所认识，考虑一下我们通常关心的美国经济中的某些时间序列会很有好处。我们所考虑的时间序列包括：(1) GDP（国内生产总值），(2) PDI（个人可支配收人），(3) PCE（个人消费支出），(4) 利润（公司税后利润），及 (5) 股息（即公司净红利）；所有数据都以 1987 年的 10 亿美元为单位，1970—1991 年期间每个季度都有一次观测，共有 88 个季度观测。原始数据在表 21.1 中给出。

795 图 21.1 描绘出表 21.1 中 GDP、PDI 和 PCE 数据的图形，图 21.2 则给出另外两个时间序列的图形。在做任何时间数列的分析时，通常第一步工作是先看看数据的图形。我们从图 21.1 和 21.2 所示时间序列得到的第一印象是，它们都有一个上升的“趋势”，尽管有所波动。假设我们琢磨这些曲线在 1992 年第 1 季度至 1996 年第 4 季度期间的形状。[2]我们能否在脑海中简单地延伸图中所示的曲线呢？如果我们知道生成这些曲线的统计或随机机制，或者说**数据生成过程**（DGP）(data generating process)，那我们就有可能做到。但这个机制是什么呢？为了回答这个及相关问题，我们需要研究一些
794 由时间序列分析家所提出的新“词汇”，我们现在就立即转到这个方面。

表 21.1　　1970 年第 1 季度至 1991 年第 4 度美国宏观经济数据

季度	GDP	PDI	PCE	利润	股息	季度	GDP	PDI	PCE	利润	股息
1970－Ⅰ	2 872.8	1 990.6	1 800.5	44.7	24.5	1981－Ⅰ	3 860.5	2 783.7	2 475.5	159.5	64.0
1970－Ⅱ	2 860.3	2 020.1	1 807.5	44.4	23.9	1981－Ⅱ	3 844.4	2 776.7	2 476.1	143.7	68.4
1970－Ⅲ	2 896.6	2 045.3	1 824.7	44.9	23.3	1981－Ⅲ	3 864.5	2 814.1	2 487.4	147.6	71.9
1970－Ⅳ	2 873.7	2 045.2	1 821.2	42.1	23.1	1981－Ⅳ	3 803.1	2 808.8	2 468.6	140.3	72.4
1971－Ⅰ	2 942.9	2 073.9	1 849.9	48.8	23.8	1982－Ⅰ	3 756.1	2 795.0	2 484.0	114.4	70.0
1971－Ⅱ	2 947.4	2 098.0	1 863.5	50.7	23.7	1982－Ⅱ	3 771.1	2 824.8	2 488.9	114.0	68.4
1971－Ⅲ	2 966.0	2 106.6	1 876.9	54.2	23.8	1982－Ⅲ	3 754.4	2 829.0	2 502.5	114.6	69.2

1971－Ⅳ	2 980.8	2 121.1	1 904.6	55.7	23.7	1982－Ⅳ	3 759.6	2 832.6	2 539.3	109.9	72.5
1972－Ⅰ	3 037.3	2 129.7	1 929.3	59.4	25.0	1983－Ⅰ	3 783.5	2 843.6	2 556.5	113.6	77.0
1972－Ⅱ	3 089.7	2 149.1	1 963.3	60.1	25.5	1983－Ⅱ	3 886.5	2 867.0	2 604.0	133.0	80.5
1972－Ⅲ	3 125.8	2 193.9	1 989.1	62.8	26.1	1983－Ⅲ	3 944.4	2 903.0	2 639.0	145.7	83.1
1972－Ⅳ	3 175.5	2 272.0	2 032.1	68.3	26.5	1983－Ⅳ	4 012.1	2 960.6	2 678.2	141.6	84.2
1973－Ⅰ	3 253.3	2 300.7	2 063.9	79.1	27.0	1984－Ⅰ	4 089.5	3 033.2	2 703.8	155.1	83.3
1973－Ⅱ	3 267.6	2 315.2	2 062.0	81.2	27.8	1984－Ⅱ	4 144.0	3 065.9	2 741.1	152.6	82.2
1973－Ⅲ	3 264.3	2 337.9	2 073.7	81.3	28.3	1984－Ⅲ	4 166.4	3 102.7	2 754.6	141.8	81.7
1973－Ⅳ	3 289.1	2 382.7	2 067.4	85.0	29.4	1984－Ⅳ	4 194.2	3 118.5	2 784.8	136.3	83.4
1974－Ⅰ	3 259.4	2 334.7	2 050.8	89.0	29.8	1985－Ⅰ	4 221.8	3 123.6	2 824.9	125.2	87.2
1974－Ⅱ	3 267.6	2 304.5	2 059.0	91.2	30.4	1985－Ⅱ	4 254.8	3 189.6	2 849.7	124.8	90.8
1974－Ⅲ	3 239.1	2 315.0	2 065.5	97.1	30.9	1985－Ⅲ	4 309.0	3 156.5	2 893.3	129.8	94.1
1974－Ⅳ	3 226.4	2 313.7	2 039.9	86.8	30.5	1985－Ⅳ	4 333.5	3 178.7	2 895.3	134.2	97.4
1975－Ⅰ	3 154.0	2 282.5	2 051.8	75.8	30.0	1986－Ⅰ	4 390.5	3 227.5	2 922.4	109.2	105.1
1975－Ⅱ	3 190.4	2 390.3	2 086.9	81.0	29.7	1986－Ⅱ	4 387.7	3 281.4	2 947.9	106.0	110.7
1975－Ⅲ	3 249.9	2 354.4	2 114.4	97.8	30.1	1986－Ⅲ	4 412.6	3 272.6	2 993.7	111.0	112.3
1975－Ⅳ	3 292.5	2 389.4	2 137.0	103.4	30.6	1986－Ⅳ	4 427.1	3 266.2	3 012.5	119.2	111.0
1976－Ⅰ	3 356.7	2 424.5	2 179.3	108.4	32.6	1987－Ⅰ	4 460.0	3 295.2	3 011.5	140.2	108.0
1976－Ⅱ	3 369.2	2 434.9	2 194.7	109.2	35.0	1987－Ⅱ	4 515.3	3 241.7	3 046.8	157.9	105.5
1976－Ⅲ	3 381.0	2 444.7	2 213.0	110.0	36.6	1987－Ⅲ	4 559.3	3 285.7	3 075.8	169.1	105.1
1976－Ⅳ	3 416.3	2 459.5	2 242.0	110.3	38.3	1987－Ⅳ	4 625.5	3 335.8	3 074.6	176.0	106.3
1977－Ⅰ	3 466.4	2 463.0	2 271.3	121.5	39.2	1988－Ⅰ	4 655.3	3 380.1	3 128.2	195.5	109.6
1977－Ⅱ	3 525.0	2 490.3	2 280.8	129.7	40.0	1988－Ⅱ	4 704.8	3 386.3	3 147.8	207.2	113.3
1977－Ⅲ	3 574.4	2 541.0	2 302.6	135.1	41.4	1988－Ⅲ	4 734.5	3 407.5	3 170.6	213.4	117.5
1977－Ⅳ	3 567.2	2 556.2	2 331.6	134.8	42.4	1988－Ⅳ	4 779.7	3 443.1	3 202.9	226.0	121.0
1978－Ⅰ	3 591.8	2 587.3	2 347.1	137.5	43.5	1989－Ⅰ	4 809.8	3 473.9	3 200.9	221.3	124.6
1978－Ⅱ	3 707.0	2 631.9	2 394.0	154.0	44.5	1989－Ⅱ	4 832.4	3 450.9	3 208.6	206.2	127.1
1978－Ⅲ	3 735.6	2 653.2	2 404.5	158.0	46.6	1989－Ⅲ	4 845.6	3 466.9	3 241.1	195.7	129.1
1978－Ⅳ	3 779.6	2 680.9	2 421.6	167.8	48.9	1989－Ⅳ	4 859.7	3 493.0	3 241.6	203.0	130.7
1979－Ⅰ	3 780.8	2 699.2	2 437.9	168.2	50.5	1990－Ⅰ	4 880.8	3 531.4	3 258.8	199.1	132.3
1979－Ⅱ	3 784.3	2 697.6	2 435.4	174.1	51.8	1990－Ⅱ	4 900.3	3 545.3	3 258.6	193.7	132.5
1979－Ⅲ	3 807.5	2 715.3	2 454.7	178.1	52.7	1990－Ⅲ	4 903.3	3 547.0	3 281.2	196.3	133.8
1979－Ⅳ	3 814.6	2 728.1	2 465.4	173.4	54.5	1990－Ⅳ	4 855.1	3 529.5	3 251.8	199.0	136.2
1980－Ⅰ	3 830.8	2 742.9	2 464.6	174.3	57.6	1991－Ⅰ	4 824.0	3 514.8	3 241.1	189.7	137.8
1980－Ⅱ	3 732.6	2 692.0	2 414.2	144.5	58.7	1991－Ⅱ	4 840.7	3 537.4	3 252.4	182.7	136.7
1980－Ⅲ	3 733.5	2 722.5	2 440.3	151.0	59.3	1991－Ⅲ	4 862.7	3 539.9	3 271.2	189.6	138.1
1980－Ⅳ	3 808.5	2 777.0	2 469.2	154.6	60.5	1991－Ⅳ	4 868.0	3 547.5	3 271.1	190.3	138.5

注：GDP(国内生产总值)，1987 年 10 亿美元，第 A－96 页。PDI(个人可支配收入)，1987 年 10 亿美元，第 A－112 页。PCE(个人消费支出)，1987 年 10 亿美元，第 A－96 页。利润(税后公司利润)，10 亿美元，第 A－110 页。股息(净公司股支付)，10 亿美元，第 A－110 页。

资料来源：U. S. Department of Commerce, Bureau of Economic Analysis, *Business Statistics, 1963－1991*, June 1992.

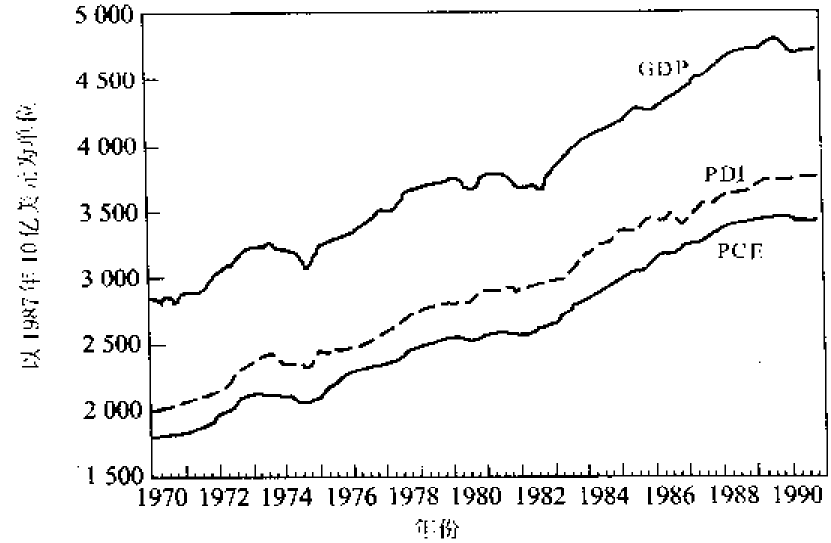


图 21.1 1970—1991 年美国 GDP、PDI 和 PCE（季度数据）

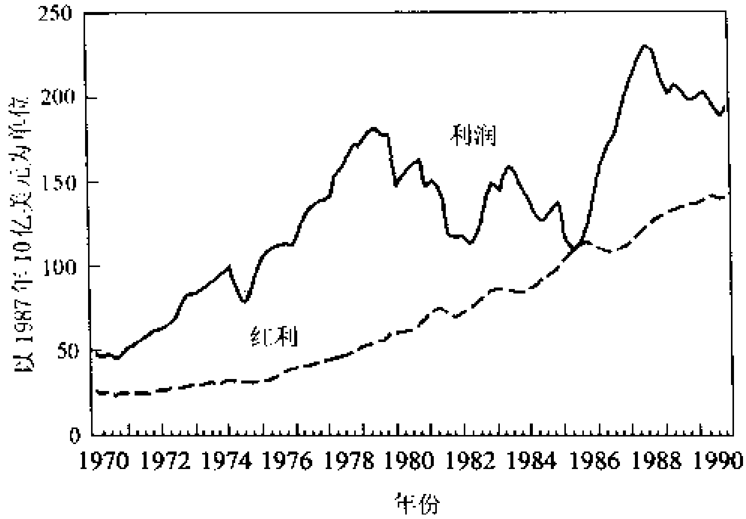


图 21.2 1970—1991 年美国利润与红利（季度数据）

§21.2 主要概念[3]

这些词汇是什么呢？它包括如下概念：

1. 随机过程
2. 平稳过程
3. 纯随机过程
4. 非平稳过程
5. 单积（单整）变量
6. 随机步游模型

7. 协积

8. 确定性和随机性趋势

9. 单位根检验

接下来我们将对每个概念分别加以讨论。我们的讨论通常是探索性的，只要有可能或有用处，我们都会举出适当的例子。

§21.3 随机过程

一个随机过程就是随机变量按时间编排的集合。[4]如果我们令 Y 表示一个随机变量，而且它是连续的，那么我们就记之为 $Y(t)$，但若它是离散的，则记之为 Y_t。前者之一例是心电图，后者的例子有 GDP、PDI 等。由于大多数经济数据都是在离散的时点上搜集的，所以我们总是用符号 Y_t 而非 $Y(t)$。若我们用 Y 表示 GDP，对我们的数据而言，则有 Y_1，Y_2，Y_3，…，Y_{86}，Y_{87}，Y_{88}，其中下标 1 表示第 1 次观测（即 1970 年第 1 季度的 GDP），下标 88 表示最后一次观测（即 1991 年第 4 季度的 GDP）。记住，这些 Y 中的每一个都是一个随机变量。

797 我们在何种意义上能说 GDP 是一个随机过程呢？比如考虑 1970 年第 1 季度的 GDP 为 28 728 亿美元。理论上讲，1970 年第 1 季度的 GDP 数字可能是任何一个数字，取决于当时的政治与经济环境。数字 28 728 只是所有这些可能性中的一个特定的**实现**（realization）。[5]因此，我们可以说，GDP 是一个随机过程，而我们在 1970 年第 1 季度至 1991 年第 4 季度期间所观测到的实际值只是这个过程的一个特定实现（即样本）。随机过程及其实现之间的区别恰似横截面数据中总体与样本之间的区别。与我们利用样本数据对总体进行推断一样，在时间序列中，我们利用这些实现对其背后的随机过程加以推断。

平稳随机过程

受到时间序列分析家大量注意和细致考察的一类随机过程就是所谓的**平稳随机过程**（stationary stochastic process）。广泛地讲，若一个随机过程的均值和方差在时间过程上保持是常数，并且在任何两时期之间的协方差值仅依赖于该两时期间的距离或滞后，而不依赖于计算这个协方差的实际时间，则称之为平稳随机过程。在时间序列文献中，这种随机过程被称为**弱平稳**（weakly stationary）、**协方差平稳**（covariance stationary）、**二阶平稳**（second-order stationary）或**广义**（wide sense）随机过程。就本章的论述而言，以及在多数实践中，考虑这种类型的平稳性足矣。[6]

为了解释弱平稳性，令随机时间序列 Y_t 有如下的性质：

$$均值: E(Y_t)=\mu \tag{21.3.1}$$

$$方差: \mathrm{var}(Y_t)=E(Y_t-\mu)^2=\sigma^2 \tag{21.3.2}$$

$$协方差: \gamma_k=E[(Y_t-\mu)(Y_{t+k}-\mu)] \tag{21.3.3}$$

其中 γ_k 即滞后 k 的协方差［或自（身）协方差］，是 Y_t 和 Y_{t+k}，也就是相隔 k 期的两个 Y 值之间的协方差。如果 $k=0$ 就得到 γ_0，这无非就是 Y 的方差（$=\sigma^2$）；如果 $k=1$，1 就是 Y 的两相邻值之间的协方差，这是我们在第 12 章讨论自相关时遇到过的一类协方差（回想马尔可夫一阶自回归模式）。

假使我们把 Y 的原点从 Y_t 移到 Y_{t+m}（比方对我们的 GDP 数据而言，从 1970 年第 1 季度移到 1975 年第 1 季度）。那么若 Y_t 是平稳的，则 Y_{t+m}
798 的均值、方差和自协方差必须和 Y_t 的一样。简言之，如果一个时间序列是平稳的，就不管在什么时间测量，它的均值、方差和（各种滞后的）自协方差都保持不变；即它们都不随时间而变化。这种时间序列有回到其均值的趋势［即**均值复原**（mean reversion）］，而且围绕其均值的波动具有大致恒定的振幅。[7]

如果一个时间序列按上述定义不是平稳的，则称之为**非平稳时间序列**（nonstationary time series）（记住我们只是在讨论弱平稳性）。换言之，一个非平稳时间序列指要么均值随时间而变化，要么方差随时间而变化，或者二者同时在发生变化。

为什么平稳的时间序列如此重要呢？因为若一个时间序列是非平稳的，则我们只能研究其在研究期间的行为。因此，每个时间序列数据集都是特定的一幕。结果，就无法把它推广到其他期间。因此，从预测角度看，这种（非平稳）时间序列没有什么太大的实际价值。

我们怎么知道某个特定的时间序列是平稳的呢？具体而言，图 21.1 和 21.2 所示的时间序列是平稳的吗？我们会在第 21.8 节和第 21.9 节讨论这个问题，在那里，我们会考虑几个平稳性检验。但若依赖直觉，图 21.1 和 21.2 所示的时间序列看起来像是非平稳的，至少均值在变化。以后我们会更详细地考虑这个问题。

在继续讨论下去之前，我们先强调一种特殊类型的随机过程，即**纯随机或白噪音**（purely random or white noise）过程。若一个随机过程的均值为 0，不变方差为 σ^2，而且不存在序列相关，那我们就称之为纯随机（时间序列）。[8] 你或许记得，我们在本书第 1 篇中讨论经典正态线性回归模型时，假定引入的误差项 u_t 为白噪音过程，并记为 $u_t \sim \mathrm{IIDN}(0, \sigma^2)$；即 u_t 是独立同分布的，而且服从 0 均值和常方差的正态分布。

非平稳随机过程

尽管我们感兴趣的是平稳时间序列，但也经常会遇到一些非平稳的时间

序列，经典的例子就是**随机步游模型**（random walk model, RWM）。[9] 通常认为诸如股票价格和汇率之类的资产价格服从随机步游；即是非平稳的。我们把随机步游分为两类：（1）带漂移的随机步游（即不存在常数项或截距项）和（2）不带漂移的随机步游（即出现常数项）。

799 **不带漂移的随机步游**。假设 u_t 是均值为 0 和方差为 σ^2 的白噪音误差项。若：

$$Y_t = Y_{t-1} + u_t \tag{21.3.4}$$

则称 Y_t 序列为随机步游。在如（21.3.4）所示的随机步游模型中，Y 在 t 时期的值等于其在（$t-1$）期的值加上一个随机冲击；因此，按照第 12 章和第 17 章的说法，它是一个 AR(1) 模型。我们可以把（21.3.4）看成第 t 期的 Y 对其一期滞后值的回归。**有效资本市场假设**（efficient capital market hypothesis）的信仰者认为，股票价格本质上是随机的，因此股市上不存在有利可图的投机空间：如果一个人能基于股票今天的价格预测明天的价格，那我们早就都是百万富翁了。

现在，我们从（21.3.4）可以写出：

$$Y_1 = Y_0 + u_1$$
$$Y_2 = Y_1 + u_2 = Y_0 + u_1 + u_2$$
$$Y_3 = Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3$$

一般地，若这个过程从第 0 期的 Y_0 开始，我们就有：

$$Y_t = Y_0 + \sum u_t \tag{21.3.5}$$

因此，

$$E(Y_t) = E(Y_0 + \sum u_t) = Y_0 \quad (为什么?) \tag{21.3.6}$$

同理，可以证明：

$$\operatorname{var}(Y_t) = t\sigma^2 \tag{21.3.7}$$

上式表明，Y 的均值等于其初始或起始值（一个常数），但随着 t 的增加，其方差无限增大，因此违背了平稳性条件。简言之，不带漂移的 RWM 是一个非平稳的随机过程。实践中通常设定 Y_0 为 0，此时 $E(Y_t) = 0$。

RWM 的一个有趣特征是，随机冲击（即随机误差项）的持久性，从（21.3.5）中明显可见：Y_t 等于初始的 Y_0 加上各期随机冲击项之和。结果是，一个特定的冲击永远也不会消失。比如，若 $u_2 = 2$ 而非 $u_2 = 0$，则从 Y_2 开始所有的 Y_t 都将提高两个单位，而且这个冲击的影响永远也不会消失。这正是为什么说随机步游具有无限记忆的原因。如帕特森（Kerry Patterson）所指出的那样，随机步游会永远记住每次冲击[10]；即具有无限记忆。

800 有趣的是，若将（21.3.4）写成：

$$(Y_t - Y_{t-1}) = \Delta Y_t = u_t \tag{21.3.8}$$

其中 Δ 为我们在第 12 章讨论过的一阶差分算子。容易证明，尽管 Y_t 是非平

稳的，但其一阶差分却是平稳的。换言之，一个随机步游时间序列的一阶差分是平稳的。但我们以后还要详细讨论这个问题。

带漂移的随机步游。让我们把（21.3.4）改写成：

$$Y_t = \delta + Y_{t-1} + u_t \tag{21.3.9}$$

其中 δ 被称为**漂移参数**（drift parameter）。"漂移"一词得自如下事实：若将上述方程写成：

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \tag{21.3.10}$$

则表明 Y_t 根据 δ 为正或负而向上或向下漂移。注意，模型（21.3.9）也是一个 AR(1) 模型。

根据讨论不带漂移随机步游的程序，可以证明，对于带漂移的随机步游模型（21.3.9），

$$E(Y_t) = Y_0 + t \cdot \delta \tag{21.3.11}$$

$$\operatorname{var}(Y_t) = t\sigma^2 \tag{21.3.12}$$

如你所见，带漂移的 RWM 的均值和方差都随着时间而递增，同样违背了（弱）平稳性条件。简言之，带不带漂移的 RWM，都是一个非平稳的时间序列。

为了看一下带和不带漂移的随机步游，我们进行如下两个模拟：

$$Y_t = Y_0 + u_t \tag{21.3.13}$$

其中 u_t 为满足 $u_t \sim N(0, 1)$的白噪音误差项；即每个都服从标准正态分布。我们从一个随机数字生成器中得到 u 的 500 次观测值，并如（21.3.13）中那样生成 Y_t。我们假定 $Y_0 = 0$。因此，（21.3.13）是一个不带漂移的 RWM。

现在考虑一个带漂移的 RWM：

$$Y_t = \delta + Y_0 + u_t \tag{21.3.14}$$

我们假定 u_t 和 Y_0 都如（21.3.13）中所示，并假定 $\delta = 2$。

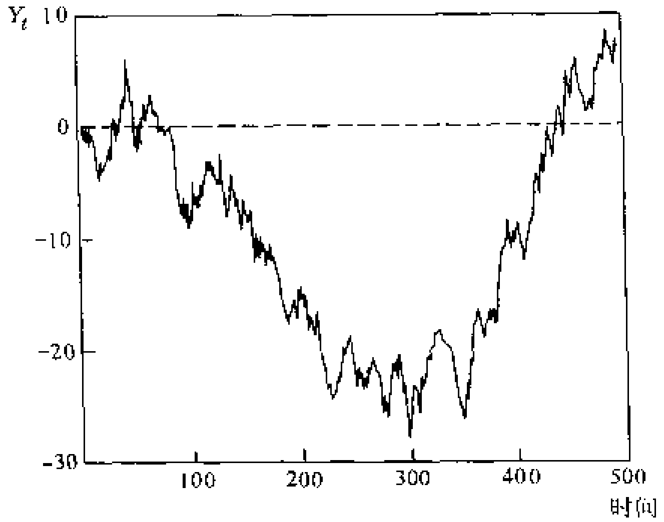
801 图 21.3 和 21.4 分别是模型的图示。读者可以根据我们对带漂移和不带漂移 RWM 的讨论，对这两个图进行比较。

随机步游模型是文献中所谓**单位根过程**（unit root process）之一例。由于"单位根"一词在时间序列文献中极为通用，所以我们现在来解释什么是单位根过程。

§21.4 单位根随机过程

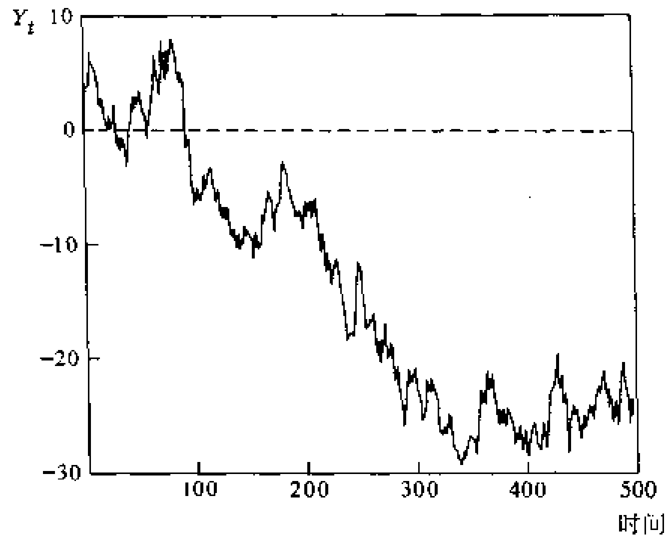
让我们把 RWM（21.3.4）写成

$$Y_t = \rho Y_{t-1} + u_t \qquad -1 \leqslant \rho \leqslant 1 \tag{21.4.1}$$



$$Y_t = Y_{t-1} + u_t$$

图 21.3　不带漂移的随机步游



$$Y_t = 2 + Y_{t-1} + u_t$$

图 21.4　带漂移的随机步游

此模型与我们在自相关一章中所讨论的马尔可夫一阶自回归模型很相似。若 $\rho=1$，则（21.4.1）就成为一个（不带漂移的）RWM。若 ρ 事实上为 1，则我们面临着所谓**单位根问题**，即非平稳性情况；我们已经知道，Y_t 的方差此时不是平稳的。单位根的名称正是源于 $\rho=1$ 这个事实。[11]因此，非平稳性、随机步游和单位根这三个术语可以看成是同义词。

但若$|\rho|\leqslant 1$，即 ρ 的绝对值小于 1，则可以证明，时间序列 Y_t 在我们所定义的意义上是是平稳的。[12]

于是，实践中，弄清楚一个时间序列是否具有一个单位根很重要。[13]我们在第 21.9 节讨论了几个单位根检验，即几个平稳性检验。我们在那一节还会判定图 21.1 和 21.2 中所示的时间序列是否平稳。或许读者猜测它们不是，但我们到那时将会明白。

§ 21.5 趋势平稳和差分平稳随机过程

平稳和非平稳随机过程（或时间序列）之间的区别，对图 21.3 和 21.4 中构造的时间序列或图 21.1 和 21.2 中的实际经济时间序列所表现出的趋势（所考虑的时间序列缓慢的长期演化结果）是确定性的还是随机性的具有关键意义。大致说来，若一个时间序列的趋势完全可以预测而且不变，我们则称之为确定性趋势，而若不能预测，则称之为随机性趋势。为了使定义更加规范，考虑时间序列的如下模型：

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \tag{21.5.1}$$

其中 u_t 为白噪音误差项，t 为按年月顺序度量的时间。现在，我们有如下可能性：

802 **纯随机步游**：若在（21.5.1）中 $\beta_1 = \beta_2 = 0$ 和 $\beta_3 = 1$，则我们得到：

$$Y_t = Y_{t-1} + u_t \tag{21.5.2}$$

它无非就是一个不带漂移的 RWM，并因此是非平稳的。但注意，若我们把写（21.5.2）成：

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \tag{21.3.8}$$

前面曾指出，这就变成平稳的随机过程。因此，一个不带漂移的 RWM 就是一个**差分平稳过程**（difference stationary process，DSP）。

带漂移的随机步游：若在（21.5.1）中 $\beta_1 \neq 0$，$\beta_2 = 0$ 和 $\beta_3 = 1$，则我们得到：

$$Y_t = \beta_1 + Y_{t-1} + u_t \tag{21.5.3}$$

它是一个带漂移的随机步游并因此是非平稳的。若我们把它写成：

$$(Y_t - Y_{t-1}) = \Delta Y_t = \beta_1 + u_t \tag{21.5.3a}$$

这就意味着 Y_t 将表现出一个带正的（$\beta_1 > 0$）或负的（$\beta_1 < 0$）趋势（见图 21.4）。这种趋势被称为**随机趋势**（stochastic trend）。由于通过对时间序列取一阶差分便可消除 Y_t 中的非平稳性，所以方程（21.5.3a）是一个 DSP 过程。

确定性趋势：若在（21.5.1）中 $\beta_1 \neq 0$，$\beta_2 \neq 0$ 和 $\beta_3 = 0$，则我们得到：

$$Y_t = \beta_1 + \beta_2 t + u_t \tag{21.5.4}$$

即所谓**趋势平稳过程**（trend stationary process，TSP）。尽管 Y_t 的均值 $\beta_1 +$

$\beta_2 t$ 不是常数，但其方差（$=\sigma^2$）是常数。一旦知道了 β_1 和 β_2 的值，就完全能预测其均值。因此，如果我们从 Y_t 中减去其均值，所得到的序列将是平稳的，因而得名**趋势平稳**（trend stationary）。这种去除确定性趋势的过程被称为**除趋势**（detrending）。

带漂移和确定性趋势的随机步游：若在（21.5.1）中 $\beta_1 \neq 0$，$\beta_2 \neq 0$ 和 $\beta_3 = 1$，则我们得到：

$$Y_t = \beta_1 + \beta_2 t + Y_{t-1} + u_t \tag{21.5.5}$$

即同时带有漂移和确定性趋势的随机步游，若将此方程写成：

$$\Delta Y_t = \beta_1 + \beta_2 t + u_t \tag{21.5.5a}$$

这就意味着 Y_t 是非平稳的。

804 **含平稳 AR（1）成分的确定性趋势**：若在（21.5.1）中 $\beta_1 \neq 0$，$\beta_2 \neq 0$ 和 $\beta_3 < 1$，则我们得到：

$$Y_t = \beta_1 + \beta_2 t + \beta_3 Y_{t-1} + u_t \tag{21.5.6}$$

它在确定性趋势周围是平稳的。

为了看出确定性和随机性趋势的区别，考虑图 21.5。[14] 此图中名曰"随机性"的序列由 RWM：$Y_t = 0.5 + Y_{t-1} + u_t$ 生成，其中 u_t 的 500 个值由一个标准正态分布生成，Y 的初始值设定为 1。名曰"确定性"的序列由 $Y_t = 0.5t + u_t$ 生成，其中 u_t 生成如上，而 t 则是按年月顺序度量的时间。

如你从图 21.5 中所见，在确定性趋势的情况下，对趋势线（代表着非平稳的均值）的偏离完全是随机的，并很快就会消逝；它们对时间序列由趋势成分 $0.5t$ 所决定的长期发展没有影响。而另一方面，在随机性趋势的情况下，随机成分 u_t 影响着序列 Y_t 的长期进展。

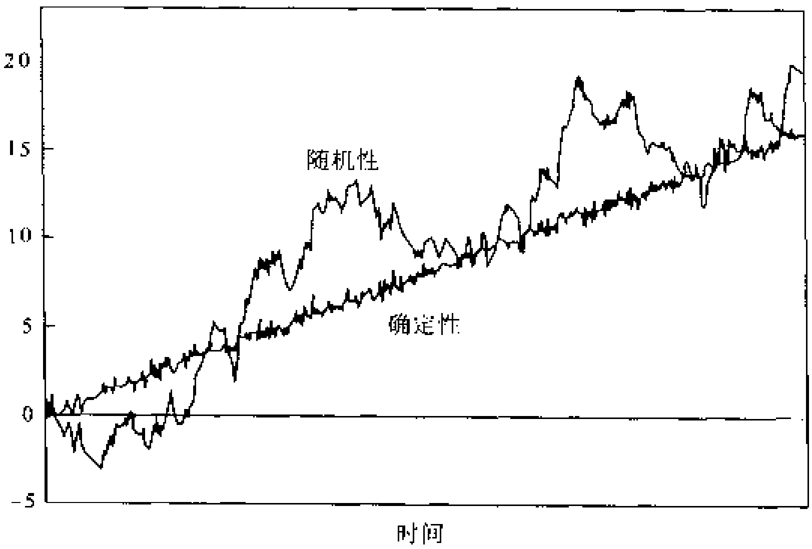


图 21.5　确定性与随机性趋势

资料来源：Charemza 等人的前引文献，第 91 页。

§21.6 单积随机过程

805 随机步游模型无非只是一类被称为**单积过程**（integrated processes）的随机过程的特殊情形。回忆一下，不带漂移的 RWM 是非平稳的，但如(21.3.8)所示，其一阶差分则是平稳的。因此，我们称不带漂移的 RWM 为**一阶单积**（integrated of order 1）序列，记为 $I(1)$。类似地，若使一个时间序列变成平稳序列需对其进行两次差分（即对一阶差分再取一阶差分），则称之为**二阶单积**（integrated of order 2）序列。[15] 一般地，若一个（非平稳的）时间序列只有经过 d 次差分才能变成平稳序列，则称之为 **d 阶单积**（integrated of order d）序列，并记为 $Y_t \sim I(d)$。若一个时间序列 Y_t 一开始就是平稳的（即不需要进行任何差分），则称之为 0 阶单积序列，并记之为 $Y_t \sim I(0)$。因此，我们使用术语“平稳时间序列”和“0 阶单积时间序列”时表示的是同一个意思。

大多数经济时间序列通常都是 $I(1)$；即只需取一阶差分便变成平稳序列。图 21.1 和 21.2 中所示的时间序列是 $I(1)$ 或者更高阶单积序列吗？我们在第 21.8 和 21.9 节将考察这一点。

单积序列的性质

单积时间序列有如下性质值得注意：令 X_t、Y_t 和 Z_t 为三个时间序列。

1. 若 $X_t \sim I(0)$ 和 $Y_t \sim I(1)$，则 $Z_t = (X_t + Y_t) \sim I(1)$；即平稳和非平稳时间序列的线性组合或之和是非平稳的。

2. 若 $X_t \sim I(d)$，则 $Z_t = (a + bX_t) \sim I(d)$，其中 a 和 b 为常数。即一个 $I(d)$ 序列的线性函数仍是 $I(d)$。因此，若 $X_t \sim I(0)$，则 $Z_t = (a + bX_t) \sim I(0)$。

3. 若 $X_t \sim I(d_1)$ 和 $Y_t \sim I(d_2)$，其中 $d_1 < d_2$，则 $Z_t = (aX_t + bY_t) \sim I(d_2)$。

4. 若 $X_t \sim I(d)$ 和 $Y_t \sim I(d)$，则 $Z_t = (aX_t + bY_t) \sim I(d^*)$；$d^*$ 通常都等于 d，但在某些情况下 $d^* < d$（参见第 21.11 节中对协积这一专题的探讨）。

如你从以上命题中所见，在合并两个或多个不同阶单积时间序列时必须小心。

为了看出这一点为什么重要，考虑在第 3 章中讨论过的双变量回归模型，即 $Y_t = \beta_1 + \beta_2 X_t + u_t$。在经典 OLS 假定之下，我们知道：

$$\hat{\beta}_2=\frac{\sum x_t y_t}{\sum x_t^2} \tag{21.6.1}$$

其中小写字母如平常一样表示对均值的离差。假设 Y_t 为 $I(0)$，但 X_t 为 $I(1)$；即前者是平稳的，而后者是非平稳的。由于 X_t 是非平稳的，所以其
806 方差无限增大，因而（21.6.1）中的分母支配着它的分子项，导致渐近地（即在大样本中）收敛于0，甚至没有一个渐近的分布。[16]

§ 21.7 谬误回归现象

为了看出平稳的时间序列为什么如此重要，考虑如下两个随机步游模型：

$$Y_t=Y_{t-1}+u_t \tag{21.7.1}$$

$$X_t=X_{t-1}+v_t \tag{21.7.2}$$

其中，我们从 $u_t \sim N(0,1)$ 中生成了 u_t 的500次观测，从 $v_t \sim N(0,1)$ 中生成了 v_t 的500次观测，并假定 Y 和 X 的初始值都为零。我们还假定 u_t 和 v_t 都不存在序列相关，而且彼此间也不存在相关关系。就你目前所知，这两个时间序列都是非平稳的；即它们都是 $I(1)$ 或表现出随机趋势。

假设我们将 Y_t 对 X_t 回归。由于 Y_t 和 X_t 是不相关的 I（1）过程，所以 Y 对 X 的回归中所得到的 R^2 应该趋于0；即这两个变量之间不应该有任何关系。但请你先看一下回归结果：

变量	系数	标准误	t 统计量
C	−13.255 6	0.620 3	−21.368 56
X	0.337 6	0.044 3	7.612 23
	$R^2=0.104\ 4$	$d=0.012\ 1$	

如你所见，X 的系数是高度统计显著的，尽管 R^2 值有些低，但它在统计上显著异于零。基于这些结论，你可能得出 Y 和 X 之间存在显著统计关系的结论，尽管先验假定它们之间没有任何关系。这就是对尤尔（G.U.Yule）首次发现的**谬误或无调回归**（phenomenon of spurious or nonsense regression）的简单概括。[17]尤尔指出，即便在样本很大时，（谬误）相关在非平稳时间序列中也可能持续存在。极低的德宾-沃森 d 值表明存在着很强的一阶自相
807 关，从而暗示着上述回归有些问题。根据格兰杰和纽博尔德（Newbold）的分析，**$R^2>d$ 就是怀疑所估计的回归是谬误回归的一个很好的经验法则**，上例正是如此。

通过将 Y_t 的一阶差分（$=\Delta Y_t$）对 X_t 的一阶差分（$=\Delta X_t$）进行回归

很容易看出，以上给出的回归结果是没有什么意义的；记住，尽管 Y_t 和 X_t 是非平稳的，但其一阶差分却是平稳的。在这样一个回归中你会发现，本该为 0 的 R^2 实际上正是 0，德宾-沃森 d 约为 2。在习题 21.24 中，要求你做这个回归并验证刚刚得到的命题。

尽管富于戏剧性，但这个例子还是强烈地提醒我们，基于表现出随机趋势的时间序列做回归分析时应该高度警惕。因此，在阅读大量基于 $I(1)$ 变量所得到的回归结果时也要极为谨慎。作为一个例子，参见习题 21.26。在某种程度上，对确定性趋势的时间序列也是如此，习题 21.25 就给出了这样的一个例子。

§21.8 平稳性的检验

到目前为止，读者可能对平稳随机过程及其重要性有了很好的了解。实践中，我们面临两个重要问题：(1) 我们如何发现一个给定的时间序列是否平稳？(2) 如果我们发现一个给定的时间序列不是平稳的，有什么办法使之变成平稳的呢？我们在本节讨论第一个问题，并在第 21.10 节讨论第二个问题。

在讲下去之前，记住我们主要考虑的是弱平稳性或协方差平稳性。

尽管有几种平稳性检验的方法，但我们在本节只讨论在文献中广泛讨论的两种：(1) 图示分析，(2) 相关图检验。由于单位根检验在最近时期尤为重要，所以我们在下一节讨论它。我们以适当的例子来解释这些检验。

1. 图形分析

前面曾指出，在进行规范的检验之前，像我们对表 21.1 中数据描点成图 21.1 和 21.2 一样，将所研究的时间序列描点总是明智之举。这种描点图对时间序列的可能性质给出初步线索。以图 21.1 中所示的 GDP 时间序列为例。你将看到，GDP 在研究期中不断增加，表现出上升趋势，从而表明 GDP 的均值在发生变化。这可能说明，GDP 序列不是平稳的。图 21.2 中所示的其他美国经济时间序列多少也有些类似情况。这种直观感受是更规范的平稳性检验的起点。

2. 自相关函数和相关图

808 有一种平稳性检验，它基于所谓的**自相关函数（ACF）**（autocorrelation function）。记滞后 k 阶的 ACF 为 ρ_k，其定义是：

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{k\text{ 阶滞后的协方差}}{\text{方差}} \tag{21.8.1}$$

其中 k 阶滞后的协方差和从前的定义一样。注：若 $k=0$，则 $\rho_0=1$。（为什么？）

由于协方差和方差都以相同的度量单位度量，所以 ρ_k 是没有度量单位的数字，或者说是纯数字。和任何一个相关系数一样，它介于 -1 和 1 之间。若将 ρ_k 对 k 描点，则所得到的图被称为**总体相关图**（population correlogram）。

由于实际上我们只有随机过程的一个实现（即样本），所以我们只能计算出**样本自相关函数（SACF）**（sample autocorrelation function），$\hat{\rho}_k$。为了计算它，我们必须首先计算中 k 阶滞后的**样本协方差**（sample covariance）$\hat{\gamma}_k$ 和**样本方差**（sample variance）$\hat{\gamma}_0$，其定义分别为[18]：

$$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n} \tag{21.8.2}$$

$$\hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y})^2}{n} \tag{21.8.3}$$

其中 n 为样本容量，$\bar{Y}$ 为样本均值。

因此，k 阶样本自相关函数就是：

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \tag{21.8.4}$$

它无非就是 k 阶样本协方差与样本方差之比。将 $\hat{\rho}_k$ 对 k 描点则称之为**样本相关图**（sample correlogram）。

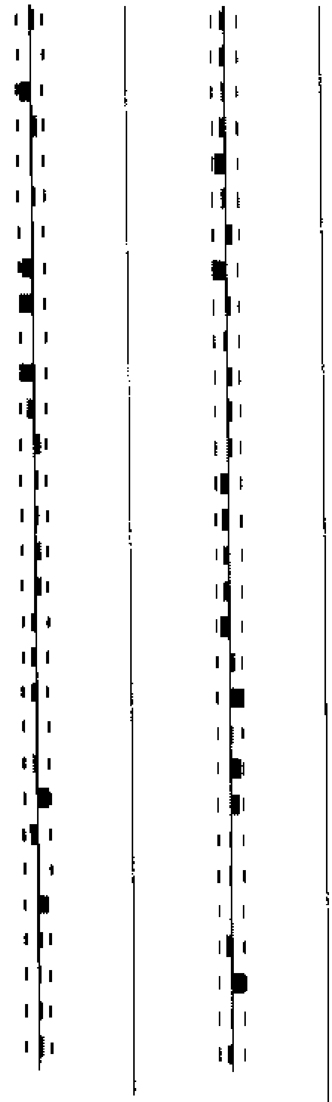
一个样本相关图如何能使我们发现一个特定的时间序列是否平稳呢？为此，我们首先给出一个纯粹白噪音随机过程和一个随机步游过程的样本相关图。回想不带漂移的 RWM（21.3.13）。在那里，我们从标准正态分布中生成了 500 个误差项 u 的一个样本。图 21.6 中给出了这 500 个纯随机误差项的相关图；我们只给出了 30 阶滞后。稍后我们会就如何选择滞后长度进行评论。

现在，只看 AC 列中的样本自相关函数和左边第一个标为自相关的图。
809 图中的实线表示零轴；此线以上的观测为正值，以下为负值。从此图中清晰可见，纯白噪音过程的各阶自相关都在零附近徘徊。这就是平稳时间序列相关图的图形。因此，如果一个实际（经济）时间序列的相关图与白噪音时间序列的相关图很相像，那我们就能说，这个时间序列很可能是平稳的。

810 再来看一下由（比方说）（21.3.13）生成的随机步游序列的相关图。图形如图 21.7 所示。此相关图最显著的特征是，各阶滞后的自相关系数都很高，甚至到 33 个季度的滞后仍居高不下。事实上，如果我们考虑 60 个季度的滞后，自相关系数仍相当高；在 60 阶滞后的系数约为 0.7。图 21.7 是典
811 型的非平稳时间序列相关图：自相关系数从一个很高的值开始，随着滞后长度的增加而缓慢地向零下降。

样本：2 500

所包含观测：499



自相关	偏相关		AC	PAC	Q-Stat	Prob
		1	−0.022	−0.022	0.233 5	0.629
		2	−0.019	−0.020	0.424 7	0.809
		3	−0.009	−0.010	0.464 0	0.927
		4	−0.031	−0.031	0.937 2	0.919
		5	−0.070	−0.072	3.418 6	0.636
		6	−0.008	−0.013	3.449 3	0.751
		7	0.048	0.045	4.641 1	0.704
		8	−0.069	−0.070	7.038 5	0.532
		9	0.022	0.017	7.295 6	0.606
		10	−0.004	−0.011	7.305 9	0.696
		11	0.024	0.025	7.610 2	0.748
		12	0.024	0.027	7.899 3	0.793
		13	0.026	0.021	8.250 2	0.827
		14	−0.047	−0.046	9.372 6	0.806
		15	−0.037	−0.030	10.074	0.815
		16	−0.026	−0.031	10.429	0.843
		17	−0.029	−0.024	10.865	0.863
		18	−0.043	−0.050	11.807	0.857
		19	0.038	0.028	12.575	0.860
		20	0.099	0.093	17.739	0.605
		21	0.001	0.007	17.739	0.665
		22	0.065	0.060	19.923	0.588
		23	0.053	0.055	21.404	0.556
		24	−0.017	−0.004	21.553	0.606
		25	−0.024	−0.005	21.850	0.644
		26	−0.008	−0.008	21.885	0.695
		27	−0.036	−0.027	22.587	0.707
		28	0.053	0.072	24.068	0.678
		29	−0.004	−0.011	24.077	0.725
		30	−0.026	−0.025	24.445	0.752

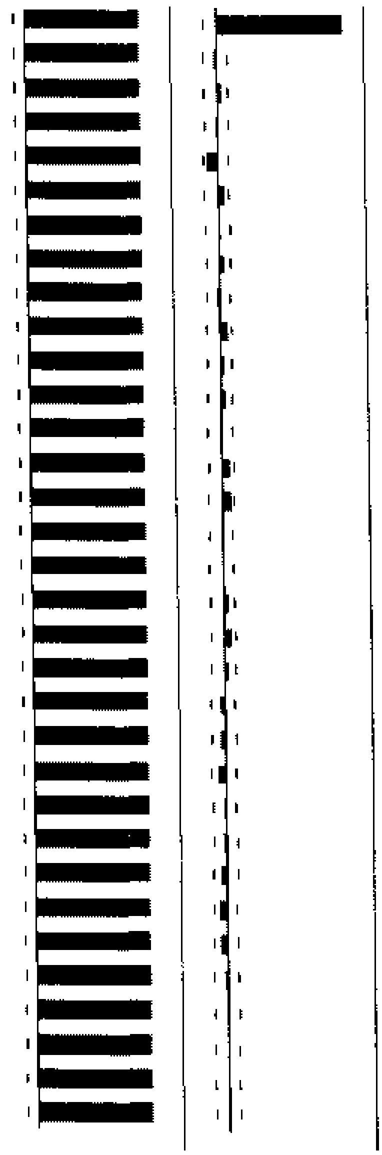
图 21.6　自噪音误差项 u 的相关图

注：AC＝自相关，PAC＝偏相关，Q－Stat＝Q 统计量，Prob ＝概率。

现在，让我们考虑一个具体的经济例子。我们考察一下表 21.1 中所给 GDP 时间序列的相关图。图 21.8 中给出了直至 25 阶滞后的相关图。这 25 阶滞后的 GDP 相关图的表现，与图 21.7 中随机步游模型的相关图很相似。自相关系数从 1 阶滞后的很高值（0.969）开始，并极其缓慢地下降。由此看来，GDP 时间序列是非平稳的。你若将图 21.1 和 21.2 中所示的美国其他经济时间序列都描出相关图，那你会看到类似形态，从而得到所有这些时间序列都非平稳的结论；它们的均值或方差或二者是非平稳的。

样本：2 500

所包含观测：499



自相关 偏相关		AC	PAC	Q-Stat	Prob
	1	0.992	0.992	493.86	0.000
	2	0.984	0.000	980.68	0.000
	3	0.976	0.030	1 461.1	0.000
	4	0.969	0.005	1 935.1	0.000
	5	0.961	−0.059	2 402.0	0.000
	6	0.953	0.050	2 862.7	0.000
	7	0.946	0.004	3 317.3	0.000
	8	0.939	0.040	3 766.4	0.000
	9	0.932	−0.009	4 210.1	0.000
	10	0.927	0.055	4 649.1	0.000
	11	0.921	0.018	5 083.9	0.000
	12	0.916	0.039	5 514.9	0.000
	13	0.912	0.002	5 942.4	0.000
	14	0.908	0.056	6 367.0	0.000
	15	0.905	0.061	6 789.8	0.000
	16	0.902	0.000	7 210.6	0.000
	17	0.899	0.006	7 629.4	0.000
	18	0.896	0.030	8 046.7	0.000
	19	0.894	0.053	8 463.1	0.000
	20	0.892	0.013	8 878.7	0.000
	21	0.890	−0.041	9 292.6	0.000
	22	0.886	−0.040	9 704.1	0.000
	23	0.882	−0.044	10 113.	0.000
	24	0.878	−0.012	10 518.	0.000
	25	0.873	−0.023	10 920.	0.000
	26	0.867	−0.041	11 317.	0.000
	27	0.860	−0.055	11 709.	0.000
	28	0.853	−0.045	12 095.	0.000
	29	0.846	−0.010	12 476.	0.000
	30	0.839	0.008	12 851.	0.000
	31	0.832	−0.006	13 221.	0.000
	32	0.825	0.003	13 586.	0.000
	33	0.819	−0.006	13 946.	0.000

图 21.7　一个随机步游时间序列的相关图

注：变量定义见图 21.6。

812 这里可能要提出两个实际问题。首先，我们如何选择滞后长度来计算 ACF? 其次，你如何判定一个相关系数在特定滞后长度下是否统计显著？答案如下。

滞后长度的选择。这基本上是个经验问题。一个经验法则是，计算 ACF 通常要用到时间序列 1/3～1/4 长度的滞后。对我们的经济数据而言，共有 88 个季度的观测，根据这个法则应选择 22～29 个季度的滞后。最佳的

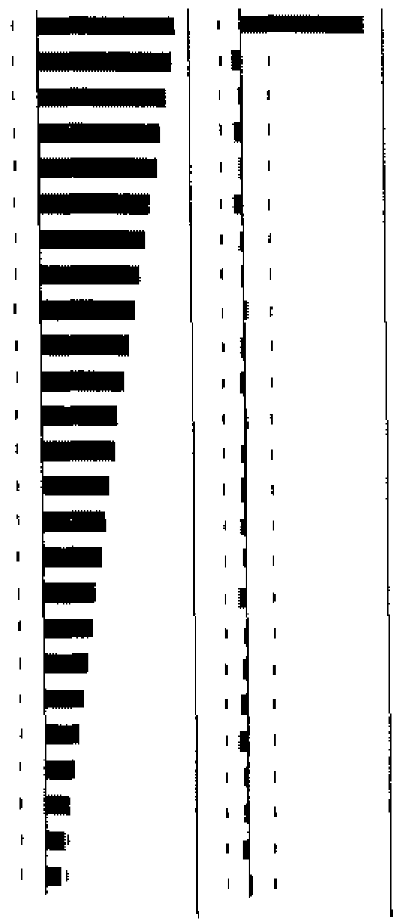
实际建议是，从足够大的滞后开始，然后利用某种统计准则（如我们在第13章中讨论过的赤池或施瓦茨信息准则）使之减小。否则，可以利用如下统计检验。

自相关系数的统计显著性

比如，考虑图21.8中所给GDP时间序列的相关图。我们如何判定10季度滞后的相关系数0.638是否统计显著呢？任何一个 $\hat{\rho}$ 的统计显著性都可由其标准误来判断。巴特利已经证明，若一个时间序列是纯随机的，即表现出白噪音性状（见图21.6），则样本自相关系数 $\hat{\rho}_k$ 近似服从如下分布[19]：

样本：1970－1　1991－4

所包含观测：88



自相关	偏相关		AC	PAC	Q-Stat	Prob
		1	0.969	0.969	86.462	0.000
		2	0.935	−0.058	166.02	0.000
		3	0.901	−0.020	241.72	0.000
		4	0.866	−0.045	312.39	0.000
		5	0.830	−0.024	378.10	0.000
		6	0.791	−0.062	438.57	0.000
		7	0.752	0.029	493.85	0.000
		8	0.713	−0.024	544.11	0.000
		9	0.675	0.009	589.77	0.000
		10	0.638	−0.010	631.12	0.000
		11	0.601	−0.020	668.33	0.000
		12	0.565	−0.012	701.65	0.000
		13	0.532	0.020	731.56	0.000
		14	0.500	−0.012	758.29	0.000
		15	0.468	−0.021	782.02	0.000
		16	0.437	−0.001	803.03	0.000
		17	0.405	−0.041	821.35	0.000
		18	0.375	−0.005	837.24	0.000
		19	0.344	−0.038	850.79	0.000
		20	0.313	−0.017	862.17	0.000
		21	0.279	−0.066	871.39	0.000
		22	0.246	−0.019	878.65	0.000
		23	0.214	−0.008	884.22	0.000
		24	0.182	−0.018	888.31	0.000
		25	0.153	0.017	891.25	0.000

图21.8　美国1970年第1季度至1991年第4季度的GDP相关图

注：变量定义见图21.6。

$$\hat{\rho}_k \sim N(0, 1/n) \tag{21.8.5}$$

即在大样本中，样本自相关系数服从均值为0和方差等于样本容量之倒数的

正态分布。既然我们有 88 个观测，所以方差就是 $1/88=0.011\ 36$，其标准误就是 $\sqrt{0.011\ 36}=0.106\ 6$。然后根据标准正态分布的性质，任何一个（总体）$\rho_k$ 的 95% 的置信区间就是：

$$\hat{\rho}_k \pm 1.96\,(0.106\ 6) \tag{21.8.6}$$

换言之，

$$\text{Prob}(\hat{\rho}_k-0.208\ 9 \leqslant \hat{\rho}_k \leqslant \hat{\rho}_k+0.208\ 9)=0.95 \tag{21.8.7}$$

若上述区间包括了零值，则我们不能拒绝真实 ρ_k 为零的假设，但若这个区间没有包括 0，则我们就拒绝真实 ρ_k 为零的假设。应用于 $\hat{\rho}_{10}=0.638$，读者可以验证真实 $\hat{\rho}_{10}$的 95% 置信区间为$(0.638\pm0.208\ 9)$或$(0.429\ 1,\ 0.846\ 9)$。[20]
813 显然，0 并不包含其中，这就表明，我们有 95% 的把握认为真实的 ρ_{10}显著异于 0。[21] 你可以验证，即便在 20 阶滞后的情况下，估计的 ρ_{20}在 5% 的显著性水平上也是统计显著的。

不用检验某个别自相关系数的统计显著性，我们可以检验所有 ρ_k（至某个滞后）都同时为 0 的联合假设。这可利用由博克斯（G.E.P.Box）和皮尔斯（D.A.Pierce）提出的 **Q 统计量**（*Q* statistic）来进行，其定义为[22]：

$$Q=n\sum_{k=1}^{m}\hat{\rho}_k^2 \tag{21.8.8}$$

其中 n 为样本容量，m 为滞后长度。Q 统计量通常用于检验一个时间序列是否白噪音。在大样本中，它近似服从自由度为 m 的 χ^2 分布。在实际应用中，若计算出来的 Q 大于在选定显著性水平下从 χ^2 分布表中查出的临界 Q 值，则拒绝所有真实的（ρ_k 都为 0）虚拟假设；至少它们中有某些一定非零。

博克斯-皮尔斯 Q 统计量的一个变形就是**扬－博克斯（LB）统计量**（Ljung-Box Statistic），其定义为[23]

$$\text{LB}=n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right)\sim\chi^2 m \tag{21.8.9}$$

尽管在大样本中，Q 和 LB 统计量都服从自由度为 m 的 χ^2 分布，但我们已经发现，LB 统计量比 Q 统计量具有更好的小样本性质（即在统计意义上更有效）。[24]

回到图 21.8 中给出的 GDP 例子中来，直至 25 阶滞后的 LB 统计量的值约为 891.25。在 25 个估计自相关系数的平方和为 0 的虚拟假设下，得到这样一个 LB 值的概率实际上为 0，如图中最后一列所示。因此，结论是 GDP 时间序列是非平稳的，从而加强了我们从图 21.1 中对 GDP 序列非平稳性的预感。在习题 21.16 中，要求你证实其他四个美国经济的时间序列也都是非平稳的。

§21.9 单位根检验

814 在过去几年变得广受欢迎的一种平稳性（或非平稳性）检验是单位根检

验。我们首先解释其概念，然后阐述其步骤，最后考虑其某些局限。

首先从我们在第21.4节中讨论的单位根（随机）过程开始：

$$Y_t = \rho Y_{t-1} + u_t \qquad -1 \leqslant \rho \leqslant 1 \tag{21.4.1}$$

其中 u_t 为白噪音误差项。

我们知道，若 $\rho = 1$，即在单位根情形下，则（21.4.1）就变成一个不带漂移的随机步游模型，我们知道这种模型是非平稳的随机过程。因此，为什么不简单地将 Y_t 对其（一期）滞后值 Y_{t-1} 回归，并搞清楚所估计的 ρ 在统计上是否等于1？若是，则 Y_t 是非平稳的。这正是平稳性的单位根检验背后的一般思想。

出于理论上的原因，我们对（21.4.1）做如下变化：从（21.4.1）的两边同时减去 Y_{t-1} 得到：

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t = (\rho - 1) Y_{t-1} + u_t \tag{21.9.1}$$

进而可写成：

$$\Delta Y_t = \delta Y_{t-1} + u_t \tag{21.9.2}$$

其中 $\delta = (\rho - 1)$，而 Δ 和平常一样表示一阶差分算子。

因此，在实践中，不用估计（21.4.1），我们估计（21.9.2）并检验 $\delta = 0$ 的虚拟假设。若 $\delta = 0$，则 $\rho = 1$，即存在单位根，从而意味着所检验的时间序列是非平稳的。

在继续估计（21.9.2）之前，注意到，若 $\delta = 0$，则（21.9.2）变成

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \tag{21.9.3}$$

由于 u_t 是白噪音误差项，所以它是平稳的，这意味着一个随机步游时间序列的一阶差分是平稳的，我们以前已经得到过这一结论。

现在转向对（21.9.2）的估计。这个估计十分简单；我们所要做的就是取 Y_t 的一阶差分，并将它们对 Y_{t-1} 回归，看回归中估计的斜率系数（$=\delta$）
815 是否为零。若为零，则断定 Y_t 是非平稳的，但若为负，则断定 Y_t 是平稳的。[25]惟一的问题在于，在判断（21.9.2）中 Y_{t-1} 的估计系数是否为零时该采用哪种检验。你可能禁不住认为，为什么不用通常的 t 检验呢？不幸的是，在虚拟假设 $\delta = 0$（即 $\rho = 1$）下，Y_{t-1} 估计系数的 t 值即便在大样本下也不服从 t 分布；即它不是渐近正态分布。

还有什么方法可用呢？迪基（Dickey）和富勒（Fuller）已经证明，在虚拟假设 $\delta = 0$ 下，（21.9.2）中 Y_{t-1} 系数的估计 t 值服从 **τ 统计量**（τ statistic）。[26]这些作者已经基于蒙特卡罗模拟运算计算出了 τ 统计量的临界值。附录D的表D.7中给出了这些临界值的一个样本。这个表很有限，但麦金农已经准备了一些更全面的表，现在几个计量经济软件包中都包含了这些表。[27]在文献中，为了纪念其发现者，τ 统计量或检验又被称为迪基–富勒（**DF**）检验（Dickey-Fuller test）。有趣的是，若假设 $\delta = 0$ 被拒绝（即时间序列是平稳的），我们就可以使用通常的（学生）t 检验。

实施DF检验的实际程序涉及几个决策。在第21.4和21.5节讨论单位根过程的性质时，我们曾指出，一个随机步游过程或不含漂移、或含有漂移、或者同时具有确定性和随机性趋势。为容许各种可能性，DF检验在三种不同的形式即三种不同的虚拟假设下进行估计。

Y_t 是一个随机步游： $\Delta Y_t = \delta Y_{t-1} + u_t$ (21.9.2)

Y_t 是一个带漂移的随机步游： $\Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$ (21.9.4)

Y_t 是一个带漂移和确定性趋势的随机步游： $\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$

(21.9.5)

其中 t 为时间或趋势变量。在每种情形中，虚拟假设都是 $\delta = 0$；即存在一个单位根——时间序列是非平稳的。对立假设是 δ 小于0；即时间序列是平稳的。[28] 若虚拟假设被拒绝，在（21.9.2）的情况下意味着 Y_t 是一个平稳的时间序列，在（21.9.4）的情况下意味着 Y_t 是一个均值非零 $[=\beta_1/(1-\rho)]$ 的平稳的时间序列，在（21.9.5）的情况下意味着 Y_t 在一个确定性趋势附近是平稳的。

816 检验假设 $\delta=0$ 的 τ 检验与上述DF检验的三种设定都不同，注意到这一点极为重要，从附录D的表D.7明显可以看出来。而且，比方说，若（21.9.4）是正确的，但我们估计了（21.9.2），那我们就遇到设定误差的问题，我们在第13章已经知道了其后果。若正确的模型是（21.9.5），但我们估计了（21.9.4），也会出现同样的情况。当然，没有办法一开始就知道哪个设定是正确的，所以尽管会遇到数据开采的问题，但在一定程度上应用试错法总不可避免。

实际的估计程序如下：用OLS估计（21.9.2）或（21.9.3）或（21.9.4）；将每种情况下得到的 Y_{t-1} 的估计系数除以其标准误来计算 τ 统计量；参考DF表（或任何一个统计软件）。若计算出来的 τ 统计量的绝对值（$|\tau|$）超过了DF或麦金农的临界 τ 值，则拒绝 $\delta=0$ 的虚拟假设，此时时间序列是平稳的。另一方面，若计算的 $|\tau|$ 没有超过临界 τ 值，则不能拒绝虚拟假设，此时时间序列就是非平稳的。确定你使用了适当的临界 τ 值。

让我们回到美国GDP时间序列。对此序列，（21.9.2）、（21.9.4）和（21.9.5）三个回归的结果如下：在每种情况下的因变量都是 $\Delta Y_t = \Delta GDP_t$

$$\widehat{\Delta GDP}_t = 0.00576\,GDP_{t-1}$$
$$t = (5.7980) \qquad R^2 = -0.0152 \qquad d = 1.34 \quad (21.9.6)$$

$$\widehat{\Delta GDP}_t = 28.2054 - 0.00136\,GDP_{t-1}$$
$$t = (1.1576)(-0.2191) \qquad R^2 = 0.00056 \qquad d = 1.35 \quad (21.9.7)$$

$$\widehat{\Delta GDP}_t = 190.3857 + 1.4776t - 0.0603\,GDP_{t-1}$$
$$t = (1.8389)(1.6109)(-1.6252) \qquad R^2 = 0.0305 \qquad d = 1.31$$

(21.9.8)

我们在这里主要感兴趣的是 GDP_{t-1} 系数的 t（$=\tau$）值。对模型

(21.9.6) 而言，显著性为 1%、5% 和 10% 的临界 τ 值分别是 -2.5897、-1.9439 和 -1.6177；对模型 (21.9.7) 而言，显著性为 1%、5% 和 10% 的临界 τ 值分别是 -3.5064、-2.8947 和 -2.5842；对模型 (21.9.8) 而言，显著性为 1%、5% 和 10% 的临界 τ 值分别是 -4.0661、-3.4614 和 -3.1567。前面曾指出，这些临界值对三个模型是不同的。

在我们考察结论之前，必须决定这三个模型中哪一个合适。我们应该排除模型 (21.9.6)，因为 GDP_{t-1} 的系数 δ 为正。但因为 $\delta=(\rho-1)$，正的 δ
817 就意味着 $\rho>1$。尽管在理论上有这种可能性，但我们还是把这种情况排除掉，因为在这种情况下，GDP 时间序列将急剧扩大。[29] 于是只剩下模型 (21.9.7) 和 (21.9.8)。在这两种情况下，所估计的 δ 系数都为负，意味着所估计的 ρ 小于 1。对这两个模型而言，所估计的 ρ 分别为 0.9986 和 0.9397。现在惟一的问题是，这些值在统计上是否显著低于 1，使我们能宣布 GDP 时间序列是平稳的。

模型 (21.9.7) 所估计的 τ 值为 -0.2191，在绝对值上甚至低于显著性水平为 10% 时的临界 τ 值 -2.5842。既然前者在绝对值上低于后者，那我们的结论就是 GDP 时间序列不是平稳的。[30]

对模型 (21.9.8) 是同样情况。所计算出来的 τ 值 -1.6252 在绝对值上甚至低于显著性水平为 10% 时的临界 τ 值 -3.1567。

因此，基于图示法、相关图和迪基-富勒检验，在 1970—1991 年按照季度划分的期间，美国 GDP 时间序列是非平稳的；即包含一个单位根。

增广迪基-富勒检验

在进行 (21.9.2)、(21.9.4) 或 (21.9.5) 中的 DF 检验时，假定误差项 u_t 是不相关的。但在 u_t 相关时，迪基和富勒又提出了一个被称为**增广迪基-富勒 (ADF) 检验** (augmented Dickey-Fuller test) 的方法。这一检验通过在上述三个方程中增加因变量 Y_t 的滞后值来进行。具体而言，假设我们使用 (21.9.5)。这里的 ADF 检验由估计如下回归而构成：

$$\Delta Y_t=\beta_1+\beta_2 t+\delta Y_{t-1}+\alpha_i\cdot\sum_{i=1}^{m}\Delta Y_{t-i}+\varepsilon_t \tag{21.9.9}$$

其中 ε_t 为纯粹白噪音误差项，而 $\Delta Y_{t-1}=(Y_{t-1}-Y_{t-2})$，$\Delta Y_{t-2}=(Y_{t-2}-Y_{t-3})$ 等等。所包含滞后差分项的数目通常由实证研究决定，包含足够多的滞后项的思想就是使 (21.9.9) 中的误差项序列不相关。在 ADF 中，我们仍检验 $\delta=0$，而且 ADF 检验服从与 DF 统计量一样的渐近分布，所以可以使用相同的临界值。

818 为了对此程序有个粗略的了解，我们用 GDP 的一个滞后差分对 GDP 序列估计 (21.9.9)；结果如下[31]：

$$\Delta GDP_t=234.9729+1.8921t-0.0786GDP_{t-1}+0.3557\Delta GDP_{t-1}$$

$$t = (2.3833) \quad (2.1522) \quad (-2.2152) \qquad (3.4647) \tag{21.9.10}$$

$$R^2 = -0.1526 \qquad d = 2.0858$$

GDP_{t-1}系数（$=\delta$）的 t（$=\tau$）值为 -2.2152，但这个值在绝对值上甚至远低于显著性水平为 10% 时的临界 τ 值 -3.1570，再次表明，即便考虑了误差项中可能出现的自相关，GDP 序列仍是非平稳的。

对不止一个系数的显著性进行检验：*F* 检验

假设我们估计模型（21.9.5）并检验假设 $\beta_1 = \beta_2 = 0$，即此模型是不带漂移和趋势的 RWM。为了检验这个联合假设，我们可以使用第 8 章中讨论过的约束 F 检验。即先估计（21.9.5）（无约束回归），再估计去掉截距项和趋势项的（21.9.5）。于是我们就使用方程（8.7.9）中所示的约束 F 检验，只是我们不能通过惯常使用的 F 表得到临界 F 值。和 τ 统计量一样，迪基和富勒已经给出了此情形下的临界 F 值，附录 D 的表 D.7 给出了一个样本。习题 21.27 给出了一个例子。

菲利普斯－佩龙单位根检验[32]

DF 检验的一个重要假定是误差项独立同分布。ADF 检验则通过增加回归子差分项的滞后值使 DF 检验考虑了误差项中可能的序列相关。菲利普斯和佩龙（Phillips-Perron，PP）在考虑误差项的序列相关时，没有添加回归子的滞后差分项，而是使用了**非参数统计方法**（nonparametric statistical methods）。由于 PP 检验的渐近分布与 ADF 检验统计量的渐近分布相同，所以我们在此就不深究这个问题了。

对单位根检验的批评[33]

819 我们已经讨论了几个单位根检验，而且还有其他几个。问题是，为什么有这么多的单位根检验？答案在于这些检验的**尺度**（size）与**功效**（power）。一个检验的能力指的是显著性水平（即犯第Ⅰ种类型错误的概率），而一个检验的功效则指虚拟假设是错误的情况下拒绝它的概率。检验功效的计算是用 1 减去犯第Ⅱ种类型错误的概率，后者指接受一个错误虚拟假设的概率。最大的功效就是 1。大多数单位根检验都是基于所研究时间序列有一个单位根（即是非平稳的）的虚拟假设而做出的。对立假设是，这个时间序列是平稳的。

检验的尺度。记得在第 13 章中，我们对名义显著性水平和真实显著性水

平做了区分。DF 检验对其进行的方式很敏感。记住，我们讨论了 DF 检验的三种变化形式：(1) 纯粹的随机步游，(2) 带漂移的随机步游，和 (3) 带漂移和趋势的随机步游。比如，若真实模型是 (1)，而我们估计了 (2)，并断言，在 5%的显著性水平上，这个时间序列是平稳的。但由于这种情况下真实的显著性水平远大于 5%，所以，这个结论可能是错误的。[34] 从模型中排除掉移动平均成分也可能导致尺度扭曲（关于移动平均，参见第 22 章）。

检验的功效。 大多数 DF 类型的检验的功效都很低；即它们倾向于比所能保证的概率更频繁地接受存在单位根的虚拟假设。这有几个方面的原因。第一，检验功效不仅仅取决于样本容量，还取决于数据的（时间）跨度。对给定样本容量 n，时间跨度大的话，功效也更大。因此，基于 30 年中 30 次观测的单位根检验可能比基于 100 天内 100 次观测进行检验的功效更大。第二，若 $\rho\approx 1$ 但不等于 1，则单位根检验会宣布这种时间序列是非平稳的。第三，这些检验都假定了惟一一个单位根；即假定给定时间序列是 $I(1)$。但若一个时间序列整合了更高阶［比方说 $I(2)$］的成分，则可能有不止一个单位根。在后面这种情形下，你或许可以使用**迪基-潘图拉检验**（Dickey-Pantula test）。[35] 第四，若一个时间序列中因 OPEC 的石油禁运等原因而出现了结构性转折，则单位根检验不能捕捉这些转折。

820 因此，在应用单位根检验时，应该牢记这些检验的局限性。当然，佩龙和恩（Ng）、埃利奥特（Elliot）、罗滕伯格（Rothenberg）和斯托克（stock）、富勒及列邦（Leybovenre）等人也对这些检验做过一些修改。[36] 正因为如此，曼德拉和金（Kim）才提议，应该放弃传统的 DF、ADF 和 PP 检验。随着计量经济软件包中越来越多地包含这些新检验，以新代旧的情况很可能会发生，但应该补充说明一点，到目前为止，对单位根假设仍没有一个一贯有效的检验。

§21.10 对非平稳时间序列进行变换

现在我们知道了与非平稳时间序列相关的问题之后，一个实际的问题就是，我们该怎么办？为了避免将一个非平稳时间序列对一或多个非平稳时间序列回归所导致的谬误回归问题，我们必须对非平稳时间序列进行变换，使之变成平稳序列。变换的方法取决于这个时间序列是差分平稳过程（DSP）还是趋势平稳过程（TSP）。我们依次对每种方法展开讨论。

差分平稳过程

若一个时间序列具有一个单位根，则这种时间序列的一阶差分就是平稳

的。[37] 因此，这里的解决办法就是对时间序列取一阶差分。

回到我们考虑的美国 GDP 时间序列中，我们已经看到，它具有一个单位根。现在我们看对 GDP 序列取一阶差分会怎么样。

令 $\Delta GDP_t=(GDP_t-GDP_{t-1})$。为方便起见，记 $D_t=\Delta GDP_t$。现在考虑如下回归：

$$\begin{aligned}\Delta\hat{D}_t&=16.0049-0.06827D_{t-1}\\ t&=(3.6402)(-6.6303)\\ R^2&=0.3435\qquad d=2.0344\end{aligned}\tag{21.10.1}$$

在显著性水平为 1%时，DF 临界 τ 值 −3.507 3。由于计算出来的 τ 值比这个临界值负得更多，所以我们断定一阶差分后的 GDP 是平稳的；即是 $I(0)$。如图 21.9 所示，若将此图与图 21.1 相比，你会看出二者之间明显的差别。

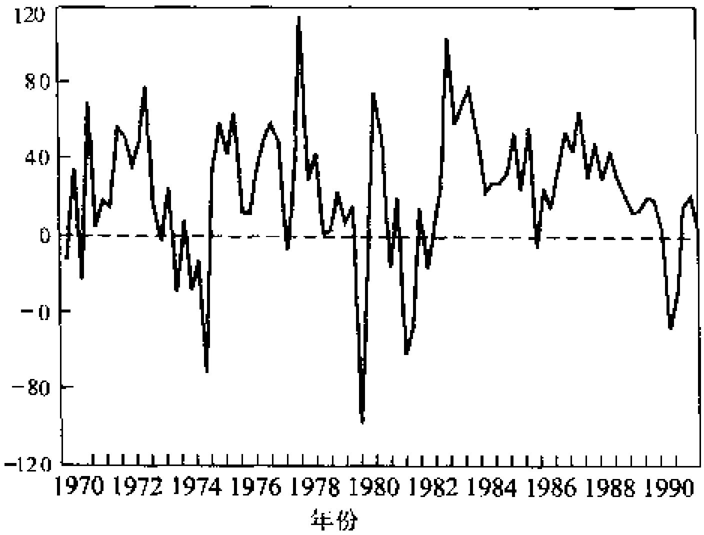


图 21.9　美国 GDP 1970—1991 年季度数据的一阶差分

趋势平稳过程

821 如我们在图 21.5 中所见，一个 TSP 在其趋势线附近是平稳的。因此，使这种时间序列变平稳的最简单办法就是将它对时间做回归，从此回归中所得到的残差将是平稳的。换言之，做如下回归：

$$Y_t=\beta+\beta_2 t+u_t\tag{21.10.2}$$

其中 Y_t 为所考虑的时间序列，t 为按年月度量的趋势变量。

现在

$$\hat{u}_t=(Y_t-\hat{\beta}_1-\hat{\beta}_2 t)\tag{21.10.3}$$

将是平稳的。$\hat{u}_t$ 被称为（线性）**除趋势时间序列**（detrended time series）。

指出趋势有可能是非线性的也很重要。比如，它可能是：

$$Y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + u_t \tag{21.10.4}$$

这是一个二次趋势序列。若果真如此，则从（21.10.4）得到的残差就是（二次）除趋势时间序列。

应该指出，若一个时间序列是 DSP 而我们把它当作 TSP 来处理，这种情况被称为**差分不足**（underdifferencing）。另一方面，若一个时间序列是 TSP 而我们把它当作 DSP 处理，则称之为**过度差分**（overdifferencing）。这种类型的设定误差所造成的后果可能很严重，取决于我们如何处理由此带来的误差项序列相关的性质。[38]

顺便指出，大多数宏观经济时间序列都是 DSP 而不是 TSP。

§21.11 协积：将一个单位根时间序列对另一个单位根时间序列进行回归

822 我们已经被警告过，将一个非平稳的时间序列对另一个非平稳的时间序列进行回归可能导致谬误回归。假设我们考虑的是表 21.1 中所给出的 PCE 和 PDI 时间序列。对这些时间序列单个地进行单位根分析，你会发现它们都是 $I(1)$；即它们都包含一个单位根。于是，假设我们将 PCE 对 PDI 做如下回归：

$$\mathrm{PCE}_t = \beta_1 + \beta_2 \mathrm{PDI}_t + u_t \tag{21.11.1}$$

让我们把它写成：

$$u_t = \mathrm{PCE}_t - \beta_1 - \beta_2 \mathrm{PDI}_t \tag{21.11.2}$$

假设我们现在对 u_t 做单位根分析，并发现它是平稳的；即它是 $I(0)$。这是一个有意思的情况，尽管 PCE_t 和 PDI_t 单个都是 $I(1)$，即它们都具有随机趋势，但它们的线性组合（21.11.2）却是 $I(0)$。也可以说，线性组合抵消了两个时间序列中的随机趋势。若你认为消费和收入是两个 $I(1)$ 变量，则定义为（收入—消费）的储蓄将是 $I(0)$。结果，如（21.11.1）一样将消费对收入做回归将是有意义的（不是谬误回归）。此时我们就说这两个变量是**协积的**（cointegrated）。从经济学上讲，若两个变量之间具有长期或均衡关系，那它们可能会是协积的。经济理论通常用均衡加以表述，比如费雪的货币数量论或购买力平价理论（PPP）等，这里只列出几个。

简言之，若我们验证了从（21.11.1）这种回归中所得到的残差是 $I(0)$ 或平稳序列，则我们曾全面考虑过的传统回归方法论（包括 t 和 F 检验）对涉及（非平稳）时间序列的数据仍可适用。单位根、协积等概念的价值所在，是迫使我们弄清楚回归的残差是否平稳。格兰杰指出："对协积的检验

可看成为避免‘谬误回归’情形而进行的预检验。”[39]

用协积理论的语言来说，一个像（21.11.1）这样的回归可称为**协积回归**（cointegrating regression），斜率参数 β_2 可称为**协积参数**（cointegrating parameter）。协积的概念可推广至含有 k 个回归元的回归模型，此时我们便有 k 个协积参数。

对协积的检验

文献中已经给出几种协积检验的方法。我们这里只考虑两种相对简单的
823 方法：(1) 对从协积回归中估计出来的残差进行 DF 或 ADF 单位根检验，(2) 协积回归德宾-沃森（CRDW）检验。[40]

恩格尔-格兰杰（EG）或增广恩格尔-格兰杰（AEG）检验。我们已经知道如何进行 DF 或 ADF 单位根检验。我们所要做的就是，估计一个像（21.11.1）这样的回归，得到残差，并用 DF 或 ADF 检验。[41]但要给一个警告。由于所估计的 u_t 以所估计的协积参数 β_2 为基础，所以 DF 和 ADF 的临界显著值就不是很合适。恩格尔和格兰杰已经计算了这些值，在参考文献中可以见到。[42]因此，目前所进行的 DF 和 ADF 检验又被称为恩格尔－格兰杰检验（Engle-Granger test）和增广恩格尔-格兰杰检验（augmented Engle-Granger test）。然而，现在有几个软件包已经与其他结果一起给出这些临界值。

让我们现在来阐释这些检验。我们首先将 PCE 对 PDI 回归，并得到如下回归结果：

$$\widehat{\text{PCE}}_t = -171.4412 + 0.9672\text{PDI}_t$$
$$t = (-7.4808)\quad(119.8712)\qquad R^2 = 0.9940 \quad d = 0.5316 \tag{21.11.3}$$

由于 PCE 和 PDI 个别地看都是非平稳序列，因此这个回归有可能是谬误回归。但当我们对从（21.11.3）中得到的残差进行单位根检验时，又得到如下结果：

$$\Delta\hat{u}_t = -0.2753u_{t-1}$$
$$t = (-3.7791)\qquad R^2 = 0.1422 \qquad d = 2.2775 \tag{21.11.4}$$

恩格尔-格兰杰 1%临界 τ 值为 -2.5899。由于计算出来的 $\tau(=t)$ 值在绝对值上比这个临界值大得多，因此我们的结论是，PCE 对 PDI 回归的残差是 $I(0)$；即它们是平稳的。因此，尽管这两个变量个别地看都是非平稳的，但（21.11.3）是一个协积回归而不是谬误回归。你可以称（21.11.3）为**静态**
824 (static) 或**长期**（long run）的消费函数，并把其参数解释为长期参数。因此，0.9672 代表着长期或均衡的边际消费倾向。

协积回归德宾-沃森（CRDW）检验。弄清楚 PCE 和 PDI 是否协积的另一个更快的方法就是 CRDW 检验，此检验的临界值最早由萨根和巴加瓦（Bhargara）首次提供。[43]在 CRDW 中，我们使用从协积回归中得到的德宾-沃森 d 统计量，如在（21.11.3）中给出的 $d=0.5316$。但现在的虚拟假设是 $d=0$ 而不是 $d=2$。这是因为，在第 12 章中我们观察到，$d\approx 2(1-\hat{\rho})$，故若存在一个单位根，则估计的 ρ 将约为 1，这就意味着 d 约为 0。

基于对 100 次观测中每个观测的 10 000 次模拟，检验真实的 $d=0$ 这个假设的 1%、5%、10%临界值分别是 0.511、0.386 和 0.322。因而，若计算的 d 值小于 0.511，我们就在 1%的显著性水平上拒绝协积的虚拟假设。在我们的例子中，$d=0.5316$ 高于此临界值，表明 PCE 和 PDI 是协积的，由此加强了我们基于 EG 检验所得到的结论。[44]

总之，我们基于 EG 和 CRDW 检验的结论都是 PCE 和 PDI 是协积的。[45]尽管它们个别地表现出随机步游，但看来它们之间有一种稳定的长期关系；它们彼此间不会游离太远，图 21.1 便是佐证。

协积与误差纠正机制

我们刚刚证明了 PCE 和 PDI 是协积的，即二者之间有一种长期或均衡的关系。当然，在短期中，有可能会偏离均衡。因此，你可以把（21.11.2）中的误差项视为“均衡误差”。而我们也可以利用这个误差项把 PCE 的短期行为与其长期值联系起来。最早由萨根提出[46]并经恩格尔和格兰杰加以推广的**误差纠正机制**（ECM）（error correction mechanism）就是对失衡状况进行纠正。一个被称为**格兰杰表述定理**（Granger representation theorem）的重要定理表明，若两个变量 Y 和 X 是协积的，则二者之间的关系可由 ECM 表述。为了看出其含义，让我们回到我们的 PCE－PDI 例子中。现在考虑如下模型：

$$\Delta \mathrm{PCE}_t=\alpha_0+\alpha_1\Delta \mathrm{PDI}_t+\alpha_2 u_{t-1}+\varepsilon_t \qquad (21.11.5)$$

其中 Δ 和往常一样表示一阶差分算子，ε_t 为随机误差项，$u_{t-1}=\mathrm{PCE}_{t-1}-\beta_1-\beta_2\mathrm{PDI}_{t-1}$，即从协积回归（21.11.1）中得到的误差的一期滞后值。

ECM 方程（21.11.5）表明，ΔPCE 取决于 ΔPDI 和均衡误差项。[47]若后者非零，则模型就偏离了均衡。假设 ΔPDI 为零而 u_{t-1} 为正。这意味着 PCE_{t-1} 太高而失衡，即 PCE_{t-1} 高于其均衡值 $(\alpha_0+\alpha_1\mathrm{PDI}_{t-1})$。由于预期 α_2 为负，所以 α_2u_{t-1} 这一项就为负，因此为了恢复均衡，$\Delta\mathrm{PCE}_t$ 就必须为负。也就是说，若 PCE_t 高于其均衡值，那么它在下一期就开始下降以纠正均衡误差；由此得名 ECM。同理，若 u_{t-1} 为负（即 PCE 低于其均衡值），则 α_2u_{t-1} 将为正，使得 $\Delta\mathrm{PCE}_t$ 为正，从而导致 PCE_t 在第 t 期上升。因此，α_2 的绝对值决定了均衡恢复的速度有多快。实践中，我们用 $\hat{u}_{t-1}=(\mathrm{PCE}_t-\hat{\beta}_1-\hat{\beta}_2\mathrm{PDI}_t)$ 来估计 u_{t-1}。

回到我们说明性的例子，(21.11.5) 的实证结果为：

$$\begin{aligned}\widehat{\Delta \mathrm{PCE}}_t &= 11.691\,8 + 0.290\,6\Delta \mathrm{PDI}_t - 0.086\,7\hat{u}_{t-1}\\ t &= (5.324\,9)(4.171\,7) \qquad (-1.600\,3)\\ R^2 &= 0.171\,7 \qquad\qquad d = 1.923\,3\end{aligned} \tag{21.11.6}$$

从统计上讲，均衡误差项为0，表明PCE对PDI的变化在同一时期就立即进行调整。(21.11.6) 表明，PDI的短期变化对个人消费的短期变化有正的影响。你可以把0.290 6解释为短期边际消费倾向；长期MPC由所估计的（静态）均衡关系（21.11.3）给出，为0.967 2。

在本节结束之前，S.G. 霍尔（Hall）做出的一些警语值得提请注意：

826 虽然协积概念是误差纠正模型的重要理论基础，但在实际应用方面仍有许多问题。对很大范围内的模型来说，许多检验的临界值和小样本表现都是未知的，所以通过相关图的检验而获得信息依然是一种重要的手段。[48]

§21.12 在经济学中的一些应用

我们以一些简明的例子来结束本章。

例 21.1 美国 M1 货币供给的月度数据：1951 年 1 月至 1999 年 9 月 30 日

图21.10给出了美国从1951年1月至1999年9月30日M1货币供给的月度数据。就我们对平稳性的知识而言，M1货币供给时间序列看起来是非平稳的，用单位根分析可以确定。（注：为节省篇幅，我们没有给出实际数据，但从联邦储备委员会或圣路易斯联邦储备银行可以得到这些数据。）

827

$$\begin{aligned}\widehat{\Delta M}_t &= \ \ 0.261\,8 + \ \ 0.015\,9t - \ \ 0.004\,4M_{t-1}\\ t &= \ (0.791\,9) \quad (4.422\,7)\ (-3.004\,6)\\ R^2 &= 0.067\,0 \qquad d = 0.717\,2\end{aligned} \tag{21.12.1}$$

由于1%、5%和10%的临界 τ 值分别为 $-3.981\,1$、$-3.421\,0$ 和 $-3.132\,9$，而 t 值 $-3.004\,6$ 在绝对值上比这三个临界值中的任何一个都小，所以结论就是，M1时间序列是非平稳的；即它包含一个单位根或是 $I(1)$。即便（按照ADF的方式）引入 ΔM_t 的几个滞后值，结论也没有什么变化。另一方面，我们发现M1货币供给的一阶差分是平稳的（验证之）。

例 21.2 美国和英国货币的汇率：1973 年 1 月 1 日至 1996 年 10 月 10 日

图21.11给出了从1973年1月1日至1996年10月10日（美元/英镑）汇率走势图，共286个观测。到现在，你应该能辨认出这个时间序列是非平稳的了吧。通过进行单位根检验，我们得到如下 τ 统计量：$-1.274\,9$（既无截距又无趋势）、$-1.771\,0$（有截距无趋势）和 $-1.626\,9$（既有截距又有

趋势)。这些统计量中的每一个，在绝对值上都比相应 DF 表中的临界 τ 值小，因而肯定了我们从图上得出美英汇率时间序列非平稳的印象。

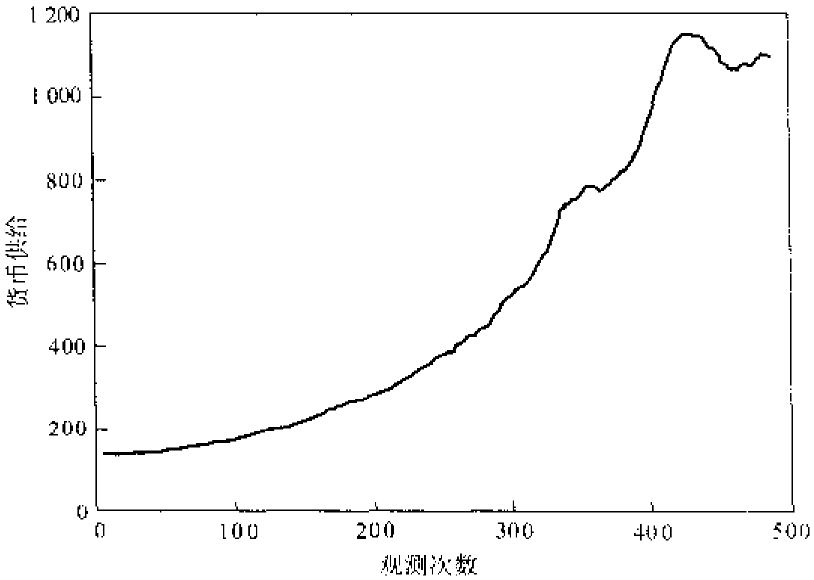


图 21.10　美国货币供给：1951 年 1 月—1999 年 9 月

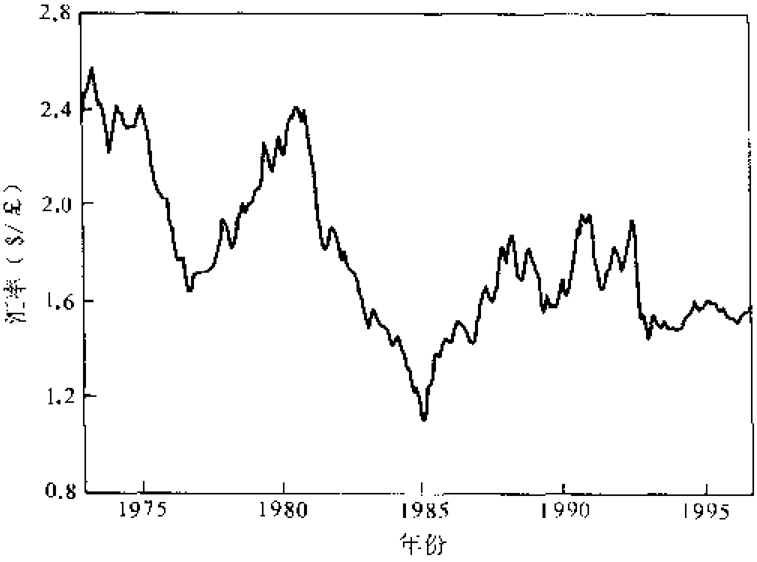


图 21.11　美英汇率：1973 年 1 月—1996 年 10 月

828 **例 21.3　美国消费者价格指数：1947 年 1 月至 2000 年 1 月**

图 21.12 给出了从 1947 年 1 月至 2000 年 1 月美国 CPI 共 649 个观测数据。CPI 序列与前面考虑过的 M1 序列一样，表现出上扬的趋势。单位根检验给出如下结果：

$$\begin{aligned}\Delta \mathrm{CPI}_t &= -0.009\,4+0.000\,51t-0.000\,66\mathrm{CPI}_{t-1}+0.547\,3\Delta \mathrm{CPI}_{t-1}\\ t &= (-0.653\,8)(4.343\,1)(-1.547\,2)\qquad(16.444\,8)\end{aligned}\tag{21.12.2}$$

$R^2 = 0.517\ 7 \qquad d = 2.141\ 0$

CPI_{t-1}的 $t\ (=\tau)$ 值为 $-1.547\ 2$。10%临界值为 $-3.131\ 7$。由于计算出来的 τ 值在绝对值上比临界 τ 值还小，所以我们的结论是，CPI 不是一个平稳的时间序列。我们可以用随机趋势来刻画它。（为什么？）但你若对 CPI 序列取一阶差分，你将发现它们是平稳的。因此，CPI 是一个差分平稳（DS）的时间序列。

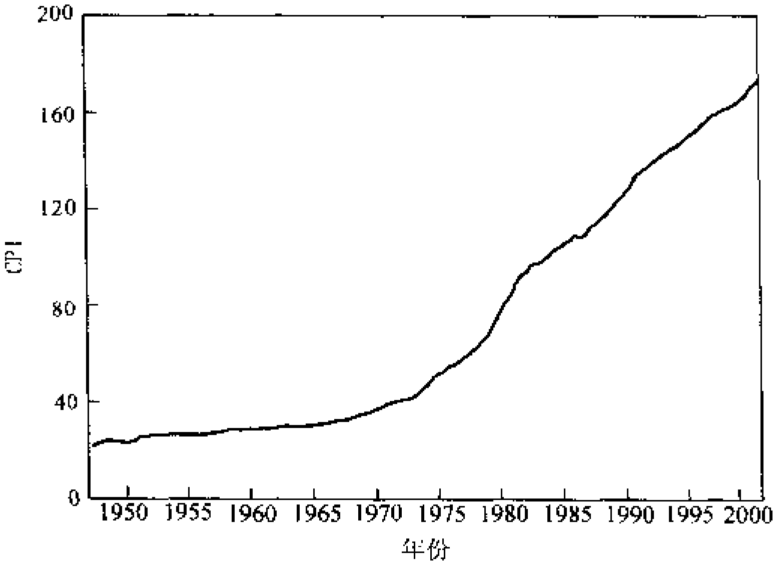


图 21.12 美国 1947 年 1 月—2000 年 1 月的 CPI

例 21.4 3 月期和 6 月期国债利率协积吗?

图 21.13 画出了从 1982 年 1 月至 2001 年 6 月美国 3 月期和 6 月期（固定期）国债的利率，共 234 个观测。此图能显示这两个利率是协积的吗？即二者之间存在一种均衡关系吗？从金融理论来看，我们预期会是这样，否则套利者将利用短期与长期利率间的出入来获利。首先，让我们看这两个时间序列是否平稳。

829 基于随机步游模型（既无截距又无趋势），这两个利率都是平稳的。包含截距、趋势和滞后差分，结果表明这两个利率是趋势平稳的；在这两种情况下，趋势系数都为负，并在约 7%的水平上显著。所以，基于这些结论，我们接受这两个利率要么平稳要么趋势平稳的结论。

将 6 月期国债利率对 3 月期国债利率回归，我们得到如下回归：

$$\begin{aligned}\widehat{TB6}_t &= \quad -0.045\ 6 + \quad 1.046\ 6TB3_t \\ t &= \ (-1.120\ 7)(171.623\ 9) \qquad\qquad (21.12.3)\\ &R^2 = 0.992\ 1 \quad d = 0.405\ 5\end{aligned}$$

对上述回归的残差应用单位根检验，我们发现残差是平稳的，这就表明 3 月期和 6 月期国债利率是协积的。利用这些信息，我们得到如下误差纠正模型（ECM）：

$$\Delta \widehat{TB6}_t = \quad -0.0067 + \quad 0.9360\Delta TB3_t - 0.2030\hat{u}_{t-1}$$

$$t = \ (-0.8662)\ (41.9592) \qquad (-5.3837) \qquad\qquad (21.12.4)$$

$$R^2 = 0.8852 \qquad d = 1.5604$$

其中 $\hat{u}_{t-1}$为上一期误差纠正项的滞后值。如这些结论所示，这两个利率上个月的差异中，有20%在这个月消除掉了。[49]此外，3月期国债利率的短期变化很快就能反映在6月期国债上，因为二者之间的斜率系数为0.9360。鉴于美国货币市场的有效性，这一结论无足为奇。

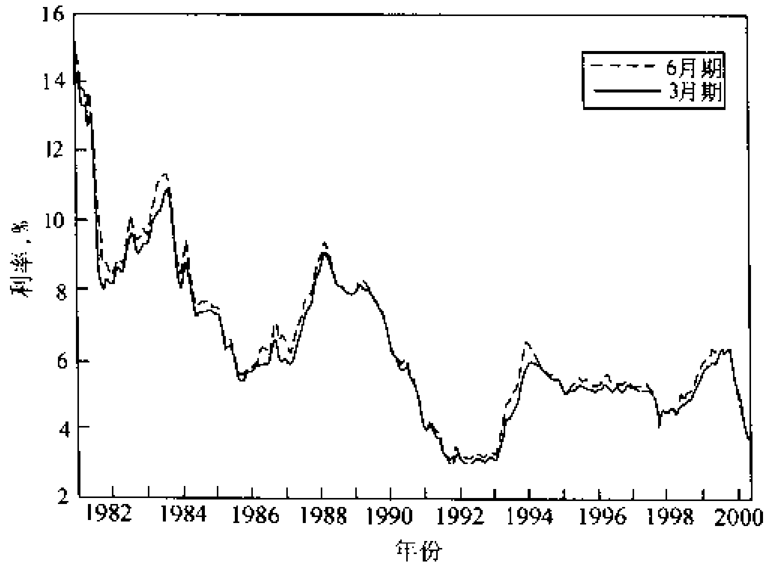


图 21.13　3 月期和 6 月期国债利率（到期日相同）

§21.13　要点与结论

830 1. 依据时间序列数据的回归分析，隐含地假定了所依据的时间序列是平稳的。经典的 t 检验、F 检验等等均以此假定作为依据。

2. 实践中大多数经济时间序列都是非平稳的。

3. 如果一个随机过程的均值、方差和自协方差在时间上是恒定的［即它们是历时不变的（time-invariant）］，那它就是**弱平稳的**。

4. 在一个非正式的（判别）水准上，弱平稳性可通过时间序列的相关图即各种滞后的自相关图形来检验。对于平稳时间序列来说，相关图很快会变平，而对非平稳时间序列来说，它则消失得很缓慢。对于一个纯随机序列，所有滞后1以上的自相关均为零。

5. 在一个正式的（判别）水准上，平稳性可通过时间序列是否含有单位根来检查。为此可利用**迪基－富勒**（DF）或**增广迪基－富勒**（ADF）检验。

6. 一个经济时间序列可以是**趋势平稳**（TS）或差分平稳（DS）的。一个 TS 时间序列有一确定的趋向。而一个 DS 时间序列则有着可变的或随机的趋向。通常在回归模型中引进一个时间或趋势变量的做法，仅对 TS 时间序列是合理的。DF 和 ADF 检验可用于判定一个时间序列是 TS 还是 DS。

7. 一个时间序列变量对另一个或多个时间序列变量做回归，常常会引出无意义的或谬误的结果。这种现象被称为**谬误回归**，提防它的一个方法是判明这些时间序列是否有协积关系。

8. **协积**是指尽管两个或多个时间序列个别而论是非平稳的，但它们的线性组合则可以是平稳的。EG、AEG 和 CRDW 检验可用来判明两个或多个时间序列是否有协积关系。

9. 两个（或多个）时间序列的协积（关系）表明它们之间有一种长期或均衡关系。

10. 由恩格尔和格兰杰研究出来的**误差纠正机制**（ECM）是协调经济变量短期行为及其长期行为的一种手段。

11. 时间序列计量经济学领域正在扩展中，已建立的一些结果和检验在某些情形中仍是尝试性的，还有许多工作要做。一个需要回答的重要问题是，为什么一些经济时间序列是平稳的，而另一些又是非平稳的。

习　题

问答题

21.1　什么是弱平稳性？

21.2　什么是单积时间序列？

831 **21.3**　单位根的意义何在？

21.4　如果时间序列是 $I(3)$，你要对它取多少次差分才使它变为平稳的？

21.5　什么是迪基-富勒和 ADF 检验？

21.6　什么是恩格尔-格兰杰和 AEG 检验？

21.7　协积的意义何在？

21.8　单位根检验与协积之间是否有差别，如果有，差别何在？

21.9　什么是谬误回归？

21.10　协积与谬误回归之间有何联系？

21.11　确定性趋势与随机性趋势之间的差别何在？

21.12　什么是趋势平稳过程？什么是差分平稳过程？

21.13　什么是随机步游（模型）？

21.14　“对随机步游式的随机过程来说，方差是无限大的。”你同意吗？为什么？

21.15　什么是误差纠正机制？它和协积有什么关系？

解答题

21.16 利用表21.1的数据，做出时间序列PCE、PDI、利润和股息直至25阶滞后的样本相关图，你看到了什么一般性的模式？凭直觉，哪些时间序列像是平稳的？

21.17 对习题21.16中的每一时间序列用DF检验去判明这些序列是否含有单位根，如果有单位根存在，你又将怎样刻画这样一个时间序列？

21.18 继续做习题21.17。你怎样决定ADF检验是否比DF检验更合适？

21.19 考虑表21.1中的股息和利润时间序列。由于股息依赖于利润，考虑以下的简单模型：

$$股息_t = \beta_1 + \beta_2\ 利润_t + u_t$$

a. 你预料此回归会受谬误回归现象的影响吗？为什么？

b. 股息和利润两时间序列是不是协积的？你怎样对此做出明显的检验？如果经过检验你发现它们是协积的，你会改变你对（a）的回答吗？

c. 利用误差纠正机制（ECM）去研究股息与利润的关系中的短期和长期行为。

d. 如果你个别地分析股息与利润序列，它们呈现随机性的抑或确定性的趋势？你使用什么检验？

*e. 假定股息与利润是协积的，那么，你用利润对股息的回归代替股息对利润的回归，这样的回归是否有效？

832 **21.20** 取表21.1所给时间序列的一阶差分并描图。仍然对每一时间序列做出直至滞后25期的相关图。这些相关图有些什么可引起你注意的地方吗？

21.21 假设你不去做股息对利润的水平形式的回归，而代之以股息的一阶差分对利润的一阶差分的回归。你会在这个回归中引进截距项吗？为什么或为什么不？说明你的计算。

21.22 继续上题。你会怎样检验一阶差分回归的平稳性呢？在本例中，你会有什么事先的预期且为什么？说明全部计算。

21.23 根据1948—1984年期间英国私有部门的私房动工数（X），米尔斯（Terence Mills）得到如下回归结果[1]：

$$\Delta X_t = 31.03 - 0.188 X_{t-1}$$
$$\text{se} = (12.50)(0.080)$$
$$(t=)\ \tau\ (-2.35)$$

注：5%临界τ值是-2.95，10%临界τ值是-2.60。

a. 根据这些结果，新房动工时间序列是平稳的，还是非平稳的？或者问，在此时间序列中有没有单位根？你怎样知道的？

b. 如果你用了平常的 t 检验，那么所测的 t 值是不是统计上显著的？根据这点你会做出结论说此时间序列是平稳的吗？

c. 现在考虑如下回归结果：

$$\Delta^2 X_t = 4.76 - 1.39\Delta X_{t-1} + 0.313\Delta^2 X_{t-1}$$
$$\text{se} = (5.06)(0.236) \quad (0.163)$$
$$(t=)\tau \quad (-5.89)$$

其中 Δ^2 是二阶差分运算子，也就是一阶差分的一阶差分。现在所估 τ 值是统计上显著的。那么你能对所考虑的时间序列的平稳性说些什么？

注：上述回归的目的是为了找出该时间序列是否有第二个单位根。

21.24 如（21.7.1）和（21.7.2）所示，生成两个随机步游序列，并将一个对另一个回归。用它们的一阶差分重做这个练习，并验证此回归中的 R^2 约为 0，而德宾-沃森 d 接近于 2。

833 **21.25** 为了说明各含确定性趋势的两个变量可能导致谬误回归，Charemza 等人基于 30 次观测得到如下回归[2]：

$$\hat{Y}_t = 5.92 + 0.030X_t$$
$$t = (9.9) \quad (21.2)$$
$$R^2 = 0.92 \qquad d = 0.06$$

其中 $Y_1=1$，$Y_2=2$，…，$Y_n=n$ 和 $X_1=1$，$X_2=4$，…，$X_n=n^2$。

a. Y 表现出什么趋势？X 又表现出什么趋势？

b. 描出这两个变量并画出回归线。从描点图中你能得到什么一般性结论？

21.26 利用加拿大 1971 年第 1 季度至 1988 年第 4 季度期间的数据，得到如下回归结果：

1. $$\ln \text{M1}_t = -10.257\,1 + 1.597\,5\ln\text{GDP}_t$$
$$t = (-12.942\,2)\ (25.886\,5)$$
$$R^2 = 0.946\,3 \qquad d = 0.325\,4$$

2. $$\Delta\ln \text{M1}_t = 0.009\,5 + 0.583\,3\Delta\ln\text{GDP}_t$$
$$t = (2.495\,7)(1.895\,8)$$
$$R^2 = 0.088\,5 \qquad d = 1.739\,9$$

3. $$\Delta\hat{u}_t = -0.195\,8\hat{u}_{t-1}$$
$$(t=\tau)\ (-2.252\,1)$$
$$R^2 = 0.111\,8 \qquad d = 1.476\,7$$

其中 M1 = M1 货币供给，GDP = 国内生产总值，均以 10 亿加元度量，ln 为自然对数，而 $\hat{u}_t$ 表示从回归 1 中得到的残差。

a. 解释回归 1 和 2。

b. 你怀疑回归 1 是谬误回归吗？为什么？

c. 回归 2 是谬误回归吗？你如何知道？

d. 利用回归 3 的结果，你会改变你在 b 部分的结论吗？为什么？

e. 现在考虑如下回归：

$$\Delta\ln M1_t = 0.0084 + 0.7340\Delta\ln GDP_t - 0.0811\hat{u}_{t-1}$$
$$t = (2.0496)(2.0636) \qquad (-0.8537)$$
$$R^2 = 0.1066 \qquad d = 1.6697$$

此回归告诉你什么信息？它能帮助你决定回归 1 是否为谬误回归吗？

834 **21.27** 如下回归是基于美国1960—1999年期间共 40 个年度观测的 CPI 数据而得出的：

1. $\Delta CPI_t = 0.0372 CPI_{t-1}$
$$t = (9.6427)$$
$$R^2 = 0.0304 \qquad d = 0.5259 \qquad RSS = 203.6222$$

2. $\Delta CPI_t = 1.8052 + 0.0208 CPI_{t-1}$
$$t = (2.5000)(2.7583)$$
$$R^2 = 0.1705 \qquad d = 0.6030 \qquad RSS = 174.1966$$

3. $\Delta CPI_t = 1.8790 + 0.5706t - 0.1158 CPI_{t-1}$
$$t = (3.1460)(4.2576)(-3.5443)$$
$$R^2 = 0.4483 \qquad d = 0.7969 \qquad RSS = 115.8579$$

其中 RSS= 残差平方和。

a. 考察上述回归，你对 CPI 时间序列的平稳性有何看法？

b. 你如何在这三个模型中做出选择？

c. 方程（1）比方程（3）缺少截距项和趋势项。为了判定模型 1 所隐含的约束是否成立，你将使用哪个检验？（提示：利用迪基－富勒 t 和 F 检验。并使用附录 D 中表 D.7 所给出的近似值。）

【习题注释】

[1] Terence C. Mills 的前引文献，第 127 页。符号略有改变。

[2] Charemza 等人的前引文献，第 93 页。

【注释】

[1] 在初级层次上，如下参考资料对你有所帮助：Gary Koop, *Analysis of Econometric Data*, John Wiley & Sons, New York, 2000; Jeff B. Cromwell, Walter C. Labys, and Michel Terraza, *Univariate Tests for Time Series Models*, Sage Publications, California, Ansbury Park, 1994; H. R. Seddighi, K. A. Lawler, and A. V. Katos, *Econometrics: A Practical Approach*, Routledge, New York, 2000. 中级水平，可参见 Walter Enders, *Applied Econometric Time Series*, John Wiley & Sons, New York, 1995; Kerry Patterson, *An Introduction to Applied Econometrics: A Time Series Ap-*

proach, St. Martin' Press, New York, 2000; T. C. Mills, *The Econometric Modelling of Financial Time Series*, 2d ed., Cambridge University Press, New York, 1999; Marno Verbeek, *A Guide to Modern Econometrics*, John Wiley & Sons, New York, 2000; Wojciech W. Charemza and Derek F. Deadman, *New Directions in Econometric Practice: General to Specific Modelling and Vector Autoregression*, 2d ed., Edward Elgar Publisher, New York, 1997; 高级水平，可参见 Hamilton, J.D., *Time Series Analysis*, Princeton University Press, Princeton, N.J., 1994; G.S. Maddala and In-Moo Kim, *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, 1998. 在应用层次上，可参见 B. Bhaskara Rao, ed., *Cointegration for the Applied Economist*, St. Martin's Press, New York, 1994; Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, New York, 1998.

[2] 当然，我们现在已经有了这一期间的实际数据，并将实际数据与基于前一期间"预测的"数据相比较。

[3] 如下讨论基于曼达拉等人的前引文献和查伦扎等人的前引文献。

[4] "随机"一语源自希腊语"stokhos"一词，意思是靶子或靶心。如果你曾经在一个圆靶上玩过掷飞镖游戏并试图击中靶心的话，你能击中靶心的机会有多大？在扔 100 次飞镖中，你可能有幸几次射中靶心；其他时候，飞镖将随机地散布在靶心周围。

[5] 你可以把28 728亿美元看成 1970 年第 1 季度 GDP 所有可能值的均值。

[6] 如果一个时间序列概率分布的所有阶矩［而不仅仅是一阶矩和二阶矩（即均值和方差）］都不随时间变化，那它就是严格平稳的。但如果这个平稳过程是正态的，那么弱平稳过程也是严格平稳的，因为正态的随机过程完全可由其均值和方差这两个矩来确定。

[7] 这一论点见 Keith Cuthbertson, Stephen G. Hall, and Mark P. Taylor, *Applied Econometric Techniques*, The University of Michigan Press, 1995, p.130。

[8] 如果它还是独立的，则这种过程可称为**严格白噪音**（strictly white noise）。

[9] 随机步游常比做一个醉汉的步游。醉汉离开酒吧后在时刻 t 移动一个随机的距离 u_t，如果他无限制地继续步游下去，他将最终漂移到离酒吧越来越远的地方。股票的价格也是这样，今天的股价等于昨天的股价加上一个随机冲击。

[10] Kerry Patterson 的前引文献第 6 章。

[11] 一个技术性注释：若 $\rho=1$，则我们可把（21.4.1）写成 $Y_t-Y_{t-1}=u_t$。现在利用**滞后算子**（lag operator）L，所以 $LY_t=Y_{t-1}$，$L^2Y_t=Y_{t-2}$，如此等等，我们可以把（21.4.1）写成 $(1-L)Y_t=u_t$。"单位根"一词指的是滞后算子多项式的根。若你令 $(1-L)=0$，则得到 $L=1$，由此得名单位根。

[12] 若在（21.4.1）中假定：Y 的初始值（$=Y_0$）为 0，$|\rho|\leqslant 1$，而且 u_t 是白噪音并服从零均值和单位方差的正态分布，则得到 $E(Y_t)=0$ 和 $\operatorname{var}(Y_t)=1/(1-\rho^2)$。由于它们都是常数，所以根据弱平稳性的定义，$Y_t$ 就是平稳的。另一方面，如我们前面所见，若 $\rho=1$，则 Y_t 是一个随机步游或非平稳序列。

[13] 一个时间序列可能包含不止一个单位根。但我们在本章后面再讨论这种情况。

[14] 以讨论都基于 Wojciech W. Charemza 等人的前引文献，第 89－91 页。

[15] 例如，若 Y_t 是 $I(2)$，则 $\Delta\Delta Y_t=\Delta(Y_t-Y_{t-1})=\Delta Y_t-\Delta Y_{t-1}=Y_t-$

$2Y_{t-1}+Y_{t-2}$就变成平稳的了。但注意 $\Delta\Delta Y_t=\Delta^2 Y_t\neq Y_t-Y_{t-2}$。

[16] 这一点源于 Maddala 等人的前引文献，第 26 页。

[17] Yule, G.U., "Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series," *Journal of the Royal Statistical Society*, vol.89, 1926, pp.1－64. 至于谬误回归方面大量的蒙特卡罗模拟，参见 C.W.J.Granger and P.Newbold, "Spurious Regressions in Econometrics," *Journal of Econometrics*, vol.2, 1974, pp.111－120。

[18] 严格地讲，我们应该将滞后 k 阶的样本协方差除以（$n-k$），并将样本方差除以（$n-1$）而不是 n（为什么?），其中 n 为样本容量。

[19] M.S.Bartlett, "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series," *Journal of the Royal Statistical Society*, Series B, vol.27, 1946, pp.27－41.

[20] 我们的样本容量是 88 次观测，尽管不是很大，但就使用正态近似而言也足够了。

[21] 换言之，若你将任何一个 ρ_k 的估计值除以标准误（$\sqrt{1/n}$），对充分大的 n，你会得到标准的 Z 值，其概率很容易从正态分布表中查到。因此，对估计的 $\rho_{10}=0.638$，Z 值为 $0.638/0.1066=5.98$（近似）。若真实的 ρ_{10}实际上为 0，则得到一个不小于 5.98 的 Z 值的概率就很小，因此拒绝 ρ_{10}为 0 的虚拟假设。

[22] G.E.P.Box and D.A.Pierce, "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, vol.65, 1970, pp.1509－1526.

[23] G.M.Ljung and G.P.E.Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, vol.66, 1978, pp.66－72.

[24] Q 和 LB 统计量并不是在每种情况下都适当。一种批评意见，可参见 Maddala 等人的前引文献，第 19 页。

[25] 由于 $\delta=(\rho-1)$，所以平稳性要求 ρ 必须小于 1。若然，则 δ 一定为负。

[26] D.A.Dickey and W.A.Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, vol.74, 1979, pp.427－431. See also W.A.Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, New York, 1976.

[27] J.G.MacKinnon, "Critical Values of Cointegration Tests," in R.E.Engle and C.W.J.Granger, *eds.*, *Long-Run Economic Relationships: Readings in Cointegration*, Chap. 13, Oxford University Press, New York, 1991.

[28] 我们排除了 $\delta>0$ 的可能性，因为在那种情况下有 $\rho>1$，此时的时间序列将急剧扩大。

[29] 更技术性地讲，由于（21.9.2）是一阶差分方程，所以所谓的稳定性条件就要求 $|\rho|<1$。

[30] 对此的另一种陈述方法是，计算出来的 τ 值应该比临界 τ 值负得更多，这里不是这种情况。因此结论成立。由于一般预期 δ 为负，所以估计的 τ 统计量具有负号。因此，一个绝对值很大的负 τ 值通常是平稳性的一个迹象。

[31] 更高阶滞后差分也考虑过，但它们都不显著。

[32] P.C.B.Phillips and P.Perron, "Testing for a Unit Root in Time Series Regres-

sion," *Biometrika*, vol.75, 1988, pp.335－346. 如今有几个软件包已经包含了 PP 检验。

[33] 详细讨论参见 Terrence C.Mills 的前引文献，第 87 88 页。

[34] 对此的一个蒙特卡罗实验，参见 Charemza 等人的前引文献，第 114 页。

[35] D.A.Dickey and S.Pantula, "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics*, vol.5, 1987, pp.455－461.

[36] 对这些检验的讨论可见于 Maddala 等人的前引文献，第 4 章。

[37] 若一个时间序列是 $I(2)$，则它将包含两个单位根，此时我们必须对它两次差分。若它是 $I(d)$，则必须对它 d 次差分，其中 d 为任意整数。

[38] 对此的详细讨论参见 Maddala 等人的前引文献第 2.7 节。

[39] C.W.J.Granger, "Developments in the Study of Co-Integrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, vol.48, 1986, p.226.

[40] 单位根检验与协积检验之间有这种差别。如 David A.Dickey, Dennis W.Jansen, and Daniel I.Thornton 观察到："单位根检验是针对单变量（惟一）时间序列，而相比之下，协积对付的是一组变量之间的关系，其中（无条件地）每个变量都有一个单位根。"参见他们的论文："A Primer on Cointegration with an Application to Money and Income," *Economic Review*, Federal Reserve Bank of St.Louis, March-April 1991, p.59。顾名思义，这篇文章是对协积检验的精彩介绍。

[41] 若 PCE 和 PDI 不是协积的，则它们的任意线性组合都将是非平稳的，u_t 也因此是非平稳的。

[42] R.F.Engle and C.W.Granger, "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, vol.55, 1987, pp.251－276.

[43] J.D.Sargan and A.S.Bhargava, "Testing Residuals from Least-Squares Regression for being Generated by the Gaussian Random Walk," *Econometrica*, vol.51, 1983, pp.153－174.

[44] 关于 CRDW 优于 DF 的问题有大量的争论，在参考文献中可以见到，争辩围绕两个统计量的功效即不犯第 II 类错误（指接受了错误的虚拟假设）的概率而进行。比如说，恩格尔和格兰杰就认为 ADF 比 CRDW 检验要好。

[45] EG 和 CRDW 检验现在又被功效更大的、由 Johansen 推出的检验所补充（取代），但由于其中涉及的数学相当复杂，所以 Johansen 方法的讨论超出了本书的范围，但有几个软件包现在使用了这种方法。

[46] J.D.Sargan, "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in K.F.Wallis and D.F.Hendry, eds., *Quantitative Economics and Econometric Analysis*, Basil Blackwell, Oxford, U.K., 1984.

[47] 以下讨论基于 Gary Koop 的前引文献，第 159－160 页，以及 Kerry Peterson 的前引文献，第 8.5 节。

[48] S.G.Hall, "An Application of the Granger and Engle Two-Step Estimation Procedure to the United Kingdom Aggregate Wage Data," *Oxford Bulletin of Economics and Statistics*, vol.48, no.3, August 1986, pp.238. See also John Y.Campbell and Pierre Perron, "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," NBER (National Bureau of Economic Research) *Macroeconomics Annual 1991*, pp.141－219.

[49] 由于两个国库券的利率都以百分比形式表示，所以这就表明，若 6 月期国库券利率与 3 月期国库券利率之差高于上个月先验预期的大小，则这个月末预期到的利率差别将会减少 0.20 个百分点以恢复二者之间的长期关系。至于短期与长期利率关系的基础理论，可参见任何一本货币银行学教材，或研读利率的期限结构理论。

第 22 章　时间序列计量经济学：预测

835 我们在引言中曾指出，预测是计量经济分析的重要部分，对某些人来说可能是最重要的部分。我们如何预测诸如 GDP、通货膨胀、汇率、股票价格、失业率及其他各种各样的经济变量呢？本章将讨论已经相当流行的两种预测方法：（1）**自回归求积移动平均**（autoregressive integrated moving average，ARIMA）法，普遍称它为**博克斯-詹金斯**（Box-Jenkins）方法论[1]；(2)**向量自回归**（vector autoregression，VAR）。

我们在本章还要讨论与预测金融资产价格（比如股票价格和汇率等）相关的一些特殊问题。这些资产价格可用**群集波动**（volatility clustering）的现象来刻画，即在相当长的时期内表现出急剧波动，而在接下来的一段时期内却又相对平静。你只须看一下近来的道·琼斯指数就明白了。所谓的**自回归条件异方差**（autoregressive conditional heteroscedasticity，ARCH）或**广义自回归条件异方差**（generalized autoregressive conditional heteroscedasticity，GARCH）模型就能刻画这种群集波动。经济预测的主题十分广泛，已有一些这方面的专著出版。本章的目标只是给读者留下这方面的粗略印象。感兴趣的读者可查阅参考文献以做进一步研究。幸运的是，为了方便使用者，大多数现代计量软件包都对本章所讨论的几个方法加以介绍。

836 本章与前一章的联系在于，以下讨论的预测方法都假定所用的时间序列是平稳的，或者可以通过适当变换而使之平稳。随着本章的推进，你将看到

我们在上一章引入的几个概念的用处。

§22.1 经济预测方法

宽泛地说，依据时间序列数据进行经济预测的方法有五种：(1) 指数平滑法，(2) 单一方程回归模型，(3) 联立方程回归模型，(4) 自回归求积移动平均模型，及 (5) 向量自回归模型。

指数平滑法[2]

它们都是一些针对给定时间序列的历史数据拟合出一条适当曲线的基本方法。有一系列这种方法，如单指数平滑法（single exponential smoothing）、豪尔特线性法（Holt's linear method）、豪尔特-温特斯方法（Holt-Winters' method）及其各种变形。尽管它们在商业和经济预测等领域仍在使用，但现在已被前面提到的其他四种方法所补充（取代?）。为了不离题太远，我们在本章就不再讨论它们。

单方程回归模型

本书的绝大部分都在讨论单方程回归模型。作为单方程回归模型的一个例子，且考虑对汽车的需求函数。根据经济理论，我们假定汽车需求是汽车价格、广告费用、消费者收入、利率（作为借款成本的一个衡量）以及其他有关变量（如家庭规模、到工作地点的距离）的函数。我们从时间序列数据估计一个适当的汽车需求模型（或者线性、或者对数线性、或者非线性），以期能用于预测将来对汽车的需求。当然，如第 5 章所指出的，如果我们眺望过于遥远的将来，预测误差会迅速增加。

联立方程回归模型[3]

在第 18、19 和 20 章里，我们曾考虑过联立方程模型。在这种模型全盛
837 时期的 20 世纪 60 和 70 年代里，基于它而精心制作的美国经济模型曾支配着经济预测的整个领域。但近年来，由于 1973 年和 1979 年的油价冲击（因为 OPEC 的石油禁运），也由于**卢卡斯批判**[4]，联立方程预测的昔日辉煌已转入低潮。卢卡斯批判的锋芒在于：所估计的计量经济模型的参数乃依赖于模型被估时所奉行的政策。若政策有所改变，参数亦将随之改变。简言之，

当政策改变时，所估参数并非不变。

例如，1979 年 10 月联邦储备银行突然改变其货币政策，宣布今后政策不再瞄准利率，而是监控货币供给的增长率。随着这种明显的变化，根据估计的计量经济模型在新的治理时期里就不会有什么预测价值。现如今，联储又把重点从控制货币供给转向控制短期利率（联邦基金利率）。

ARIMA 模型

博克斯与詹金斯所著《时间序列分析：预测与控制》一书的问世，带来了新一代的预测工具。这种普遍称之为博克斯－詹金斯（BJ）方法论或技术性地称之为 ARIMA 方法论的新预测方法，在“让数据自己说话”的哲理的指引下，着重于分析经济时间序列本身的概率或随机性质，而不在意于构造单一方程抑或联立方程模型。在 BJ 型时间序列模型中，Y_t 可由其自身的过去或滞后值以及随机误差项来解释，而不像回归模型那样，用 k 个回归元 X_1，X_2，…，X_k 去解释 Y_t。正因为这样，ARIMA 模型不是从任何经济理论推演出来的，所以有时被称为乏理论（atheoretic）模型，另一方面，联立方程模型却常常以经济理论为其基础。

顺便提一句，注意到，我们在本章中强调单变量 ARIMA 模型，即只包含一个时间序列的 ARIMA 模型。但这一分析可推广到多变量 ARIMA 模型。

VAR 模型

VAR 方法论同时考虑几个内生变量，就此而言，它看起来类似于联立方程模型。但是，在 VAR 模型中，每一内生变量都由它的滞后或过去值以及模型中所有的其他内生变量的滞后或过去值来解释。通常，模型中没有任何外生变量。

838 在本章的其余部分里，我们讨论经济预测的博克斯-詹金斯以及 VAR 方法的基本内容。我们的讨论是初等的和直觉的。想进一步探讨这些问题的读者，建议他们阅读参考文献。[5]

§22.2 时间序列数据的 AR、MA 和 ARIMA 建模

我们通过表 21.1 中的美国 GDP 时间序列数据介绍几个概念。其中有一些是旧的，而另一些则是新的。该时间序列的图形已由图 21.1（未经差分的 GDP）和图 21.9（一阶差分后的 GDP）给出；记得水平值形式的 GDP 是非平稳的，但其（一阶）差分形式则是平稳的。

如果一个时间序列是平稳的，则有多种方法建立它的模型。

自回归过程

令 Y_t 代表时期 t 的 GDP。如果我们把 Y_t 的模型写为：

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \tag{22.2.1}$$

其中 δ 是 Y 的均值，而 u_t 是有零均值和恒定方差 σ^2 的不相关随机误差项（即 u_t 是白噪音），则我们说 Y_t 遵循一个**一阶自回归**（first-order autoregressive）或 AR（1）随机过程。这个过程我们曾在第 12 章里遇见过。这里，Y_t 在时期 t 的值依赖于它在前一时期的值和一个随机项，并且将 Y 值表示为对其均值的离差。换句话说，此模型表明 Y 在 t 时期的预测值，不外是它在（$t-1$）期的值的一个比例部分加上在 t 时期的一个随机冲击或干扰；Y 仍然被表示为离开其均值的周围数值。

但如果我们考虑这样的模型：

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t \tag{22.2.2}$$

我们就说 Y_t 遵循一个**二阶自回归**（second-order autoregressive）或 AR（2）过程。就是说，t 时期的 Y 值依赖于它在先前两个时期的值，Y 仍被表示为对其均值 δ 的离差。

一般，我们有：

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \cdots + \alpha_p(Y_{t-p} - \delta) + u_t \tag{22.2.3}$$

这时 Y_t 是一个 p **阶自回归**（pth-order autoregressive）或 AR(p)过程。

839 注意，所有上述模型仅涉及现期和前期的 Y 值，再没有其他的回归元。在这个意义上，我们说"让数据自己说话"。它们是我们讨论联立方程模型时遇到过的一种诱导型模型。

移动平均过程

刚才讨论的 AR 过程并非产生 Y 的惟一可能的机制。假令我们把 Y 的模型描述为：

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \tag{22.2.4}$$

其中 μ 是常数，并且 u 和前面一样，是白噪音随机误差项。t 时期的 Y 等于一个常数加上现在和过去误差项的一个移动平均值。因此，像这样的情形，我们就说 Y 遵循一个**一阶移动平均**或 MA(1)过程。

但如果 Y 的表达式为：

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} \tag{22.2.5}$$

则它是一个 MA (2) 过程，更一般地：

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \cdots + \beta_q u_{t-q} \tag{22.2.6}$$

是一个 MA (q) 过程。总之，移动平均过程不外是一些白噪音误差项的一个线性组合。

自回归与移动平均过程

当然，Y 很可能兼有 AR 和 MA 的特性，从而它是 ARMA。比如说，如果 Y_t 可以写为：

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \tag{22.2.7}$$

其中有一自回归项和一移动平均项，那么它就是一个 **ARMA** (1，1) 过程。(22.2.7) 中的 θ 代表一个常数项。

一般，在一个 **ARMA** (***p***，***q***) 过程中将有 p 个自回归和 q 个移动平均项。

自回归求积移动平均过程

以上所讨论的时间序列模型建立在如下假定的基础上：所考虑的时间序列是在第 21 章的定义下（弱）平稳的。简单地说，一个弱平稳时间序列的均值和方差都是常数，并且它的协方差有时间上的不变性。但是我们知道许多经济时间序列是非平稳的，即它们是经过**单积的** (integrated)；例如，表 21.1 中的经济时间序列就是单积的。

但我们也在第 21 章中看到，如果一个时间序列是 1 阶单积的［即它是
840 I(1)］，那么它的一阶差分就是 I(0)，即平稳的。类似地，如果一个时间序列是 I(2)，则它的 2 阶差分就是 I(0)。一般地，如果一个时间序列是 $I(d)$，那么将它差分 d 次就得到一个 I(0) 序列。

因此，如果我们必须将一个时间序列差分 d 次，把它变为平稳的，然后用 ARMA (p，q) 作为它的模型，那么，我们就说那个原始的时间序列是 ARIMA (p，d，q)，也就是说它是一个**自回归求积移动平均** (autoregressive integrated moving average) 时间序列。其中 p 指自回归项数，d 指序列成为平稳之前必须取其差分的次数，而 q 指移动平均项数。例如，一个 ARIMA (2，1，2) 时间序列在它成为平稳序列之前必先差分一次 ($d=1$)，然后方可用一个 ARMA (2，2) 过程作为这个（一阶差分）平稳时间序列的模型，使它有 2 个 AR 和 2 个 MA 项。当然，如果 $d=0$（即开始便有一个平稳序列），则有 ARIMA(p，$d=0$，q) = ARMA(p，q)。注意，一个 ARIMA(p，0，0) 过程意味着一个纯 AR(p) 平稳过程；一个 ARIMA(0，0，q) 则意味着一个纯 MA(q) 平稳过程。给定 p、d 和 q 的值，我们就能说出模型是怎样一个过程。

应用博克斯-詹金斯（BJ）方法论时，要注意的一个重要问题是，我们必须有一平稳的时间序列，或者是经过一次或多次差分而变为平稳的时间序列。假定平稳性的原因，可解释如下：

> BJ（博克斯-詹金斯）的目的，是要辨别并估计一个可解释为产生现有样本数据的统计模型。如果现在要把所估计的模型用于预测，我们必须假定该模型的特征在不同时期里特别是在将来的时期里保持不变。因此，要求有平稳数据的简单理由是，从这些数据推测出来的任何模型本身就可解释为平稳的或稳定的，从而为预测奠定有效的基础。[6]

§22.3 博克斯-詹金斯方法论

一个一问值万金的显然问题是：面对一个时间序列，例如图 21.1 中的美国 GDP 序列，我们怎样知道它是遵循纯 AR 过程（若然的话，p 取什么值)、纯 MA 过程（若然的话，q 又取什么值)、ARMA（若然的话，p 和 q 各取什么值）还是 ARIMA 过程，这时我们必须知道 p、d 和 q 的值。在回答上述问题时，BJ 方法论是迟早要用到的。此方法有四个步骤：

步骤 1. 识别（identification)。就是找出适当的 p、d 和 q 值。我们即将说明**相关图**（correlogram）和**偏相关图**（partial correlogram）怎样能用来帮助解决此问题。

步骤 2. 估计（estimation)。一旦辨识适当的 p 和 q 值，下一步便是估计模型中所含自回归和移动平均项的参数。有时可用简单的最小二乘法完成这
841 一计算，但有时则有必要寻求（对参数）非线性估计方法。由于当今的一些统计软件包都能按例行程序做好这一工作，我们就不必为估计中所遇到的数学而烦恼，有兴趣的读者可询查有关参考文献。

步骤 3. 诊断（diagnostic checking)。选定 ARIMA 模型并估计其参数之后，下一步就要看所选的模型对数据拟合得是否够好。因为有可能，另外的一个 ARIMA 模型也会做得同样好。这就是为什么博克斯-詹金斯 ARIMA 建模方法与其说是一门科学，毋宁说是一门艺术；为了选取正确的 ARIMA 模型，需要有高度的技巧。对所选模型的一个简单的检验，是看从该模型估计算出来的残差是不是白噪音；如果是，就可接受这个具体的拟合；如果不是，我们必须重新再做。由此可见，BJ 方法论是一个反复过程（见图 22.1)。

步骤 4. 预测（forecasting)。ARIMA 建模方法之所以得以普及，理由之一是它在预测方面的成功。有许多事例用这个方法做出的预报比用传统的计量经济建模方法做出的预报更为可靠，特别是在短期预测方面。当然，每一事例都必须加以核实。

有了这些一般性讨论在前，我们再来看每一步骤中的一些细节。我们仍

将利用表 21.1 所给的 GDP 数据去说明种种问题。

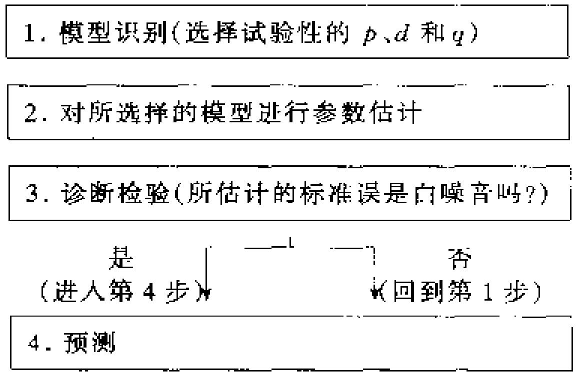


图 22.1　博克斯-詹金斯方法论

§22.4　识　　别

识别的主要工具是**自相关函数**(autocorrelation function, ACF),**偏自相关函数**(partial autocorrelation function, PACF)以及由此而得的**相关图**。后者只不过是将 ACF 和 PACF 相对于滞后长度描图而已。

842 在前一章中,我们曾定义(总体)ACF(ρ_k)和样本 ACF($\hat{\rho}$)。偏自相关的概念可类比于偏回归系数的概念。在 k 变量复回归模型中,第 k 个回归系数 β_k 度量着当所有的其他回归元的影响保持不变时第 k 个回归元 X_k 的每单位变化所引起的回归子的平均变化率。

类似地,**偏自相关** ρ_{kk} 度量着在控制对滞后小于 k 的相关下,相隔 k 个时期的(时间序列)观测值之间的相关。换言之,偏相关就是 Y_t 和 Y_{t-k} 之间的、除去居中的诸 Y(即 Y_{t-1}, Y_{t-2}, …, Y_{t-k+1})的影响后的相关。[7] 在第 7.11 节中,我们曾在回归的论述中介绍过偏相关的概念,并且表明了它与简单相关的关系。现在大多数统计软件包都把这种偏相关作为一种例行程序来计算。

在图 22.2 中,我们给出了 GDP 序列的相关图和偏相关图。该图有两个
843 凸显的事实:第一,ACF 非常缓慢地下降;如图 21.8 所示,直至 23 期滞后 ACF 逐期地看都是统计上显著地异于零的;它们都超越了 95% 置信限。第二,PACF 在第一个滞后之后急剧地下降,并且在 1 阶滞后之后的全部 PACF 都是统计上不显著的。

因美国 GDP 时间序列不是平稳的,故在我们应用博克斯-詹金斯方法论之前,必须把它变为平稳的。在图 21.9 中,我们画出 GDP 的一阶差分图。不像图 21.1 那样,我们没有看到序列中的任何趋势,这也许表明,取一阶差分后的 GDP 时间序列是平稳的。[8] 迪基-富勒单位根检验的一次正式的应用表明,情况的确如此。我们还可从图 22.3 所估计的 ACF 和 PACF 相关图

中观察到这点。现在我们有一个非常不同的 ACF 和 PACF 模式。看来 ACF 在滞后 1，8 和 12 处在统计上异于零；回顾第 21 章，ρ_k 的 95% 置信限约为
844 −0.208 9 和 +0.208 9。（注：如在第 21 章中所讨论的，这些置信限是渐近性质的，所以可把它们看作是近似的。）但在其他所有的滞后处，ρ_k 都不是统计上异于零的。这些对偏相关来说也是对的。

滞后	样本 ACF($\hat{\rho}_k$)	样本 PACF($\hat{\rho}_{kk}$)
1	0.969	0.969
2	0.935	−0.058
3	0.901	−0.020
4	0.866	−0.045
5	0.830	−0.024
6	0.791	−0.062
7	0.752	−0.029
8	0.713	−0.024
9	0.675	−0.009
10	0.638	−0.010
11	0.601	−0.020
12	0.565	−0.012
13	0.532	−0.020
14	0.500	−0.012
15	0.468	0.021
16	0.437	−0.001
17	0.405	−0.041
18	0.375	−0.005
19	0.344	−0.038
20	0.313	−0.017
21	0.279	−0.066
22	0.246	−0.019
23	0.214	−0.008
24	0.182	−0.018
25	0.153	0.017

95%置信域　　95%置信域

图 22.2　美国 GDP 的相关图和偏相关图：1970 年第 1 季度至 1991 年第 4 季度

现在，图 22.3 所给的相关图又怎样能帮助我们找出 GDP 时间序列的 ARMA 模式呢？（注：因取一阶差分后的 GDP 序列是平稳的，故我们仅考虑它。）为达到此目的，方法之一是考虑 ACF 和 PACF 以及与一些选定的 ARMA 过程如 AR(1)、AR(2)、MA(1)、MA(2)、ARMA(1, 1)、ARMA(2, 2) 等相对应的相关图。因为每一随机过程都有它典型的 ACF 和 PACF 式样，如果所研究的时间序列适合于其中的一个式样，我们就能辨识该时间序列符合这个过程。当然，我们仍有必要利用诊断性检验以判明所选的 ARMA 模型是否足够精确。

滞后	样本 ACF($\hat{\rho}_k$)	样本 PACF($\hat{\rho}_{kk}$)
1	0.316	0.316
2	0.186	0.095
3	0.049	−0.038
4	0.051	0.033
5	−0.007	−0.032
6	−0.019	−0.020
7	−0.073	−0.062
8	−0.289	−0.280
9	−0.067	0.218
10	0.019	0.100
11	0.037	−0.008
12	−0.239	−0.311
13	−0.117	0.011
14	−0.204	−0.114
15	−0.128	−0.051
16	−0.035	−0.021
17	−0.056	−0.019
18	0.009	0.122
19	−0.045	−0.071
20	0.066	−0.126
21	0.084	0.089
22	0.039	−0.060
23	−0.068	−0.121
24	−0.032	−0.041
25	0.013	0.092

95%置信域　　　　95%置信域

图 22.3　美国 GDP 一阶差分的相关图和偏相关图：1970 年第 1 季度至 1991 年第 4 季度

为了研究各种标准的 ARIMA 过程的性质，将要花费大量的篇幅。我们只打算提供一般性的指引（参看表 22.1）；有关参考资料能给出各种随机过程的细节。

表 22.1　ACF 与 PACF 的理论模式

模型种类	ACF 的典型模式	PACF 的典型模式
AR(p)	指数衰减或衰减的正弦波或两者	显著的直至滞后 p 的尖柱
MA(q)	显著的直至滞后 q 的尖柱	指数下降
ARMA(p,q)	指数衰减	指数衰减

注：指数衰减和几何衰减两名词的意义相同（回顾我们对考伊克分布滞后的讨论）。

845 注意，AR(p)过程的 ACF 和 PACF，和 MA(q) 过程的 ACF 和 PACF 相比，有相反的模式；对于 AR(p) 情形，AC 按几何或指数规律下降（常描写为拖尾者），而 PACF 则在一定的滞后次数之后忽然截断（常描写为断尾

——译者注）。但对于 MA（q）情况适得其反。

从几何图形看，这些模式有如图 22.4 所示。

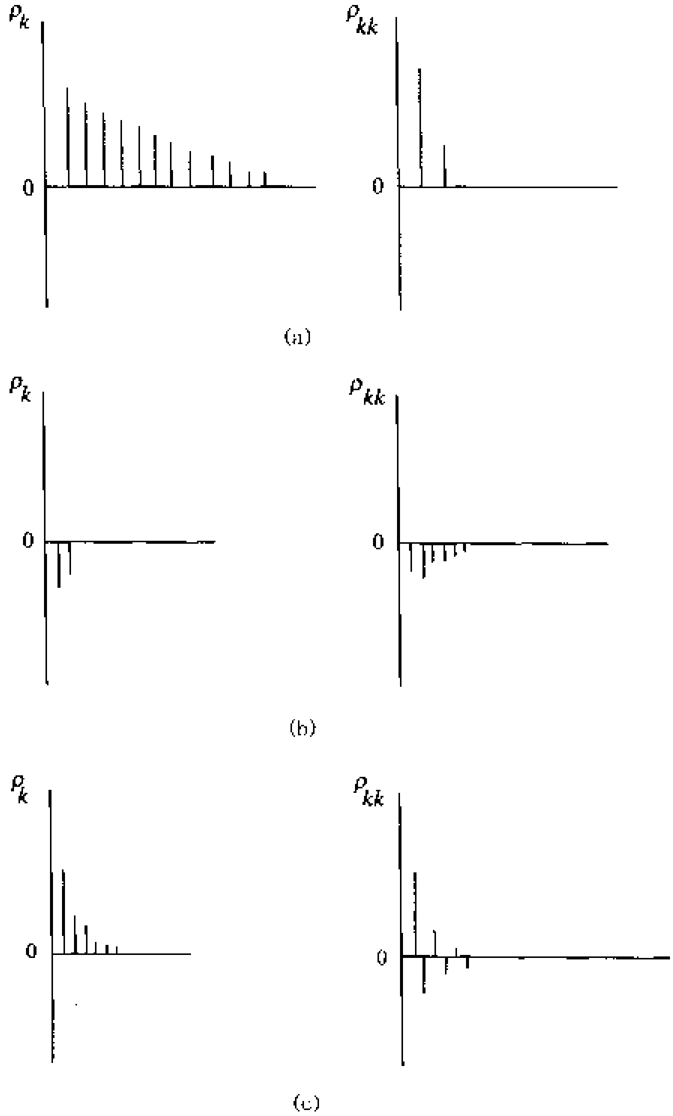


图 22.4 某些选定的随机过程的 ACF 和 PACF

注：(a)AR(2)：$\alpha_1=0.5,\alpha_2=0.3$；(b)MA(2)：$\beta_1=0.5,\beta_2=0.3$；(c)ARMA(1，1)；$\alpha_1=0.5$，$\beta_1=0.5$。

一点告诫。实际上我们无从观测理论（即总体）ACF 和 PACF 而有赖于它们的样本函数，故所估 ACF 和 PACF 将不会和理论函数相一致。我们所寻求的是理论与样本 ACF 和 PACF 之间的类似性，以便指引我们朝着正确的方向去建立 ARIMA 模型。这就是为什么 ARIMA 的应用要求有巧妙的技能，而这种技能只能来自实践。

美国 GDP 的 ARIMA 辨识。回到表 22.2 中所给 1970 年第 1 季度至 1991 年第 4 季度美国 GDP（取一阶差分后的）平稳序列的相关图和偏相关图，看我们有些什么发现？

须知图中所示的 ACF 和 PACF 是一些样本数量，并不像是表 22.1 所提示的那种干净利落的模式。自相关一直到滞后 4 都是下降的，然后，除了在滞后 8 和 12 两处外，其余的自相关都不是统计上异于零的（图中所画的实线给出约为 95%置信限）。偏自相关在滞后 1、8 和 12 处看来冒出了统计上显著的尖柱，而在其余地方则均不显著；假如偏相关系数只在滞后 1 处是显著的，我们就可认定这是一个 AR(1) 模型。因此，且假定产生这个（一阶差分）GDP 的模型最多是一个 AR(12) 过程。当然，由于在偏相关图中我们看到 AR 项仅在滞后 1、8 和 12 处显著，我们没有必要把直至滞后 12 的全部 AR 项都包含进来。

§22.5 ARIMA 模型的估计

用 Y_t^* 表示美国 GDP 的一阶差分。那么，我们所辨识的一个尝试性的 AR 模型是：

$$Y_t^* = \delta + \alpha_1 Y_{t-1}^* + \alpha_8 Y_{t-8}^* + \alpha_{12} Y_{t-12}^* \tag{22.5.1}$$

846 利用 Eviews，我们得到如下估计结果：

$$\begin{aligned}
\hat{Y}_t^* &= 23.0894 + 0.3428\,Y_{t-1}^* - 0.2994\,Y_{t-8}^* - 0.2644\,Y_{t-12}^* \\
\text{se} &= (2.9774)\quad(0.0987)\qquad(0.1016)\qquad(0.0986) \\
t &= (7.7547)\quad(3.4695)\qquad(-2.9475)\qquad(-2.6817) \\
& \qquad\qquad\qquad R^2 = 0.2931 \qquad d = 1.7663
\end{aligned} \tag{22.5.2}$$

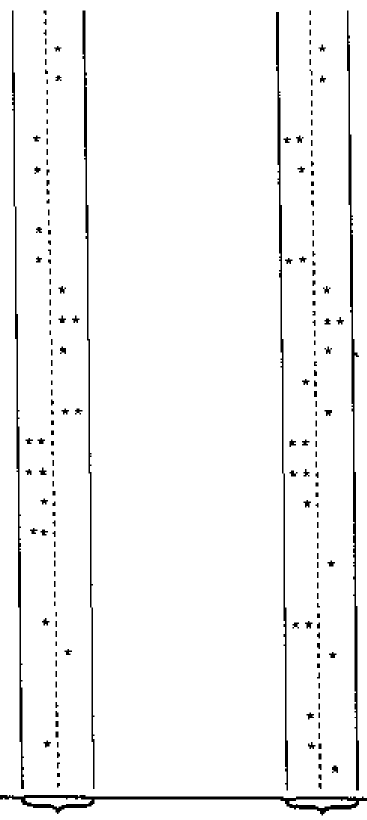
作为习题，请读者估计仅含 Y_{t-1}^* 的一个模型和兼含 Y_{t-1}^* 和 Y_{t-8}^* 两项的一个模型，并将所得结果同 (22.5.2) 所给的相比较。

§22.6 诊断检查

我们怎样知道模型 (22.5.2) 对数据是一个良好的拟合呢？一种简易的诊断是求出 (22.5.2) 中的残差并计算这些残差的直至滞后 25 的 ACF 和
847 PACF。图 22.5 给出了所估计的 AC 和 PACF。如图所示，没有任何自相关和偏自相关是个别地在统计上显著的。而且按照博克斯-皮尔斯 Q（Box-Pierce Q）和扬-博克斯（Ljung-Box）LB 统计量（见第 21 章），25 个自回归平方和也不是统计上显著的。换句话说，自相关和偏自相关的相关图给我们的印象都表明，从 (22.5.2) 估计出来的残差是纯随机的。因此，似无必要再去寻觅其他的 ARIMA 模型了。

§ 22.7 预 测

记住 GDP 数据是从 1970 年第 1 季度到 1991 年第 4 季度的数据。假使我们想根据模型（22.5.2）预测 1992 年头 4 个季度的 GDP。但在（22.5.2）中，因变量是 GDP 相对于前一季度的变化。因此，我们若使用（22.5.2），则我们所能得到的是 1992 年第 1 季度和 1991 年第 4 季度之间的 GDP 变化，1992 年第 2 季度相对于 1992 年第 1 季度的 GDP 变化等等的预报值。



滞后	ACF ($\hat{\rho}_k$)	PACF ($\hat{\rho}_{kk}$)
1	0.043	0.043
2	0.113	0.112
3	0.020	0.012
4	−0.100	−0.116
5	−0.068	−0.065
6	−0.029	0.001
7	−0.040	−0.019
8	−0.112	−0.118
9	0.065	0.069
10	0.126	0.151
11	0.099	0.076
12	−0.026	−0.106
13	0.120	0.102
14	−0.181	−0.150
15	−0.128	−0.131
16	−0.073	−0.050
17	−0.121	−0.038
18	0.017	0.059
19	−0.017	−0.027
20	−0.085	−0.163
21	0.055	0.059
22	0.010	−0.016
23	−0.038	−0.103
24	−0.053	−0.072
25	−0.002	0.100

博克斯-皮尔斯 Q 统计量　14.42　概率　0.954 0　相关的标准误　0.110

扬-博克斯 LB 统计量　17.63　概率　0.857 8

图 22.5　从 ARIMA 模型（22.5.2）中所得到残差的相关图

为了得到 GDP 的水平值而不是它的变化的预报值，我们可以做曾经用来获得变化值的一阶差分变换的反变换。说得更技术性些，我们对一阶差分

序列单积（integrate）（或求积）。这样，为了得到 1992 年第 1 季度的 GDP（而不是 ΔGDP），我们将模型（22.5.1）重写为：

$$
\begin{aligned}
Y_{1992-\mathrm{I}} - Y_{1991-\mathrm{IV}} = \delta &+ \alpha_1[Y_{1991-\mathrm{IV}} - Y_{1991-\mathrm{III}}] \\
&+ \alpha_8[Y_{1989-\mathrm{IV}} - Y_{1989-\mathrm{III}}] \\
&+ \alpha_{12}[Y_{1988-\mathrm{IV}} - Y_{1988-\mathrm{III}}] + u_{1992-\mathrm{I}}
\end{aligned} \tag{22.7.1}
$$

即：

$$
\begin{aligned}
Y_{1992-\mathrm{I}} = \delta &+ (1+\alpha_1)Y_{1991-\mathrm{IV}} - \alpha_1 Y_{1991-\mathrm{III}} + \alpha_8 Y_{1989-\mathrm{IV}} - \alpha_8 Y_{1989-\mathrm{III}} \\
&+ \alpha_{12} Y_{1988-\mathrm{IV}} - \alpha_{12} Y_{1988-\mathrm{III}} + u_{1992-\mathrm{I}}
\end{aligned} \tag{22.7.2}
$$

从所估计的回归（22.5.1）中已经获知 δ、α_1、α_8 和 α_{12}的值。假定 $u_{1992-\mathrm{I}}$ 为零。（为什么？）我们便容易获得 $Y_{1992-\mathrm{I}}$ 的预测值。此预测值的数值估计是[9]：

$$
\begin{aligned}
\hat{Y}_{1992-\mathrm{I}} &= 23.089\,4 + (1+0.342\,8)Y_{1991-\mathrm{IV}} - 0.342\,8Y_{1991-\mathrm{III}} \\
&\quad + (-0.299\,4)Y_{1989-\mathrm{IV}} - (-0.299\,4)Y_{1989-\mathrm{III}} \\
&\quad + (-0.264\,4)Y_{1988-\mathrm{IV}} - (-0.264\,4)Y_{1988-\mathrm{III}} \\
&= 23.089\,4 + 1.342\,8(4\,868) - 0.342\,8(4\,862.7) - 0.299\,4(4\,859.7) \\
&\quad + 0.299\,4(4\,845.6) - 0.264\,4(4\,779.7) + 0.264\,4(4\,734.5) \\
&= 4\,876.7(\text{近似地})
\end{aligned}
$$

848 就是说，1992 年第 1 季度的 GDP 的预报值约为 48 770 亿（1987 年美元）。顺便提一提，1992 年第 1 季度的实际 GDP 的实测值是 48 737 亿；预报误差是过高估计了 33 亿。

注意，如果你想用（22.5.2）去计算从 1991 年第 4 季度到 1992 年第 1 季度 GDP 的变化的预测值，你就会得到 −42.5 亿这个数字。

§22.8 BJ 方法论的其他方面

前几节中我们仅对 BJ 建模做一简略的介绍。这一方法论还有许多方面的问题，例如**季节性**（seasonality），由于篇幅所限而未能考虑。有许多时间序列展现出季节性态。百货商店主要节日中的销售量就是个例子。还有季节性的雪糕消费。公众节日的旅游，等等。如果，比方说，我们拥有百货商店的季度销售数据，则这些数据必定在第 4 季度冒出尖柱。这时，可取销售数字的第 4 季差分，把季节性影响消除掉，再决定用哪一种 ARIMA 模型进行拟合。

以上我们限于每次分析单一个时间序列。但没有任何事情要妨碍我们把 BJ 方法论加以推广同时研究两个或多个时间序列。对这个问题展开讨论会使我们走得太远。有兴趣的读者可参阅有关文献。[10]然而，在下一节中，我们将在向量自回归的名义下讨论这个问题。

§ 22.9 向量自回归

在第 18～20 章里，我们曾考虑联立或结构方程模型。在这些模型中，我们把一些变量看作内生的，而另一些变量看作外生的或前定的（外生的和滞后内生的）。在估计这些模型之前，还必须肯定方程组中的方程是可识别的（恰好识别或过度识别）。而为达到识别的目的，常常要假定某些前定变量仅出现在某些方程之中。这种决定往往是主观的，并且受到 C.A. 西姆斯（Christopher Sims）的严厉批判。[11]

根据西姆斯的理论，如果在一组变量之中有真实的联立性，那么，这些变量就应平等地加以对待，而不应该事先区分内生和外生变量。正是本着这一精神，西姆斯推出他的 VAR 模型。

第 17 章中讨论的格兰杰因果检验，就已播下了这种模型的种子。在用
849 滞后货币供给和滞后 GDP 解释当前 GDP 的方程（17.14.1）以及当前货币供给的方程（17.14.2）中，我们实质上是把 GDP 和货币供给看作一对内生变量。在这个方程组中没有外生变量。

类似地，在例题 17.13 中我们检验了加拿大货币和利率间的因果关系，在货币方程中，只出现了滞后的货币值和利率，同样，在利率方程中，也只出现了滞后的利率值和货币数。

这两个例子都是说明向量自回归模型的例子；自回归一词的使用是因为方程的右端出现有因变量的滞后值，而“向量”一词的使用是因为我们在同含有两个（或多个）变量的一个向量打交道。

VAR 的估计

回到加拿大货币与利率关系的例子。我们曾看到，当我们引进每一变量的 6 个滞后项作为回归元时，我们无法拒绝货币（M1）与利率 R（90 天公司利率）之间有双向因果关系的假设。就是说，M1 影响 R，而 R 反过来又影响 M1。这种情况是应用 VAR 的理想情形。

为了说明怎样估计一个 VAR 模型，我们仍以这个模型为例。为简单起见，假定每个方程都含有 M(M1)和 R 的 k 个滞后值作为回归元，这时，每个方程都可用 OLS 去估计。[12]我们所估计的实际模型是：

$$\mathrm{M1}_t = \alpha + \sum_{j=1}^{k} \beta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{1t} \tag{22.9.1}$$

$$R_t = \alpha' + \sum_{j=1}^{k} \theta_j M_{t-j} + \sum_{j=1}^{k} \gamma_j R_{t-j} + u_{2t} \tag{22.9.2}$$

其中诸 u 是随机误差项，在 VAR 术语中称之为**脉冲值**（impulses）或**新生值**

(innovations) 或冲击值 (shocks)。

我们在估计 (22.9.1) 和 (22.9.2) 之前，必须先决定最大滞后长度 k。这是一个经验问题。我们共有 40 个观测。包括过多的滞后项将消耗自由度，更不用说会引入多重共线性的可能性。而包括过少的滞后值则导致设定误差。解决这个问题的办法之一，就是使用赤池、施瓦茨或诸如此类的某个准则，并选择这些准则最低值的模型。无疑，某些试错法就不可避免。

为了说明这个机制，我们首先使用每个变量的 4 阶滞后 ($k=4$)，使用
850 Eviews 4 我们得到上述两个方程中参数的估计值，并示于表 22.2。注意，尽管我们的样本是从 1979 年第 1 季度至 1988 年第 4 季度，但我们只以 1979 年第 1 季度至 1987 年第 4 季度为样本期间，而把最后 4 个观测留待检查所拟合 VAR 的预测准确性。

表 22.2　　基于 4 阶滞后的向量自回归估计值

样本时期:1980.Ⅰ—1987.Ⅳ

观测次数:在进行端点调整后共 32 个

( )中为标准误　[ ]中为 t 统计量

	M1	R
M1(−1)	1.076 737(0.201 74)[5.337 33]	0.001 282(0.000 67)[1.900 83]
M1(−2)	0.173 433(0.314 44)[0.551 57]	0.002 140(0.001 05)[−2.035 84]
M1(−3)	−0.366 465(0.346 87)[−1.056 48]	0.002 176(0.001 16)[1.876 99]
M1(−4)	0.077 602(0.207 89)[0.373 29]	−0.001 479(0.000 69)[−2.128 55]
$R(-1)$	−275.029 3(57.217 4)[−4.806 75]	1.139 310(0.191 27)[5.956 70]
$R(-2)$	227.175 0(95.394 7)[2.381 42]	−0.309 053(0.318 88)[−0.969 17]
$R(-3)$	8.511 851(96.917 6)[0.087 83]	0.052 361(0.323 97)[0.161 62]
$R(-4)$	−50.199 26(64.755 4)[−0.775 21]	0.001 076(0.216 46)[0.004 97]
C	2 413.827(1 622.65)[1.487 59]	4.919 000(5.424 16)[0.906 87]
R^2	0.988 154	0.852 89 0
校正 R^2	0.984 034	0.801 721
残差平方和	4 820 241	53.862 33
SE 方程	457.794 4	1.530 307
F 统计量	239.831 5	16.668 15
对数似然比	−236.167 6	−53.737 16
赤池 A/C	15.322 98	3.921 073
施瓦茨 SC	15.735 21	4.333 311
因变量的均值	28 514.53	11.672 92
因变量的标准差	3 623.058	3.436 688
残差协方差矩阵的行列式	490 782.3	
对数似然比(对 df 进行了调整)	−300.472 2	
赤池信息准则	19.904 51	
施瓦茨准则	20.728 99	

由于上述方程都是 OLS 方程，所以表 22.2 给出的回归结果可如平常一样解释。当然，同时引入同一变量的几个滞后项，可能会因多重共线性而使每个估计系数在统计上都不显著。但基于标准的 F 检验，它们可能是联合显著的。

让我们检查一下表22.2中给出的结果。首先考虑 M1 回归。个别地看，
只有 M1 在 1 阶滞后、R 在第 1 和 2 阶滞后处是统计显著的。但由于 F 值很
851 高，所以我们不能拒绝所有滞后项联合统计显著的假设。转而从利率回归来
看，我们发现 4 个货币滞后项都是统计显著的（在 10%或更好的显著性水
平上），而利率变量只有一期滞后项才是显著的。

为便于比较，我们在表 22.3 中仅基于每个内生变量的 2 阶滞后给出了 VAR 结果。你在此将会看到，在货币回归中，货币变量的一期滞后和利率变量的 1 和 2 阶滞后都是个别统计显著的。而在利率回归中，两个货币变量的滞后项（约在 5%的显著性水平上）都是个别显著的，利率滞后项则只有一个是显著的。

表 22.3　　基于 2 阶滞后的向量自回归估计值

样本时期：1979.Ⅲ－1987.Ⅳ

观测次数：在进行端点调整后共 34 个

（ ）中为标准误　［ ］中为 t 统计量

	M1	R
M1(−1)	1.037 537(0.160 48)[6.465 09]	0.001 091(0.000 59)[1.858 25]
M1(−2)	−0.044 661(0.155 91)[−0.286 46]	−0.001 255(0.000 57)[−2.198 71]
R(−1)	−234.885 0(45.522 4)[−5.159 77]	1.069 081(0.166 60)[6.417 08]
R(−2)	160.156 0(48.528 3)[3.300 26]	−0.223 364(0.177 60)[−1.257 68]
C	1 451.977(1 185.59)[1.224 68]	5.796 434(4.338 94)[1.335 91]
R^2	0.988 198	0.806 660
校正 R^2	0.986 571	0.779 993
残差平方和	5 373 510.	71.970 54
SE 方程	430.457 3	1.575 355
F 统计量	607.072 0	30.248 78
对数似然比	−251.744 6	−60.992 15
赤池 A/C	15.102 63	3.881 891
施瓦茨 SC	15.327 09	4.106 356
因变量的均值	28 216.26	11.750 49
因变量的标准差	3 714.506	3.358 613
残差协方差矩阵的行列式	458 485.4	
对数似然比（对 df 进行了调整）	−318.094 4	
赤池信息准则	19.299 67	
施瓦茨准则	19.748 60	

如果我们必须在表 22.2 和表 22.3 所给出的模型之间做出选择，我们会选择哪一个呢？表 22.2 中模型的赤池和施瓦茨信息值分别是 15.32 和 15.73，而表 22.3 中模型的对应信息值分别为 15.10 和 15.33。由于赤池和施瓦茨统计量的值越低，模型就越好，由此看来，表 22.3 中给出的较节省的模型更好。我们还考虑了每个内生变量的 6 阶滞后，并发现其赤池和施瓦茨统计量分别为 15.37 和 15.98。同样，表 22.3 中的只含有每个内生变量 2 阶滞后的模型看来才是我们正确的选择。

用 VAR 做预测

852 假设我们选择了表 22.3 中给出的模型。我们就可以用它预测 M1 和 R 的值。记住，尽管我们的数据涵盖了从 1979 年第 1 季度至 1988 年第 4 季度整个期间，但我们在估计 VAR 模型时并没有使用 1988 年的数据。现在假设我们想预测 1988 年第 1 季度即 1988 年第 1 季度 M1 的值。可得到 1988 年第 1 季度的预测值如下：

$$\hat{M}_{1988-\mathrm{I}} = 1\ 451.977 + 1.037\ 5M_{1987-\mathrm{IV}} - 0.044\ 6M_{1987-\mathrm{III}}$$
$$234.885\ 0R_{1987-\mathrm{IV}} + 160.156\ 0Y_{1987-\mathrm{III}}$$

其中的系数值都是从表 22.3 中得到的。现在代入表 17.3 中 M 和 R 的相应值，可以看出 1988 年第 1 季度货币的预测值为 36 996（百万加元）。1988 年第 1 季度 M 的实际值为 36 480，这就意味着我们的模型预测值比实际值高 516（百万加元），约占 1988 年第 1 季度 M 实际值的 1.4%。当然，这些估计值将随着我们在模型中考虑滞后项的多少而改变。作为一个练习，请读者预测 1988 年第 1 季度的 R 值并与其实际值相比较。

VAR 与因果性

你或许记得，我们在第 17 章讨论过因果性的话题。我们在那里讨论了格兰杰因果性检验和西姆斯因果性检验。VAR 与因果性之间有什么联系吗？我们在第 17 章（第 14 节）看到，直至 2、4 和 6 阶滞后，M1 和 R 之间还有双向因果关系，但到 8 阶滞后时，这两个变量之间就没有因果关系了。因此，结论是含糊的。你现在或许还记得第 21 章的格兰杰表述定理。这个定理的引申含义之一便是，若两个变量（X_t 和 Y_t）是协积的，而且每个都是一阶单积序列 $I(1)$（即都是不平稳的），那么，要么 X_t 一定是 Y_t 的格兰杰原因，要么 Y_t 一定是 X_t 的格兰杰原因。

在我们说明性的例子中，这就意味着，若 M1 和 R 都是单积序列，但是协积的，则 M1 一定格兰杰导致 R 或 R 一定格兰杰导致 M1。这就要求我们必须首先弄清楚这两个变量都是 $I(1)$，并看它们是否协积。如果不是这样，那么整个因果性问题也就变成纯学术上的讨论，没有任何实际意义了。习题 22.22 要求读者弄清楚其中的两个变量是否非平稳但协积。你若做此

题，你会发现 M1 和 R 之间存在着协积的弱证据，这正是第 17.14 节中讨论因果性检验时模棱两可的原因。

VAR 建模的一些问题

853 VAR 的倡导者强调此法有如下的优点：(1) 方法简单：无须决定哪些变量是内生的，哪些变量是外生的。VAR 中的全部变量都是内生的。[13] (2) 估计简单：常用的 OLS 法可用于逐个地估计每一方程。(3) 在许多案例中，用此法得到预报优于用更复杂的联立方程模型得到的预报。[14]

但 VAR 建模的批评者指出如下的一些问题：

1. 不同于联立方程模型，VAR 利用较少的先验信息，所以是乏理论的 (a-theoretic)。须知在联立方程中排除或包含某些变量，对模型的识别起到关键性的作用。

2. 由于重点放到预测，VAR 模型较不适合于政策分析。

3. 实际上，对 VAR 建模最大的挑战在于选择适当滞后长度。假令你有一个 3 变量 VAR 模型，并且你决定每个方程含有每个变量的 8 个滞后值，你在每一方程中将有 24 个滞后参数，加上一个常数就共有 25 个参数。除非样本很大，估计如此多的参数将消耗大量自由度并带来所有随之而来的种种问题。[15]

4. 严格地说，在一个 m 变量 VAR 模型中，所有的 m 个变量都应该是(联合地) 平稳的。如果不是这样，则有必要适当变换数据 (例如，通过一阶差分)。如哈维所指出的，由变换数据得到的结果未必令人满意。他进一步指出："因此，VAR 狂热者通常采取的策略是利用水平值进行工作，即使其中的一些序列是非平稳的。似此情形，认识到单位根对估计量分布的影响就是重要的。"[16] 更糟糕的情况则是，模型中掺杂有 $I(0)$ 和 $I(1)$ 变量，也就是它是一个平稳和非平稳变量的混合体，如何变换数据将不是容易的事。

5. 由于所估计的模型中的系数往往难于逐一地加以解释，VAR 技术的操作人员常估计一种所谓的**脉冲响应函数** (impulse response function, IRF)。
854 IRF 描绘 VAR 系数中的因变量如何响应于诸如方程 (22.9.1) 和 (22.9.2) 中的误差项 u_1 和 u_2 的冲击。假使在 M1 方程中的 u_1 值增加一个标准差，这样的一个冲击或变化将会改变现期以及今后时期里的 M1。但因 M1 出现在 R 的回归中，u_1 的变化将影响到 R。类似地，R 方程中的 u_2 的一个标准差变化将影响到 M1，IRF 跟踪这种冲击在将来若干个时期里所起的影响。尽管这种 IRF 分析的效用已受到研究人员的质疑，但它仍是 VAR 分析的代表作。[17]

为了比较 VAR 和其他预测技术的优缺点，可参考有关文献。[18]

VAR 的一个应用：得克萨斯州经济的一个 VAR 模型

为了检验谚语："油业兴旺，得克萨斯经济也就兴旺。" T. 冯拜（Thomas Fomby）和 J. 赫希伯格（Joseph Hirschberg）对 1974 年第 1 季度至 1988 年第 1 季度的得克萨斯经济研制了一个 3 变量 VAR 模型。[19]所考虑的 3 个变量是：（1）实际油价的百分率变化，（2）得克萨斯非农业（部门）就业的百分率变化和（3）美国其他地区的非农业就业的百分率变化。作者们在每个方程中引进一个常数项和每一变量的两个滞后值。因此，每个方程有 7 个待估的参数。表 22.4 给出此 VAR 模型的 OLS 估计结果。表中所给的 F 统计量用以检验各种滞后系数集为零的（联立）假设。于是，对 x 变量（实际油价的百分率变化）的 F 检验表明，x 的两个滞后项都在统计上异于零；在它们同时为零的虚拟假设下，得到一个等于 12.553 6 的 F 值的概率是很低的，约为 0.000 04。另一方面，用以解释 x 的两个滞后 y 值（得克萨斯非农业就业的百分率变化），集体地看，则不是显著地异于零的；其 F 值仅为 1.36。对其余的 F 统计量可作类似的解释。

冯拜和赫希伯格根据他们论文中的这些和其他结果，得到的结论是：上述关于得克萨斯经济的谚语并没有很准确的意义。因为在经历石油输出国组织（OPEC）的石油禁运冲击所造成的初始不稳定之后，得克萨斯经济现在已较少地依赖于油价的波动了。

表 22.4　　二阶* 得克萨斯 VAR 系统的估计结果：1974 年第 1 季度至 1988 年第 1 季度

因变量：x（实际油价的百分率变化）

变量	滞后	系数	标准误	显著性水平
x	1	0.705 4	0.140 9	0.830 5E－5
x	2	－0.335 1	0.150 0	0.302 7E－1
y	1	－1.352 5	2.701 3	0.618 9
y	2	3.437 1	2.434 4	0.164 5
z	1	3.456 6	2.804 8	0.223 9
z	2	－4.870 3	2.750 0	0.830 4E－1
常数	0	－0.998 3E－2	0.169 6E－1	0.558 9

$\bar{R}^2=0.298\ 2$；$Q\ (21)\ =8.261\ 8\ (P=0.993\ 9)$

对联合显著性的检验，因变量 = x

变量	F-统计量	显著性水平
x	12.553 6	0.428 3E－4
y	1.364 6	0.265 4
z	1.569 3	0.218 8

因变量：y（得克萨斯州非农业就业的百分率变化）

变量	滞后	系数	标准误	显著性水平
x	1	0.222 8E-1	0.875 9E-2	0.143 0E-1
x	2	-0.188 3E-2	0.932 2E-2	0.840 7
y	1	0.646 2	0.167 8	0.355 4E-3
y	2	0.423 4E-1	0.151 2	0.780 7
z	1	0.265 5	0.174 2	0.134 2
z	2	-0.171 5	0.170 8	0.320 5
常数	0	-0.160 2E-2	0.105 3E-1	0.135 1

$\bar{R}^2=0.631\ 6$; $Q\ (21)\ =21.590\ 0\ (P=0.423\ 4)$

对联合显著性的检验，因变量 $=y$

变量	F-统计量	显著性水平
x	3.628 3	0.342 4E-4
y	19.144 0	0.828 7E-6
z	1.168 4	0.319 7

因变量：z（美国其他地区非农业就业百分率变化）

变量	滞后	系数	标准误	显著性水平
x	1	-0.833 0E-2	0.684 9E-2	0.229 9
x	2	0.363 5E-2	0.728 9E-2	0.620 2
y	1	0.384 9	0.131 2	0.517 0E-2
y	2	-0.480 5	0.118 2	0.182 8E-2
z	1	0.722 6	0.136 2	0.300 4E-5
z	2	-0.136 6E-1	0.133 6	0.919 0
常数	0	-0.238 7E-2	0.824 1E-3	0.570 1E-2

$R=0.650\ 3$; $Q\ (21)\ =15.618\ 2\ (P=0.790\ 7)$

对联合显著性的检验，因变量 $=z$

变量	F-统计量	显著性水平
x	0.739 6	0.482 7
y	8.271 4	0.836 0E-3
z	27.960 9	0.100 0E-7

资料来源：*Economic Review*, Federal Reserve Bank of Dallas, January 1989, p. 21.

§22.10 度量金融时间序列中的波动性：ARCH 和 GARCH 模型

856 本章引言曾指出，诸如股票价格、汇率、通货膨胀率等金融时间序列通常表现出**群集波动**（volatility clustering）的现象，即在相当长一段时期，其价格表现出大幅波动，然后又会在下一段时期内保持相对稳定。菲利普·弗兰西斯（Philip Franses）指出：

> 由于这种（金融时间序列）数据反应了（比方说）股票市场上买卖双方交易的结果，各种信息来源及其他外生经济事件都有可能对资产价格的时间

序列模式产生影响。由于对信息有各种不同的解释，而且诸如石油冲击等特定经济事件可能持续一段时间，所以我们通常会观察到，金融时间序列中较大的正观测值和负观测值都倾向于群集出现。[20]

波动性方面的知识在许多领域都至关重要。比如，在研究通货膨胀随时间的变化方面，已经做了大量的宏观计量工作。对某些决策者而言，通货膨胀本身或许不是一件坏事情，但其波动性使得金融计划很难做好，从而对决策者不利。

外汇市场上的进口商、出口商和交易者也是一样，因为汇率的波动性意味着巨大的损失或利润。股票市场上的投资者显然对股票市场上的波动性很感兴趣，因为很大的波动性就意味着很大的损失或收益，也就意味着巨大的不确定性。在波动的市场上，公司很难通过资本市场来筹集资本。

我们如何模型化可能存在这种波动性的金融时间序列呢？比如，我们如何模型化股票价格、汇率和通货膨胀等序列呢？这些金融时间序列多数都具有这样一个特征：它们的水平值为随机步游；即是非平稳的。但另一方面，它们的一阶差分形式则通常都是平稳的，比如我们在上一章中所见到的GDP序列便是如此，尽管GDP并非严格的金融时间序列。

因此，为什么不去模型化金融时间序列的一阶差分而要去模型化其水平值呢？但这些一阶差分通常都表现出大幅摆动或变动，说明金融时间序列的方差也在随着时间而变化。我们也能模型化这种“变动着的方差”吗？这就使得最早由恩格尔提出的所谓**自回归条件异方差**模型派上了用场。

857 顾名思义，由于不同时期所观测到的异方差（或者不相等的方差）也可能自相关，所以这种异方差便可能具有自回归的结构。为看出其全部含义，让我们先考虑两个简明的例子。

美英汇率：一个例子

图22.6给出了美英汇率（美元/英镑）1973—1995年共276个月度数据对数。从此图中可以看出，汇率在样本期间有大量起伏。为了更清楚地看到这一点，我们用图22.7描出了汇率对数的变化图；注意，一个变量对数的变化表示相对变化，乘以100便得到变化的百分数。你可以看到，美英汇率的相对变化在某些时期表现出大幅摆动，而在其他时期则只有适度的变化，从而为群集波动现象提供了例证。现在的实际问题是：我们如何在统计上度量波动性？让我们用汇率的例子来说明它。

令 Y_t = 美英汇率

Y_t^* = Y_t 的对数

$dY_t^* = Y_t^* - Y_{t-1}^*$ = 汇率的相对变化

$d\bar{Y}_t^*$ = dY_t^* 的均值

$X_t = dY_t^* - d\bar{Y}_t^*$

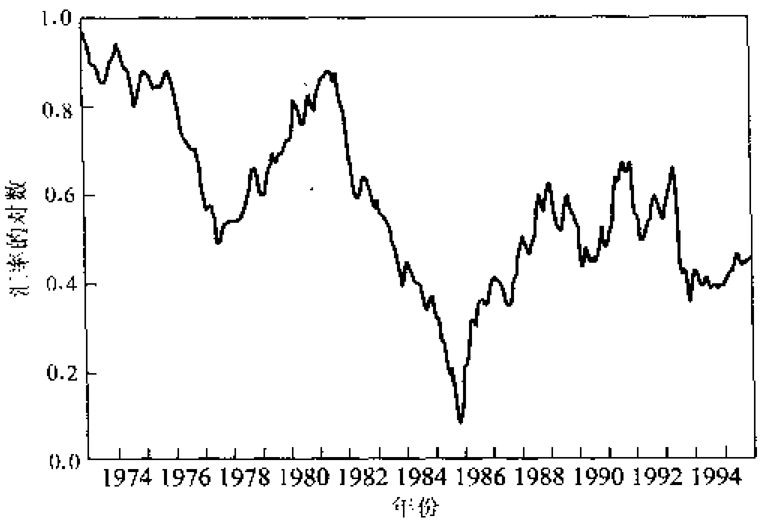


图 22.6　1973—1995 年（月度）期间美英汇率的对数

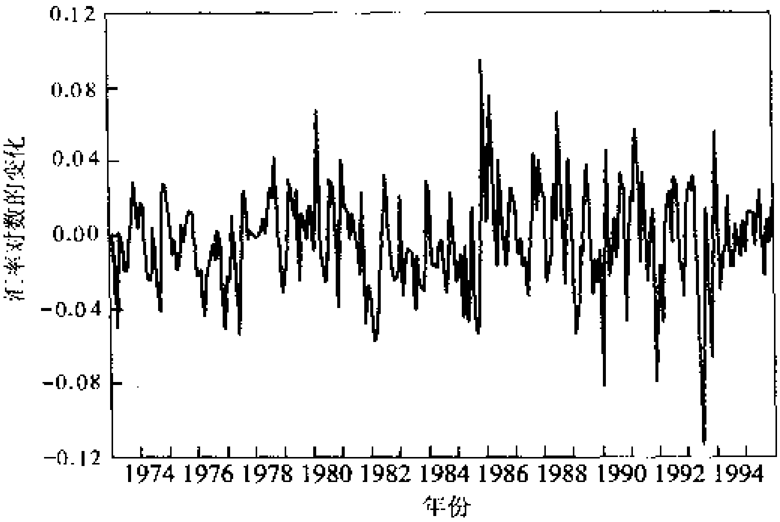


图 22.7　美英汇率对数的变化

因此，X_t 为进行均值调整后的汇率的相对变化。现在，我们可以用 X_t^2 作为一个度量波动性的工具。作为一个平方量，当金融资产价格变化大时，它的值就大；而当金融资产价格变化温和时，它的值就小。[22]

858　接受 X_t^2 作为对波动性的一个度量之后，我们如何知道它是否随时间而变化呢？假设我们考虑如下 AR(1)，或 ARIMA(1，0，0) 模型：

$$X_t^2 = \beta_0 + \beta_1 X_{t-1}^2 + u_t \tag{22.10.1}$$

此模型表明，当期波动性与其上一期波动性和白噪音误差项有关系。若 β_1 为正，则表明上一期较高的波动性将导致下一期的波动性继续高，即标志着群集波动。若 β_1 为 0，则表明不存在群集波动的情况。用通常的 t 检验即可判断所估计的 β_1 的统计显著性。

我们自然想到考虑波动性的 AR(p)模型：

$$X_t^2 = \beta_0 + \beta_1 X_{t-1}^2 + \beta_2 X_{t-2}^2 + \cdots + \beta_p X_{t-p}^2 + u_t \tag{22.10.2}$$

此模型表明任一期的波动性与其前面 p 期的波动性相关，p 的值只是一个经验问题。这个经验问题可用我们在第 13 章讨论的那些模型选择准则中的一或多个来解决（比如赤池信息准则）。我们可以用 t 检验来检验任何一个 β 系数的显著性、或用通常的 F 检验来检验两或多个系数的联合显著性。

模型（22.10.1）是 ARCH (1) 模型的一个例子，模型（22.10.2）就被称为 ARCH (p) 模型，其中 p 代表模型中自回归项的个数。

在做进一步解释之前，让我们先用美英汇率数据来说明 ARCH 模型。ARCH（1）模型的估计结果如下：

$$\begin{aligned} X_t^2 &= 0.000\,6 + 0.164\,9X_{t-1}^2 \\ t &= (6.783\,1)(2.835\,5) \qquad R^2 = 0.028\,7 \qquad d = 1.997\,2 \end{aligned} \tag{22.10.3}$$

其中 X_t^2 的定义如前。

859 由于滞后项系数的高度显著性（p 值约为 0.005）看来在本例中出现了群集波动的特征。我们也试了更高阶的 ARCH 模型，但最终只有 AR(1) 模型是显著的。

我们如何检验基于时间序列数据的一般回归模型中的 ARCH 模型呢？更具体而言，让我们考虑 k 变量线性回归模型：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \tag{22.10.4}$$

并假定以（$t-1$）期可利用的信息为条件，误差项服从如下分布：

$$u_t \sim N[0, (\alpha_0 + \alpha_1 u_{t-1}^2)] \tag{22.10.5}$$

的正态分布，而 u_t 服从均值为 0，方差为：

$$\text{var}(u_t) = (\alpha_0 + \alpha_1 u_{t-1}^2) \tag{22.10.6}$$

即 u_t 的方差服从一个 ARCH(1)过程。

u_t 的正态性质对我们而言并不新鲜。新鲜的是，u 在 t 时期的方差取决于（$t-1$）期的干扰平方，从而给出序列相关的现象。[23] 当然，误差方差可能不仅取决于误差方差项的一阶滞后，还取决于几个滞后平方项：

$$\text{var}(u_t) = \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2 \tag{22.10.7}$$

若误差方差不存在自相关，我们有

$$H_0:\ \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0 \tag{22.10.8}$$

在这种情况下，$\text{var}(u_t) = \alpha_0$，即不具有 ARCH 效应。

由于我们不能直接观测到 σ_t^2，所以恩格尔证明了，做如下回归很容易检验上述虚拟假设：

$$\hat{u}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{u}_{t-1}^2 + \hat{\alpha}_2 \hat{u}_{t-2}^2 + \cdots + \hat{\alpha}_p \hat{u}_{t-p}^2 \tag{22.10.9}$$

其中 $\hat{u}_t$ 和平常一样表示从原回归模型（22.10.4）中得到的OLS残差。

你可以使用通常的 F 检验来检验虚拟假设 H_0，也可以通过计算 nR^2 来检验。其中 R^2 为辅助回归（22.10.9）的判定系数。可以证明：

$$nR^2_{\text{asy}} \sim \chi^2_p \tag{22.10.10}$$

即在大样本中，nR^2 服从自由度等于辅助回归中自相关回归元个数的 χ^2 分布。

在继续讲解下去之前，确信你没有把第12章中讨论的误差项的自相关和ARCH模型搞混淆。在ARCH模型中，u_t 的（条件）方差取决于先前误差项的平方，所以给你留下了自相关的印象。

纽约证券交易所的价格变化

860 作为对ARCH效应的进一步说明，图22.8给出了NYSE（纽约证券交易所）价格指数在1952—1995期间月度百分比变化。[24] 此图表明NYSE价格指数的百分比变化有相当可观的波动性。特别注意在1987年股市暴跌附近的大幅摆动。

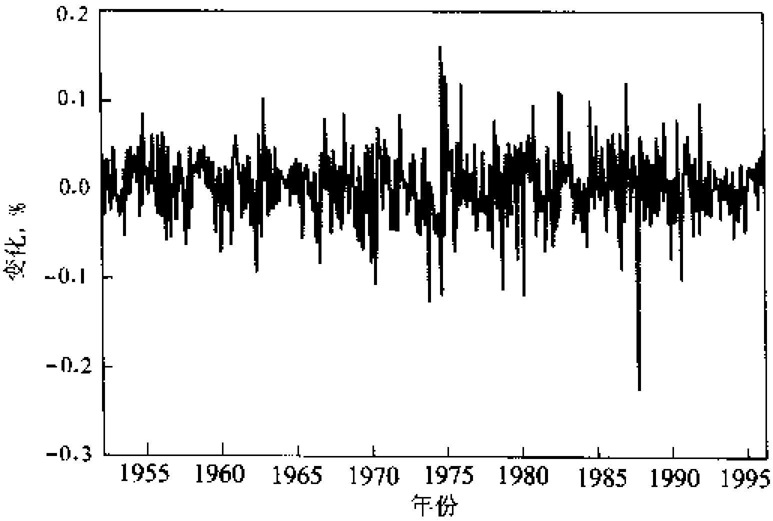


图22.8　NYSE价格指数的月度百分比变化：1952—1995

为了刻画图中所见股票收益的波动性，让我们考虑一个十分简单的模型：

$$Y_t = \beta_1 + u_t \tag{22.10.11}$$

其中，Y_t = NYSE股票价格指数的百分比变化，u_t = 随机误差项。

注意，除了截距项外，模型中没有其他的解释变量。我们从数据中得到如下OLS回归：

$$\hat{Y}_t = 0.006\,86$$
$$t = (3.883\,5) \qquad (22.10.12)$$
$$d = 1.921\,5$$

这个截距项表示了什么呢? 它无非就是NYSE指数的平均百分比回报率，或 Y_t 的均值（你能验证吗)。因此，在整个样本期间，NYSE指数的月平均回报率约为0.006 9%。

861 我们现在从上述回归得到残差，并估计ARCH(1)模型，结果如下：

$$\hat{u}_t^2 = 0.001\,45 + 0.116\,7\hat{u}_{t-1}^2 \qquad (22.10.13)$$
$$t = (8.892\,9)\ (2.693\,4) \qquad R^2 = 0.013\,6 \qquad d = 2.012\,1$$

其中 $\hat{u}_t$ 为从回归（22.10.12）中估计的残差。由于滞后误差项的平方在统计上是显著的（p 值约为0.007)，所以误差方差看来是相关的；即存在ARCH效应。我们也试了更高阶的ARCH模型，但只有ARCH(1)才是统计显著的。

出现ARCH时怎么办

记得我们曾讨论了几种纠正异方差性的方法，基本上都是应用OLS去变换数据。记住，用于变换数据的OLS就是广义最小二乘（GLS)。若发现存在ARCH效应，我们就要使用GLS。我们不再深究其技术上的细节，因为这些都超出了本书的范围。[25]幸运的是，诸如Eviews，Shazam，Microfit和Pc-Give等软件包现在都为了方便读者而例行估计这种模型了。

对德宾-沃森 d 和ARCH效应的一句忠告

我们已经几次提醒读者，显著的 d 统计量并不总是意味着所处理数据中存在着显著的自相关。一个显著的 d 值时常预示着我们在第13章中讨论的模型设定误差。我们现在又有了一种因ARCH效应而导致的设定误差。因此，在一个时间序列回归中，若得到一个显著的 d 值时，在以其面值接受这个 d 统计量之前，我们应该检验ARCH效应。习题22.23给出了一个例子。

对GARCH模型的一个注解

自1982年“发现”ARCH模型以来，ARCH建模已经成为一个蒸蒸日上的产业，对原模型提出了各种各样的变形。其中广为流传的一个，便是由
862 波勒斯列夫（Bollerslev）最早提出的**广义自回归条件异方差**（GARCH）模

型。最简单的 GARCH 模型便是 GARCH(1, 1)，可写作：

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \tag{22.10.14}$$

即 t 时期 u 的条件方差不仅取决于上一时期误差项的平方［如在 ARCH(1) 中一样］，还取决于上一时期的条件方差。这个模型可以推广至GARCH(p, q)，即模型中有误差平方项的 p 阶滞后和条件方差的 q 阶滞后。

因为这些模型太复杂，所以我们仍不深究其技术细节，只是指出一个 GARCH(1, 1) 模型等价于一个 ARCH(2) 模型，而一个 GARCH(p, q) 模型就等价于一个 ARCH($p+q$) 模型。[27]

就我们的美英汇率和 NYSE 股票回报的例子来看，我们已经说过，一个 ARCH(2) 模型并不显著，这表明一个 GARCH(1, 1) 模型在这些例子中可能也不适合。

§22.11 总结性例子

我们再给出几个例子来结束本章，以解说我们在本章中已经得到的一些要点。

1969 年 1 月—2000 年 1 月招聘指数（HWI）和失业率（UN）之间的关系

为了研究美国的两个劳动市场情况指标 HWI 和 UN 之间的因果关系，贾莫托（Marc A.Giammatteo）考虑了如下回归模型[28]：

$$\mathrm{HWI}_t = \alpha_0 + \sum_{i=1}^{25} \alpha_i \mathrm{UN}_{t-i} + \sum_{j=1}^{25} \beta_j \mathrm{HWI}_{t-j} \tag{22.11.1}$$

$$\mathrm{UN}_t = \alpha_0 + \sum_{i=1}^{25} \lambda_i \mathrm{UN}_{t-i} + \sum_{j=1}^{25} \delta_j \mathrm{HWI}_{t-j} \tag{22.11.2}$$

为了节省篇幅，我们不再给出实际回归结果，但此研究所突现的主要结论是，这两个劳动市场指标之间有双向因果关系，而且这个结论不随滞后长度的变化而改变。HWI 和 UN 的数据在数据盘中给出。

868 **日美汇率的 ARIMA 建模：1971 年 1 月—1998 年 12 月**[29]

日美汇率（¥/$）是一个关键汇率。从¥/$ 月度数据的对数发现，这个汇率的水平值表现出典型的非平稳时间序列类型。但考察一阶差分则发现，它们是平稳的；这里的图与图 22.8 极其相似。

单位根检验证实了¥/$ 对数的一阶差分是平稳的。在考察了对数一阶差分的相关图之后，我们估计了如下 ARIMA (1, 0, 2) 模型：

$$\begin{aligned} \hat{Y}_t = & -0.0034 + 0.9678\hat{Y}_{t-1} - 0.5866u_{t-1} - 0.4057u_{t-2} \\ t = & (-4.3638) \quad (67.5439) \quad (-11.4361) \quad (-7.9532) \end{aligned} \tag{22.11.3}$$

$$R^2 = 0.1454 \qquad d = 1.9803$$

其中 Y_t 表示￥/$ 对数的一阶差分，u 表示一个白噪音误差项。

为节省篇幅，我们把上述分析背后的数据放到数据盘中。我们鼓励读者利用这些数据去尝试其他的模型，并比较它们的预测表现。

美国通货膨胀率的 ARCH 模型：1947 年 1 月—2001 年 1 月

为了看出以 CPI 度量的美国通货膨胀率中是否存在 ARCH 效应，我们得到了从 1947 年 1 月至 2001 年 1 月的 CPI 数据。CPI 对数的描点图表明，这个时间序列是非平稳的。但图 22.9 所示的 CPI 对数的一阶差分描点图则表明，尽管它是平稳的，但也出现了可观的波动性。

864 根据回归（22.10.12）和（22.10.13）所勾勒的程序，我们首先估计 CPI 一阶差分的对数对一个常数项的回归，并得到此方程的残差。将这些残差平方，我们得到如下 ARCH(3) 模型：

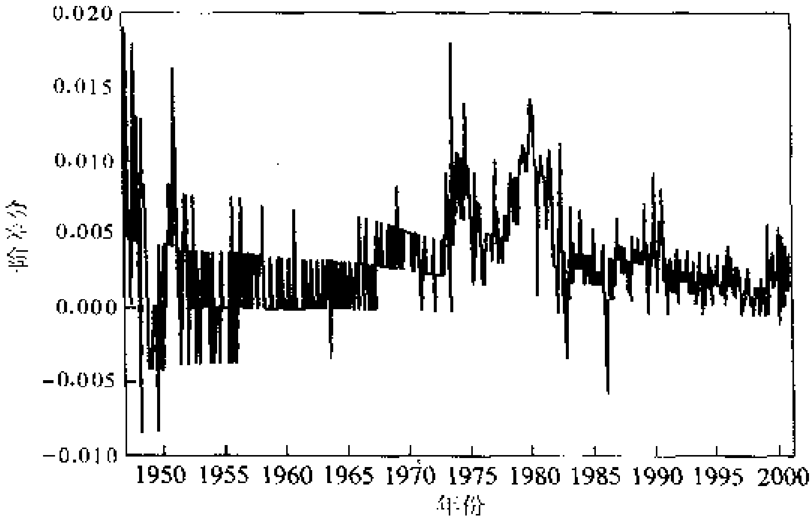


图 22.9　CPI 对数的一阶差分

$$
\begin{aligned}
\hat{u}_t^2 &= 0.000\,052 + 0.339\,9\hat{u}_{t-1}^2 + 0.133\,8\hat{u}_{t-2}^2 + 0.092\,0\hat{u}_{t-3}^2 \\
t &= (5.189\,3) \quad (8.727\,0) \quad (3.562\,0) \quad (2.538\,7) \\
& R^2 = 0.215\,3 \qquad d = 2.033\,4
\end{aligned}
\tag{22.11.4}
$$

如你所见，由于当月的波动性取决于前三个月的波动性，所以通货膨胀的变动有相当大的持久性。建议读者从政府资源中获得 CPI 的数据，并看是否有其他更好的模型，特别是 GARCH 模型。

§22.12　要点与结论

1. 博克斯-詹金斯和 VAR 经济预测方法是与传统的单一方程和联立方程模型相对立的。

2. 对于一个时间序列的预测，基本的博克斯-詹金斯策略如下：

a. 首先检验序列的平稳性。可通过自相关函数（ACF）和偏自相关函数（PACF）的计算或者通过正式的单位根分析来完成这一步骤。对应于ACF和PACF的相关图往往是一种良好的视觉诊断工具。

b. 如果时间序列不是平稳的，将它差分一次或多次，以获得平稳性。

c. 然后计算此平稳时间序列的ACF和PACF，以判明序列是纯自回归的或纯移动平均类型的，或这两者的一种混合体。从表22.1所提供的概略性指引，我们将能决定有待拟合的ARMA过程中的 p 和 q 值。在此阶段中所选的ARMA(p, q) 模型是尝试性质的。

d. 然后估计此尝试性模型。

e. 分析尝试性模型的残差，看这些残差是不是白噪音。如果是，则此尝试性模型也许是所依据的随机过程的一个良好的迫近；如果不是，则整个程序要从头做起，因此，博克斯-詹金斯方法是一反复过程。

f. 最后选定的模型便可用于预测。

3. VAR预测方法同时考虑多个时间序列，其特点如下：

a. 所有的变量都被看作是内生的，在这个意义上VAR是一个真正的联立方程组。

865 b. 在VAR建模中，一个变量的值被表达为该变量和模型中所含有的全部其他变量的过去或滞后值的一个线性函数。

c. 如果每一方程都含有同样个数的系统中的滞后变量，它就可以用OLS来估计，而无须求助于诸如二阶最小二乘（2SLS）或似无关回归（SURE）等任何系统方法。

d. VAR建模的简单性，也许就是它的欠缺性。鉴于在大多数的经济分析中一般能获得的观测值的个数是有限的，引进每一变量的多个滞后会耗费掉大量的自由度。[30]

e. 如果每一方程都有多个滞后，要解释每一个系数，特别是当系数的符号正负交替时，就不是容易的事。正因为这个缘故，人们转而分析VAR建模中的脉冲响应函数（IRF），看因变量是怎样在系统中的一个或多个方程受到冲击时做出响应的。

f. 关于各种预测方法的优越性有大量的争执和辩论。单方程、联立方程、博克斯-詹金斯和VAR预测方法各有其仰慕者和诋毁者。我们所能说的是，没有哪一个方法能适用于所有的情形。如果真有这样一种方法，我们就没有必要讨论各种不同的方法了。有一点是肯定的：现在，博克斯-詹金斯和VAR方法论已成为计量经济学的一个组成部分。

4. 我们在本章还考虑了一类特殊的模型，ARCH和GARCH模型，它们在分析诸如股票价格、通货膨胀率和汇率等金融时间序列时特别有用。这些模型的一个明显特征是，由于群集波动现象，所以不同时间的误差方差可能相关。我们还指出，在许多情况下，一个显著的德宾-沃森 d 可能事实上起因于ARCH或GARCH效应。

习 题

问答题

22.1 什么是经济预测的主要方法?

22.2 联立方程和博克斯-詹金斯两经济预测方法的主要差别何在?

22.3 略述应用博克斯-詹金斯预测方法的主要步骤。

22.4 如果把博克斯-詹金斯技术应用于非平稳时间序列，会出现什么情况?

866 **22.5** 博克斯-詹金斯和 VAR 两经济预测方法的差别何在?

22.6 在什么意义下 VAR 是乏理论的?

22.7 “如果主要目的在于预测，则 VAR 是够好的。”从严评议这一陈述。

22.8 既然在一个 VAR 中引进滞后的个数可以是一个主观的问题，那么在一个具体的应用中，怎样决定引进多少个滞后呢?

22.9 “博克斯-詹金斯和 VAR 是凭测量而无理论的典范。”试加评论这一陈述。

22.10 格兰杰因果检验和 VAR 建模如果有关系的话，是些什么关系?

解答题

22.11 考虑表 21.1 所给的 PDI（个人可支配收入）数据。假如你要对这些数据拟合一个适当的 ARIMA 模型。略述完成这一工作的步骤。

22.12 对表 21.1 所给的 PCE（个人消费支出）数据重复做习题 22.11。

22.13 对表 21.1 所给的利润数据重做习题 22.11。

22.14 对表 21.1 所给的股利数据重做习题 22.11。

22.15 在第 13.9 节中，我们向你介绍了决定滞后长度的施瓦茨准则，你怎样利用这个准则去决定一个 VAR 模型中的适当滞后长度?

22.16 利用表 21.1 所给的 PCE 和 PDI 数据构造 1990 年第 1 季度至 1990 年第 4 季度这个时期的一个二维 VAR 模型。利用此模型预测这两个变量在 1991 年的 4 个季度里的值。并将预测值同表 21.1 所给的实测值做比较。

22.17 利用股息和利润数据，重做习题 22.16。

***22.18** 利用任何一个统计软件包，对你在习题 22.16 中构造的 VAR 模型估计一个直至 8 期滞后的脉冲响应函数。

***22.19** 用你在习题 22.18 中构造的 VAR 模型重做习题 22.17。

22.20 参照表 22.4 所给的 VAR 回归结果。根据那里报道的三个回归

的各种 F 检验，你能对三个变量的因果性质说些什么？

22.21 继续考虑习题 20.20，你能猜测为什么论文的作者们选择百分率变化的形式，而不是选择水平值的形式来表达模型中的三个变量吗？（提示：平稳性。）

22.22 利用表 17.3 给出的加拿大数据，看 M1 和 R 是否为平稳的随机变量？若不是，它们是否协积？给出必要的计算。

22.23 继续使用表 17.3 中的数据。现在考虑加拿大货币需求的如下简单模型：

$$\ln \mathrm{M1}_t = \beta_1 + \beta_2 \ln \mathrm{GDP}_t + \ln R_t + u_t$$

a. 你如何解释此模型的参数？

b. 从模型中得到残差，并看是否存在 ARCH 效应。

867 **22.24** 参照（22.11.4）中给出的 ARCH（3）模型。利用同样的数据，我们估计了如下 ARCH（1）模型：

$$\hat{u}_t^2 = 0.000\,000\,78 + 0.373\,7\hat{u}_{t-1}^2$$
$$t = \quad (7.584\,3) \qquad (10.235\,1)$$
$$R^2 = 0.139\,7 \qquad d = 1.989\,6$$

你如何在这两个模型之间做出选择？给出必要的计算。

【注释】

[1] G.P.E.Box and G.M.Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Holden Day, San Francisco, 1978.

[2] 至于对这些方法相对简单的说明，参见 Spyros Makridakis, Steven C.Wheelwright, and Rob J.Hyndman, *Forecasting Methods and Applications*, 3d ed., John Wiley & Sons, New York, 1998。

[3] 作为联立方程模型用于预测的教材，参见 Robert S.Pindyck and Daniel L.Rubinfeld, *Econometric Models & Economic Forecasts*, 4d ed.McGraw-Hill, New York, 1998, Part Ⅲ。

[4] Robert E.Lucas, "Econometric Policy Evaluation: A Critique," in Carnegie-Rochester Conference Series, *The Phillips Curve*, North-Holland, Amsterdam, 1976, pp.19–46. 此文与其他论文使得卢卡斯获得诺贝尔经济学奖。

[5] 参看 Pindyck and Rubinfeld，前引文献，第 3 篇；Alan Pankratz, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, New York, 1991（这是一本应用方面的书）；以及 Andrew Harvey, *Econometric Analysis of Time Series*, The MIT Press, 2d ed., Cambridge, Mass., 1990（这是一本较高深的书）。一本讨论透彻然而易读的读物还见于 Terence C.Mills, *Time Series Techniques for Economists*, Cambridge University Press, New York, 1990。

[6] Michael Pokorny, *An Introduction to Econometrics*, Basil Blackwell, New York, 1987, p.343.

---

* 选作题。

[7] 对时间序列数据来说，Y_t 和 Y_{t-k}之间的相关大部分是由于它们和介于 t 和 $t-k$ 之间的滞后值 Y_{t-1}，Y_{t-2}，…，Y_{t-k-1}的相关。偏相关 ρ_{kk}将除掉这些中介变量的影响。

[8] 要说清楚这个序列的方差是否平稳，特别是在 1979—1980 年前后，是不容易的。1979 年石油禁运以及 1979 年联邦储备银行货币政策的显著变化也许和这种困难有关。

[9] 虽然标准的计算机软件包都把这些计算作为例行程序，但我们通过这些详细的计算得以说明所涉及的操作步骤。

[10] 关于这个问题的一个易读的论述，参看 Terence C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, New York, 1990, part Ⅲ。

[11] C. A. Sims, "Macroeconomics and Reality," *Econometrica*, vol.48, 1980, pp.1 – 48.

[12] 可以利用 **SURE**（似无关回归）技术同时估计两方程。然而，由于每个回归都含有同样多个滞后内生变量，每个方程的 OLS 估计将各自产生相同且（有效）的估计。

[13] 有时考虑到趋势和季节因素而包含有一些纯外生的变量。

[14] 例如，参看 T. Kinal and J. B. Ratner, "Regional Forecasting Models with Vector Autoregression: The Case of New York State," *Discussion* Paper #155, Department of Economics, State University of New York at Albany, 1982。

[15] 对于一个有 m 个变量、p 个滞后值的 m 方程 VAR 模型，总共有（$m + pm^2$）个参数有待估计。

[16] Andrew Harvey, *The Econometric Analysis of Time Series*, The MIT Press, 2d ed., Cambridge, Mass., 1990, p.83.

[17] D. E. Runkle, "Vector Autoregression and Reality," *Journal of Business and Economic Statistics*, vol.5, 1987, pp.437 – 454.

[18] S. McNees, "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, vol.4, 1986, pp.5 – 15; E. Mahmoud, "Accuracy in Forecasting: A Survey," *Journal of Forecasting*, vol.3, 1984, pp.139 – 159.

[19] Thomas B. Fomby and Joseph G. Hirschberg, "Texas in Transition: Dependence on Oil and the National Economy," *Economic Review*, Federal Reserve Bank of Dallas, January 1989, pp.11 – 28.

[20] Philip Hans Franses, *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, New York, 1998, p.155.

[21] R. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, vol.50. no.1, 1982, pp.987 – 1007. 也可参见 A. Bera and M. Higgins, "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, vol.7, 1993, pp.305 – 366.

[22] 你可能想知道我们为什么不用 X_t 的方差 $\sum X_t^2/n$ 来度量波动性。这是因为我们想考虑资产价格波动性随着时间的变化。如果我们使用 X_t 的方差，那它对每个给定的数据集都只有一个值。

[23] 一个技术性注解：记得对经典线性回归模型而言，u_t 的方差假定为 σ^2，在目前的背景下，它是无条件方差。若 $\alpha_1 < 1$，则稳定性条件可以写成 $\sigma^2 = \alpha_0 +$

$\alpha_1\sigma^2$；即 $\sigma^2 = \alpha_0 / (1-\alpha_1)$。这就表明，$u$ 的无条件方差并不取决于 t，但取决于 ARCH 参数 α_1。

[24] 此图和下面给出的回归结果所采用的数据都来自 Gary Koop, *Analysis of Economic Data*, John Wiley & Sons, New York, 2000（见数据盘）。股票价格指数的月百分比变化可视为指数的回报率。

[25] 参阅 Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, New York, 1993, Sec.16.4 and William H. Greene, *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, Sec.18.5。

[26] T. Bollerslev, "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, vol.31, 1986, pp.307－326.

[27] 详细情况参见 Davidson and MacKinnon 的前引文献，第 558－560 页。

[28] Marc A. Giammatteo (West Point, Class of 2000), "The Relationship between the Help Wanted Index and the Unemployment Rate," unpublished term paper.（为了与我们的符号保持一致而对原来的符号作了改变。）

[29] 我想感谢 Gregory M. Ogborn and Marc C. Ogborn (West Point, Class of 2001) 收集并分析了这些数据。

[30] 贝叶斯统计学的信徒相信这个问题可大大得到缓解。参见 R. Litterman, "A Statistical Approach to Economic Forecasting," *Journal of Business and Economic Statistics*, vol.4, 1986, pp.1－4。

附录 A　统计学中的若干概念复习

869　本附录对本书中遇到的一些统计学概念做一个非常简略的介绍。讨论是非严格的，而且不加证明，原因是已有多种统计学书籍出色地完成了这一工作。本附录末列出一些书籍。

§A.1　总和与乘积运算子

希腊大写字母 $\sum$(sigma) 表示总和。例如，

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

总和运算子 $\sum$ 的一些重要性质是：

1. $\sum_{i=k}^{n} k = nk$，其中 k 是常数。例如，$\sum_{i=1}^{4} 3 = 4 \cdot 3 = 12$。

2. $\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i$，其中 k 是常数。

3. $\sum_{i=1}^{n}(a + bx_i) = na + b\sum_{i=1}^{n} x_i$，其中 a 和 b 是常数，并且这里利用了上面的性质 1 和 2。

4. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n}x_i + \sum_{i=1}^{n}y_i$。总和运算子还可推广到多重总和。例如，双重总和运算子$\sum\sum$的定义是：

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij} &= \sum_{i=1}^{n}(x_{i1} + x_{i2} + \cdots + x_{im}) \\ &= (x_{11} + x_{21} + \cdots + x_{n1}) + (x_{12} + x_{22} + \cdots + x_{n2}) \\ &\quad + \cdots + (x_{1m} + x_{2m} + \cdots + x_{nm})\end{aligned}$$

870 $\sum\sum$的一些性质是：

1. $\sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n}x_{ij}$；就是说，双重总和的运算次序是可交换的。

2. $\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j = \sum_{i=1}^{n}x_i\sum_{j=1}^{m}y_j$。

3. $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} + y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}$。

4. $[\sum_{i=1}^{n}x_i]^2 = \sum_{i=1}^{n}x_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_i x_j$
$= \sum_{i=1}^{n}x_i^2 + 2\sum_{i<j}x_i x_j$。

乘积运算子Π定义为：

$$\prod_{i=1}^{n}x_i = x_1 . x_2 \ldots x_n$$

例如， $\prod_{i=1}^{3}x_i = x_1 . x_2 . x_3$

§A.2 样本空间、样本点与事件

一个随机或机遇试验的所有可能结果的集合叫做**总体**（population）或**样本空间**（sample space），而此样本空间的每一元素叫做一个**样本点**（sample point）。例如，在抛掷两枚硬币的试验中，样本空间由 *HH*、*HT*、*TH* 和 *TT* 四个可能结果构成。其中 *HH* 表示第一次抛掷出现正面，第二次抛掷也出现正面；*HT* 表示第一次抛掷出现正面，第二次抛掷出现反面，等等。上述每一种结果构成一个样本点。

一个**事件**（event）就是样本空间的一个子集。例如，令 *A* 表示出现一个正面和一个反面，那么，在上述可能结果中，只有 *HT* 和 *TH* 两个结果属于*A*，而 *A* 就是一个事件，类似地，在抛掷两枚硬币的试验中，出现两个正面是一个事件。如果一个事件的出现排斥另一事件的出现，我们说事件是**互斥的**（mutually exclusive），在上述试验中，如果 *HH* 出现，事件 *HT* 就不可能同时出现。我们说事件是**穷举的**（cxhaustive），如果它们举尽了一个试验的全部可能结果。例如，在这个例子中，事件（a）两个正面，（b）两个

反面和（c）一正一反举尽了试验的全部结果，因而它们是（集体地）穷举事件。

§A.3　概率与随机变量概率

871 令 A 为样本空间中的一个事件。事件 A 的概率，记为 $P(A)$，是指在重复试验中事件 A 将出现的次数的比例。另一提法，在总共 n 个等可能的试验结果中，如果有 m 个有利于事件 A 的出现，我们就定义比率 m/n 为 A 的**相对频率**（relative frequency）。当 n 很大时，这个相对频率就是 A 的概率的一个很好的近似值。(在等可能的前提下，这个频率就可理解为概率。——译者注)

概率的性质。 $P(A)$ 是一个实值函数，并且有如下的性质：

1. 对每个 A 有 $0 \leqslant P(A) \leqslant 1$。

2. 如果 A、B、C，…构成事件的一个穷举集，则 $P(A+B+C+\cdots)=1$，其中 $A+B+C$ 表示 A 或 B 或 C，如此等等。

3. 如果 A、B、C…是互斥事件，则

$$P(A+B+C+\cdots)=P(A)+P(B)+P(C)+\cdots$$

例 1.

考虑投掷一颗有 1 到 6 点的骰子的试验，样本空间由结果 1，2，3，4，5 和 6 构成。这 6 个事件因此穷举了整个样本空间。因为共有 6 个等可能结果，而任一结果都有同等的机会出现，故出现任一结果的概率都是 1/6。既然 1，2，3，4，5 和 6 构成事件的穷举集，故 $P(1+2+3+4+5+6)=1$，其中 1，2，3，…指点 1 或点 2 或点 3 等等的概率，又因任何两点都不能同时出现，即 1，2，…，6 是互斥事件，故 $P(1+2+3+4+5+6)=P(1)+P(2)+\cdots+P(6)=1$。

随机变量

一个变量如果它的值由随机试验的结果决定，就叫做**随机变量**（random variable，rv）。通常随机变量用大写字母 X、Y、Z 等等表示，而它的值由小写字母 x、y、z 等等表示。

随机变量可以是**离散的**或**连续的**。一个离散 rv 只取有限（或可数无穷）多个值。[2]例如，投掷两颗骰子，各有数码 1 至 6，如果我们定义随机变量 X 为两骰子出现的数码之和，则 X 将取如下数码之一：2，3，4，5，6，7，8，9，10，11 或 12。从而它是一离散随机变量。另一方面，一个连续随机变量可以取某一区间的任何值。例如，某人的身高是一连续变量，它可以取
872 某个范围内，比方说 60～65 英寸之间的任何值，这个值的读数还有赖于测

量的精度。

§A.4 概率密度函数

离散随机变量的概率密度函数

令 X 为取相异值 x_1，x_2，…，x_n，…的一个离散 rv，则函数：

$$\begin{aligned} f(x) &= P(X=x_i) \quad \text{当 } i=1,2\cdots,n,\cdots \\ &= 0 \quad\quad\quad\quad\quad\ \text{当 } x \neq x_i \end{aligned}$$

叫做 X 的**离散概率密度函数**（PDF），其中 $P(X=x_i)$ 表示离散 rv X 取值 x_i 的概率。

例 2.

在两颗骰子的投掷中，两骰子所出现的数码之和这一随机变量 X，可取所示的 11 个数值之一。此变量的 PDF 可表示如下（还可参看图 A.1）：

$$\begin{array}{cccccccccccc} x= & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ f(x)= & \left(\frac{1}{36}\right) & \left(\frac{2}{36}\right) & \left(\frac{3}{36}\right) & \left(\frac{4}{36}\right) & \left(\frac{5}{36}\right) & \left(\frac{6}{36}\right) & \left(\frac{5}{36}\right) & \left(\frac{4}{36}\right) & \left(\frac{3}{36}\right) & \left(\frac{2}{36}\right) & \left(\frac{1}{36}\right) \end{array}$$

这些概率容易加以证实。在全部 36 个可能结果中，有一个有利于（总和）数 2，有两个有利于数 3（总和 3 的出现或者因为第一个骰子出现 1，第二个骰子出现 2，或者因为第一个骰子出现 2，第二个骰子出现 1），余下类推。

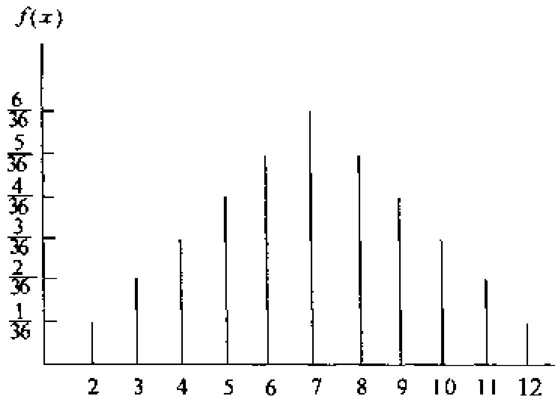


图 A.1　例 2 的离散随机变量的概率密度函数

连续随机变量的概率密度函数

873　令 X 为一连续 rv。我们说 $f(x)$ 是 X 的 PDF 如果下述条件成立：

$$f(x) \geqslant 0$$

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

$$\int_{a}^{b} f(x)\mathrm{d}x = P(a \leqslant x \leqslant b)$$

其中 $f(x)\mathrm{d}x$ 称概率元素（与一连续变量的一个微小区间相对应的概率），而 $P(a\leqslant x\leqslant b)$ 指 x 落在 a 至 b 区间上的概率，用几何图形表示，我们有图 A.2。

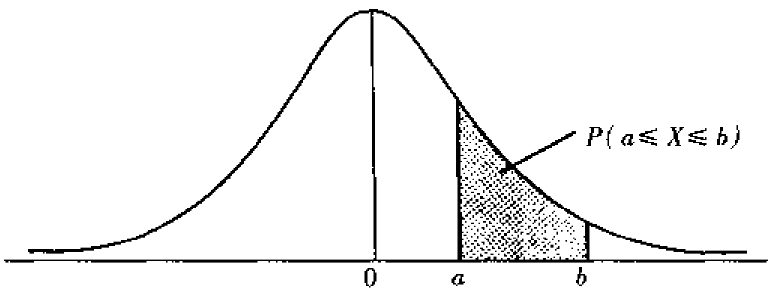


图 A.2　连续随机变量的概率密度函数

与离散 rv 相对照，对于一个连续 rv，X 取某一特定值的概率为零[3]；对于这样的一个变量，概率仅对一个给定的范围或区间，诸如图 A.2 的 (a, b)，才是可测的。

例 3.

考虑如下密度函数：

$$f(x) = \frac{1}{9}x^2 \quad 0 \leqslant x \leqslant 3$$

容易证实，对所有从 0 到 3 的 x，$f(x) \geqslant 0$，并且 $\int_0^3 \frac{1}{9}x^2\mathrm{d}x = 1$。（注：这个积分是 $\left(\frac{1}{27}x^3\Big|_0^3\right) = 1$。）如果我们想估计上述 PDF 比方说在 0 与 1 之间的值，我们就得到 $\int_0^1 \frac{1}{9}x^2\mathrm{d}x = \left(\frac{1}{27}x^3\Big|_0^1\right) = \frac{1}{27}$；就是说，$x$ 落在 0 和 1 之间的概率是 1/27。

联合概率密度

874 **离散联合 PDF。**令 X 和 Y 为两个离散随机变量，则函数：

$$\begin{aligned} f(x, y) &= P(X = x \text{ 且 } Y = y) \\ &= 0 \qquad \text{当 } X \neq x \text{ 且 } Y \neq y \end{aligned}$$

称**离散联合概率密度函数**，并给出 X 取值 x 和 Y 取值 y 的概率。

例 4.

下表给出离散随机变量 X 和 Y 的联合 PDF。

		X			
		−2	0	2	3
Y	3	0.27	0.08	0.16	0
	6	0	0.04	0.10	0.35

此表告诉我们 X 取值 -2 的同时 Y 取值 3 的概率是 0.27；X 取值 3 的同时 Y 取值 6 的概率是 0.35；等等。

边际概率密度函数

相对于 $f(x, y)$ 来说，$f(x)$ 和 $f(y)$ 称为个别或**边际**（marginal）**概率密度函数**。这些边际 PDF 的推导如下：

$$f(x) = \sum_y f(x,y) \qquad X \text{ 的边际 PDF}$$

$$f(y) = \sum_x f(x,y) \qquad Y \text{ 的边际 PDF}$$

其中，比如说，$\sum_y$ 表示对所有的 Y 值求和，而 $\sum_x$ 表示对所有的 X 值求和。

例 5.

考虑例 4 中的数据，X 的边际 PDF 可求得如下：

$$f(x=-2) = \sum_y f(x,y) = 0.27 + 0 = 0.27$$

$$f(x=0) = \sum_y f(x,y) = 0.08 + 0.04 = 0.12$$

$$f(x=2) = \sum_y f(x,y) = 0.16 + 0.10 = 0.26$$

$$f(x=3) = \sum_y f(x,y) = 0 + 0.35 = 0.35$$

875 同理，求得 Y 的边际 PDF 为：

$$f(y=3) = \sum_x f(x,y) = 0.27 + 0.08 + 0.16 + 0 = 0.51$$

$$f(y=6) = \sum_x f(x,y) = 0 + 0.04 + 0.10 + 0.35 = 0.49$$

如本例所表明的，我们把列的数值相加而得 X 的边际 PDF，把行的数值相加而得 Y 的边际 PDF。注意，对所有的 X 值取 $\sum_x f(x)$ 就等于 1，对所有的 Y 值取 $\sum_y f(y)$ 也等于 1。（为什么？）

条件 PDF。如第 2 章所表明的，在回归分析中我们常感兴趣的是，研究一个变量在另一（些）变量的给定值的条件下的行为。这可通过条件 PDF 来做到。函数：

$$f(x \mid y) = P(X = x \mid Y = y)$$

叫做 X 的**条件**（conditional）**PDF**；它给出 Y 取给定值 y 的条件下 X 取值 x

的概率。类似地，

$$f(y|x)=P(Y=y|X=x)$$

给出 Y 的条件 PDF。

这些条件 PDF 可求得如下：

$$f(x|y)=\frac{f(x,y)}{f(y)} \qquad X \text{ 的条件 PDF}$$

$$f(y|x)=\frac{f(x,y)}{f(x)} \qquad Y \text{ 的条件 PDF}$$

以上表达式表明，一个变量的条件 PDF 可表达为联合 PDF 和另一变量的边际 PDF 之比。

例 6.

继续用例 4 和 5，让我们计算以下条件概率：

$$f(X=-2|Y=3)=\frac{f(X=-2,\ Y=3)}{f(Y=3)}=0.27/0.51=0.53$$

注意，无条件概率 $f(X=-2)$ 是 0.27，但若 Y 已取定 3，则 X 取值 -2 的概率是 0.53。

$$f(X=2|Y=6)=\frac{f(X=2,\ Y=6)}{f(Y=6)}=0.10/0.49=0.20$$

再次注意到 X 取值 2 的无条件概率是 0.26，而不同于 Y 取定 6 时的概率 0.20。

统计独立性

两个随机变量 X 和 Y 是统计上独立的，当且仅当：

$$f(x,y)=f(x)f(y)$$

就是联合 PDF 可表达为诸边际 PDF 的乘积。

例 7.

袋中装有编号为 1，2 和 3 的三个球，从中有回置地随机抽取两个（就是，第一次抽出的球被回置后再抽第二次。）令 X 表示第一次抽出的球的号码，而 Y 表示第二次抽出的球的号码，下表给出 X 和 Y 的联合 PDF：

		X		
		1	2	3
Y	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
	2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
	3	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

现在 $f(X=1,X=1)=\frac{1}{9}$，$f(X=1)=\frac{1}{3}$（将第 1 列相加得到），并且

$f(Y=1)=\frac{1}{3}$（将第1行相加得到）。因 $f(X,Y)=f(X)f(Y)$，故在本例中我们说两个变量在统计上独立。容易验证，对上表所给 X 和 Y 值的任意其他组合，联合PDF都可分解为个别PDF因子。

可以证明，例4中所给的 X 和 Y 变量由于两边际PDF的乘积不等于联合PDF，所以不是统计上独立的。[注：如果两变量是统计上独立的，则必须对 X 和 Y 的一切组合都有 $f(x, y)=f(x)f(y)$。]

连续联合PDF。两个连续变量 X 和 Y 的PDF $f(x, y)$ 是指

$$f(x, y) \geqslant 0$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$$

$$\int_{c}^{d}\int_{a}^{b} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = P(a \leqslant x \leqslant b, c \leqslant y \leqslant d)$$

877 **例8.**

考虑如下PDF：

$$f(x,y)=2-x-y\quad 0\leqslant x\leqslant 1;\ 0\leqslant y\leqslant 1$$

显然，$f(x,y)\geqslant 0$，此外[4]，

$$\int_{0}^{1}\int_{0}^{1}(2-x-y)\,\mathrm{d}x\mathrm{d}y = 1$$

X 和 Y 的边际PDF可获得如下：

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y \qquad X\text{ 的边际PDF}$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \qquad Y\text{ 的边际PDF}$$

例9.

例8所给的联合PDF的两个边际PDF有如下列：

$$f(x) = \int_{0}^{1} f(x,y)\,\mathrm{d}y = \int_{0}^{1}(2-x-y)\,\mathrm{d}y$$

$$= \left(2y - xy - \frac{y^2}{2}\right)\Big|_{0}^{1} = \frac{3}{2} - x \quad 0 \leqslant x \leqslant 1$$

$$f(y) = \int_{0}^{1}(2-x-y)\,\mathrm{d}x$$

$$= \left(2x - xy - \frac{x^2}{2}\right)\Big|_{0}^{1} = \frac{3}{2} - y \quad 0 \leqslant y \leqslant 1$$

为了看出例8的两个变量是否统计上独立的，我们需要判明 $f(x, y)=f(x)f(y)$ 是否成立。由于

$$(2-x-y)\neq\left(\frac{3}{2}-x\right)\left(\frac{3}{2}-y\right),$$

我们可以说这两个变量不是统计上独立的。

§A.5 概率分布的特征值

878 一个概率分布常常能用少数几个它的特征值——称之为分布的**矩**（moments）来概括它，用得最广的一些矩是**均值**（mean）或**期望值**（expected value）和**方差**（variance）。

期 望 值

一个离散 rv X 的期望值，记为 $E(X)$，定义如下：

$$E(X)=\sum_x xf(x)$$

其中 $\sum_x$ 表示对所有的 X 值求和，而 $f(x)$ 为（离散）变量 X 的 PDF。

例 10.

考虑例 2 中投掷两个骰子出现的两个数码之和的概率分布（见图 A.1）。将那里给出的各个 X 值乘以它们的概率并对所有的观测值求和，便得：

$$E(X)=2\left(\frac{1}{36}\right)+3\left(\frac{2}{36}\right)+4\left(\frac{3}{36}\right)+\cdots+12\left(\frac{1}{36}\right)=7$$

这就是一次投掷两颗骰子所观测的数码和的平均值。

例 11.

估计例 4 所给数据的 $E(X)$ 和 $E(Y)$。我们曾看到：

x	-2	0	2	3
$f(x)$	0.27	0.12	0.26	0.35

因此，

$$\begin{aligned}E(X)&=\sum_x xf(x)\\&=(-2)(0.27)+(0)(0.12)+(2)(0.26)+(3)(0.35)\\&=1.03\end{aligned}$$

类似地，

y	3	6
$f(y)$	0.51	0.49

$$\begin{aligned}E(Y)&=\sum_y yf(y)\\&=(3)(0.51)+(6)(0.49)\end{aligned}$$

$$= 4.47$$

879 一个连续 rv 的期望值定义为：

$$E(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$$

它和离散 rv 的期望值的惟一差别是这里我们用积分符号代替了总和符号。

例 12.

让我们来求例 3 中所给连续 PDF 的期望值：

$$\begin{aligned} E(X) &= \int_0^3 x\left(\frac{x^2}{9}\right)\mathrm{d}x \\ &= \frac{1}{9}\left[\left(\frac{x^4}{4}\right)\right]_0^3 \\ &= \frac{9}{4} \\ &= 2.25 \end{aligned}$$

期望值的性质

1. 一个常数的期望值是该常数本身。例如，若 b 是一常数，则 $E(b)=b$。

2. 如果 a 和 b 是常数，则：

$$E(aX+b)=aE(X)+b$$

这可加以推广，如果 X_1、X_2，…，X_N 是 N 个随机变量，并且 a_1、a_2，…，a_N 和 b 是常数，则：

$$\begin{aligned} &E(a_1X_1+a_2X_2+\cdots+a_NX_N+b) \\ &=a_1E(X_1)+a_2E(X_2)+\cdots+a_NE(X_N)+b \end{aligned}$$

3. 如果 X 和 Y 是独立随机变量，则：

$$E(XY)=E(X)E(Y)$$

意谓乘积 XY 的期望值等于 X 和 Y 的各自期望值的乘积。

4. 如果 X 是以 $f(x)$ 为其概率密度函数的一个随机变量，而 $g(X)$ 是 X 的任一函数，则：

$$\begin{aligned} E[g(X)] &= \sum_x g(X)f(x) \qquad &\text{如果 } X \text{ 是离散的} \\ &= \int_{-\infty}^{\infty} g(X)f(x)\mathrm{d}x \qquad &\text{如果 } X \text{ 是连续的} \end{aligned}$$

880 例如，若 $g(X)=X^2$ 则：

$$E(X^2) = \sum_x x^2 f(X) \qquad \text{如果 } X \text{ 是离散的}$$
$$= \int_{-\infty}^{\infty} X^2 f(X)\,dx \qquad \text{如果 } X \text{ 是连续的}$$

例 13.

考虑如下 PDF:

x	-2	1	2
$f(x)$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{2}{8}$

于是:

$$E(X) = -2\left(\frac{5}{8}\right) + 1\left(\frac{1}{8}\right) + 2\left(\frac{2}{8}\right)$$
$$= -\frac{5}{8}$$

及

$$E(X^2) = 4\left(\frac{5}{8}\right) + 1\left(\frac{1}{8}\right) + 4\left(\frac{2}{8}\right)$$
$$= \frac{29}{8}$$

方　　差

令 X 为一随机变量并令 $E(X) = \mu$，围绕期望值的 X 值的分布或散布可由方差来度量．方差的定义为：

$$\text{var}(X) = \sigma_X^2 = E(X-\mu)^2$$

σ_X^2 的正的平方根 σ_X 被定义为 X 的**标准差**（standard deviation），方差或标准差标志着个别的 X 值围绕它们的均值散布的远近。

上面定义的方差可计算如下：

$$\text{var}(X) = \sum_x (X-\mu)^2 f(x) \qquad \text{如果 } X \text{ 是一离散 rv}$$
$$= \int_{-\infty}^{\infty} (X-\mu)^2 f(x)\,dx \qquad \text{如果 } X \text{ 是一连续 rv}$$

为了计算上的方便，上面给出的方差公式还可表达成：

$$\begin{aligned}\text{var}(X) = \sigma_x^2 &= E(X-\mu)^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2\end{aligned}$$

利用这一公式，可以看到例 13 所给随机变量的方差是

$$\frac{29}{8} - \left(-\frac{5}{8}\right)^2 = \frac{207}{64} = 3.23$$

881 **例 14.**

让我们求例 3 所给随机变量的方差：

$$\mathrm{var}(X)=E(X^2)-[E(X)]^2$$

现在

$$\begin{aligned}E(X^2) &= \int_0^3 x^2\left(\frac{x^2}{9}\right)\mathrm{d}x\\ &= \int_0^3 \frac{x^4}{9}\mathrm{d}x = \frac{1}{9}\left[\frac{x^5}{5}\right]_0^3\\ &= 243/45\\ &= 27/5\end{aligned}$$

由于 $E(X)=\frac{9}{4}$（见例 12），我们最后得到：

$$\begin{aligned}\mathrm{var}(X) &= 243/45-\left(\frac{9}{4}\right)^2\\ &= 243/720 = 0.34\end{aligned}$$

方差的性质

1. 如上面所提到的，$E(X-\mu)^2=E(X^2)-\mu^2$。
2. 一个常数的方差是零。
3. 如果 a 和 b 是常数，则：

$$\mathrm{var}(aX+b)=a^2\mathrm{var}(X)$$

4. 如果 X 和 Y 是独立随机变量，则：

$$\mathrm{var}(X+Y)=\mathrm{var}(X)+\mathrm{var}(Y)$$
$$\mathrm{var}(X-Y)=\mathrm{var}(X)+\mathrm{var}(Y)$$

这可推广到多于两个变量的情形。

5. 如果 X 和 Y 是独立 rv，而 a 和 b 是常数，则：

$$\mathrm{var}(aX+bY)=a^2\mathrm{var}(X)+b^2\mathrm{var}(Y)$$

协 方 差

令 X 和 Y 为两个 rv，其均值分别为 μ_x 和 μ_y。于是该两变量的**协方差**定义为：

$$\mathrm{cov}(X,Y)=E\{(X-\mu_x)(Y-\mu_y)\}=E(XY)-\mu_x\mu_y$$

易见，一个变量的方差就是这个变量和它自身的协方差。

882 协方差可计算如下：如果 X 和 Y 是离散随机变量，则：

$$\begin{aligned}\operatorname{cov}(X,Y) &= \sum_y\sum_x (X-\mu_x)(Y-\mu_y)f(x,y)\\ &= \sum_y\sum_x XYf(x,y)-\mu_x\mu_y\end{aligned}$$

如果 X 和 Y 是连续随机变量，则：

$$\begin{aligned}\operatorname{cov}(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_x)(Y-\mu_y)f(x,y)\mathrm{d}x\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)\mathrm{d}x\,\mathrm{d}y-\mu_x\mu_y\end{aligned}$$

协方差的性质

1. 如果 X 和 Y 是独立的，则它们的协方差是零。因为，

$$\begin{aligned}\operatorname{cov}(X,Y) &= E(XY)-\mu_x\mu_y\\ &= \mu_x\mu_y-\mu_x\mu_y \qquad \text{由于 } E(XY)=E(X)E(Y)=\mu_x\mu_y\\ &= 0 \qquad \text{如果 } X \text{ 和 } Y \text{ 独立}\end{aligned}$$

2. $\operatorname{cov}(a+bX, c+\mathrm{d}Y)=bd\operatorname{cov}(X,Y)$

其中 a、b、c 和 d 是常数。

例 15.

例 4 中给出了离散随机变量 X 和 Y 的一个联合 PDF，让我们来求 X 和 Y 的协方差，由例 11 我们已知道，$\mu_x=E(X)=1.03$ 以及 $\mu_y=E(Y)=4.47$。

$$\begin{aligned}E(XY) &= \sum_y\sum_x XYf(x,y)\\ &= (-2)(3)(0.27)+(0)(3)(0.08)+(2)(3)(0.16)+(3)(3)(0)\\ &\quad +(-2)(6)(0)+(0)(6)(0.04)+(2)(6)(0.10)+(3)(6)(0.35)\\ &= 6.84\end{aligned}$$

因此，

$$\begin{aligned}\operatorname{cov}(X,Y) &= E(XY)-\mu_x\mu_y\\ &= 6.84-(1.03)(4.47)\\ &= 2.24\end{aligned}$$

相关系数

883 （总体）相关系数 ρ 的定义是：

$$\rho=\frac{\operatorname{cov}(X,Y)}{\sqrt{\{\operatorname{var}(X)\operatorname{var}(Y)\}}}=\frac{\operatorname{cov}(X,Y)}{\sigma_x\sigma_y}$$

如此定义的 ρ，将是两变量之间的线性关联（linear association）的一个度量，它落在 -1 与 $+1$ 之间，-1 表示完全负关联，而 $+1$ 表示完全正关联。

由上述公式可见：

$$\operatorname{cov}(X,Y)=\rho\sigma_x\sigma_y$$

例 16.

对例 4 中的数据估计相关系数。

从例 11 给出的 PDF 容易算出 $\sigma_x=2.05$ 和 $\sigma_y=1.50$。我们曾经得出 cov（X，Y）$=2.24$。因此，应用上述公式，我们估计 ρ 为 $2.24/(2.05)(1.50)=0.73$。

相关变量的方差　令 X 和 Y 为两随机变量，于是有：

$$\begin{aligned}\operatorname{var}(X+Y)&=\operatorname{var}(X)+\operatorname{var}(Y)+2\operatorname{cov}(X,Y)\\&=\operatorname{var}(X)+\operatorname{var}(Y)+2\rho\sigma_x\sigma_y\\\operatorname{var}(X-Y)&=\operatorname{var}(X)+\operatorname{var}(Y)-2\operatorname{cov}(X,Y)\\&=\operatorname{var}(X)+\operatorname{var}(Y)-2\rho\sigma_x\sigma_y\end{aligned}$$

然而，如果 X 和 Y 独立，则 cov(X，Y）为零。这时，如前所示，var($X+Y$）和 var($X-Y$）两者都等于 var(X)+var(Y)。

上述结果可推广如下，令 $\sum_{i=1}^{n}X_i=X_1+X_2+\cdots+X_n$，则线性组合 $\sum_i X_i$ 的方差是：

$$\begin{aligned}\operatorname{var}\Big(\sum_{i=1}^{n}x_i\Big)&=\sum_{i=1}^{n}\operatorname{var}X_i+2\sum_{i<j}\sum\operatorname{cov}(X_i,X_j)\\&=\sum_{i=1}^{n}\operatorname{var}X_i+2\sum_{i<j}\sum\rho_{ij}\sigma_i\sigma_j\end{aligned}$$

其中 ρ_{ij} 是 X_i 和 X_j 的相关系数，而 σ_i 和 σ_j 是 X_i 和 X_j 的标准差。

884 于是，

$$\begin{aligned}\operatorname{var}(X_1+X_2+X_3)&=\operatorname{var}X_1+\operatorname{var}X_2+\operatorname{var}X_3+2\operatorname{cov}(X_1,X_2)\\&\quad+2\operatorname{cov}(X_1,X_3)+2\operatorname{cov}(X_2,X_3)\\&=\operatorname{var}X_1+\operatorname{var}X_2+\operatorname{var}X_3+2\rho_{12}\sigma_1\sigma_2\\&\quad+2\rho_{13}\sigma_1\sigma_3+2\rho_{23}\sigma_2\sigma_3\end{aligned}$$

其中 σ_1、σ_2 和 σ_3 分别是 X_1、X_2 和 X_3 的标准差，而 ρ_{12} 是 X_1 和 X_2 之间的相关系数，ρ_{13} 是 X_1 和 X_3 之间的相关系数，ρ_{23} 是 X_2 和 X_3 之间的相关系数。

条件期望与条件方差

令 $f(x,y)$ 为随机变量 X 和 Y 的联合 PDF。那么，给定 $Y=y$，X 的条件期望（值）的定义为：

$$
\begin{aligned}
E(X \mid Y=y) &= \sum_x x f(x \mid Y=y) && \text{如果 } X \text{ 是离散的} \\
&= \int_{-\infty}^{\infty} x f(x \mid Y=y)\,\mathrm{d}x && \text{如果 } X \text{ 是连续的}
\end{aligned}
$$

其中 $E(X \mid Y=y)$ 表示给定 $Y=y$ 的 X 的条件期望，而 $f(X \mid Y=y)$ 为 X 的条件 PDF。类似地定义 Y 的条件期望 $E(Y \mid X=x)$。

条件期望 注意，$E(X \mid Y)$ 是条件变量 Y 的一个函数，所以是一个随机变量。然而，在 $E(X \mid Y=y)$ 中的 y 是 Y 的一个特定值，所以 $E(X \mid Y=y)$ 是一常数。

条件方差 给定 $Y=y$ 的 X 的条件方差定义为：

$$
\begin{aligned}
\operatorname{var}(X \mid Y=y) &= E\{[X-E(X \mid Y=y)]^2 \mid Y=y\} \\
&= \sum_x [X-E(X \mid Y=y)]^2 f(x \mid Y=y) && \text{如果 } X \text{ 是离散的} \\
&= \int_{-\infty}^{\infty} [X-E(X \mid Y=y)]^2 f(x \mid Y=y)\mathrm{d}x && \text{如果 } X \text{ 是连续的}
\end{aligned}
$$

885 **例 17.**

对例 4 的数据计算 $E(Y \mid X=2)$ 和 $\operatorname{var}(Y \mid X=2)$

$$
\begin{aligned}
E(Y \mid X=2) &= \sum_y y f(Y=y \mid X=2) \\
&= 3f(Y=3 \mid X=2) + 6f(Y=6 \mid X=2) \\
&= 3(0.16/0.26) + 6(0.10/0.26) \\
&= 4.15
\end{aligned}
$$

注：$f(Y=3 \mid X=2) = f(Y=3, X=2)/f(X=2) = 0.16/0.26$，而 $f(Y=6 \mid X=2) = f(Y=6, X=2)/f(X=2) = 0.10/0.26$，所以：

$$
\begin{aligned}
\operatorname{var}(Y \mid X=2) &= \sum_Y [Y-E(Y \mid X=2)]^2 f(Y \mid X=2) \\
&= (3-4.15)^2(0.16/0.26) + (6-4.15)^2(0.10/0.26) \\
&= 2.13
\end{aligned}
$$

条件期望和条件方差的性质

1. 若 $f(X)$ 是 X 的函数，则 $E[f(X) \mid X] = f(X)$，即在计算以 X 为条件的 $f(X)$ 的期望时，$f(X)$ 就像一个常数一样。因此 $[E(X^3 \mid X)] = E(X^3)$；因为，若知道了 X，则 X^3 也就知道了。

2. 若 $f(X)$ 和 $g(X)$ 是 X 的函数，则

$$E[f(X)Y+g(X)|X]=f(X)E(Y|X)+g(X)$$

比如，$E(XY+cX^2|X)=XE(Y|X)+cX^2$，其中 c 为常数。

3. 若 X 和 Y 独立，则 $E(Y|X)=E(Y)$。也就是说，若 X 和 Y 为独立随机变量，则给定 X 下 Y 的条件期望等同于 Y 的无条件期望。

4. **迭代期望法则**（the law of iterated expectations）。一个随机变量 Y 的无条件期望 $E(Y)$ 与其基于另一个随机变量 X 的条件期望 $E(Y|X)$ 之间的关系是：

$$E(Y)=E_X[E(Y|X)]$$

注意到这种关系很有意思，这就是迭代期望法则，它在这里说明，Y 的边缘或无条件分布等于其条件期望的期望，符号 E_X 表示对 X 的值求期望。简言之，这一法则说明，如果我们首先得到作为 X 的函数的 $E(Y|X)$，然后对 X 值的分布求期望，那就最终得到 Y 的无条件期望 $E(Y)$。读者可能用例 4 中给出的数据来验证这一关系。

886 5. 若 X 和 Y 独立，则 $\mathrm{var}(Y|X)=\mathrm{var}(Y)$。

6. $\mathrm{var}(Y)=E[\mathrm{var}(Y|X)]+\mathrm{var}[E(Y|X)]$；即 Y 的（无条件）方差等于 Y 的条件方差的期望与 Y 的条件期望的方差之和。

概率分布的高阶矩

虽然均值、方差和协方差是最常用的单元和多元 PDF 的摘要度量，但有时我们仍需要考虑 PDF 的更高阶矩，诸如第 3 和第 4 阶矩。围绕着均值（μ）的单元 PDF $f(x)$ 的第 3 和第 4 阶矩定义为：

第 3 阶矩：$E(X-\mu)^3$

第 4 阶矩：$E(X-\mu)^4$

一般，围绕均值的第 r 阶矩定义为：

第 r 阶矩：$E(X-\mu)^r$

一个分布的 3 阶和 4 阶矩常用来研究一个概率分布的"形状"，特别是它的**偏态**（skewness）S（即不对称性）和**峰态**（kurtosis）K（指高尖或平扁），如图 A.3 所示。

偏态的一个度量定义为：

$$S=\frac{E(X-\mu)^3}{\sigma^3}=\frac{\text{围绕均值的 3 阶矩}}{\text{标准差立方}}$$

注：围绕均值的 2 阶矩就是方差。

常用的一个峰态度量是：

$$K=\frac{E(X-\mu)^4}{[E(X-\mu)^2]^2}$$

$$=\frac{\text{围绕均值的 4 阶矩}}{\text{2 阶矩的平方}}$$

K 值小于 3 的 PDF 叫做**扁峰态**（platykurtic），有肥而短的尾部；K 值大于 3 的 PDF 则称**尖峰态**（leptokurtic），有细而长的尾部。峰态值为 3 的 PDF 称为常峰态（mesokurtic）。正态分布是常峰态的典型例子，见图 A.3。（参看 A.6 节中关于正态分布的讨论。）

我们即将表明如何能并用偏态和峰态两个度量来决定一个随机变量是否遵从正态分布。回想一下，我们的假设检验程序诸如 t 和 F 检验，是基于这样一个假定（至少对小样本或有限样本是如此），即所依据的变量（或样本统计量）的分布是正态的，因此，在具体应用中明确这一假定是否成立，就是非常重要的。

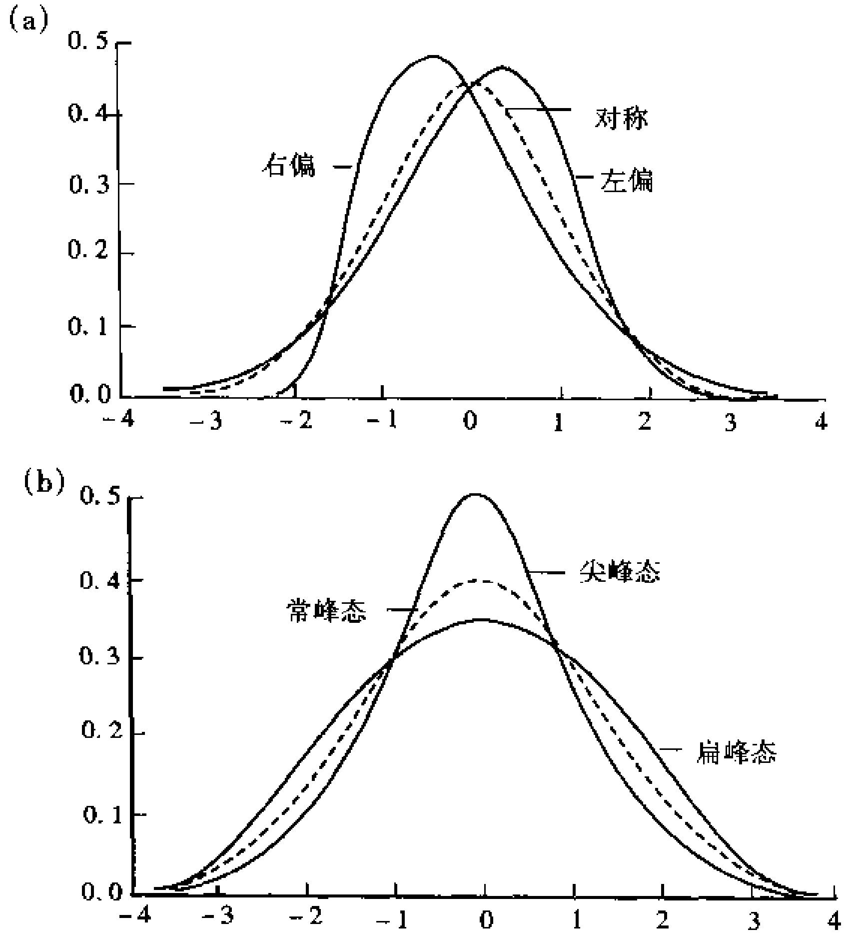


图 A.3　(a) 偏态　　(b) 峰态

§A.6 若干重要的理论概率分布

本书中广泛地利用了如下的概率分布。

正态分布

最著名的理论概率分布莫过于正态分布，其钟形图像已为稍具统计学知识的人所熟悉。

888 一个（连续）随机变量是正态分布的，如果它的 PDF 有如下形式：

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)\qquad -\infty<x<\infty$$

其中 μ 和 σ^2，称为分布的参数，分别是分布均值和方差。此分布有以下性质：

1. 它围绕着它的均值对称分布。

2. 正态曲线下的面积约有 68% 位于 $\mu\pm\sigma$ 两值之间；约有 95% 的面积位于 $\mu\pm2\sigma$ 之间；而约有 99.7% 的面积位于 $\mu\pm3\sigma$ 之间，如图 A.4 所示。

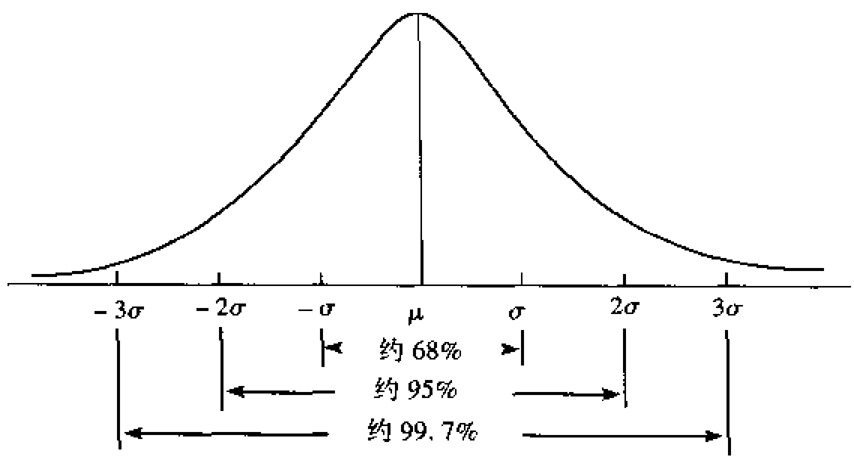


图 A.4 正态曲线下的面积

3. 正态分布依赖于 μ 和 σ^2 两个参数，一旦给定了这两个参数值，就可利用正态分布的 PDF 找出 X 将落入某一区间的概率。但这一工作任务可因附录 D 表 D.1 的引用而大为减轻。为了使用此表，我们通过下列变换，把给定均值 μ 和方差 σ^2 的正态分布变量 X 转换成**标准（化）正态变量** Z：

$$Z = \frac{x - \mu}{\sigma}$$

任何标准化变量都有均值为零，方差为 1 的一个重要性质。例如 Z 有零均值和单位方差，将 Z 代入前面的正态 PDF，我们得到：

$$f(Z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Z^2\right)$$

889 这就是标准正态变量的 PDF。附录 D 表 D.1 所给的概率是根据这个标准化正态变量计算的。

按照惯例，我们把一个正态分布的变量表示为：

$$X \sim N(\mu, \sigma^2)$$

其中～表示“分布”，如 N 代表正态分布，而括号中的量为正态分布的两个参数，即均值与方差，按此惯例

$$X \sim N(0, 1)$$

意谓 X 是一个有零均值和单位方差的正态分布变量。换言之，它是一个标准化正态变量 Z。

例 18.

假定 $X \sim N(8, 4)$，问 X 取的值将落在 $X_1 = 4$ 和 $X_2 = 12$ 之间的概率是什么？为了计算所求的概率，我们把 Z 值计算为：

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{4-8}{2} = -2$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12-8}{2} = +2$$

现在从表 D.1 中我们查出 $\Pr(0 \leqslant Z \leqslant 2) = 0.477\,2$。于是，根据对称性我们有 $\Pr(-2 \leqslant Z \leqslant 0) = 0.477\,2$。因此，所求概率是 $0.477\,2 + 0.477\,2 = 0.954\,4$。（参看图 A.4。）

例 19.

问在上例中 X 超过 12 的概率是什么？X 超过 12 就是 Z 超过 2，其概率由表 D.1 显见是 $(0.5 - 0.477\,2)$ 或 $0.022\,8$。

4. 令 $X_1 \sim N(\mu_1, \sigma_1^2)$ 和 $X_2 \sim N(\mu_2, \sigma_2^2)$，并假定它们是独立的。现考虑线性组合：

$$Y = aX_1 + bX_2$$

其中 a 和 b 是常数。于是可以证明：

$$Y \sim N[(a\mu_1 + b\mu_2), (a^2\sigma_1^2 + b^2\sigma_2^2)]$$

这个结果是说，正态分布变量的线性组合本身是正态分布的。这一结果容易推广到多于两个正态分布变量的线性组合上。

890 5. **中心极限定理**。令 $X_1, X_2, \cdots, X_n$ 为 n 个独立的、有均值 $= \mu$ 和

方差 $=\sigma^2$ 的相同 PDF 的随机变量。令 $\bar{X}=\sum X_i/n$（即样本均值），那么随着 n 无限地增大（即 $n\to\infty$），

$$\bar{X}_{n\to\infty}\sim N\left(\mu,\ \frac{\sigma^2}{n}\right)$$

就是说，$\bar{X}$ 趋于均值为 μ、方差为 σ^2/n 的正态分布。注意，这一结果的成立与 PDF 的形式无关。从而推知：

$$z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{\sqrt{n}(\bar{X}-u)}{\sigma}\sim N(0,1)$$

也就是，Z 是一标准化正态变量。

6. 围绕均值而正态分布的 3 阶和 4 阶矩是：

第 3 阶矩：$E(X-\mu)^3=0$

第 4 阶矩：$E(X-\mu)^4=3\sigma^4$

注：所有围绕均值的正态分布变量的奇数阶矩都等于零。

7. 于是，按照前面讨论的偏态和峰态的度量，对于一个正态 PDF，偏态 $=0$ 而峰态 $=3$，就是说，正态分布是对称的和常峰态的。因此，正态性的一个简单的检验是判明偏态和峰态的计算值是否离开了标准的 0 和 3，事实上这就是本书讨论的**雅克-贝拉（JB）正态性检验**的逻辑基础：

$$\mathrm{JB}=n\left[\frac{S^2}{6}+\frac{(K-3)^2}{24}\right] \tag{5.12.1}$$

其中 S 代表偏态而 K 代表峰态。在正态性的虚拟假设，JB 遵从两个自由度的 χ^2 统计量的分布。

8. 一个正态分布变量的均值和方差相互独立，因为它们都不是对方的函数。

χ^2 分布

令 Z_1，Z_2，…，Z_k 为**独立**标准化正态变量（即零均值、单位方差的正态变量），则量：

$$Z=\sum_{i=1}^{k}Z_i^2$$

891 遵从 k 个自由度（df）的 χ^2 分布，这里 df 一词指上述总和中独立的量的个数。一个 χ^2 分布变量用 χ_k^2 来表示，其中下标 k 指 df，其几何图形见于图 A.5。

χ^2 分布有如下性质：

1. 如图 A.5 所示，χ^2 分布是一有偏倚的分布，其偏倚程度与 df 有关。当 df 较小时，该分布高度向右偏倚；但随着 df 的个数增大，分布变得越来

越对称。事实上，当 df 超过 100 时，变量：

$$\sqrt{2\chi^2} - \sqrt{(2k-1)}$$

可视同标准化正态变量，其中 k 是 df。

2. χ^2 分布的均值为 k，方差为 $2k$，其中 k 是 df。

3. 如果 Z_1 和 Z_2 是两个有 k_1 和 k_2 个自由度的独立 χ^2 变量，则 Z_1+Z_2 也是 χ^2 变量，其自由度为 $df=k_1+k_2$。

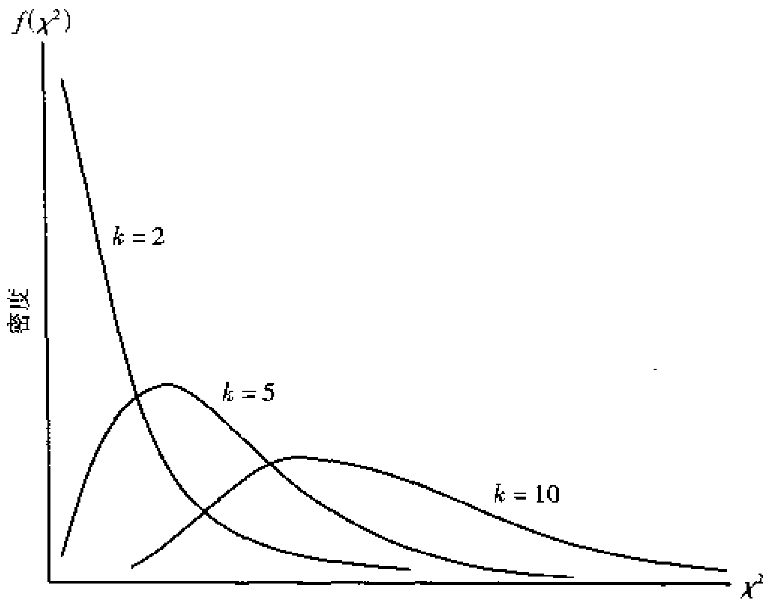


图 A.5　χ^2 变量的密度函数

例 20.

给定自由度为 20，问获得 40 或更大的一个 χ^2 值的概率为何？

查表 D.4，获得一个（大于）39.996 8 的 χ^2 值的概率（20 个自由度）是 0.005。因此，获得一个（大于）40 的 χ^2 值的概率少于 0.005，这是一个相当小的概率。

“学生” t 分布

892 如果 Z_1 是一标准化正态变量［即 $Z_1 \sim N(0, 1)$］而另一变量 Z_2 遵从自由度为 k 的 χ^2 分布且独立于 Z_1，则如下定义的变量：

$$t = \frac{Z_1}{\sqrt{(Z_2/k)}} = \frac{Z_1\sqrt{k}}{\sqrt{Z_2}}$$

服从自由度为 k 的“学生” t 分布，一个 t 分布变量常记为 t_k，以便用下标来指明自由度。t 分布的几何形状如图 A.6 所示。

“学生” t 分布有如下性质：

1. 如图 A.6 所示，t 分布像正态分布那样是对称的，但比正态分布要扁平些，然而随着 df 增加，t 分布迫近于正态分布。

2. t 分布的均值为零，方差为 $k/(k-2)$。t 分布已列表显示，见表 D.2。

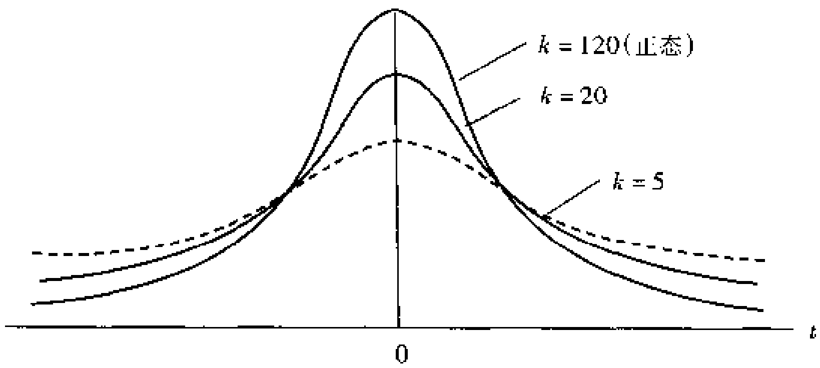


图 A.6 对选定自由度的“学生” t 分布

例 21.

给定 df = 13，求以下概率：(a) 获得约为 3 或更大的一个 t 值，(b) 获得约为 −3 或更小的一个 t 值，以及 (c) 获得约为 3 或更大的 $|t|$，这里 $|t|$ 表示 t 的绝对值（即不计符号的 t 值）。

从表 D.2 得到答案是 (a) 约为 0.005，(b) 约为 0.005，因为 t 的分布是对称的，(c) 约 0.01 = 2(0.005)。

F 分布

893 如果 Z_1 和 Z_2 是自由度分别为 k_1 和 k_2 的独立分布 χ^2 变量，则变量：

$$F=\frac{Z_1/k_1}{Z_2/k_2}$$

遵从（费希尔的）自由度为 k_1 和 k_2 的 F 分布，一个 F 分布的变量，记为 F_{k_1,k_2}，其中下标用以表明与这两个 Z 变量相应的自由度，k_1 被称为**分子**自由度，而 k_2 称为**分母**自由度，F 分布的几何形状见于图 A.7。

1. 像 χ^2 平方分布那样，F 分布向右偏倚。但可以证明，随着 k_1 和 k_2 的变大，F 分布趋向于正态分布。

2. F 分布变量的均值是 $k_2/(k_2-2)$，其定义域是 $k_2>2$，而它的方差是：

$$\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$$

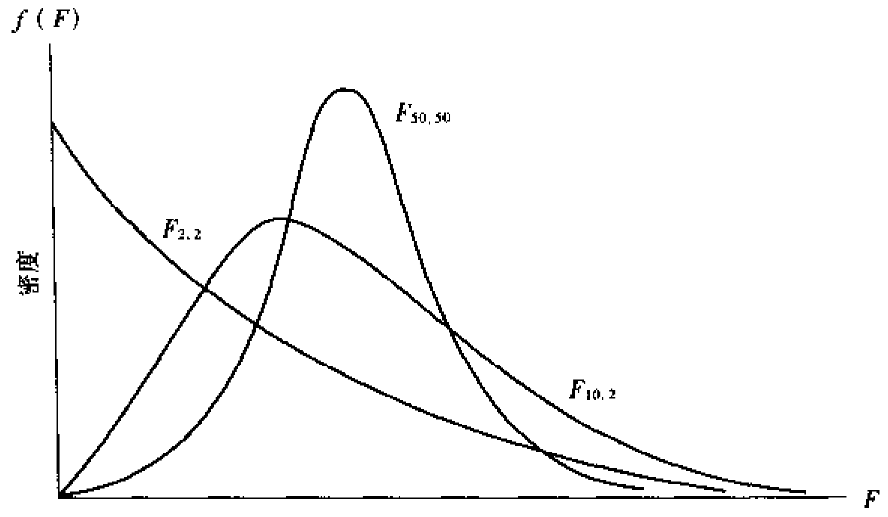


图 A.7 对不同自由度的 F 分布

其定义域是 $k_2>4$。

3. 一个有 k 个自由度的 t 分布随机变量的平方是一个有 1 和 k 个自由度的 F 分布变量，用符号表示：

$$t_k^2=F_{1,k}$$

894 **例 22.**

给定 $k_1=10$ 和 $k_2=8$，分别求获得（a）3.4 或更大以及（b）5.8 或更大的 F 值的概率。

查表 D.3，这些概率约为（a）0.05 和（b）0.01。

4. 如果分母自由度相当大，将成立有如下的 F 和 χ^2 分布之间的关系式：

$$k_1F\sim\chi^2_{k_1}$$

就是说，对于较大的分母自由度，F 值乘以分子自由度差不多相当于一个有分子自由度的 χ^2 值。

例 23.

令 $k_1=20$ 和 $k_2=120$。对这些自由度的 5%临界 F 值是 1.48。因此，$k_1F=(20)(1.48)=29.6$。而由自由度为 20 的 χ^2 分布，5%临界 χ^2 值约为 31.41。

顺便提一提，因为对于大的 df，t、χ^2 和 F 分布都趋于正态分布，所以把这三个分布都称做**正态分布的相关分布**。

贝努里二项式分布

如果一个随机变量 X 的概率密度（或质量）函数（PDF）为：

$$P(X=0)=1-p$$
$$P(X=1)=p$$

其中 $p(0\leqslant p\leqslant 1)$为某一事件“成功”的概率（如掷硬币时得到头像的概率），那么，就称它服从以贝努里（瑞士数学家）命名的分布。对于这样的变量，

$$E(X)=[1\times p(X=1)+0\times p(X=0)]=p$$
$$\operatorname{var}(X)=pq$$

其中 $q=(1-p)$，即“失败”的概率。

二项式分布

895 二项式分布是贝努里分布的推广。令 n 表示独立试验的次数，每次试验结果“成功”的概率都是 p，而“失败”的概率都是$q=(1-p)$。若X表示 n 次试验中成功的次数，则 X 服从二项式分布，其 PDF 为：

$$f(X)=\binom{n}{x}p^{x}(1-p)^{n-x}$$

其中 x 表示 n 次试验中成功的次数，而且，

$$\binom{n}{x}=\frac{n!}{x!\ (n-x)!}$$

这里 $n!$ 读作“n 的阶乘”，意味着 $n(n-1)(n-2)\cdots 1$。二项式分布是有两个参数 n 和 p 的分布，对于这个分布，

$$E(X)=np$$
$$\operatorname{var}(X)=np(1-p)=npq$$

比如你掷 100 次硬币，想求出得到 60 次头像的概率，那你就将 $p=0.5$，$n=100$ 和 $x=60$ 代入上式，有计算机例行为你计算出结果。

你可以看出，二项式分布如何推广了贝努里分布。

泊松分布

如果一个随机变量 X 的 PDF 为：

$$f(X)=\frac{e^{-\lambda}\lambda^{x}}{x!}\qquad x=0,1,2,\cdots\cdots,\lambda>0$$

那么它就服从泊松分布。泊松分布只取决于一个参数 λ。泊松分布的一个明显特征是，其方差等于其期望值 λ。即，

$$E(X)=\operatorname{var}(X)=\lambda$$

如我们在非线性回归模型那一章所见，泊松模型只用于极少见或不经常发生的现象，比如一个时间段（比方说5分钟）内接到电话的次数，或者一小时内收到违章超速驾驶传票的次数，或者一个企业一年内得到的专利个数。

§A.7 统计推断

896 在第A.6节中，我们考虑了若干理论概率分布。往往我们知道或愿意假定一个随机变量 X 遵从某一概率分布，但不知道该分布的参数值。例如，假定 X 遵从正态分布，而想知道它的两个参数值，即均值和方差，为了估计这些未知数，通常的程序是假定我们有一个来自已知概率分布的、大小为 n 的**随机样本**（random sample），并用这些样本数据去估计未知的参数。[5]这就是所谓**估计问题**（problem of estimation）。在本节中我们对这个问题做一较严密的观察。估计问题可划分为两类：点估计和区间估计。

点 估 计

为了建立概念，令 X 为有概率密度 $f(x;\theta)$ 的一个随机变量，其中 θ 是分布的参数（为讨论上简单起见，且假定只有一个未知参数；我们的讨论是容易加以推广的）。假定我们知道了函数形式，即我们知道理论PDF，比如说 t 分布，但不知道 θ 值，于是，我们从这个已知的PDF抽取大小为 n 的一个随机样本，并做出这样的一个样本值的函数：

$$\hat{\theta}=f(x_1,\ x_2,\ \cdots,\ x_n)$$

以提供真 θ 的一个估计值。$\hat{\theta}$ 叫做一个**统计量**（statistic）或一个**估计量**（estimator），而此估计量所取的一个特殊的或具体的数值则叫做一个**估计值**（estimate）或**估计**。注意，因为 $\hat{\theta}$ 是样本数据的一个函数，故可把它看作一个**随机变量**。$\hat{\theta}$ 为我们提供了一个规则或公式，告诉我们怎样去估计真 θ。比如说，如果令：

$$\hat{\theta}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\bar{X}$$

其中 $\bar{X}$ 是样本均值，那么 $\bar{X}$ 就是真均值 μ（比方说）的一个估计量。如果在一具体例子里 $\bar{X}=50$，这就为 μ 提供一个估计（值），以上方式获得的估计量 $\hat{\theta}$ 由于仅提供 θ 的单一（一点）估计值，故称**点估计量**（point estimator）。

区间估计

假如我们不仅仅是获得 θ 的单一估计值，而是通过构造两个估计量 $\hat{\theta}_1(x_1, x_2, \cdots, x_n)$和 $\hat{\theta}_2(x_1, x_2, \cdots, x_n)$而获得 θ 的两个估计值，并且声称在 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 之间的这个区间里包含着真 θ 有一定的可信度（即概率）。可见，与点估计相对照，在区间估计中，我们提供真 θ 将落入其间的一个**可能值域**。

897 区间估计所依据的主要概念是**估计量的抽样**或**概率分布**。例如，可以证明，如果变量 X 是正态分布的，则样本均值 $\bar{X}$ 也是正态分布的，且其均值 = μ（真均值），方差 = σ^2/n，其中 n 是样本大小。换句话说，估计量 $\bar{X}$ 的抽样或概率分布是$\bar{X} \sim N(\mu, \sigma^2/n)$。因此，如果我们构造区间：

$$\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$$

并声称类似这样的许许多多区间将有约 0.95 或 95% 的概率包含着真 μ，那么我们事实上正在构造着 μ 的一个区间估计。注意上面所给的区间依据于从一个样本变到另一个样本的 $\bar{X}$，所以是随机的。

更一般地，在区间估计中，我们构造两个估计量 $\hat{\theta}_1$ 和 $\hat{\theta}_2$，两者都是样本 X 值的函数，使得，

$$\Pr(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2) = 1 - \alpha \qquad 0 < \alpha < 1$$

就是说，我们可以断言，从 $\hat{\theta}_1$ 到 $\hat{\theta}_2$ 的区间里含有真 θ 的概率是 $1-\alpha$。此区间被称为 θ 的置信度为 $1-\alpha$ 的**置信区间**（*confidence interval*），而 $1-\alpha$ 称为**置信系数**（*confidence coefficient*）。例如 $\alpha = 0.05$，则 $1-\alpha = 0.95$，意谓如果我们构造一个置信系数为 0.95 的置信区间，则在（重复）这样的得自重复抽样的构造中，当我们坚持认为所构区间含有真 θ 时，我们将在每 100 例的 95 例中是对的。当置信系数是 0.95 时，我们常说我们有了一个 95% 置信域。一般地，如果置信系数是 $1-\alpha$，就说我们有了一个$100(1-\alpha)$% 置信区间，注意，α 就是我们所知的**显著（性）水平**（level of significance）或犯第Ⅰ类错误的概率。第 A.8 节将讨论问题。

例 24.

假定总体中男子身高是正态分布的，其均值 = μ 英寸且 $\sigma = 2.5$ 英寸。从总体取一个 100 人的随机样本，其平均身高为 67 英寸，求总体平均身高（= μ）的一个 95% 置信区间。

由上述可知，$\bar{X} \sim N(\mu, \sigma^2/n)$。在本例中这将是 $\bar{X} \sim N(\mu, 2.5^2/100)$。查表 D.1 可见：

$$\bar{X} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leqslant \mu \leqslant \bar{X} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

包含正态曲线下的面积的95%。因此，这个区间给出μ的一个95%置信区间。将给定的$\bar{X}$，σ和n值代入，就得到这个95%置信区间为：

$$66.51 \leqslant \mu \leqslant 67.49$$

在重复同样的测量中，这样建立起来的区间将有95%的可信度包含有真均值μ，这里不妨指出一个技术性问题，就是，虽然我们可以说随机区间$[\bar{X} + 1.96(\sigma/\sqrt{n})]$包含$\mu$的概率是95%，却**不可以**说某一具体区间（66.51，67.49）包括μ的概率是95%，一旦这个区间被固定了，它包含μ的概率不是0就是1。我们所能说的，只是如此构造的区间，每100个中将有95个含有真μ；我们不能保证某一区间必定含有μ。

估计方法

898 宽泛地讲，有三种参数估计方法：(1) 最小二乘法（LS），(2) 最大似然（ML）法和（3）矩法（MOM）及其推广——广义矩法（GMM）。我们已经花了相当多的时间来说明LS法。在第4章，我们又在回归的背景下介绍了ML法，但这种方法的应用要广泛得多。

ML背后的关键思想是**似然函数**（likelihood function）。为说明这一点，假设随机变量X的PDF$f(X,\theta)$只取决于一个参数θ。我们知道PDF（比如贝努里或二项式分布），但我们不知道参数值。假设我们得到n个X值的一个随机样本。这n个值的联合PDF为：

$$g(x_1, x_2, \cdots, x_n; \theta)$$

由于它是一个随机样本，所以我们可以把前面的联合PDF写成每个PDF的积：

$$g(x_1, x_2, \cdots, x_n; \theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$$

这个联合PDF具有双重解释。若θ已知，则我们可以把它理解为观测到给定样本值的联合概率。另一方面，我们可以把它看成给定x_1，x_2，…，x_n时θ的一个函数。按后一种解释，我们称联合PDF为**似然函数**（LF）并记作：

$$L(\theta; x_1, x_2, \cdots, x_n) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$$

注意θ在联合概率密度函数和似然函数中的角色转换。

θ的ML估计量是最大化（样本）似然函数L的θ值。为了数学上的方便，我们通常将似然函数取对数，称为**对数似然函数**（$\log L$）。根据最大化的微积分法则，我们将对数似然函数对未知参数微分，并令导数等于零。由此得到估计量的值被称为**最大似然估计量**，还可以用最大化的二阶条件来保证所得到的值确实是最大值。

在不止一个未知参数的情况下，我们将对数似然函数分别对每个未知参

数微分，并令由此得到的表达式等于零，然后联立求解得到未知参数的值。我们已对多元回归模型说明过这一点（见第4章的附录）。

899 **例25.**

假定随机变量 X 服从均值为 λ 的泊松分布。假设 x_1，x_2，…，x_n 都是均值为 λ 的独立泊松随机变量。若我们想求出 λ 的ML估计量，似然函数就是：

$$
\begin{aligned}
L(x_1,x_2,\cdots,x_n;\lambda) &= \frac{e^{-\lambda}\lambda^{x_1}e^{-\lambda}\lambda^{x_2}}{x_1!\ x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\
&= \frac{e^{-n\lambda}\lambda^{\sum x_i}}{x_1!\ x_2!\cdots x_n!}
\end{aligned}
$$

这是一个相当庞大的表达式，但若取对数，则变成：

$$
\log(x_1,x_2,\cdots,x_n;\lambda) = -n\lambda + \sum x_i\log\lambda - \log c
$$

其中 $\log c = \Pi x_i!$，将上述表达式对 λ 微分，我们得到 $[-n + (\sum x_i)/\lambda]$。通过令最后一个表达式为零，我们得到 $\lambda_{ml} = (\sum x_i)/n = \bar{X}$，这就是未知参数 λ 的ML估计量。

矩法。在习题3.4中，在试图以样本矩得到总体矩特征的所谓类比原理的地方，我们已经粗略地了解了一下MOM。作为MOM的推广，GMM目前越来越受到欢迎，但在初级教材中无法详细介绍，这里就不做深究。

理想的统计性质分为两类：小样本或有限样本性质和大样本或渐近性质。在这两组性质的背后，都有估计量具有抽样或概率分布的概念。

小样本性质

无偏性。如果一个估计量 $\hat{\theta}$ 的期望值等于真 θ，即

$$E(\hat{\theta}) = \theta$$

或

$$E(\hat{\theta}) - \theta = 0$$

就说 $\hat{\theta}$ 是 θ 的一个无偏估计量，如果这个等式不成立，则说估计量是有偏误的，且偏误的计算如下：

$$偏误(\hat{\theta}) = E(\hat{\theta}) - \theta$$

当然，如果 $E(\hat{\theta}) = \theta$ 即 $\hat{\theta}$ 是无偏的，则偏误为零。

900 在几何上，这种情形可描述为图A.8，顺便指出，无偏性是一个重复抽样的性质，而不是任一给定样本的性质：固定样本大小，抽取多个样本，每次得到未知参数的一个估计值，如果估计量是无偏的，这些估计值的平均值就可望等于真值。

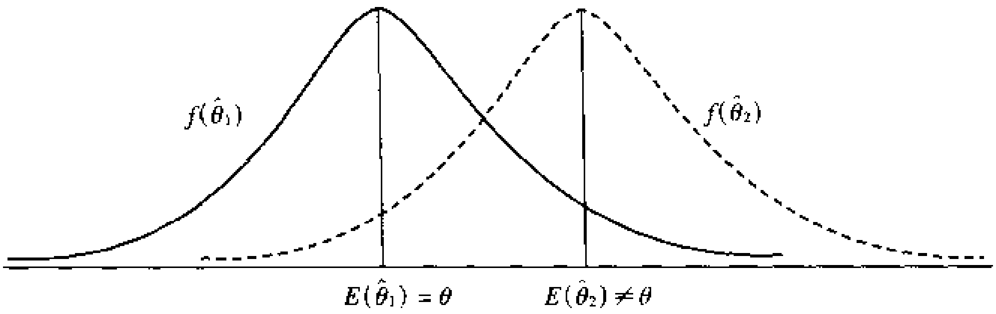


图 A.8　有偏与无偏估计量

最小方差（性）。如果 θ 的估计量 $\hat{\theta}_1$ 的方差比 θ 的任何其他估计量 $\hat{\theta}_2$ 的方差都小或最多相等，则说 $\hat{\theta}_1$ 是 θ 的最小方差估计量。图 A.9 从几何上展现 θ 的三个估计量 $\hat{\theta}_1$、$\hat{\theta}_2$ 和 $\hat{\theta}_3$ 以及它们的概率分布。如图所示，$\hat{\theta}_3$ 的方差既小于 $\hat{\theta}_1$ 也小于 $\hat{\theta}_2$ 的。因而，假定只有三个可能的估计量的话，$\hat{\theta}_3$ 就是最小方差的。但注意 $\hat{\theta}_3$ 却是有偏误的估计量。(为什么?)

最优无偏或有效估计量。如果 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是 θ 的两个无偏估计量，而且 $\hat{\theta}_1$ 的方差小于或最多等于 $\hat{\theta}_2$ 的方差，则 $\hat{\theta}_1$ 是**最小方差无偏**或**最优无偏**或**有效**估计量。这样，对于图 A.9 中的两个无偏估计量 $\hat{\theta}_1$ 和 $\hat{\theta}_2$，$\hat{\theta}_1$ 就是最优无偏的或有效的。

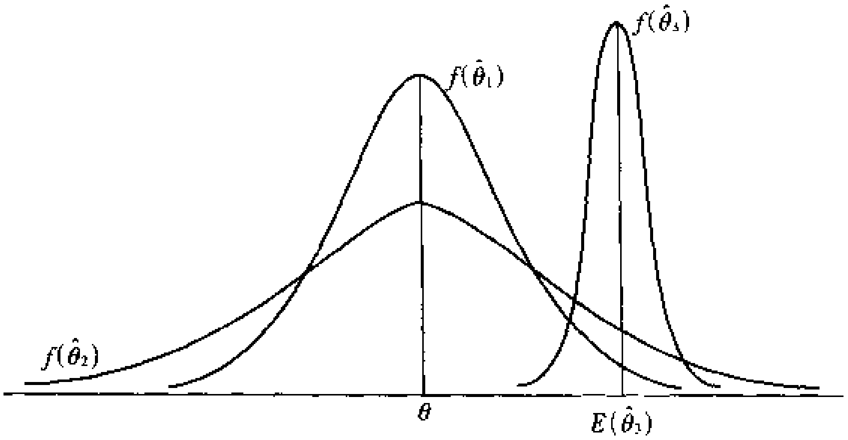


图 A.9　θ 的三种估计量的分布

901

线性性质。如果 θ 的一个估计量 $\hat{\theta}$ 是样本观测值的一个线性函数，就说它是 θ 的一个线性估计量。例如，定义如下的样本均值：

$$\bar{X}=\frac{1}{n}\sum X_i=\frac{1}{n}(X_1+X_2+\cdots+X_n)$$

由于它是诸 X 值的一个线性函数因而是一个线性估计量。

最优线性无偏估计量（BLUE）。如果 θ 的估计量 $\hat{\theta}$ 是线性、无偏的，并

且在 θ 的所有线性无偏估计类中有最小方差，就称它为最优线性无偏估计量，或简记为 BLUE。

最小均方误（MSE）估计量。 一个估计量 $\hat{\theta}$ 的均方误（MSE）定义为：

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

它不同于 $\hat{\theta}$ 的方差，后者的定义是：

$$\text{var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$$

两者的差别在于：$\text{var}(\hat{\theta})$衡量着 $\hat{\theta}$ 围绕其均值或期望值而分布的散度，而 $\text{MSE}(\hat{\theta})$则衡量着 $\hat{\theta}$ 围绕参数的真值而散布的程度，两者的关系如下：

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 + 2E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2 \qquad \text{因最后一项为零}^{[6]} \\ &= \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2 \\ &= \hat{\theta}\ \text{的方差加偏误的平方}\end{aligned}$$

当然，如果偏误为零，则 $\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta})$。

所谓最小 MSE 准则，就是在相互媲美的一个估计量集合中挑选 MSE 最小的一个估计量。但应看到，即使找到了这样的一个估计量，也将涉及得失方面的权衡。须知：为了得到最小方差，不免要承受一些偏误。这种情况可由图 A.10 做几何上的说明。图中 $\hat{\theta}_2$ 稍有偏误，但它的方差小于无偏估计量 $\hat{\theta}_1$ 的方差，然而，在实践中，当最优无偏准则不能给出有较小方差的估计量时，就会用到最小 MSE 准则。

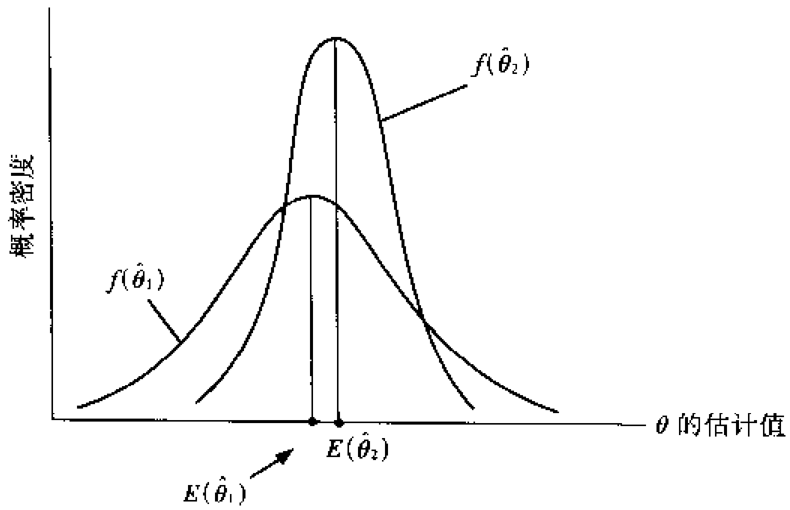


图 A.10　偏误与方差之间的权衡

大样本性质

往往一个估计量不具备小样本中的一或多种优良统计性质，但随着样本无限地增大，该估计量却具有一些称之为**大样本**或**渐近性质**的优良统计性质。

渐近无偏性。我们说估计量 $\hat{\theta}$ 是 θ 的渐近无偏估计量，如果

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

其中 $\hat{\theta}_n$ 表示估计量以样本大小 n 为基础，lim 表示极限，而 $n\to\infty$ 表示 n 无限地增加。也就是说，如果随着样本变得越来越大，$\hat{\theta}_n$ 的期望值或均值趋于真值，则它是 θ 的一个渐近无偏估计量。作为例子，考虑随机变量 X 的样本方差的如下度量：

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

可以证明：

$$E(S^2) = \sigma^2\left(1 - \frac{1}{n}\right)$$

其中 σ^2 是真方差。显见，在小样本中 S^2 是有偏误的，但随着 n 无限地增大，$E(S^2)$ 趋向于真 σ^2；从而它是渐近无偏的。

903 **一致性。**如果随着样本变得越来越大，$\hat{\theta}$ 趋于真值 θ，就说 $\hat{\theta}$ 是一个一致性估计量。图 A.11 展示了这一性质。

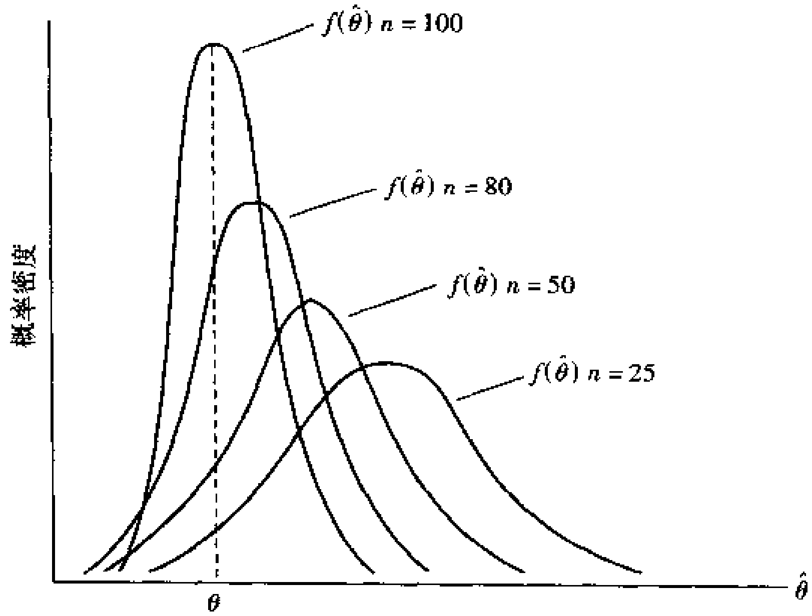


图 A.11　随样本增大而变化的 $\hat{\theta}$ 的分布

图中我们有基于样本大小为 25、50、80 和 100 的 $\hat{\theta}$ 的分布。如图所示，基于 $n=25$ 的 $\hat{\theta}$ 由于它的抽样分布不以真 θ 为中心，故而是偏误的。但随着 n 的增大，不仅 $\hat{\theta}$ 趋于更紧密地围绕 θ 而分布（即 $\hat{\theta}$ 的偏误在减少），而且它的方差也在变小。如果取极限时（即当 n 无限增大时），$\hat{\theta}$ 的分布收缩到单一个点 θ，即如果 $\hat{\theta}$ 的分布散度或方差为零，则我们说 $\hat{\theta}$ 是 θ 的一个**一致性**估计量。

更正式地说，如果 $\hat{\theta}$ 与 θ 之差的绝对值小于一个任意小的正数 δ 的概率趋于 1，估计量 $\hat{\theta}$ 是 θ 的一个一致性估计量。用符号表示，

$$\lim_{n\to\infty} P\{|\hat{\theta}-\theta|<\delta\}=1 \quad \delta>0$$

其中 P 代表概率。这个表达式又常写为：

$$\operatorname*{plim}_{n\to\infty}\hat{\theta}=\theta$$

其中 plim 表示概率极限。

注意无偏性和一致性两个性质在概念上是迥然不同的。无偏性可以对任何样本大小都成立，而一致性则仅仅是一大样本性质。

904 一致性的一个充分条件是随着样本无限地增大，偏误和方差都趋于零。[7] 一致性的另一充分条件是随着 n 无限地增大 MSE($\hat{\theta}$) 趋于零。[关于 MSE($\hat{\theta}$)，参看前面的讨论。]

例 26.

令 $X_1, X_2, \cdots, X_n$ 为来自有均值 μ 和方差 σ^2 的一个分布的随机样本。说明样本均值 $\bar{X}$ 是 μ 的一个一致性估计量。

由初等统计学可知 $E(\bar{X})=\mu$ 和 $\operatorname{var}(\bar{X})=\sigma^2/n$。因 $E(\bar{X})=\mu$ 而无论样本的大小，故 $\bar{X}$ 是无偏的。再则，随着 n 无限增大，var($\bar{X}$) 趋于零。从而 $\bar{X}$ 是 μ 的一个一致性估计量。

概率极限有如下值得注意的规则：

1. 不变性（Slutsky）性质。如果 $\hat{\theta}$ 是 θ 的一个一致性估计量，并且如果 $h(\hat{\theta})$ 是 $\hat{\theta}$ 的任何连续函数，则：

$$\operatorname*{plim}_{n\to\infty} h(\hat{\theta})=h(\theta)$$

据此，如果 $\hat{\theta}$ 是 θ 的一个一致性估计量，则 $1/\hat{\theta}$ 也是 $1/\theta$ 的一个一致性估计量；$\log(\hat{\theta})$ 也是 $\log(\theta)$ 的一个一致性估计量。注意这个性质对期望值运算子 E 来说却不成立；就是说，如果 $\hat{\theta}$ 是 θ 的一个无偏估计量[即 $E(\hat{\theta})=\theta$]，那么说 $1/\hat{\theta}$ 是 $1/\theta$ 的一个无偏估计量并**不真实**；即 $E(1/\hat{\theta})\neq 1/E(\hat{\theta})$ 或 $1/\theta$。

2. 如果 b 是一常数，则：

$$\operatorname*{plim}_{n\to\infty} b=b$$

就是说，一个常数的概率极限就是这个常数。

3. 如果 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是一致性估计量，则：

$$\operatorname{plim}(\hat{\theta}_1+\hat{\theta}_2)=\operatorname{plim}\hat{\theta}_1+\operatorname{plim}\hat{\theta}_2$$

$$\text{plim}(\hat{\theta}_1\hat{\theta}_2) = \text{plim}\hat{\theta}_1\text{plim}\hat{\theta}_2$$
$$\text{plim}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) = \frac{\text{plim}\hat{\theta}_1}{\text{plim}\hat{\theta}_2}$$

最后两个性质一般地说对期望运算子 E 并不成立。例如，$E(\hat{\theta}_1/\hat{\theta}_2) \neq E(\hat{\theta}_1)/E(\hat{\theta}_2)$。类似地，$E(\hat{\theta}_1\hat{\theta}_2) \neq E(\hat{\theta}_1)E(\hat{\theta}_2)$。然而，如果 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 是独立分布的，则如前面所看到的，$E(\hat{\theta}_1\hat{\theta}_2) = E(\hat{\theta}_1)E(\hat{\theta}_2)$。

905 **渐近有效性**。设 $\hat{\theta}$ 为 θ 的一个估计量。$\hat{\theta}$ 的渐近分布的方差叫做 $\hat{\theta}$ 的**渐近方差**（asymptotic variance）。如果 $\hat{\theta}$ 是一致性的并且它的渐近方差小于 θ 的任何其它一致性估计量的渐近方差，则称 $\hat{\theta}$ 为**渐近有效的**（asymptotically efficient）。

渐近正态性。如果一个估计量 $\hat{\theta}$ 的抽样分布随着样本大小的无限增大而趋于正态分布，就说它是渐近正态分布的。例如，统计理论表明，如果 X_1，X_2，…，X_n 是有相同均值 μ 和相同方差 σ^2 的独立正态分布变量，则样本均值 $\bar{X}$ 无论在小样本或大样本中都是以 μ 为均值、σ^2/n 为方差的正态分布变量。但若 X_i 是以 μ 为均值，并以 σ^2 为方差的独立但不一定正态分布的变量，则样本均值 $\bar{X}$ 是以 μ 为均值、σ^2/n 为方差的渐近正态分布变量；就是说，随着样本大小 n 无限地增大，样本均值趋于以 μ 为均值、σ^2/n 为方差的正态分布。事实上，这就是前面讨论过的中心极限定理。

§A.8 统计推断：假设检验

估计与假设检验是经典统计推断的一对孪生分支。既然已分析过估计的问题，现在我们简略地审视一下检验统计假设的问题。

假设检验的问题可叙述如下。假定随机变量 X 有一已知的概率密度函数$f(x;\ \theta)$，其中 θ 是分布的参数，在取得一个大小为 n 的随机样本之后，我们得到点估计量 $\hat{\theta}$，由于真 θ 鲜为人知，我们提问：这个估计量 $\hat{\theta}$ 是否与某个假设的 θ 值"相符"？比方说，$\theta = \theta^*$，这里 θ^* 是一个特定的（假设的）θ 数值，换句话说，我们的样本会来自 PDF$f(x;\ \theta = \theta^*)$ 吗？在假设检验的术语中，$\theta = \theta^*$ 称**虚拟**（或维持）**假设**［null (or maintained) hypothesis］并通常记以 H_0。虚拟假设是相对于一个记之为 H_1 的**对立假设**（alternative hypothesis）而检验的。例如，H_1 可叙述为 $\theta \neq \theta^*$。（注：在某些教科书中把 H_0 和 H_1 分别记为 H_1 和 H_2。）

虚拟假设和对立假设都可以是**简单的**或**复合的**。一个假设被称为**简单**的，如果它确定了分布的（诸）参数的（各）一个值；否则就称它为**复合假设**，例如，如果 $X \sim N(\mu,\ \sigma^2)$，并且我们声称：

$$H_0:\ \mu = 15 \quad 和 \quad \sigma = 2$$

这就是一个简单假设；而

$$H_0: \mu - 15 \quad 和 \quad \sigma > 2$$

则因 σ 值未予确定，则是一个复合假设。

906 为了检验虚拟假设（即检验其真实性），我们利用样本信息以获得所谓的**检验统计量**（test statistic）。这个检验统计量常常就是未知参数的点估计量。然后我们试图找出检验统计量的抽样或概率分布，并利用**置信区间**或**显著性检验**方法去检验虚拟假设。现将其操作步骤说明如下：

为了建立概念，让我们回到例 23 所考虑的一个总体中的男子身高（X），我们被告知：

$$X_i \sim N(\mu, \sigma^2) = N(\mu, 2.5^2)$$
$$\bar{X} = 67 \qquad n = 100$$

现假设

$$H_0: \mu = \mu^* = 69$$
$$H_1: \mu \neq 69$$

问题是：这个检验统计量为 $\bar{X} = 67$ 的样本会来自均值为 69 的总体吗？直觉上，如果 $\bar{X}$ “足够接近” μ^*，我们也许不会拒绝虚拟假设；否则我们宁可拒绝它而接受对立假设。但怎样决定 $\bar{X}$ 是否足够接近“μ^*”呢？可以采取两种方法之一：（1）置信区间法和（2）显著性检验法。在任一具体应用中两种方法都将导致同一结论。

置信区间法

因为 $X_i \sim N(\mu, \sigma^2)$，所以我们知道检验统计量 $\bar{X}$ 的分布是：

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

既然知道了 $\bar{X}$ 的概率分布，为什么不根据 $\bar{X}$ 建立 μ 的一个 $100(1-\alpha)$ 置信区间，然后看此置信区间是否包含 $\mu = \mu^*$ 呢？如果包含，我们就不拒绝虚拟假设；如果不包含，就可拒绝虚拟假设。例如，取 $\alpha = 0.05$，我们将有一个 95% 置信区间。如果此区间包含 μ^*，由于这样建立起来的区间每 100 个中有 95 个会含有 μ^*，我们就不拒绝虚拟假设。

实际操作步骤如下：因 $\bar{X} \sim N(\mu, \sigma^2/n)$，从而

$$Z_i = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

就是说，这是一个标准正态变量，于是由正态分布表知：

$$\Pr(-1.96 \leqslant Z_i \leqslant 1.96) = 0.95$$

907 即

$$\Pr\left(-1.96 \leqslant \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leqslant 1.96\right) = 0.95$$

重新整理得

$$\Pr\left[\bar{X}-1.96\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right]=0.95$$

这就是 μ 的一个 95%置信区间。一旦建立了这个区间，虚拟假设的检验就是简单的。我们所要做的不外是看 $\mu=\mu^*$ 是否落入此区间而已。如果落入，就不拒绝虚拟假设；如果不落入则拒绝之。

回到我们的例子，我们已建立 μ 的一个 95%置信区间，就是

$$66.51\leqslant\mu\leqslant67.49$$

此区间显然不包含 $\mu=69$，因此我们能以 95%置信系数拒绝真 μ 是 69 的虚拟假设。图 A.12 描绘了这一情况的几何意义。

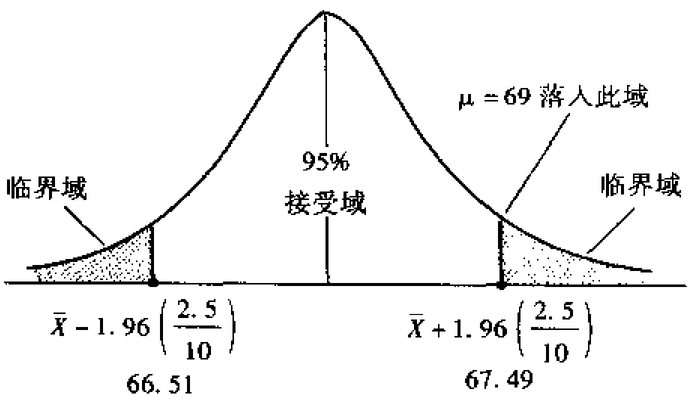


图 A.12　μ 的 95%置信区间

用假设检验的语言说，我们所建立的置信区间叫做**接受域**（acceptance region）。接受域以外的区域叫做虚拟假设的**临界域**［critical region（s）］或**拒绝域**［region(s) of rejection］。接受域的上下限（与拒绝域的分界线）叫做**临界值**（critical values）。那么，用假设检验的语言说，如果假设值落入接受区间，就不可拒绝虚拟假设；否则可以拒绝。

重要的是要看到，在决定拒绝或不拒绝 H_0 时，我们可能犯两类错误：
908 （1）我们也许会拒绝一个事实上是真的 H_0。这叫做（第）**Ⅰ类错误**（type Ⅰ error）。例如，在上例中 $\bar{X}=67$ 有可能来自均值是 69 的总体。或者（2）我们也许没有拒绝一个事实上是不真的 H_0。这叫做（第）**Ⅱ类错误**（type Ⅱ error），因此，假设检验并不确立真 μ 的值，而只提供一种手段，以便决定我们可不可以按照 $\mu=\mu^*$ 行事。

Ⅰ类和Ⅱ类错误。可系统地表示为

	自然状态－	
决策	H_0 是对的	H_0 是错的
拒绝	Ⅰ类错误	无错
不拒绝	无错	Ⅱ类错误

理想上，我们希望Ⅰ类和Ⅱ类错误都能最小化。可是，对任一给定的样本大小，要同时最小化两类错误是不可能的。解决此问题的经典方法已体现于内曼和皮尔逊的工作中，即假定Ⅰ类错误实际上比Ⅱ类错误很可能更严重。因此，人们应把犯Ⅰ类错误的概率定在一个相当低的水平上，比如说0.01或0.05，然后试图使出现Ⅱ类错误的概率达到尽可能小。

文献中把Ⅰ类错误的概率记为 α，并称为**显著性水平**（level of significance），而Ⅱ类错误的概率记为 β，并把不犯Ⅱ类错误的概率 $1-\beta$ 称为**检验的功效**（power of the test）。换句话说，检验的功效就是它拒绝一个错误假设的能力。假设检验的经典方法是把 α 固定在诸如0.01（或1%）或0.05（5%）的水平上，然后试图把检验的功效最大化；也就是使 β 最小化。

重要的是，读者要理解检验功效的概念，最好用一个例子加以理解。[8]。

令 $X \sim N(\mu, 100)$；即 X 服从均值为 μ 和方差为100的正态分布。假定 $\alpha = 0.05$。假设我们有一个含有25次观测的样本，样本均值为 $\bar{X}$。进而假设我们对假设 $H_0: \mu = 50$ 有兴趣。既然 X 服从正态分布，那我们就知道样本均值也是正态分布的：$\bar{X} \sim N(\mu, 100/25)$。因此，在所述 $\mu = 50$ 的虚拟假设下，$\bar{X}$ 的95%的置信区间是 $(\mu \pm 1.96\sqrt{100/25}) = \mu \pm 3.92$，即（46.08，53.92）。于是所有小于46.08或大于53.92的 $\bar{X}$ 就构成了临界区域。也就是说，如果发现样本均值低于46.08或大于53.92，那我们就拒绝真实均值为50的虚拟假设。

909 但如果真实的 μ 不等于50，$\bar{X}$ 落在上述临界区域的概率是多少？假设有三个对立假设：$\mu = 48$，$\mu = 52$，和 $\mu = 56$。如果这些对立假设中有任何一个成立，那它将是 $\bar{X}$ 分布的实际均值。由于仍假定 $\sigma^2 = 100$，所以标准误对这三个对立假设而言都是不变的。

图A.13中的阴影区域表明了每个对立假设为真实 $\bar{X}$ 落入拒绝区域的概率。如你可以验证的那样，这些概率分别是0.17（对 $\mu = 48$），0.05（对 $\mu = 50$），0.17（对 $\mu = 52$）和0.85（对 $\mu = 56$）。从数字可以看出，只要 μ 的真实值与考虑中的假设（这里是 $\mu = 50$）值显著不同，那么拒绝这个假设的概率就很高，但当真实值与虚拟假设的值相差不大，拒绝的概率就很小。从直觉上讲，如果虚拟假设和对立假设紧密地捆在一起，就会出现很难拒绝的情况。

考察被称为**功效函数图**（power function graph）的图A.14，可以进一步看出来这一点，那里所示的曲线被称为**功效曲线**（power curve）。

读者现在必定发觉，前面讨论的置信系数 $(1-\alpha)$ 不外是1减去犯Ⅰ类错误的概率。例如一个95%置信系数是说我们准备接受一个最多是5%的犯Ⅰ类错误的概率——我们不想在100次中有多于5次拒绝一个真的假设。

***p* 值或准确显著性水平。** 除预选某个任定的 α 水平外，还可求出一个检验统计量的 ***p*（概率）值** 或准确的显著性水平。p 值被定义为虚拟假设可被拒绝时所看到的最低显著性水平。

910

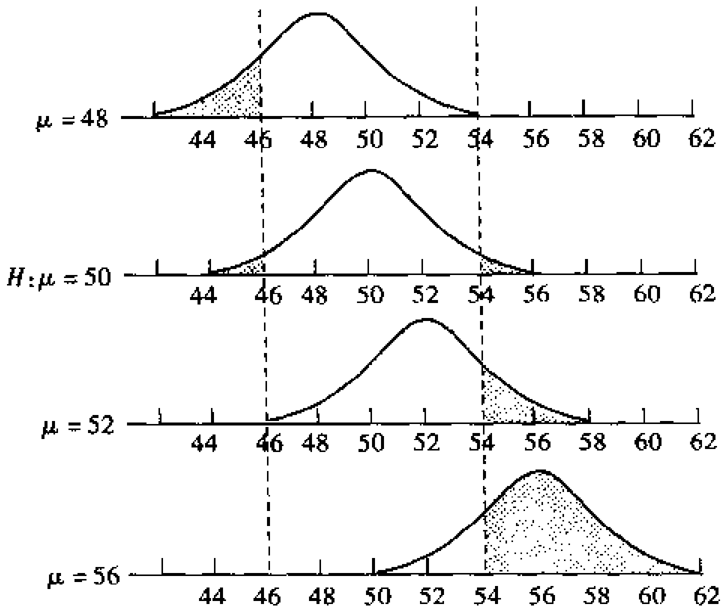


图 A.13　在 $N=10$，$\sigma=25$，和 $\mu=48$，50，52 或 56 时 X 的分布。在 H：$\mu=50$ 下，$\alpha=0.05$ 的拒绝区域是 $\bar{X}<46.1$ 或 $\bar{X}>53.9$。阴影区域表示了 $\bar{X}$ 将落入拒绝区域的概率。这些概率分别是：

0.17（$\mu=48$）　0.17（$\mu=52$）

0.05（$\mu=50$）　0.85（$\mu=56$）

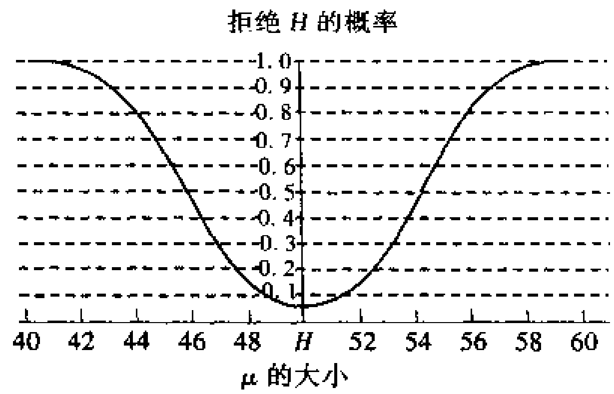


图 A.14　在 $N=25$，$\sigma=10$ 和 $\alpha=0.05$ 时，检验假设 $\mu=50$ 的功效函数

假使在一项应用中我们得到一个自由度为 20 的 t 值为 3.552。从表 D.2 我们能看到，获得一个等于或大于 3.552 的 t 值的 p 值或准确概率是 0.001（单尾）或 0.002（双尾）。我们说所观测的 t 值 3.552 是在 0.001 或 0.002 水平上统计上显著的，究竟 0.001 还是 0.002，视我们使用单尾或双尾检验而定。

现在一些统计软件包例行程序般地把所估计的检验统计量的 p 值打印出来。因此，只要可能，读者最好给出这个 p 值。

显著性检验方法

回顾

$$Z_i=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

在任一给定的应用中，$\bar{X}$ 和 n 是已知的（或可以估计的），而真 μ 和 σ 是未知的。但如果我们规定 σ 并在 H_0 下假定 $\mu=\mu^*$，我们就能直接算出 Z_i，然后查正态分布表以得到所算的 Z 值的概率。如果这是一个小的概率，比方
911 说小于 5% 或 1%，就可拒绝虚拟假设——如果假设真实，那么获得所算的 Z 值的机会就应该很大，这是假设检验的显著性检验法所持的一般逻辑思维。这里重要的概念是检验统计量（即 Z 统计量）以及它在假设值 $\mu=\mu^*$ 下的概率分布。在本例中由于我们使用了 Z 变量，故称此检验为 Z 检验是适宜的。

回到我们的例子，如果 $\mu=\mu^*=69$，则 Z 统计量变为：

$$\begin{aligned}Z&=\frac{\bar{X}-\mu^*}{\sigma/\sqrt{n}}\\&=\frac{67-69}{2.5/\sqrt{100}}\\&=-2/0.25=-8\end{aligned}$$

查一下正态分布表 D.1，便知得到如此一个 Z 值的概率是极其小的。（注：超出 3 或 -3 的 Z 值的概率约为 0.001。因此，超过 8 的 Z 值的概率就更小了。）因此，可拒绝 $\mu=69$ 这个虚拟假设；给定此值而得到 $\bar{X}$ 为 67 的机会是微乎其微的。于是我们怀疑我们的样本是来自均值为 69 的一个总体。图 A.13 对这种情况做了一个图解。

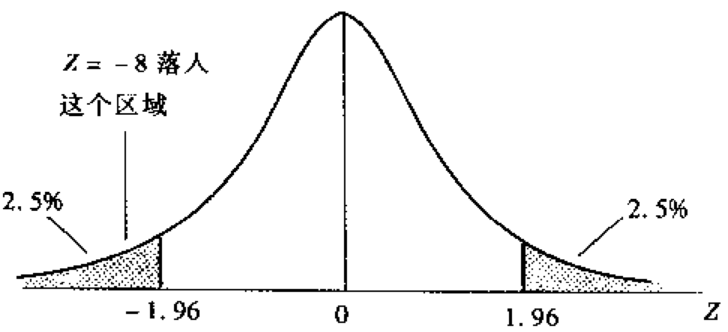


图 A.15 Z 统计量的分布

按照显著性检验的术语，当我们说一个检验（统计量）是显著的，我们通常的意思是我们可以拒绝虚拟假设。而一个检验统计量被认为是显著的，如果得到它的概率等于或小于犯 I 类错误的概率 α。例如，取 $\alpha=0.05$，我

们知道得到一个等于 -1.96 或 1.96 的 Z 值的概率是 5%（或标准正态分布每侧 2.5%）。在我们的说明性例子中 Z 是 -8。由此知，得到这样一个 Z 值
912 的概率要比 2.5% 小得多，大大低于我们预定的犯Ⅰ类错误的概率。这就说明为什么所算的 $Z=-8$ 这个值是统计上显著的；也就是为什么我们拒绝 μ^* 等于 69 这个虚拟假设。当然，我们得到了用假设检验的置信区间法的同样结论。

现在我们把检验统计假设的步骤归纳如下：

步骤 1. 叙述虚拟假设 H_0 和对立假设 H_1（例如，H_0：$\mu=69$ 和 H_1：$\mu\neq69$）。

步骤 2. 选择检验统计量（例如，$\bar{X}$）。

步骤 3. 确定检验统计量的概率分布，例如，$\bar{X}\sim N(\mu,\ \sigma^2/n)$。

步骤 4. 选定显著性水平（即犯Ⅰ类错误的概率）α。

步骤 5. 利用检验统计量的概率分布，建立一个 $100(1-\alpha)\%$ 置信区间。如果虚拟假设下的参数值（例如 $\mu=\mu^*=69$）落入此置信域即接受域，则不要拒绝虚拟假设。但如果它落在此区间之外（即落入拒绝域），就可拒绝虚拟假设。记住，当你拒绝一个虚拟假设时，你正在冒着 $100\%\alpha$ 次的犯错误风险。

参考文献

关于本附录所含内容的细节，读者可参阅以下参考文献：

Hoel Paul G.: *Introduction to Mathematical Statistics*,, 4th ed., John Wiley & Sons, New York, 1974. 本书对数理统计的各个方面提供了一个较简易的导引。

Freund, John E., and Ronald E. Walpole: *Mathematical Statistics*, 3d ed., Prentice Hall, Englewood Cliffs, N. J., 1980. 这是数理统计的另一本入门教材。

Mood, Alexander M., Franklin A. Graybill, and Duane C. Bose: *Introduction to the Theory of Statistics*, 3d ed., McGraw-Hill, New York, 1974. 这是统计学理论的一个全面的介绍，但比前两本书深些。

Newbold, Paul: *Statistics for Business and Economics*, PrenticeHall, Englewood Cliffs, N. J., 1984. 这是一本统计学的全面的非数学导论，含有大量解答问题的内容。

【附录 A 注释】

[1] 通常，一个函数的值域如果是实数的一个子集，就称这个函数为实值函数。关于细节，参看 Alpha C. Chiang, *Fundamental Methods of Mathematical Economics*, 3d ed., McGraw-Hill, 1984, Chap. 2。

[2] 关于可数无穷集的一个简单的讨论，参看 R. G. D. Allen, *Basic Mathematics*, London, 1964, p. 104。

[3] 注：$\int_a^a f(x)\mathrm{d}x = 0$。

[4] $$\int_0^1\left[\int_0^1(2-x-y)\mathrm{d}x\right]\mathrm{d}y = \int_0^1\left[\left(2x-\frac{x^2}{2}-xy\right)\Big|_0^1\right]\mathrm{d}y$$
$$= \int_0^1\left(\frac{3}{2}-y\right)\mathrm{d}y$$
$$= \left(\frac{3}{2}y-\frac{y^2}{2}\right)\Big|_0^1 = 1$$

注：表达式 $(\frac{3}{2}y-y^2/2)\Big|_0^1$ 表示在上限值为 1 和下限值为 0 处估算括号中的表达式；然后用前一估算值减去后一估算值以获得这个积分值。例如，在上例中，在 $y=1$ 处的上限值是 1，而在 $y=0$ 处的下限值是 0，从而给出积分值为 1。

[5] 令 X_1，X_2，…，X_n 为 n 个随机变量，其联合 PDF 为 $f(x_1, x_2, \cdots, x_n)$。如果我们能够写：

$$f(x_1,x_2,\cdots,x_n)=f(x_1)f(x_2)\cdots f(x_n)$$

其中 $f(x)$ 是每个 X 的共同 PDF，则说 x_1，x_2，…，x_n 构成一个以PDF $f(x_n)$为其总体的大小为 n 的随机样本。

[6] 最后一项可写为$2\{[E(\hat\theta)]^2-[E(\hat\theta)]^2-\theta E(\hat\theta)+\theta E(\hat\theta)\}=0$。还得注意，由于一个常数的期望值就是该常数本身，故$E[E(\hat\theta)-\theta]^2=[E(\hat\theta)-\theta]^2$。

[7] 更专门地，$\lim_{n\to\infty} E(\hat\theta_n)=\theta$ 和 $\lim_{n\to\infty}\mathrm{var}(\hat\theta_n)=0$。

[8] 如下讨论和数字都源自于 Helen M. Walker and Joseph lev，*Statistical Inference*，Holt，Rinehart an Winston，New York，1953，pp.161－162。

附录 B　矩阵代数初步

913　本附录提供为读懂附录 C 以及第 18 章部分内容所需的矩阵代数的基本知识。讨论是非严格的，而且不加任何证明。关于证明和更多的细节，读者可阅读参考文献。

§B.1　定义

矩　阵

矩阵乃是把一些数或元素排成行和列的一个长方形阵列。说得准确些，一个**阶数**（order）或**维数**（dimension）为 M 乘 N（写为 $M\times N$）的矩阵是指排成 M 行和 N 列的一个 $M\times N$ 元素集。例如，用粗体字母表示矩阵，一个（$M\times N$）矩阵 **A** 可表达为：

$$\mathbf{A}=[a_{ij}]=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{M1} & a_{M2} & a_{M3} & \cdots & a_{MN} \end{bmatrix}$$

其中 a_{ij} 是出现在 **A** 的第 i 行和第 j 列的元素，而 $[a_{ij}]$ 是以 a_{ij} 为其典型元素的矩阵 **A** 的缩写表达式。一个矩阵的阶或维，也就是它的行数和列数，常常写在该矩阵的下方以便于核对。

$$\underset{2\times 3}{\mathbf{A}}=\begin{bmatrix} 2 & 3 & 5 \\ 6 & 1 & 3 \end{bmatrix} \quad \underset{3\times 3}{\mathbf{B}}=\begin{bmatrix} 1 & 5 & 7 \\ -1 & 0 & 4 \\ 8 & 9 & 11 \end{bmatrix}$$

纯量。一个纯量（scalar）是指单一个（实）数。另一种说法是，一个纯量是一个 1×1 矩阵。

列 向 量

914 由 M 行和仅仅 1 列组成的矩阵叫做**列向量**（column vector）。用粗体小写字母表示向量，列向量的一个例子是：

$$\underset{4\times 1}{\mathbf{x}}=\begin{bmatrix} 3 \\ 4 \\ 5 \\ 9 \end{bmatrix}$$

行 向 量

由仅仅 1 行和 N 列组成的矩阵叫做**行向量**（row vector）。

$$\underset{1\times 4}{\mathbf{x}}=[1 \quad 2 \quad 5 \quad -4] \qquad \underset{1\times 5}{\mathbf{y}}=[0 \quad 5 \quad -9 \quad 6 \quad 10]$$

转 置

一个 $M\times N$ 矩阵 **A** 的转置（transpose），记为 **A**′（读 A 一撇或 A 转置），是将 **A** 的行和列交换后得到的一个 $N\times M$ 矩阵；也就是 **A** 的第 i 行变成了 **A**′的第 i 列。例如，

$$\underset{3\times 2}{\mathbf{A}}=\begin{bmatrix} 4 & 5 \\ 3 & 1 \\ 5 & 0 \end{bmatrix} \qquad \underset{2\times 3}{\mathbf{A}'}=\begin{bmatrix} 4 & 3 & 5 \\ 5 & 1 & 0 \end{bmatrix}$$

由于向量是矩阵的一种特殊类型，故一个行向量的转置是一个列向量，而一个列向量的转置是一个行向量。例如：

$$\mathbf{x}=\begin{bmatrix}4\\5\\6\end{bmatrix} \qquad 和 \qquad \mathbf{x}'=\begin{bmatrix}4 & 5 & 6\end{bmatrix}$$

我们将遵从用一撇来表示行向量的惯例。

子 矩 阵

给定任一 $M\times N$ 矩阵 **A**，如果除了 **A** 的 r 行和 S 列外把其余的行和列全部删除掉，则余下的 $r\times s$ 阶矩阵叫做 **A** 的一个**子矩阵**（submatrix）。例如，如果

$$\underset{3\times 3}{\mathbf{A}}=\begin{bmatrix}3 & 5 & 7\\8 & 2 & 1\\3 & 2 & 1\end{bmatrix}$$

去掉 **A** 的第 3 行和第 3 列后则得到：

$$\underset{2\times 2}{\mathbf{B}}=\begin{bmatrix}3 & 5\\8 & 2\end{bmatrix}$$

这是 **A** 的一个 2×2 阶子矩阵。

§B.2 矩阵的类型

方 阵

915 行数和列数相同的矩阵叫做**方阵**（square matrix）。

$$\mathbf{A}=\begin{bmatrix}3 & 4\\5 & 6\end{bmatrix} \qquad \mathbf{B}=\begin{bmatrix}3 & 5 & 8\\7 & 3 & 1\\4 & 5 & 0\end{bmatrix}$$

对角（矩）阵

主对角线（指从左上角到右下角的对角线）上至少有一个非零元素，而其余地方均是零的方阵叫做**对角（矩）阵**（diagonal matrix）。

$$\underset{2\times 2}{\mathbf{A}}=\begin{bmatrix}2 & 0\\ 0 & 3\end{bmatrix}\qquad \underset{3\times 3}{\mathbf{B}}=\begin{bmatrix}-2 & 0 & 0\\ 0 & 5 & 0\\ 0 & 0 & 1\end{bmatrix}$$

纯量（矩）阵

对角元素完全相等的对角（矩）阵叫做**纯量（矩）阵**（scalar matrix）。方程 C.2.3）所给经典线性回归模型的总体干扰的方差——协方差（矩）阵，即：

$$\text{var-cov}(\mathbf{u})=\begin{bmatrix}\sigma^2 & 0 & 0 & 0 & 0\\ 0 & \sigma^2 & 0 & 0 & 0\\ 0 & 0 & \sigma^2 & 0 & 0\\ 0 & 0 & 0 & \sigma^2 & 0\\ 0 & 0 & 0 & 0 & \sigma^2\end{bmatrix}$$

就是一例。

恒等或单位矩阵

对角元素全是 1 的对角阵叫做**恒等**（identity）或**单位**（unit）**矩阵**，并记为 $\mathbf{I}$。它是纯量矩阵的一个特殊类。

$$\underset{3\times 3}{\mathbf{I}}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\qquad \underset{4\times 4}{\mathbf{I}}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

对称矩阵

一个方阵其主对角线的上方元素是其下方元素的映像时叫做**对称矩阵**。另一说法，一个对称矩阵就是其转置等于其自身的矩阵；就是说 $\mathbf{A}=\mathbf{A}'$，也就是 $\mathbf{A}$ 的元素 a_{ij} 等于 $\mathbf{A}'$ 的元素 a_{ji}。一个例子是方程（C.2.2）所给的方差一协方差矩阵，另一个例子是（C.5.1）所给的相关矩阵。

零矩阵

916 全部元素为零的矩阵叫做**零**（null）**矩阵**，并记为 0。

零 向 量

全部元素为零的行或列向量叫做零向量，并且也记为0。

相等矩阵

我们说两矩阵 **A** 和 **B** 是相等的，如果它们有相同的阶并且它们的相应元素都相等；即对所有 i 和 j，$a_{ij}=b_{ij}$，例如，两矩阵：

$$\underset{3\times 3}{\mathbf{A}}=\begin{bmatrix}3&4&5\\0&-1&2\\5&1&3\end{bmatrix}\quad \text{和}\quad \underset{3\times 3}{\mathbf{B}}=\begin{bmatrix}3&4&5\\0&-1&2\\5&1&3\end{bmatrix}$$

是相等的；就是说，$\mathbf{A}=\mathbf{B}$。

§B.3 矩阵运算

矩阵加法

令 $\mathbf{A}=[a_{ij}]$和 $\mathbf{B}=[b_{ij}]$。如果 **A** 和 **B** 是同阶的，我们就定义矩阵加法为：

$$\mathbf{A}+\mathbf{B}=\mathbf{C}$$

其中 **C** 与 **A** 和 **B** 同阶，并且对所有 i 和 j 求得 $c_{ij}=a_{ij}+b_{ij}$；就是 **C** 由 **A** 和 **B** 的对应元素相加而得。如果这种加法是可以做到的，就说 **A** 和 **B** 是加法相适的（conformable for addition）。例如，

$$\mathbf{A}=\begin{bmatrix}2&3&4&5\\6&7&8&9\end{bmatrix}\text{和 }\mathbf{B}=\begin{bmatrix}1&0&-1&3\\-2&0&1&5\end{bmatrix}$$

并且 $\mathbf{C}=\mathbf{A}+\mathbf{B}$，则

$$\mathbf{C}=\begin{bmatrix}3&3&3&8\\4&7&9&14\end{bmatrix}$$

矩阵减法

矩阵减法仿照矩阵加法同样原理，只不过 $\mathbf{C}=\mathbf{A}-\mathbf{B}$；即从 **A** 的对

应元素减去 **B** 的对应元素，这里假定了 **A** 和 **B** 是同阶的。

纯量乘法

917 矩阵 **A** 乘以纯量 λ（一个实数），就是用 λ 去乘矩阵的每一个元素：

$$\lambda\mathbf{A} = [\lambda a_{ij}]$$

例如，取 $\lambda = 2$，并且

$$\mathbf{A} = \begin{bmatrix} -3 & 5 \\ 8 & 7 \end{bmatrix}$$

则

$$\lambda\mathbf{A} = \begin{bmatrix} -6 & 10 \\ 16 & 14 \end{bmatrix}$$

矩阵乘法

设 **A** 是 $M \times N$ 和 **B** 是 $N \times P$ 矩阵。那么乘积 **AB**（按照 **A** 乘 **B** 的这个顺序）按定义是一个新的 $M \times P$ 阶矩阵 **C**，其典型元素为：

$$c_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj} \qquad \begin{matrix} i = 1, 2, \cdots, M \\ j = 1, 2, \cdots, P \end{matrix}$$

即将 **A** 的第 i 行元素乘以 **B** 的第 j 列相应元素后，对所有项求总和便得到 **C** 的第 i 行第 j 列元素；这就是所谓行乘以列的乘法法则（row by column rule of multiplication）。例如，为了得到 c_{11}，即 **C** 的第 1 行第 1 列元素，我们将 **A** 的第 1 行元素乘以 **B** 的第 1 列相应元素，然后求所有项的总和。类似地，为了得到 c_{12}，我们将 **A** 的第 1 行元素乘以 **B** 的第 2 列相应元素，然后求其总和，依此类推。

注意，为了乘法可行，矩阵 **A** 和 **B** 必须是乘法相适的（conformable with respect to multiplication）（或者说 **A** 和 **B** 是可乘矩阵——译者注），就是说，**A** 的列数必须等于 **B** 的行数。例如

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 4 & 7 \\ 5 & 6 & 1 \end{bmatrix} \quad 和 \quad \underset{3\times 2}{\mathbf{B}} = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 6 & 2 \end{bmatrix}$$

$$\mathbf{AB} = \underset{2\times 2}{\mathbf{C}} = \begin{bmatrix} (3\times 2) + (4\times 3) + (7\times 6) & (3\times 1) + (4\times 5) + (7\times 2) \\ (5\times 2) + (6\times 3) + (1\times 6) & (5\times 1) + (6\times 5) + (1\times 2) \end{bmatrix}$$
$$= \begin{bmatrix} 60 & 37 \\ 34 & 37 \end{bmatrix}$$

但若

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 4 & 7 \\ 5 & 6 & 1 \end{bmatrix} \text{和} \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}$$

则因 **A** 和 **B** 不是乘法相适的，乘积 **AB** 就是没有定义的。

矩阵乘法的性质

918 1. 矩阵乘法不一定是可交换的；一般地说，$\mathbf{AB}\neq\mathbf{BA}$，因此相乘的矩阵的顺序非常重要。**AB** 是指 **A** 后乘以 **B** 或者 **B** 前乘以 **A**。

2. 即使 **AB** 和 **BA** 都存在，结果的两个（乘积）矩阵可以是不同阶的。例如，**A** 为 $M\times N$ 而 **B** 为 $N\times M$，则 **AB** 为 $M\times M$ 而 **BA** 为 $N\times N$，即各有不同的阶。

3. 即使 **A** 和 **B** 都是方阵，从而 **AB** 和 **BA** 两者都有定义，但结果所得的矩阵不一定相等。例如，如果

$$\mathbf{A} = \begin{bmatrix} 4 & 7 \\ 3 & 2 \end{bmatrix} \quad \text{和} \quad \mathbf{B} = \begin{bmatrix} 1 & 5 \\ 6 & 8 \end{bmatrix}$$

则

$$\mathbf{AB} = \begin{bmatrix} 46 & 76 \\ 15 & 31 \end{bmatrix} \quad \text{和} \quad \mathbf{BA} = \begin{bmatrix} 19 & 17 \\ 48 & 58 \end{bmatrix}$$

从而 $\mathbf{AB}\neq\mathbf{BA}$。当 **A** 和 **B** 之一为恒等矩阵时，我们将看到 $\mathbf{AB}=\mathbf{BA}$ 的一个例子。

4. 一个行向量后乘以一个列向量是一个纯量。例如，考虑普通最小二乘回归残差 $\hat{u}_1$、$\hat{u}_2$、…，$\hat{u}_n$，令 **u** 为一列向量和 **u′** 为一行向量，我们有：

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= [\hat{u}_1\hat{u}_2\hat{u}_3\cdots\hat{u}_n]\begin{bmatrix}\hat{u}_1\\ \hat{u}_2\\ \hat{u}_3\\ \vdots\\ \hat{u}_n\end{bmatrix}\\ &= \hat{u}_1^2+\hat{u}_2^2+\hat{u}_3^2+\cdots+\hat{u}_n^2\\ &= \sum\hat{u}_i^2 \qquad \text{为一纯量［参看方程 (C.3.5)］}\end{aligned}$$

5. 一个列向量后乘以一个行向量是一个矩阵。作为例子，考虑经典线性回归模型的总体干扰，即 u_1、u_2、$\cdots u_n$，令 **u** 为列向量和 **u′** 为行向量，我们得到：

$$\mathbf{uu'}=\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_n\end{bmatrix}\begin{bmatrix}u_1 u_2 u_3\cdots u_n\end{bmatrix}$$

$$=\begin{bmatrix}u_1^2 & u_1u_2 & u_1u_3 & \cdots u_1u_n\\u_2u_1 & u_2^2 & u_2u_3 & \cdots u_2u_n\\\cdots & \cdots & \cdots & \cdots\\u_nu_1 & u_nu_2 & u_nu_3 & \cdots \quad u_n^2\end{bmatrix}$$

919 这是一个 $n\times n$ 阶矩阵。注意上述矩阵是对称的。

6. 一个矩阵后乘以一个列向量是一个列向量。

7. 一个行向量后乘以一个矩阵是一个行向量。

8. 矩阵乘法是可结合的（associative）；就是，$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})$，其中 $\mathbf{A}$ 为$M\times N$，$\mathbf{B}$ 为 $N\times P$ 和 $\mathbf{C}$ 为 $P\times K$。

9. 矩阵乘法是对加法可分配的；就是，$\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$ 和$(\mathbf{B}+\mathbf{C})\mathbf{A}=\mathbf{BA}+\mathbf{CA}$。

矩阵转置

我们曾经定义矩阵转置过程为一个矩阵（或一个向量）的行和列的互换。现在我们来叙述转置的一些性质。

1. 转置矩阵的转置是原矩阵本身，即 $(\mathbf{A}')'=\mathbf{A}$。

2. 如果 $\mathbf{A}$ 和 $\mathbf{B}$ 是加法相适的，则 $\mathbf{C}=\mathbf{A}+\mathbf{B}$ 和 $\mathbf{C}'=(\mathbf{A}+\mathbf{B})'=\mathbf{A}'+\mathbf{B}'$。就是说，两矩阵之和的转置是它们的转置之和。

3. 如果 $\mathbf{AB}$ 有定义，则 $(\mathbf{AB})'=\mathbf{B}'\mathbf{A}'$，就是说，两矩阵之积的转置是它们顺序相反的转置之积。这可推广为 $(\mathbf{ABCD})'=\mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$。

4. 恒等矩阵 $\mathbf{I}$ 的转置是恒等矩阵本身；即 $\mathbf{I}'=\mathbf{I}$。

5. 纯量的转置是该纯量本身。就是说，如果 λ 是一纯量，则 $\lambda'=\lambda$。

6. $(\lambda\mathbf{A})'$的转置是 $\lambda\mathbf{A}'$，其中 λ 是一纯量。［注：$(\lambda\mathbf{A})'=\mathbf{A}'\lambda'=\mathbf{A}'\lambda=\lambda\mathbf{A}'$］

7. 如果 $\mathbf{A}$ 是这样的一个方阵：$\mathbf{A}=\mathbf{A}'$，那么 $\mathbf{A}$ 是一个对称矩阵。（参照前面给出的对称矩阵定义。）

矩阵求逆

方阵 $\mathbf{A}$ 的逆（矩）阵，记为 $\mathbf{A}^{-1}$（读 $\mathbf{A}$ 逆），如果存在的话，是满足下式的惟一方阵：

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

其中 **I** 是一个与 **A** 同阶的单位矩阵。例如

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} \quad \mathbf{A}^{-1} = \begin{bmatrix} -1 & \frac{1}{2} \\ \frac{6}{8} & -\frac{1}{4} \end{bmatrix} \quad \mathbf{AA}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

在我们学习行列式的问题后，我们就会看到 $\mathbf{A}^{-1}$是怎样计算的，这里且注意逆矩阵的如下性质。

1．$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$；即两矩阵之积的逆阵是它们的顺序相反的逆阵之积。

2．$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$；即 **A** 逆的转置是 **A** 转置的逆。

§B.4　行列式

920　对应于每一个方阵 **A**，都有一个称为矩阵的行列式的数。这个数记为 det**A** 或 $|\mathbf{A}|$，这里 $|\;|$ 表示“行列式”，注意一个矩阵本身是没有数值的，但一个矩阵的行列式则是一个数。

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -7 \\ 2 & 5 & 0 \\ 3 & 8 & 6 \end{bmatrix} \quad |\mathbf{A}| = \begin{vmatrix} 1 & 3 & -7 \\ 2 & 5 & 0 \\ 3 & 8 & 6 \end{vmatrix}$$

在本例中，$|\mathbf{A}|$ 因为对应着一个 3×3 阶矩阵，所以叫做一个 3 阶行列式。

行列式的计算

求行列式的值的过程叫做行列式计算（evaluation）展开（expansion）或化简（reduction）。如何对矩阵中的记载进行操作以完成这一过程，有明确的方法。

2×2 行列式的计算。如果，

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

则其行列式的计算如下：

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

如箭头所指，将 **A** 的主对角线上的元素相乘，减去另一（副）对角线上的元素乘积，便得此结果。

3×3 行列式的计算。如果

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

则

$$|\mathbf{A}|=a_{11}a_{22}a_{33}-a_{11}a_{23}a_{32}+a_{12}a_{23}a_{31}-a_{12}a_{21}a_{33}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}$$

对 3×3 行列式的展开，经仔细分析表明：

1. 行列式的展开式中每一项都包含每行和每列的 1 个而且仅仅一个元素。

921 2. 每项的元素个数都与矩阵的行数（或列数）相同。例如，2×2 行列式在其展开式的每一项中都有两个元素，3×3 行列式在其展开式的每一项中都有三个元素，依此类推。

3. 展开式中各项的正负号轮流出现。

4. 2×2 行列式的展开式有 2 项，而 3×3 行列式的展开式有 6 项。一般规律是：$N\times N$ 阶行列式的展开式有 $N!=N(N-1)(N-2)\cdots3\cdot2\cdot1$ 项，其中 $N!$ 读如“N 阶乘”。按照这一规律，一个 5×5 阶行列式的展开式就有 $5\cdot4\cdot3\cdot2\cdot1=120$ 项。[1]

行列式的性质

1. 行列式值为零的矩阵叫做**退化［矩］阵**（singular matrix），而有非零行列式的矩阵叫做**非退化阵**（nonsingular matrix）。前面定义的逆矩阵对退化阵来说是不存在的。

2. 如果 **A** 的任何一行元素全为零，则它的行列式为零。例如，

$$|\mathbf{A}|=\begin{vmatrix} 0 & 0 & 0 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{vmatrix}=0$$

3. $|\mathbf{A}'|=|\mathbf{A}|$；就是说，**A** 和 **A** 转置有相同的行列式。

4. 交换矩阵 **A** 的任何两行或任何两列将改变 $|\mathbf{A}|$ 的符号。

例：如果

$$\mathbf{A}=\begin{bmatrix} 6 & 9 \\ -1 & 4 \end{bmatrix} \quad 和 \quad \mathbf{B}=\begin{bmatrix} -1 & 4 \\ 6 & 9 \end{bmatrix}$$

其中 **B** 由交换 **A** 的两行而得，则：

$$|\mathbf{A}| = 24 - (-9) \qquad \text{和} |\mathbf{B}| = -9 - (24)$$
$$= 33 \qquad\qquad = -33$$

922 5. 如果用纯量 λ 乘 **A** 的某一行或一列的每一元素，则 $|\mathbf{A}|$ 就乘以 λ。

例：如果

$$\lambda = 5 \quad \text{和} \ \mathbf{A} = \begin{bmatrix} 5 & -8 \\ 2 & 4 \end{bmatrix}$$

并且用 5 乘 **A** 的第一行得：

$$\mathbf{B} = \begin{bmatrix} 25 & -40 \\ 2 & 4 \end{bmatrix}$$

我们就能看到 $|\mathbf{A}| = 36$，而 $|\mathbf{B}| = 180 = 5|\mathbf{A}|$。

6. 如果一个矩阵的两行或列相同，则其行列式为零。

7. 如果一个矩阵的一行或列是另一行或列的一个倍数，则它的行列式为零。例如，

$$\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 2 & 4 \end{bmatrix}$$

其中 **A** 的第 1 行是第 2 行的两倍，故 $|\mathbf{A}| = 0$。更一般地，如果矩阵的任一行（列）是其他行（列）的一个线性组合，则其行列式为零。

8. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$；就是说，两矩阵之积的行列式等于它们（各自）的行列式之积。

矩阵的秩

一个矩阵的秩是其行列式不为零的最大子方阵的阶。

例：

$$\mathbf{A} = \begin{bmatrix} 3 & 6 & 6 \\ 0 & 4 & 5 \\ 3 & 2 & 1 \end{bmatrix}$$

可以看出 $|\mathbf{A}| = 0$。换言之，**A** 是一个退化矩阵，因此，虽然它的阶是 3×3，但它的秩小于 3。实际上它是 2，因为可以找到一个行列式不为零的 2×2 子方阵，例如去掉 **A** 的第 1 行和第 1 列，我们得到：

$$\mathbf{B} = \begin{bmatrix} 4 & 5 \\ 2 & 1 \end{bmatrix}$$

它的行列式是 −6，这是个非零值。从而 **A** 的秩是 2，如前面已指出的，一个退化矩阵不存在有逆阵。因此，对于一个 $N \times N$ 矩阵 **A**，要它的逆阵存在，它的秩必须是 N；如果它小于 N，则 **A** 是退化的。

子　式

923 如果把 $N\times N$ 矩阵 $\mathbf{A}$ 的第 i 行和第 j 列去掉，所余下来的子矩阵的行列式就叫做元素 a_{ij}（第 i 行和第 j 列交叉处的元素）的**子式**（minor），并记以 $|\mathbf{M}_{ij}|$。

例

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

a_{11} 的子式是：

$$|\mathbf{M}_{11}|=\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}=a_{22}a_{33}-a_{23}a_{32}$$

类似地，$|\mathbf{M}_{21}|=\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix}=a_{12}a_{33}-a_{13}a_{32}$

用同样方法可求出 $\mathbf{A}$ 的其他元素的子式。

余　因　子

$N\times N$ 矩阵 $\mathbf{A}$ 的元素 a_{ij} 的余因子（cofactor），记为 c_{ij}，定义为：

$$c_{ij}=(-1)^{i+j}|\mathbf{M}_{ij}|$$

换句话说，余因子是有符号的子式，当 $i+j$ 为偶数时取正号，而当 $i+j$ 为奇数时取负号。例如，前面所给 3×3 矩阵 $\mathbf{A}$ 的元素 a_{11} 的余因子是 $a_{22}a_{33}-a_{23}a_{32}$，而 a_{21} 的余因子是 $-(a_{12}a_{33}-a_{13}a_{32})$，因为下标 2 与 1 之和 3 是一个奇数。

余因子矩阵。矩阵 $\mathbf{A}$ 的元素 a_{ij} 代之以其余因子，即给出所谓的 $\mathbf{A}$ 的余因子矩阵（cofactor matrix），记为 $(\text{cof }\mathbf{A})$。

伴随矩阵。伴随矩阵（adjoint matrix），记为 $(\text{adj }\mathbf{A})$，是余因子矩阵的转置；即 $(\text{adj }\mathbf{A})=(\text{cof }\mathbf{A})'$。

§B.5　求一个方阵的逆阵

如果 $\mathbf{A}$ 是方阵的且非退化的（即 $|\mathbf{A}|\neq0$），则其逆阵可按下式求得：

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\ (\text{adj } \mathbf{A})$$

924 所涉及的计算步骤如下：

1. 求 **A** 的行列式，如果它不为零，就进行第二步。
2. 将 **A** 的每一元素 a_{ij} 代之以它的余因子，以得到余因子矩阵。
3. 将余因子转置，以得到伴随矩阵。
4. 用 $|\mathbf{A}|$ 去除伴随矩阵中的每一元素。

例：求下列矩阵的逆阵

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 7 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

第 1 步 先求矩阵的行列式，应用前面给的 3×3 行列式的展开规则，我们求得 $|\mathbf{A}| = -24$。

第 2 步 现在求余因子矩阵，且记为 **C**

$$\mathbf{C} = \begin{bmatrix} \begin{vmatrix} 7 & 4 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 5 & 4 \\ 2 & 3 \end{vmatrix} & \begin{vmatrix} 5 & 7 \\ 2 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 7 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 5 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 5 & 7 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 17 & -7 & -9 \\ -3 & -3 & 3 \\ -13 & 11 & -3 \end{bmatrix}$$

第 3 步 将上述余因子矩阵转置，我们得到如下的伴随矩阵：

$$(\text{adj } \mathbf{A}) = \begin{bmatrix} 17 & -3 & -13 \\ -7 & -3 & 11 \\ -9 & 3 & -3 \end{bmatrix}$$

第 4 步 用行列式值 −24 遍除 (adj **A**) 的元素便得：

$$\mathbf{A}^{-1} = -\frac{1}{24}\begin{bmatrix} 17 & -3 & -13 \\ -7 & -3 & 11 \\ -9 & 3 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{17}{24} & \frac{3}{24} & \frac{13}{24} \\ \frac{7}{24} & \frac{3}{24} & -\frac{11}{24} \\ \frac{9}{24} & -\frac{3}{24} & \frac{3}{24} \end{bmatrix}$$

容易验证：

$$\mathbf{AA}^{-1}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$$

这是一个恒等矩阵。读者能验证，对第9章所给的说明性例子，**X′X** 矩阵的逆阵确是方程（C.10.5）所展示的。

§B.6 矩阵微分法

925 **要跟上第CA.2节的内容，我们需要一些关于矩阵微分法的规则。**

规则 1. 如果 $\mathbf{a}'=[a_1a_2\cdots a_n]$ 是一数值行向量，而

$$\mathbf{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

是变量 x_1、x_2、$\cdots x_n$ 的一个列向量，则

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial\mathbf{x}}=\mathbf{a}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_n\end{bmatrix}$$

规则 2. 考虑这样的矩阵 $\mathbf{x}'\mathbf{Ax}$：

$$\mathbf{x}'\mathbf{Ax}=[x_1\quad x_2\quad\cdots\quad x_n]\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\cdots&\cdots&\cdots&\cdots\\a_{n1}&a_{n2}&&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

那么

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial\mathbf{x}}=2\mathbf{Ax}$$

这是一个 n 元素的列向量，或者

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial\mathbf{x}}=2\mathbf{x}'\mathbf{A}$$

这是一个 n 元素的行向量。

参考文献

Chiang, Alpha C.: *Fundamental Methods of Mathematical Economics*, 3d ed., McGraw-Hill, New York 1984, Chaps. 4 and 5. 这是一本初等的著述。

Hadley G.: *Linear Algebra*, Addison-Wesley, Reading, Mass., 1961. 这是一本高深的著述。

【附录 B 注释】

[1] 关于 $N \times N$ 矩阵的行列式计算，请参阅文献。

附录 C 线性回归模型的矩阵方法

926 本章介绍用矩阵代数符号表示的 k 变量（Y 和 X_2，X_3，…，X_k）经典线性回归模型。概念上，k 变量模型是本书迄今讨论的二和三变量模型的逻辑推广。因此本章除矩阵符号外不涉及什么新概念。[1]

矩阵代数与标量（即纯量）代数（处理标量或实数的初等代数）相比，最大的优越性在于，它为处理涉及任意多个变量的回归模型提供了一个简洁的方法；一旦用矩阵符号建立并求解了 k 变量模型，其解式适用于一、二、三或任意多个变量的情形。

§C.1 k 变量线性回归模型

如果我们把二和三变量线性回归模型推广，则含因变量 Y 和 $k-1$ 个解释变量 X_2，X_3，…，X_k 的 k 变量总体回归模型（PRF）就可写为：

$$\text{PRF：} Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i \quad i=1,2,3,\cdots,n \tag{C.1.1}$$

其中 β_1 = 截距，β_2 至 β_k = 偏斜率系数，u = 随机干扰项，i = 第 i 次观测而 n 为总体的大小。PRF（C.1.1）可按平常那样加以解释：它给出以 X_2，X_3，…，X_k 的固定（在重复抽样中）值为条件的 Y 的均值或期望值，即 $E(Y|X_{2i}, X_{3i}, \cdots, X_{ki})$。

927 方程（C.1.1）是以下 n 个联立方程组的一个缩写表达式：

$$\begin{aligned}
Y_1 &= \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \cdots + \beta_k X_{k1} + u_1 \\
Y_2 &= \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \cdots + \beta_k X_{k2} + u_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_n &= \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \cdots + \beta_k X_{kn} + u_n
\end{aligned} \tag{C.1.2}$$

让我们把方程（C.1.2）写成另一更有启发性的方式如下：[2]

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{C.1.3}$$

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times k}{\mathbf{X}} \quad \underset{k\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{u}}$$

其中 $\mathbf{y}$ = 对应变量 Y 的观测值的 $n\times 1$ 列向量

$\mathbf{X}$ = 给出对 $k-1$ 个变量 X_2 至 X_k 的 n 次观测值的 $n\times k$ 矩阵，其全为 1 的列代表截距项。（此矩阵又名为**数据矩阵**。）

$\boldsymbol{\beta}$ = 未知参数 β_1、β_2，…，β_k 的 $k\times 1$ 列向量

$\mathbf{u}$ = n 个干扰项 u_i 的 $n\times 1$ 列向量

利用矩阵乘法和加法法则，读者应能证实方程组（C.1.2）和（C.1.3）是等价的。

方程组（C.1.3）被称为一般（k 变量）线性回归模型的矩阵表现。可以更紧凑地把它写为：

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times k}{\mathbf{X}} \quad \underset{k\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{u}} \tag{C.1.4}$$

如果在矩阵 $\mathbf{X}$ 和向量 $\mathbf{y}$，$\boldsymbol{\beta}$ 和 $\mathbf{u}$ 的维数或阶数上不会有误解，方程（C.1.4）就可简单地写为：

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{C.1.5}$$

作为矩阵表现的一个具体说明，考虑第 3 章中讨论过的双变量消费—收

928 入模型，即 $Y_i = \beta_1 + \beta_2 X_i + u_i$，其中 Y 是消费支出，而 X 是收入。利用表 3.2 的数据，可将此矩阵公式写为：

$$
\begin{bmatrix} 70 \\ 65 \\ 90 \\ 95 \\ 110 \\ 115 \\ 120 \\ 140 \\ 155 \\ 150 \end{bmatrix} = \begin{bmatrix} 1 & 80 \\ 1 & 100 \\ 1 & 120 \\ 1 & 140 \\ 1 & 160 \\ 1 & 180 \\ 1 & 200 \\ 1 & 220 \\ 1 & 240 \\ 1 & 260 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{bmatrix} \tag{C.1.6}
$$

$$
\underset{10\times 1}{\mathbf{y}} = \underset{10\times 2}{\mathbf{X}} \; \underset{2\times 1}{\boldsymbol{\beta}} + \underset{10\times 1}{\mathbf{u}}
$$

如同二和三变量情形那样，我们的目的是估计复回归(C.1.1)的参数并从所掌握的数据对它们做出推断。用矩阵符号表示，就是要估计 $\boldsymbol{\beta}$ 并对此 $\boldsymbol{\beta}$ 进行推断。为了估计，可用普通最小二乘(OLS)法或最大似然(ML)法。但如前所述，这两种方法给出回归系数的同样的估计。[3]因此，我们将限于 OLS 法的讨论。

§C.2 用矩阵表示的关于经典线性回归模型的假定

关于经典线性回归模型的基本假定，列示于表 C.1；这些假定并行地用标量符号和矩阵符号表示出来。由(C.2.1)给出的假定 1 是指干扰向量 $\mathbf{u}$ 的期望值为零，即其中的每一元素的期望值为零。更明显地表示，$E(\mathbf{u})=0$ 是指

$$
E\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{C.2.1}
$$

假定 2[方程(C.2.2)]是(3.2.5)和(3.2.2)两个用标量符号表示的假
929 定的一个简洁的表达式。为了看清楚这点，可写为：

表 C.1　　关于经典回归模型的假定

标量符号		矩阵符号
1. $E(u_i)=0$ 对每个 i	(3.2.1)	1. $E(\mathbf{u})=\mathbf{0}$ 其中 $\mathbf{u}$ 和 $\mathbf{0}$ 都是 $n\times 1$ 列向量 且 $\mathbf{0}$ 是零向量

2. $E(u_iu_j)=0 \quad i\neq j$ (3.2.5) $=\sigma^2 \quad i=j$ (3.2.2)	2. $E(\mathbf{uu}')=\sigma^2\mathbf{I}$ 其中 $\mathbf{I}$ 是 $n\times n$ 恒等矩阵
3. $X_2, X_3, \cdots, X_k$ 是非随机的或固定的	3. $n\times k$ 矩阵 $\mathbf{X}$ 是非随机的，即它由固定数的一个集合构成
4. X 诸变量之间无准确的线性关系，即无多重共线性 (7.1.7)	4. $\mathbf{X}$ 的秩是 $\rho(\mathbf{X})=k$，其中 k 是 $\mathbf{X}$ 中的列数，且 k 小于观测次数 n
5. 为了假设检验 (4.2.4) $u_i\sim N(0,\sigma^2)$	5. 向量 $\mathbf{u}$ 有一多维正态分布，即 $\mathbf{u}\sim N(\mathbf{0},\sigma^2\mathbf{I})$

$$E(\mathbf{uu}')=E\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}\begin{bmatrix}u_1u_2\cdots u_n\end{bmatrix}$$

其中 $\mathbf{u}'$ 是列向量 $\mathbf{u}$ 的转置或者是一个行向量。做向量的乘法，我们得到：

$$E(\mathbf{uu}')=E\begin{bmatrix}u_1^2 & u_1u_2\cdots & u_1u_n\\ u_2u_1 & u_2^2\cdots & u_2u_n\\ \cdots\cdots & \cdots\cdots & \cdots\cdots\\ u_nu_1 & u_nu_2\cdots & u_n^2\end{bmatrix}$$

把期望值运算子 E 应用于上述矩阵的每一元素得：

$$E(\mathbf{uu}')=\begin{bmatrix}E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n)\\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2)\end{bmatrix} \tag{C.2.2}$$

930 由于同方差性及无序列相关性假定，矩阵(C.2.2)简化为：

$$\begin{aligned}E(\mathbf{uu}')&=\begin{bmatrix}\sigma^2 & 0 & 0 & \cdots & 0\\ 0 & \sigma^2 & 0 & \cdots & 0\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ 0 & 0 & 0 & \cdots & \sigma^2\end{bmatrix}\\ &=\sigma^2\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix}\\ &=\sigma^2\mathbf{I}\end{aligned} \tag{C.2.3}$$

其中 $\mathbf{I}$ 是一个 $n\times n$ 恒等矩阵。

矩阵(C.2.2)[及其在(C.2.3)中的表现]称为干扰项 u_i 的**方差—协方差**

矩阵(variance-covariance matrix);此矩阵的主对角线(由左上角穿越到右下角)上的元素给出方差,而偏离主对角线的元素则给出协方差。[4]注意方差协方差矩阵的**对称性**:主对角线上方和下方的元素是对称的。

假定3是说 $n\times k$ 矩阵 $\mathbf{X}$ 是非随机的;就是说它由固定的数构成。如前所述,我们的回归分析是条件回归分析,以诸 X 变量的固定值作为条件。

假定4是说 $\mathbf{X}$ 矩阵是列满秩(full column rank)的,即其秩等于矩阵的列数。意思是,$\mathbf{X}$ 矩阵的列是线性独立的;就是说,在 X 诸变量之间无准确的线性关系,换言之,无多重共线性。用标量符号表示,这就等于说,不存在非全为零的一组数 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 使得[比较(7.1.8)]:

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0 \tag{C.2.4}$$

其中对一切 i 有 $X_{1i}=1$(即式中包括了 $\mathbf{X}$ 矩阵中全为1的一列)。(C.2.4)可用矩阵符号表现为:

$$\boldsymbol{\lambda}'\mathbf{x} = 0 \tag{C.2.5}$$

其中 $\boldsymbol{\lambda}'$ 为 $1\times k$ 行向量而 $\mathbf{x}$ 为 $k\times 1$ 列向量。

931 如果存在有像(C.2.4)那样的一个准确线性关系式,则说诸变量是共线的。反之,如果仅当 $\lambda_1=\lambda_2=\lambda_3=\cdots=0$ 时(C.2.4)成立,则说 X 诸变量是线性独立的。无多重共线性假定的直觉理由,已在第7章中陈述。在第10章中,我们还将对此假定做进一步的探讨。

§C.3 OLS 估计

为了求 $\boldsymbol{\beta}$ 的 OLS 估计值,首先让我们写出 k 变量的样本回归函数(SRF):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i \tag{C.3.1}$$

该式可用矩阵符号更简洁地表达为:

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}} \tag{C.3.2}$$

其矩阵形式为:

$$\underset{n\times 1}{\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{y}}} = \underset{n\times k}{\underbrace{\begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}}_{\mathbf{X}}} \underset{k\times 1}{\underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\boldsymbol{\beta}}}} + \underset{n\times 1}{\underbrace{\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}}_{\hat{\mathbf{u}}}} \tag{C.3.3}$$

其中 $\hat{\boldsymbol{\beta}}$ 是回归系数的 OLS 估计量的一个 k 元素列向量,而 $\hat{\mathbf{u}}$ 是 n 个残差的 $n\times 1$ 列向量。

如同二和三变量模型那样,k 变量情形的 OLS 估计量也是从残差平方和

$\sum \hat{u}_i^2$(RSS)：

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2 \tag{C.3.4}$$

的最小化求得。用矩阵符号表示，这就等于使 $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ 最小化，因为：

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = [\hat{u}_1 \quad \hat{u}_2 \cdots \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2 = \sum \hat{u}_i^2 \tag{C.3.5}$$

现在由(C.3.2)可得：

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \tag{C.3.6}$$

932 因此，

$$\begin{aligned} \hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{C.3.7}$$

这里我们利用了矩阵转置的一些性质：即 $(\mathbf{X}\hat{\boldsymbol{\beta}})' = \hat{\boldsymbol{\beta}}'\mathbf{X}'$；以及因 $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ 为一标量(一实数)，它的转置 $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$ 就是它本身。

方程(C.3.7)是(C.3.4)的矩阵表现。用标量表现时，OLS 法是要对 β_1、β_2，…，β_k 的估计能使 $\sum \hat{u}_i^2$ 尽可能小。要做到这点，方法是将(C.3.4)对 $\hat{\beta}_1$，$\hat{\beta}_2$，…，$\hat{\beta}_k$ 微分并令微分的结果表达式为零，以产生最小二乘理论的正规方程——k 个未知数的 k 个联立方程。如附录 CA，第 CA.1 节中所展示的，这些方程是：

$$\begin{aligned} n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki} &= \sum Y_i \\ \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i}X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i}X_{ki} &= \sum X_{2i}Y_i \\ \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{3i}X_{2i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i}X_{ki} &= \sum X_{3i}Y_i \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki}X_{2i} + \hat{\beta}_3 \sum X_{ki}X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}^2 &= \sum X_{ki}Y_i \end{aligned}$$

(C.3.8)[5]

写成矩阵形式，方程(C.3.8) 可表现为：

$$\underbrace{\begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \cdots & \sum X_{ki}^2 \end{bmatrix}}_{(\mathbf{X}'\mathbf{X})} \underbrace{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{21} & X_{22} & \cdots & X_{2n} \\ X_{31} & X_{32} & \cdots & X_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix}}_{\mathbf{X}'} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{y}} \tag{C.3.9}$$

或更简洁地写为：

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y} \tag{C.3.10}$$

933 注意$(\mathbf{X}'\mathbf{X})$矩阵的如下特点：(1)它给出 X 诸变量的粗(原生)平方和及交叉乘积和，变量之一是每次观测都取值 1 的截距项。主对角线上的元素是粗平方和，而主对角线以外的元素是粗交叉乘积和("粗"是指以原始测量单位计算)。(2)因 X_{2i}与X_{3i}之间的交叉乘积就是X_{3i}与X_{2i}之间的交叉乘积，故它是对称的。(3)它的阶数是$(k\times k)$，就是 k 行与 k 列。

在(C.3.10)中，已知量是$(\mathbf{X}'\mathbf{X})$和$(\mathbf{X}'\mathbf{y})$(X 诸变量与 y 的交叉乘积)，未知量是 $\hat{\boldsymbol{\beta}}$。由矩阵代数，如果$(\mathbf{X}'\mathbf{X})$的逆(矩阵)存在，写为$(\mathbf{X}'\mathbf{X})^{-1}$，则用此逆去前乘(C.3.10)的两边便得到：

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

但因$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})=\mathbf{I}$ 为 $k\times k$ 阶恒等矩阵，故得：

$$\mathbf{I}\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

或：

$$\underset{k\times 1}{\hat{\boldsymbol{\beta}}}=\underset{k\times k}{(\mathbf{X}'\mathbf{X})^{-1}}\ \underset{(k\times n)}{\mathbf{X}'}\ \underset{(n\times 1)}{\mathbf{y}} \tag{C.3.11}$$

方程(C.3.11)是矩阵符号表示的 OLS 理论的一个基本结果。它表明 $\hat{\boldsymbol{\beta}}$ 向量怎样能从给定的数据估计出来。虽然(C.3.11)是从(C.3.9)得来，但它能直接从(C.3.7)通过 $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ 对 $\hat{\boldsymbol{\beta}}$ 的微分直接求得。证明见附录 CA，第 CA.2 节。

一个说明

作为现在已讲到的矩阵方法的一个说明，让我们重新计算第 3 章的消费—收入例子。它的数据已在(C.1.6)中重现。对这个双变量情形我们有：

$$\hat{\boldsymbol{\beta}}=\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})=\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ X_1 & X_2 & X_3 & \cdots & X_n\end{bmatrix}\begin{bmatrix}1 & X_1\\ 1 & X_2\\ 1 & X_3\\ \cdots & \\ 1 & X_N\end{bmatrix}=\begin{bmatrix}n & \sum X_i\\ \sum X_i & \sum X_i^2\end{bmatrix}$$

934 及

$$
\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 \cdots 1 \\ X_1 & X_2 & X_3 \cdots X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}
$$

使用(C.1.6)中的数据，我们得到：

$$
\mathbf{X}'\mathbf{X} = \begin{bmatrix} 10 & 1\ 700 \\ 1\ 700 & 322\ 000 \end{bmatrix}
$$

及

$$
\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1\ 110 \\ 205\ 500 \end{bmatrix}
$$

利用附录 B 给出的矩阵求逆法则，可以得到上面($\mathbf{X}'\mathbf{X}$)矩阵的逆矩阵为：

$$
(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.975\ 76 & -0.005\ 152 \\ -0.005\ 152 & 0.000\ 030\ 3 \end{bmatrix}
$$

因此，

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \begin{bmatrix} 0.975\ 76 & -0.005\ 152 \\ -0.005\ 152 & 0.000\ 030\ 3 \end{bmatrix} \begin{bmatrix} 1\ 110 \\ 205\ 500 \end{bmatrix} \\
&= \begin{bmatrix} 24.454\ 5 \\ 0.507\ 9 \end{bmatrix}
\end{aligned}
$$

先前我们曾用计算机程序得到 $\hat{\beta}_1 = 24.454\ 5$ 和 $\hat{\beta}_2 = 0.509\ 1$。两种估计的差异来自进位误差。顺便指出，如果用台式计算器计算，则必须保留许多位有效数字，才能降低进位误差。

$\hat{\boldsymbol{\beta}}$ 的方差—协方差矩阵

矩阵方法不仅能使我们导出 $\hat{\boldsymbol{\beta}}$ 的任一元素 $\hat{\beta}_i$ 的方差公式，还求出 $\hat{\boldsymbol{\beta}}$ 的任意两元素 $\hat{\beta}_i$ 和 $\hat{\beta}_j$ 的协方差公式。我们需要用这些方差和协方差来做统计推断。

按定义，$\hat{\boldsymbol{\beta}}$ 的方差—协方差是[比较(C.2.2)]：

$$
\text{var-cov}(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]'\}
$$

可更明显地把它写为：

935

$$
\text{var-cov}\ (\hat{\boldsymbol{\beta}}) = \begin{bmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \text{cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \text{cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{var}(\hat{\beta}_2) & \cdots & \text{cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \text{cov}(\hat{\beta}_k, \hat{\beta}_1) & \text{cov}\ (\hat{\beta}_k, \hat{\beta}_2) & \cdots & \text{var}(\hat{\beta}_k) \end{bmatrix} \tag{C.3.12}
$$

在附录 CA，第 CA.3 节中，我们看到上述方差—协方差矩阵可从下述公式算得：

$$\text{var-cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{C.3.13}$$

其中 σ^2 是 u_i 的共同方差，而$(\mathbf{X}'\mathbf{X})^{-1}$就是出现在给出 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 的方程(C.3.11)中的逆矩阵。

在二和三变量线性回归模型中，σ^2 的一个无偏估计量分别由 $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$和 $\hat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$给出。在 k 变量情形中，相应的公式是：

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum \hat{u}_i^2}{n-k} \\ &= \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\end{aligned} \tag{C.3.14}$$

其中 $n-k$ 代表自由度。(为什么？)

虽然原理上 $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ 可从估计的残差中算出，但实践中可按下述方法直接得到：回顾 $\sum \hat{u}_i^2(=\text{RSS}) = \text{TSS} - \text{ESS}$，在双变量情形中可得：

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \tag{3.3.6}$$

在三变量情形中可得：

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{7.4.19}$$

可以看出，将这一原则加以推广，对于 k 变量模型则有：

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \cdots - \hat{\beta}_k \sum y_i x_{ki} \tag{C.3.15}$$

用矩阵符号表示，

$$\textbf{TSS:} \sum y_i^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.16}$$

$$\textbf{ESS:} \hat{\beta}_2 \sum y_i x_{2i} + \cdots + \hat{\beta}_k \sum y_i x_{ki} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2 \tag{C.3.17}$$

936 其中 $n\bar{Y}^2$ 一项被称为均值校正值(correction for mean)。[6]因此，

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{C.3.18}$$

一旦得到 $\hat{\mathbf{u}}'\hat{\mathbf{u}}$，$\hat{\sigma}^2$ 便容易由(C.3.14)算出。这样就能估计方差—协方差矩阵(C.3.13)。

对于我们的说明性例子，

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= 132\ 100 - [24.454\ 5 \quad 0.509\ 1]\begin{bmatrix}1\ 110 \\ 205\ 500\end{bmatrix} \\ &= 337.373\end{aligned}$$

由此知，$\hat{\sigma}^2 = (337.273/8) = 42.159\ 1$，这和先前在第 3 章中得到的值差不

多一样。

OLS 向量 $\hat{\boldsymbol{\beta}}$ 的性质

我们知道，在二和三变量情形中，OLS 估计量是线性且无偏的，并在所有线性无偏估计量一类中有最小方差(高斯-马尔可夫性质)。简言之，OLS 估计量是最优线性无偏估计量(BLUE)。此性质可推广到整个 $\hat{\boldsymbol{\beta}}$ 向量，就是说，$\hat{\boldsymbol{\beta}}$ 是线性的(其每一元素都是应变量 Y 的线性函数)。$E(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$，即 $\hat{\boldsymbol{\beta}}$ 的每一元素的期望值都等于真 $\boldsymbol{\beta}$ 的相应元素，并在 $\boldsymbol{\beta}$ 的所有线性无偏估计量一类中 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 有最小方差。

证明见于附录 CA，第 CA.4 节。如在引言中所说的，k 变量情形大多是二和三变量情形的直接推广。

§C.4 用矩阵表示的判定系数 R^2

判定系数 R^2 曾被定义为：

$$R^2=\frac{\mathrm{ESS}}{\mathrm{TSS}}$$

在双变量情形中：

$$R^2=\frac{\hat{\beta}_2^2\sum x_i^2}{\sum y_i^2} \tag{3.5.6}$$

在三变量情形中：

$$R^2=\frac{\hat{\beta}_2\sum y_i x_{2i}+\hat{\beta}_3\sum y_i x_{3i}}{\sum y_i^2} \tag{7.5.5}$$

937 推广到 k 变量情形，我们得到：

$$R^2=\frac{\hat{\beta}_2\sum y_i x_{2i}+\hat{\beta}_3\sum y_i x_{3i}+\cdots+\hat{\beta}_k\sum y_i x_{ki}}{\sum y_i^2} \tag{C.4.1}$$

利用(C.3.16)和(C.3.17)，可把(C.4.1)写为：

$$R^2=\frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}-n\bar{Y}^2}{\mathbf{y}'\mathbf{y}-n\bar{Y}^2} \tag{C.4.2}$$

这就是 R^2 的矩阵表现。

对于我们的说明性例子：

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}=[24.357\ 1\quad 0.507\ 9]\begin{bmatrix}1\ 110\\205\ 500\end{bmatrix}=131\ 409.831$$

$$\mathbf{y}'\mathbf{y}=132\ 100$$

及

$$n\bar{Y}^2=123\ 210$$

将这些值代到(C.4.2)中去，即得 $R^2=0.922\ 4$，除进位误差不计外，这和前面得到的结果差不多一样。

§C.5 相关矩阵

在前面几章里，我们遇到零阶相关或简单相关系数 r_{12}，r_{13}，r_{23}和偏相关或1阶相关系数 $r_{12.3}$，$r_{13.2}$，$r_{23.1}$及其相互关系。在 k 变量情形中，一共有 $k(k-1)/2$ 个零阶相关系数。(为什么?)这 $k(k-1)/2$ 个相关系数可排成一个矩阵(方阵)，叫做**相关矩阵 R**，如下所示：

$$\begin{aligned}R&=\begin{bmatrix}r_{11}&r_{12}&r_{13}&\cdots&r_{1k}\\r_{21}&r_{22}&r_{23}&\cdots&r_{2k}\\\hline r_{k1}&r_{k2}&r_{k3}&\cdots&r_{kk}\end{bmatrix}\\&=\begin{bmatrix}1&r_{12}&r_{13}&\cdots&r_{1k}\\r_{21}&1&r_{23}&\cdots&r_{2k}\\\hline r_{k1}&r_{k2}&r_{k3}&\cdots&1\end{bmatrix}\end{aligned}\tag{C.5.1}$$

其中下标1和前面一样表示因变量 Y(如 r_{12}指 Y 与 X_2 的相关，等等)并且利用了一个变量同它自己的相关系数恒为1这一事实($r_{11}=r_{22}=\cdots=r_{kk}=1$)。

938 从相关矩阵 **R** 可求得1阶(见第7章)和高阶(诸如 $r_{12.34\cdots k}$)(见习题C.4)相关系数。许多计算机程序都把计算 **R** 矩阵当作一种常规。我们还将在以后的工作中讨论相关矩阵(参看第10章)。

§C.6 关于个别回归系数的假设检验的矩阵表示

由于在前些章中已经一一申述过的理由，如果我们的目的既是推断又是估计，我们有必要假定干扰项 u_i 遵循某种概率分布。还由于前面已经明确了的理由，在回归分析中，我们经常假定每一 u_i 都遵循零均值和不变方差 σ^2 的正态分布。用矩阵符号表示，我们有：

$$\mathbf{u} \sim N(\mathbf{0},\ \sigma^2\mathbf{I}) \tag{C.6.1}$$

其中 $\mathbf{u}$ 和 $\mathbf{0}$ 都是 $n\times 1$ 列向量，$\mathbf{I}$ 是 $n\times n$ 恒等矩阵，而 $\mathbf{0}$ 是虚拟假设中的零向量。

给定正态性假定，我们知道，在二和三变量线性回归模型中，(1)OLS 估计量 $\hat{\beta}_i$ 和 ML 估计量 $\tilde{\beta}_i$ 相同，但 ML 估计量 $\tilde{\sigma}^2$ 有偏误，虽然这一偏误可通过用无偏 OLS 估计量 $\hat{\sigma}^2$ 而加以消除；以及(2)OLS 估计量 $\hat{\beta}_i$ 也是正态分布的。推广而言，在 k 变量情形中，我们可以证明

$$\hat{\boldsymbol{\beta}} \sim N[\boldsymbol{\beta},\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1}] \tag{C.6.2}$$

就是，$\hat{\boldsymbol{\beta}}$ 的每一元素都是正态分布的，且其均值等于真 $\boldsymbol{\beta}$ 的相应元素，而方差为逆矩阵$(\mathbf{X}'\mathbf{X})^{-1}$的主对角线上相应元素的 σ^2 倍。

由于实际上 σ^2 未知，而估计为 $\hat{\sigma}^2$，就要用到从正态到 t 分布的通常转换，随之，$\hat{\boldsymbol{\beta}}$ 的每一元素就遵循 $n-k$ 个自由度的 t 分布。符号上，

$$t=\frac{\hat{\beta}_i-\beta_i}{\mathrm{se}(\hat{\beta}_i)} \tag{C.6.3}$$

有 $n-k$ 个自由度，其中 $\hat{\beta}_i$ 是 $\hat{\boldsymbol{\beta}}$ 中的任一元素。

因此，t 分布可用来检验关于真 β_i 的假设并建立它的置信区间，其具体操作步骤已在第 5 章和第 8 章中说明。第 C.10 节将给出一个完全解答了的例子。

§C.7 检验回归的总显著性：用矩阵表示的方差分析

939 在第 8 章中，我们曾展示方差分析(ANOVA)的技术，用以(1)检验回归估计的总显著性，即检验全部真(偏)斜率系数同时为零的虚拟假设，以及(2)评价一个解释变量的增量贡献。ANOVA 技术可以容易地推广到 k 变量情形。回想一下，ANOVA 技术是要把 TSS 分解为 ESS 和 RSS。这三个平方和的矩阵表达式已分别由(C.3.17)和(C.3.18)给出。对应于这些平方和的自由度依次是 $n-1$，$k-1$ 和 $n-k$。(为什么?)于是，仿效第 8 章表 8.1，我们能制出表 C.2。

表 C.2　　k 变量线性回归模型的 ANOVA 矩阵形式

变异来源	平方和	自由度	均方和
来自回归(即来自 $X_2, X_3, \cdots, X_k$)	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2$	$k-1$	$\dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{k-1}$
来自残差	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$n-k$	$\dfrac{\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{n-k}$
总计	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2$	$n-1$	

假定干扰项 u_i 是正态分布的，并且虚拟假设是 $\beta_2 = \beta_3 = \cdots = \beta_k = 0$，那么，仿照第 8 章，可以证明：

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2)/(k-1)}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k)} \tag{C.7.1}$$

遵循 $k-1$ 和 $n-k$ 个自由度的 F 分布。

在第 8 章中我们看到，在前述假定下，F 与 R^2 之间有一紧密关系，即：

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{8.5.11}$$

因此，ANOVA 表 C.2 又可表达为表 C.3。和表 C.2 相比，表 C.3 的一个优点是全部分析都能通过 R^2 来做，而无需考虑 F 比中的已被消掉的($\mathbf{y}'\mathbf{y} - n\bar{Y}^2$)。

940

表 C.3　　由 R^2 表示的 k 变量 ANOVA 表的矩阵形式

变异来源	平方和	自由度	均方和
来自回归(即来自 $X_2, X_3, \cdots, X_k$)	$R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$	$k-1$	$\dfrac{R^2(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{k-1}$
来自残差	$(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)$	$n-k$	$\dfrac{(1-R^2)(\mathbf{y}'\mathbf{y} - n\bar{Y}^2)}{n-k}$
总计	$\mathbf{y}'\mathbf{y} - n\bar{Y}^2$	$n-1$	

§C.8　检验线性约束：用矩阵表示的一般 F 检验法

第 8.7 节中我们介绍了怎样用 F 检验去检验施加在 k 变量线性回归模型的一或多个参数上的约束的真实性。适当的检验已由(8.7.9)[或与之等效的(8.7.10)]给出。(8.7.9)矩阵对等式是不难导出的。

令：

$\hat{\mathbf{u}}_{R}$ = 受约束最小二乘回归的残差向量

$\hat{\mathbf{u}}_{UR}$ = 无约束最小二乘回归的残差向量

则：

$\hat{\mathbf{u}}'_{R}\hat{\mathbf{u}}_{R} = \sum \hat{u}_{R}^{2}$ = 受约束回归的 RSS

$\hat{\mathbf{u}}'_{UR}\hat{\mathbf{u}}_{UR} = \sum \hat{u}_{UR}^{2}$ = 无约束回归的 RSS

m = 线性约束的个数

k = 无约束回归中的参数个数（包括截距）

n = 观测次数

于是得(8.7.9)的矩阵对应式：

$$F = \frac{(\hat{\mathbf{u}}'_{R}\hat{\mathbf{u}}_{R} - \hat{\mathbf{u}}'_{UR}\hat{\mathbf{u}}_{UR})/m}{(\hat{\mathbf{u}}'_{UR}\hat{\mathbf{u}}_{UR})/(n-k)} \tag{C.8.1}$$

它遵循以$(m, n-k)$为自由度的 F 分布。按照平常的做法，如果由(C.8.1)算出的 F 值超过临界 F 值，就可拒绝受约束回归；否则不拒绝它。

§C.9 用复回归做预测：矩阵表述

在 8.9 节中，我们用标量符号讨论怎样能通过复回归的估计在给定的回归元 X 值下，预测 Y 的(1)均值和(2)个值。本节中，我们说明怎样用矩阵形式表达这些预测。我们还要介绍估计这些预测值的方差和标准差的公式；
941 在第 8 章中我们表示过最好是用矩阵符号去处理这些公式，因为这些公式的标量表达式会变得相当臃肿。

均值预测

令：
$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ X_{02} \\ X_{03} \\ \vdots \\ X_{0k} \end{bmatrix} \tag{C.9.1}$$

为我们在预测 Y 的平均预测值 $\hat{Y}_0$ 时所要取定的诸 X 变量的值的向量。

用标量形式的复回归估计式：

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + u_i \tag{C.9.2}$$

可简洁地写成矩阵形式：

$$\hat{Y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} \tag{C.9.3}$$

其中 $\mathbf{x}_i' = [1 \quad X_{2i} \quad X_{3i} \cdots X_{ki}]$和

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

方程(C.9.2)或(C.9.3)无疑是 Y_i 的对应于给定 $\mathbf{x}_i'$的平均预测值。

如果 $\mathbf{x}_i'$由(C.9.1)那样地给出，则(C.9.3)变为：

$$(\hat{Y}_i \mid \mathbf{x}_0') = \mathbf{x}_0'\hat{\boldsymbol{\beta}} \tag{C.9.4}$$

其中 $\mathbf{x}_0$ 的值自然是设定的。注意，因 $E(\mathbf{x}_0'\hat{\boldsymbol{\beta}}) = \mathbf{x}_0'\boldsymbol{\beta}$(为什么?)，故(C.9.4)给出 $E(Y_i \mid \mathbf{x}_0')$的一个无偏预测值。

均值预测的方差

估计$(\hat{Y}_0 \mid \mathbf{x}_0')$的方差的公式如下：[7]

$$\operatorname{var}(\hat{Y}_0 \mid \mathbf{x}_0') = \sigma^2\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \tag{C.9.5}$$

其中 σ^2 是 u_i 的方差，$\mathbf{x}_0'$是我们用来对 Y 做预测而给定的诸 X 变量的值。而$(\mathbf{X}'\mathbf{X})$就是(C.3.9)中的矩阵。实际上，我们用 σ^2 的无偏估计量 $\hat{\sigma}^2$ 代替了 σ^2。

我们在下一节将说明均值预测及其方差。

个值预测

942 如果像第 5 章和第 8 章所指的那样，$Y(=Y_0)$的个值预测也由(C.9.3)给出，或更具体地由(C.9.4)给出。均值预测与个值预测的区别在于它们的方差。

个值预测的方差

计算个值预测的方差的公式如下：[8]

$$\operatorname{var}(Y_0 \mid \mathbf{x}_0) = \hat{\sigma}^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0] \tag{C.9.6}$$

其中 $\operatorname{var}(Y_0 \mid \mathbf{x}_0)$代表 $E[Y_0 - \hat{Y}_0 \mid X]^2$。实际上，我们用 σ^2 的无偏估计量 $\hat{\sigma}^2$ 取代了它。我们在下一节将对这个公式加以说明。

§C.10 矩阵方法总结：一个说明性例子

考虑表 C.4 中给出的数据。这些数据涉及人均消费支出(PPCE)与人均可支配收入(PPDI)和时间或趋势变量之间的关系。通过在模型中包含趋势变量，我们试图发现 PPCE 与 PPDI 在扣除时间变量(可由技术和偏好变化等一系列其他因素来表示)影响后的关系。

因此，为了经验研究的目的，取回归模型为：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{C.10.1}$$

其中 Y = 人均消费支出，X_2 = 人均可支配收入，及 X_3 = 时间。表 C.4 给出做回归(C.10.1)所需的数据。

943 可用矩阵符号将我们的问题表述为：

$$\begin{bmatrix} 1\,673 \\ 1\,688 \\ 1\,666 \\ 1\,735 \\ 1\,749 \\ 1\,756 \\ 1\,815 \\ 1\,867 \\ 1\,948 \\ 2\,048 \\ 2\,128 \\ 2\,165 \\ 2\,257 \\ 2\,316 \\ 2\,324 \end{bmatrix} = \begin{bmatrix} 1 & 1\,839 & 1 \\ 1 & 1\,844 & 2 \\ 1 & 1\,831 & 3 \\ 1 & 1\,881 & 4 \\ 1 & 1\,883 & 5 \\ 1 & 1\,910 & 6 \\ 1 & 1\,969 & 7 \\ 1 & 2\,016 & 8 \\ 1 & 2\,126 & 9 \\ 1 & 2\,239 & 10 \\ 1 & 2\,336 & 11 \\ 1 & 2\,404 & 12 \\ 1 & 2\,487 & 13 \\ 1 & 2\,535 & 14 \\ 1 & 2\,595 & 15 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \\ \hat{u}_8 \\ \hat{u}_9 \\ \hat{u}_{10} \\ \hat{u}_{11} \\ \hat{u}_{12} \\ \hat{u}_{13} \\ \hat{u}_{14} \\ \hat{u}_{15} \end{bmatrix} \tag{C.10.2}$$

$$\underset{15\times 1}{\mathbf{y}} = \underset{15\times 3}{\mathbf{X}} \quad \underset{3\times 1}{\hat{\boldsymbol{\beta}}} + \underset{15\times 1}{\hat{\mathbf{u}}}$$

表 C.4　1956—1970 年美国人均消费支出(PPCE)及人均可支配收入(PPDI)，以 1958 年美元计

PPCE, Y	PPDI, X_2	Time, X_3	PPCE, Y	PPDI, X_2	Time, X_3
1 673	1 839	1(=1956)	1 948	2 126	9
1 688	1 844	2	2 048	2 239	10

1 666	1 831	3	2 128	2 336	11
1 735	1 881	4	2 165	2 404	12
1 749	1 883	5	2 257	2 487	13
1 756	1 910	6	2 316	2 535	14
1 815	1 969	7	2 324	2 595	15(=1970)
1 867	2 016	8			

资料来源：*Economic Report of the President*, January 1972, Table B-16.

由以上数据算得以下的一些结果：

$$\bar{Y} = 1\,942.333 \quad \bar{X}_2 = 2\,126.333 \quad \bar{X}_3 = 8.0$$

$$\sum (Y_i - \bar{Y})^2 = 830\,121.333$$

$$\sum (X_{2i} - \bar{X}_2)^2 = 1\,103\,111.333 \qquad \sum (X_{3i} - \bar{X}_3)^2 = 280.0$$

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & 1 \cdots 1 \\ X_{21} & X_{22} & X_{23} \cdots X_{2n} \\ X_{31} & X_{32} & X_{33} \cdots X_{3n} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ 1 & X_{23} & X_{33} \\ \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{2i} & \sum X_{3i} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{2i}X_{3i} & \sum X_{3i}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 15 & 31\,895 & 120 \\ 31\,895 & 68\,922.513 & 272\,144 \\ 120 & 272\,144 & 1\,240 \end{bmatrix} \tag{C.10.3}$$

$$\mathbf{X'y} = \begin{bmatrix} 29\,135 \\ 62\,905\,821 \\ 247\,934 \end{bmatrix} \tag{C.10.4}$$

944 利用附录 B 的矩阵求逆规则，可见：

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 37.232\,491 & -0.022\,508\,2 & 1.336\,707 \\ -0.022\,508\,2 & 0.000\,013\,7 & -0.000\,831\,9 \\ 1.336\,707 & -0.000\,831\,9 & 0.054\,034 \end{bmatrix} \tag{C.10.5}$$

因此，

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 300.286\ 25 \\ 0.741\ 98 \\ 8.043\ 56 \end{bmatrix} \tag{C.10.6}$$

现在可算出残差平方和为：

$$\begin{aligned} \sum \hat{u}_i^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \\ &= 57\ 420\ 003 - [300.286\ 25 \quad 0.741\ 98 \quad 8.043\ 56] \begin{bmatrix} 29\ 135 \\ 62\ 905\ 821 \\ 247\ 934 \end{bmatrix} \\ &= 1\ 976.855\ 74 \end{aligned} \tag{C.10.7}$$

由此得：

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{12} = 164.737\ 97 \tag{C.10.8}$$

因此，可得到 $\hat{\boldsymbol{\beta}}$ 的方差—协方差为：

$$\text{var-cov}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 6\ 133.650 & -3.707\ 94 & 220.206\ 34 \\ -3.707\ 94 & 0.002\ 26 & -0.137\ 05 \\ 220.206\ 34 & -0.137\ 05 & 8.901\ 55 \end{bmatrix} \tag{C.10.9}$$

此矩阵的主对角线元素分别给出 $\hat{\beta}_1$，$\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差而其正的平方根就是相应的标准误。

由上述数据能容易验算：

$$\text{ESS: } \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} - n\bar{Y}^2 = 828\ 144.477\ 86 \tag{C.10.10}$$

$$\text{TSS: } \mathbf{y}'\mathbf{y} - n\bar{Y}^2 = 830\ 121.333 \tag{C.10.11}$$

于是得：

$$\begin{aligned} R^2 &= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{Y}^2}{\mathbf{y}'\mathbf{y} - n\bar{Y}^2} \\ &= \frac{828\ 144.477\ 86}{830\ 121.333} \\ &= 0.997\ 61 \end{aligned} \tag{C.10.12}$$

945 利用(7.8.4)能看到校正判定系数是：

$$R^2 = 0.997\ 22 \tag{C.10.13}$$

收集我们的全部结果即有：

$$\begin{aligned} \hat{Y}_i &= 300.286\ 25 + 0.741\ 98X_{2i} + 8.043\ 56X_{3i} \\ &\quad\ (78.317\ 63) \quad (0.047\ 53) \quad\ (2.983\ 54) \\ t &= \ (3.834\ 21) \quad (15.609\ 56) \quad (2.695\ 98) \\ R^2 &= 0.997\ 61 \qquad \bar{R}^2 = 0.997\ 22 \qquad df = 12 \end{aligned} \tag{C.10.14}$$

对(C.10.14)的解释是：如果 X_2 和 X_3 都固定为零，则估计人均消费支出的均值约为 300 美元。按照常理，要接受这种机械式的对截距的解释，确实有难言之苦。偏回归系数 0.741 98 是说，保持所有其他变量不变，人均收入每增加 1 美元，平均地说，人均消费支出将随之增加约 74 美分。简言之，边际消费倾向估计约为 0.74 或 74%。同理，保持所有其他变量不变，在 1956—1970 年这个研究期间，人均消费支出的均值每年约增加 8 美元。R^2 值 0.997 6 表示两个解释变量在 1956—1970 年期间解释了美国人均消费支出变异的 99% 以上。$\bar{R}^2$ 虽然低下来一点，仍然是很高的。

至于估计系数的统计显著性，从(C.10.14)我们看到每一个估计系数都在(比方说)5% 显著水平上是个别地显著的：估计的系数对它们的标准误的比率(即 t 比率)分别是 3.834 21、15.610 77 和 2.695 98。使用 5% 显著水平的双尾 t 检验，我们查出自由度为 12 的临界 t 值是 2.179。现在每一个计算的 t 值都超出这个临界值，故可个别地拒绝真实的总体系数值为零的虚拟假设。

如前面已指明的，我们不能把平常的 t 检验用在同时检验假设 $\beta_2 = \beta_3 = 0$ 上。因为 t 检验方法假定了每次应用 t 检验时都用了一个独立抽取的样本，如果用同一样本同时检验关于 β_2 和 β_3 的假设，就很可能估计量 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 是相关的，从而违反了 t 检验方法所依据的假定。[19] 事实上，看一看(C.10.9)中的 $\hat{\boldsymbol{\beta}}$ 的方差—协方差矩阵，就知道估计量 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 是负相关的。(两者的协方差是 −0.137 05)。因此我们不能用 t 检验来检验 $\beta_2 = \beta_3 = 0$ 的虚拟假设。

946 然而，像 $\beta_2 = \beta_3 = 0$ 这样的联合虚拟假设，可通过第 8 章介绍的方差分析技术及其伴随的 F 检验加以检验。对于我们的问题的方差分析表，见于表 C.5。在通常的假设下，我们得到：

$$F = \frac{414\ 072.389\ 3}{164.737\ 97} = 2\ 513.52 \tag{C.10.15}$$

表 C.5　　表 C.4 中数据的 ANOVA 表

变异来源	SS	df	MSS
来自 X_2、X_3	828 144.477 86	2	414 072.389 3
来自残差	1 976.855 74	12	164.737 97
总计	830 121.333 60	14	

它遵循以 2 和 12 为自由度的 F 分布。显然 F 的计算值是高度显著的；我们可以拒绝虚拟假设：$\beta_2 = \beta_3 = 0$，即人均消费支出与人均可支配收入以及时间均无线性关系。

在 C.9 节中，我们讨论了均值和个值预报方法的具体步骤。假定 1971

年的 PPDI 数字是 2 610 美元，我们想知道对应于这一数字的 PPCE。那么，1971 年的 PPCE 的均值和个值预报值是相同的，并由下式给出：

$$
\begin{aligned}
&(\mathrm{PPCE}_{1971} \mid \mathrm{PPDI}_{1971},\ X_3 = 16) = \mathbf{x}'_{1971}\hat{\boldsymbol{\beta}} \\
&= [1 \quad 2\,610 \quad 16]\begin{bmatrix} 300.286\,25 \\ 0.741\,98 \\ 8.043\,56 \end{bmatrix} \\
&= 2\,365.55
\end{aligned}
\tag{C.10.16}
$$

这里利用了(C.9.3)。

如第 C.9 节所示可知，$\hat{Y}_{1971}$和 Y_{1971}的方差却不相同。现分述如下：

$$
\begin{aligned}
\operatorname{var}(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) &= \hat{\sigma}^2[\mathbf{x}'_{1971}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{1971}] \\
&= 164.737\,97[1 \quad 2\,610 \quad 16](\mathbf{X}'\mathbf{X})^{-1}\begin{bmatrix} 1 \\ 2\,610 \\ 16 \end{bmatrix}
\end{aligned}
\tag{C.10.17}
$$

其中$(\mathbf{X}'\mathbf{X})^{-1}$已见于(C.10.5)。将它代入(C.10.17)，即可证实：

$$\operatorname{var}(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) = 48.642\,6 \tag{C.10.18}$$

947 并因而：

$$\operatorname{se}(\hat{Y}_{1971} \mid \mathbf{x}'_{1971}) = 6.974\,4$$

我们留给读者[利用(C.9.6)]去证实：

$$\operatorname{var}(Y_{1971} \mid \mathbf{x}'_{1971}) = 213.380\,6 \tag{C.10.19}$$

和

$$\operatorname{sc}(Y_{1971} \mid \mathbf{x}'_{1971}) = 14.607\,6$$

注：$\operatorname{var}(Y_{1971} \mid \mathbf{x}'_{1971}) = E[Y_{1971} - \hat{Y}_{1971} \mid \mathbf{x}'_{1971}]^2$。

在 C.5 节中，我们曾介绍相关矩阵 **R**。对于我们的数据，相关矩阵是：

$$
R = \begin{array}{c} \\ Y \\ X_2 \\ X_3 \end{array}
\begin{array}{c}
\begin{array}{ccc} Y & X_2 & X_3 \end{array} \\
\begin{bmatrix} 1 & 0.998\,0 & 0.974\,3 \\ 0.998\,0 & 1 & 0.966\,4 \\ 0.974\,3 & 0.966\,4 & 1 \end{bmatrix}
\end{array}
\tag{C.10.20}
$$

注意在(C.10.20)中，我们用了模型的变量来为相关矩阵镶边，目的是为了容易辨认在相关系数的计算中所涉及的变量。例如，矩阵(C.10.12)的头一行中的系数 0.998 0 告诉我们它是 Y 与 X_2 之间的相关系数(即 γ_{12})。由相关矩阵(C.10.20)给出的零阶相关容易推出一阶相关系数。(见习题 C.7。)

§C.11 广义最小二乘法

我们曾几次偶尔提到 OLS 是 GLS 的特殊情形。为了看出这一点，回到方程(C.2.2)。为了考虑异方差[(C.2.2)主对角线上的元素]和该差项的自相关[(C.2.2)主对角线外的元素]，假定：

$$E(\mathbf{u}\,\mathbf{u}')=\sigma^2\mathbf{V} \tag{C.11.1}$$

其中 $\mathbf{V}$ 是一个已知的 $n\times n$ 矩阵。

因此，如果我们的模型是：

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$$

其中 $E(\mathbf{u})=0$ 和 var-cov$(\mathbf{u})=\sigma^2\mathbf{V}$。在 σ^2 未知的典型情形下，$\mathbf{V}$ 就表示假定的随机误差 u_t 的方差和协方差结构。

948 在明确的该差项的方差—协方差条件下，可以证明：

$$\boldsymbol{\beta}^{gls}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{C.11.2}$$

$\boldsymbol{\beta}^{gls}$就被称为 $\boldsymbol{\beta}$ 的**广义最小二乘(GLS)估计量**。

还可以看出

$$\text{var-cov}(\boldsymbol{\beta}^{gls})=\sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \tag{C.11.3}$$

可以证明，$\boldsymbol{\beta}^{gls}$是 $\boldsymbol{\beta}$ 的最优线性无偏估计量。

如果假定每个该差项的方差都是 σ^2，而且该差项彼此不相关，那么 $\mathbf{V}$ 矩阵就简化成(C.2.3)所示的单位矩阵。如果该差项彼此不相关但具有不同的方差(即异方差)，那么 $\mathbf{V}$ 矩阵就是主对角线元素具有不等方差的对角阵。当然，如果既存在异方差又存在自相关，那么 $\mathbf{V}$ 矩阵的主对角线上和主对角线外都有元素。

实践中真实的问题在于，我们并不知道 σ^2 和真实的方差—协方差(即 $\mathbf{V}$ 矩阵的结构)。作为一种解决办法，我们可以使用**估计(或可行)的广义最小二乘法**(EGLS)(estimated or feasible generalized least squares)。我们在此先不考虑异方差和/或自相关的问题，而直接用 OLS 估计我们的模型。我们从这个模型得到残差，并通过用估计的 u，即 $\hat{u}$ 取代刚才(C.2.2)表达式中的项以得到误差项的(估计)方差一协方差矩阵。可以证明，EGLS 估计量是 GLS 估计量的一致估计量。用符号表示：

$$\boldsymbol{\beta}^{egls}=(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}) \tag{C.11.4}$$

$$\text{var-cov}(\boldsymbol{\beta}^{egls})=\sigma^2(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \tag{C.11.5}$$

其中 $\hat{\mathbf{V}}$ 是 $\mathbf{V}$ 的一个估计值。

§C.12 要点与结论

本章的主要目的是介绍经典线性回归模型的矩阵方法。虽然并未涉及多少回归分析的新概念，但矩阵符号却为处理任意多个变量的线性回归模型提供了一种简洁的方法。

作为本章的收尾，我们提请注意，如果 Y 和 X 变量都以离差(指对样本均值的离差)形式测算，则在上述公式中有少数的变更。现将这些变更列成

949 表 C.6。[10]如该表所表明的，在离差形式中，均值校正项 $n\bar{Y}^2$ 将从 TSS 和 ESS 中消失。(为什么?)这一消失造成 R^2 计算公式中的一个变化。除此以外，大多数按原来测算单位导出的公式，对离差形式说，仍然是对的。

表 C.6　原始单位的和离差形式的 k 变量回归模型*

原始单位		离差形式	
$\mathbf{y}=\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\mathbf{u}}$	(C.3.2)	$\mathbf{y}=\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\mathbf{u}}$ 全为 1 的列向量已从 $\mathbf{X}$ 矩阵中消失(为什么?)	
$\hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$	(C.3.11)	相同	
$\text{var-cov}(\hat{\boldsymbol{\beta}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$	(C.3.13)	相同	
$\hat{\mathbf{u}}'\hat{\mathbf{u}}=\mathbf{y}'\mathbf{y}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	(C.3.18)	相同	
$\sum y_i^2=\mathbf{y}'\mathbf{y}-n\bar{Y}^2$	(C.3.16)	$\sum y_i^2=\mathbf{y}'\mathbf{y}$	(C.12.1)
$\text{ESS}=\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}-n\bar{Y}^2$	(C.3.17)	$\text{ESS}=\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	(C.12.2)
$R^2=\dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}-n\bar{Y}^2}{\mathbf{y}'\mathbf{y}-n\bar{Y}^2}$	(C.4.2)	$R^2=\dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}$	(C.12.3)

*注意，虽然在两种情形中矩阵和向量的符号都是一样的，但在离差形式中矩阵和向量中的元素都假定是离差而非原始数据。还要注意，离差形式中的 $\hat{\boldsymbol{\beta}}$ 的阶是 $k-1$，而 var-cov($\hat{\boldsymbol{\beta}}$)的阶是$(k-1)(k-1)$。

习　题

C.1　在 C.10 节的说明性例子中，按离差形式的数据算得 $\mathbf{X}'\mathbf{X}$ 和 $\mathbf{X}'\mathbf{y}$ 如下：

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}1\,103\,111.333 & 16\,984\\ 16\,984 & 280\end{bmatrix}$$

$$\mathbf{X}'\mathbf{y}=\begin{bmatrix}955\ 099.333\\ 14\ 854.000\end{bmatrix}$$

a. 估计 β_2 和 β_3。

b. 你将怎样估计 β_1?

c. 估计 $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 的方差和它们的协方差。

d. 求出 R^2 和 $\bar{R}^2$。

e. 将你的结果同 C.10 节所给出的相比，你发现离差形式有什么好处?

C.2 参考习题 22.23。利用那里给出的数据，建立适当的($\mathbf{X}'\mathbf{X}$)矩阵的($\mathbf{X}'\mathbf{y}$)向量，并估计参数向量 $\boldsymbol{\beta}$ 及其方差—协方差矩阵。同样得到
950
R^2。你将如何检验如下虚拟假设：M1 对 GDP 和利率 R 的弹性在数值上相等?

C.3 **检验两个回归系数是否相等。**假如给定如下的回归模型：

$$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i$$

并且要检验假设 $\beta_2=\beta_3$。如果 u_i 是常态分布的。则可证明：

$$t=\frac{\hat{\beta}_2-\hat{\beta}_3}{\sqrt{\operatorname{var}(\hat{\beta}_2)+\operatorname{var}(\hat{\beta}_3)-2\operatorname{cov}(\hat{\beta}_2,\ \hat{\beta}_3)}}$$

遵循自由度为 $n-3$ 的 t 分布(参看 8.6 节)。(一般地说，对 k 变量情形，自由度为 $n-k$。)从而上述 t 检验可用于检验虚拟假设：$\beta_2=\beta_3$。

应用上述 t 检验去检验假设：回归(C.10.4)中的 β_2 和 β_3 有相同的真值。

提示：用(C.10.9)所给 $\boldsymbol{\beta}$ 的方差协方差矩阵(简称 $\boldsymbol{\beta}$ 的协差阵——译者注)。

C.4 **用低阶相关表达高阶相关。**p 阶相关系数可通过下述**降阶公式**(reduction formula)用 $p-1$ 阶相关系数来表达：

$$r_{12.345\cdots p}=\frac{r_{12.345\cdots(p-1)}-[r_{1p.345\cdots(p-1)}r_{2p.345\cdots(p-1)}]}{\sqrt{[1-r^2_{1p.345\cdots(p-1)}]}\sqrt{[1-r^2_{2p.345\cdots(p-1)}]}}$$

因此，
$$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r^2_{13}}\sqrt{1-r^2_{23}}}$$

已见于第 7 章。

现给定以下的相关矩阵：

$$\mathbf{R}=\begin{array}{c}\\ Y\\ X_2\\ X_3\\ X_4\\ X_5\end{array}\begin{array}{c}\begin{array}{ccccc}Y & X_2 & X_3 & X_4 & X_5\end{array}\\ \begin{bmatrix}1 & 0.44 & -0.34 & -0.31 & -0.14\\ & 1 & 0.25 & -0.19 & -0.35\\ & & 1 & 0.44 & 0.33\\ & & & 1 & 0.85\\ & & & & 1\end{bmatrix}\end{array}$$

求以下各阶相关系数：

a. $r_{12.345}$ b. $r_{12.34}$ c. $r_{12.3}$

d. $r_{13.245}$ e. $r_{13.24}$ f. $r_{13.2}$

951 **C.5 用低阶回归系数表达高阶回归系数。** p 阶回归系数可通过下述降阶公式用 $p-1$ 阶回归系数来表达：

$$\hat{\beta}_{12.345\cdots p}=\frac{\hat{\beta}_{12.345\cdots(p-1)}-\left[\hat{\beta}_{1p.345\cdots(p-1)}\hat{\beta}_{p2.345\cdots(p-1)}\right]}{1-\hat{\beta}_{2p.345\cdots(p-1)}\hat{\beta}_{p2.345\cdots(p-1)}}$$

因而，

$$\hat{\beta}_{12.3}=\frac{\hat{\beta}_{12}-\hat{\beta}_{13}\hat{\beta}_{32}}{1-\hat{\beta}_{23}\hat{\beta}_{32}}$$

其中 $\beta_{12.3}$是 Y 在保持 X_3 不变的情况下对 X_2 的回归中的斜率系数。类似地，$\beta_{12.34}$是 Y 在保持 X_3 和 X_4 不变的情况下对 X_2 回归中的斜率系数，其余类推。

利用上述公式，求出用较低阶回归系数表达的如下回归系数的表达式：$\hat{\beta}_{12.3456}$，$\hat{\beta}_{12.345}$和 $\hat{\beta}_{12.34}$。

C.6 建立以下恒等式：

$$\hat{\beta}_{12.3}\hat{\beta}_{23.1}\hat{\beta}_{31.2}=r_{12.3}r_{23.1}r_{31.2}$$

C.7 对(C.10.20)所给的相关矩阵 **R**，求出所有的一阶偏相关系数。

C.8 在研究美国某些大城市的犯罪率的变异时奥格本(Ogburn)获得以下数据[1]：

$\bar{Y}=19.9$　$S_1=7.9$

$\bar{X}_2=49.2$　$S_2=1.3$

$\bar{X}_3=10.2$　$S_3=4.6$

$\bar{X}_4=481.4$　$S_4=74.4$

$\bar{X}_5=41.6$　$S_5=10.8$

$$\mathbf{R}=\begin{array}{c} \\ Y \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array}\begin{array}{c} \begin{array}{ccccc} X & X_2 & X_3 & X_4 & X_5 \end{array} \\ \begin{bmatrix} 1 & 0.44 & -0.34 & -0.31 & -0.14 \\ & 1 & 0.25 & -0.19 & -0.35 \\ & & 1 & 0.44 & 0.33 \\ & & & 1 & 0.85 \\ & & & & 1 \end{bmatrix} \end{array}$$

其中 Y = 犯罪率，每千人的已知犯法次数

X_2 = 男性居民所占百分比

X_3 = 全部外来男性居民所占百分比

X_4 = 年龄在 15～44 岁的每千个已婚妇女中拥有的 5 岁以下的儿童数

X_5 = 13 岁以上的每百人中教会会员数。S_1 至 S_5 为变量 Y，X_2，…，X_5 的样本标准差，而 **R** 为相关矩阵

a. 视 Y 为因变量，求 Y 对 4 个 X 变量的回归并解释所估计的回归。

952 b. 求 $r_{12.3}$，$r_{14.35}$和 $r_{15.34}$。

c. 求 R^2 并检验假设：全部偏斜率系数同时等于零。

C.9 下表给出短期内某商品的产出与总生产成本数据。(参看例 7.4。)

产出	总成本,美元
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

为了检验以上数据是否适合于短期内看到的 U 形平均和边际成本曲线这种典型情形,不妨利用以下的模型:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i$$

其中 Y = 总成本及 X = 产出。另加的解释变量 X^2 和 X^3 是从 X 衍生的。

a. 把数据表达为离差形式,然后求($\mathbf{X'X}$),($\mathbf{X'y}$)和$(\mathbf{X'X})^{-1}$。

b. 估计 β_2, β_3 和 β_4。

c. 估计 $\hat{\boldsymbol{\beta}}$ 的协差阵(即方差协方差矩阵)。

d. 估计 β_1。根据题意解释 $\hat{\beta}_1$。

e. 求 R^2 和 $\bar{R}^2$。

f. 先验地,β_2, β_3 和 β_4 的符号为何?为什么?

g. 从前面给的总成本函数,求边际和平均成本函数的表达式。

h. 对数据拟合平均和边际成本函数,并评论拟合的结果。

i. 如果 $\beta_3 = \beta_4 = 0$,边际成本函数将有什么性质?又怎样去检验假设:$\beta_3 = \beta_4 = 0$?

j. 你怎样从所给数据推导出总可变成本和平均可变成本函数?

C.10 为了研究城市贫困家庭(1969 年劳动收入少于 3 943 美元的家庭)的劳动参与人数,我们从 1970 年人口普查取得附表中的数据。

a. 用回归模型 $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$ 求回归系数的估计值,并解释你的结果。

b. 先验地,以上模型中的回归系数有什么预期的符号,并且为什么?

c. 你怎样检验在附表给出的普查区道(census tracts)上总失业率

对城市贫困户的劳动参与人数无影响的假设？

d. 有什么变量应从上述模型中剔除的吗？为什么？

e. 你认为有什么其他的变量可以放进模型中去？

表 C.7　　城市贫民劳动参与经验：1970 年纽约市普查区道

区道号	%劳动人数 Y^*	平均家庭收入 X_2^*	平均家庭人口 X_3	失业率 X_4^{**}
137	64.3	1 998	2.95	4.4
139	45.4	1 114	3.40	3.4
141	26.6	1 942	3.72	1.1
142	87.5	1 998	4.43	3.1
143	71.3	2 026	3.82	7.7
145	82.4	1 853	3.90	5.0
147	26.3	1 666	3.32	6.2
149	61.6	1 434	3.80	5.4
151	52.9	1 513	3.49	12.2
153	64.7	2 008	3.85	4.8
155	64.9	1 704	4.69	2.9
157	70.5	1 525	3.89	4.8
159	87.2	1 842	3.53	3.9
161	81.2	1 735	4.96	7.2
163	67.9	1 639	3.68	3.6

* Y = 限于户主年龄在 65 岁以下

* X_2 = 美元

** X_4 = 民间(非军事)劳动者失业百分数

资料来源：Census Tracts: New York, Bureau of the Census, U.S.Department of Commerce, 1970.

C.11　在一项柯布-道格拉斯生产函数的应用中，得到了如下的结果：

$$\widehat{\ln Y_i} = 2.354\,2 + 0.957\,6\ln X_{2i} + 0.824\,2\ln X_{3i}$$
$$(0.302\,2) \qquad (0.357\,1)$$
$$R^2 = 0.843\,2 \qquad \mathrm{df} = 12$$

其中 Y = 产出，X_2 = 劳动投入，及 X_3 = 资本投入。括号中的数字是估计的标准误。

a. 如第 7 章所指出的，在上述方程中，劳动和资本投入的系数表示产出对劳力和资本的弹性。检验这些弹性个别地等于 1 的假设。

b. 分别假定估计的劳力和资本系数的协方差(i)是零;(ii)是 -0.0972。检验劳力和资本两弹性相同的假设。

c. 你怎样检验上述回归方程的总显著性?

*C.12 用矩阵符号表示 k 变量回归模型的似然函数，并说明最大似然估计量的向量 $\tilde{\boldsymbol{\beta}}$ 和 k 变量回归模型的 OLS 估计量的向量 $\hat{\boldsymbol{\beta}}$ 是一样的。

C.13 **用标准化变量做回归。**考虑以下样本回归函数(SRF)：

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{1}$$

$$Y_i^* = b_1 + b_2 X_{2i}^* + b_3 X_{3i}^* + \hat{u}_i^* \tag{2}$$

其中

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_Y}$$

$$X_{2i}^* = \frac{X_{2i} - \bar{X}_2}{s_2}$$

$$X_{3i}^* = \frac{X_{3i} - \bar{X}_3}{s_3}$$

以上的诸 S 均指样本标准差。如第6章习题6.7所表明的，上面带有星号的变量均指标准化了的变量。这些变量都有零均值和单位标准差。把所有的变量都表达成离差形式，然后说明对于模型(2)有如下关系：

a. $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & r_{23} \\ r_{23} & 1 \end{bmatrix} n$

b. $\mathbf{X}'\mathbf{y} = \begin{bmatrix} r_{12} \\ r_{13} \end{bmatrix} n$

c. $(\mathbf{X}'\mathbf{X})^{-1} = \dfrac{1}{n(1-r_{23}^2)} \begin{bmatrix} 1 & -r_{23} \\ -r_{23} & 1 \end{bmatrix}$

d. $\hat{\boldsymbol{\beta}} = \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \dfrac{1}{(1-r_{23}^2)} \begin{bmatrix} r_{12} - r_{23} r_{13} \\ r_{13} - r_{23} r_{12} \end{bmatrix}$

e. $b_1 = 0$

并建立诸 b 与诸 $\hat{\beta}$ 之间的关系。

(注意，以上关系式中的 n 指样本大小；r_{12}，r_{13}和 r_{23}分别指 Y 与 X_2，Y 与 X_3 以及 X_2 与 X_3 之间的相关。)

C.14 证实方程(C.10.18)和(C.10.19)。

*C.15 **受约束最小二乘。**假定：

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1}$$

---

* 选做题。

我们要在下面的一组等式约束条件下估计上述模型：

$$\mathbf{R\beta} = \mathbf{r} \tag{2}$$

955 其中 $\mathbf{R}$ 为 qxk 阶已知矩阵（$q \leqslant k$），而 $\mathbf{r}$ 为 q 个元素的一已知向量。为说明起见，假使我们的模型是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i \tag{3}$$

并且要在约束条件：

$$\begin{aligned} &\beta_2 - \beta_3 = 0 \\ &\beta_4 + \beta_5 = 1 \end{aligned} \tag{4}$$

下估计此模型。我们可以利用第 8 章讲过的一些技术把这些约束融入模型中（例如，取 $\beta_2 = \beta_3$ 和 $\beta_4 = 1 - \beta_5$，就可把 β_2 和 β_4 从模型中消掉），然后用当时讲的 F 检验来检验这些约束的有效性。但是，有一个把约束条件(4)容纳在估计(3)的估计过程中的更为直接的方法，就是先把约束表达成(2)的形式，也就是在本例中将约束条件写成：

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5}$$

令 $\boldsymbol{\beta}^*$ 表示受约束或受限制最小二乘估计量，可以证明 $\boldsymbol{\beta}^*$ 可由如下公式估计之：[2]

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}}) \tag{6}$$

其中 $\hat{\boldsymbol{\beta}}$ 是通常的（无约束的）估计量，即由通常的公式 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ 来估计。

a.(3)中的 β 向量是什么？

b. 验证对给定的 β 向量，(5)中的 $\mathbf{R}$ 矩阵和 $\mathbf{r}$ 向量确实包含(4)中的两个约束。

c. 对以下的几种情形写出相应的 $\mathbf{R}$ 和 $\mathbf{r}$：

(i) $\beta_2 = \beta_3 = \beta_4 = 2$

(ii) $\beta_2 = \beta_3$ and $\beta_4 = \beta_5$

(iii) $\beta_2 - 3\beta_3 = 5\beta_4$

(iv) $\beta_2 + 3\beta_3 = 0$

d. 什么时候 $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$？

【习题注释】

[1] W.F.Ogburn, "Factors in the Variation of Crime among Cities," *Journal of American Statistical Association*, vol.30, 1935, p.12.

[2] 见 J.Johnston，前引著作，第 205 页。

附录 CA

CA.1 k 个正规或联立方程的推导

将

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

956 对 $\hat{\beta}_1$, $\hat{\beta}_2$, …, $\hat{\beta}_k$ 求偏微分得：

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-1)$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})(-X_{2i})$$

……………………………………………………………………

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_k} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{ki} - \cdots - \hat{\beta}_k X_{ki})(-X_{ki})$$

置这些偏导数为零，整理后即得（C.3.8）中的 k 个正规方程。

CA.2 正规方程的矩阵推导

由（C.3.7）得：

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

利用附录 B 给出的矩阵微分法可得：

$$\frac{\partial(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

置上式等于零便有：

$$(\mathbf{X}'\mathbf{X})\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

从而 $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$，如果逆矩阵存在的话。

CA.3 $\hat{\boldsymbol{\beta}}$ 的方差—协方差矩阵

由（C.3.11）得：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

将 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ 代入上式给出：

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (1)\end{aligned}$$

因此，

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (2)$$

957 按定义，

$$\begin{aligned}\text{var-cov}(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] \\ &= E\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \qquad (3)\end{aligned}$$

这里的最后一步用到了转置规则 $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$。

注意到诸 X 是非随机的，对（3）取期望值就有：

$$\begin{aligned}\text{var-cov}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

这就是（C.3.13）中所给的结果。注意在上述结果的演算中我们利用了 $E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$ 的假定。

CA.4 OLS 估计量的 BLUE 性质

由（C.3.11）可得：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (1)$$

因 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ 是一固定数矩阵，故 $\hat{\boldsymbol{\beta}}$ 是 Y 的线性函数。从而按定义它是一个线性估计量。

回顾 PRF 是：

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad (2)$$

将此代入（1），得：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \qquad (3)$$

$$= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \qquad (4)$$

这是因为 $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$。

取（4）的期望值，得：

$$E(\hat{\boldsymbol{\beta}}) = E(\boldsymbol{\beta}) + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}) = \boldsymbol{\beta} \qquad (5)$$

这是因为 $E(\boldsymbol{\beta}) = \boldsymbol{\beta}$（为什么？）并且按假定 $E(\mathbf{u}) = 0$。从而说明 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的一无

偏估计量。

958 令 $\hat{\boldsymbol{\beta}}^*$ 为 $\boldsymbol{\beta}$ 的任意其他线性估计量，可以把它写为：

$$\hat{\boldsymbol{\beta}}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}]\mathbf{y} \tag{6}$$

其中 C 是一常数矩阵。

将 (2) 代入 (6) 中的 y，可得：

$$\begin{aligned}\hat{\boldsymbol{\beta}}^* &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{C}](\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= \boldsymbol{\beta} + \mathbf{CX}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}\end{aligned} \tag{7}$$

现在，如果要求 $\hat{\boldsymbol{\beta}}^*$ 是 $\boldsymbol{\beta}$ 的一个无偏估计量，则必须有：

$$\mathbf{CX} = 0 \qquad (为什么?) \tag{8}$$

利用 (8)，就可把 (7) 写为：

$$\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu} \tag{9}$$

按定义，var-cov ($\hat{\boldsymbol{\beta}}^*$) 是：

$$E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})' = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} + \mathbf{Cu}]' \tag{10}$$

利用矩阵的求逆和转置性质并经代数简化，我们得到：

$$\begin{aligned}\text{var-cov}(\hat{\boldsymbol{\beta}}^*) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{CC}' \\ &= \text{var-cov}(\hat{\boldsymbol{\beta}}) + \sigma^2\mathbf{CC}'\end{aligned} \tag{11}$$

这表明另一无偏线性估计量 $\hat{\boldsymbol{\beta}}^*$ 的方差－协方差矩阵等于 OLS 估计量 $\hat{\boldsymbol{\beta}}$ 的方差协方差矩阵加 σ^2 倍的 CC′，后者是一半正定矩阵。由此知 $\hat{\boldsymbol{\beta}}^*$ 的一个给定元素的方差必然等于或大于 $\hat{\boldsymbol{\beta}}$ 的相应元素的方差，从而说明 $\hat{\boldsymbol{\beta}}$ 是 BLUE。当然，如果 C 是一零矩阵，即 C = 0，则 $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}$。但这不外是用另一种方式说：如果我们找到了一个 BLUE 估计量，那么它必然是最小二乘估计量。

【注释】

[1] 不熟悉矩阵代数的读者，在往下阅读之前应先复习附录 B。该附录提供了阅读本章所需的矩阵代数的要义。

[2] 仿效附录 B 中所用的符号。我们将用小写粗体字母表示向量，大写粗体字母表示矩阵。

[3] 对 k 变量情形的证明，参考第 4 章章末注中所给的参考文献。

[4] 按定义，u_i 的方差 $= E[u_i - E(u_i)]^2$ 并且 u_i 和 u_j 之间的协方差 $= E[u_i - E(u_i)][u_j - E(u_j)]$，但因对每个 i 假定了 $E(u_i) = 0$，故有方差—协方差矩阵(C.2.3)。

[5] 这些方程能容易记住。从方程 $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$ 开始，将此方程对 n 个 i 值求和，即得 (C.3.8) 的第一个方程；将它遍乘以 X_2 后再对 n 个值求和，即得第二个方程，将它遍乘以 X_3 后再求和，即得第三个方程，依此类推。顺便指出，(C.3.8) 的第一个方程立即给出 $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \cdots - \hat{\beta}_k\bar{X}_k$ [比较 (7.4.6)]。

[6] 注：$\sum y_i^2 = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 = \mathbf{y}'\mathbf{y} - n\bar{Y}^2$。因此，若无校正项，$\mathbf{y}'\mathbf{y}$ 将只给出粗平方和，而不是离差平方和。

[7] 公式推导见 J. Johnston, *Econometrics Methods*, McGraw-Hill, 3d ed., New York, 1984, pp.195－196。

[8] 同上。

[9] 详见第 8.5 节。

[10] 在这些高速计算机的日子里，也许用不着离差形式，但是当我们使用台式计算器并且处理大数时，离差形式将使公式以致计算得以简化。

附录 D 统计学用表

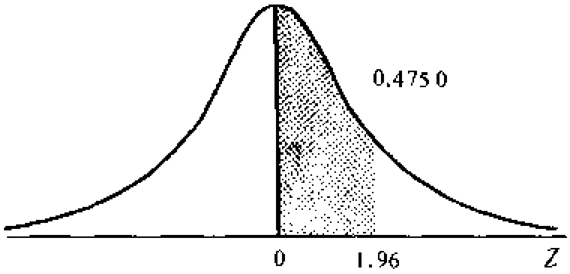
959

表 D.1　标准化正态分布下的面积

例

Pr（0≤Z≤1.96）=0.475 0

Pr（Z≥1.96）=0.5−0.475 0=0.025



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000 0	0.004 0	0.008 0	0.012 0	0.016 0	0.019 9	0.023 9	0.027 9	0.031 9	0.035 9
0.1	0.039 8	0.043 8	0.047 8	0.051 7	0.055 7	0.059 6	0.063 6	0.067 5	0.071 4	0.075 3
0.2	0.079 3	0.083 2	0.087 1	0.091 0	0.094 8	0.098 7	0.102 6	0.106 4	0.110 3	0.114 1
0.3	0.117 9	0.121 7	0.125 5	0.129 3	0.133 1	0.136 8	0.140 6	0.144 3	0.148 0	0.151 7
0.4	0.155 4	0.159 1	0.162 8	0.166 4	0.170 0	0.173 6	0.177 2	0.180 8	0.184 4	0.187 9
0.5	0.191 5	0.195 0	0.198 5	0.201 9	0.205 4	0.208 8	0.212 3	0.215 7	0.219 0	0.222 4
0.6	0.225 7	0.229 1	0.232 4	0.235 7	0.238 9	0.242 2	0.245 4	0.248 6	0.251 7	0.254 9
0.7	0.258 0	0.261 1	0.264 2	0.267 3	0.270 4	0.273 4	0.276 4	0.279 4	0.282 3	0.285 2
0.8	0.288 1	0.291 0	0.293 9	0.296 7	0.299 5	0.302 3	0.305 1	0.307 8	0.310 6	0.313 3
0.9	0.315 9	0.318 6	0.321 2	0.323 8	0.326 4	0.328 9	0.331 5	0.334 0	0.336 5	0.338 9
1.0	0.341 3	0.343 8	0.346 1	0.348 5	0.350 8	0.353 1	0.355 4	0.357 7	0.359 9	0.362 1
1.1	0.364 3	0.366 5	0.368 6	0.370 8	0.372 9	0.374 9	0.377 0	0.379 0	0.381 0	0.383 0
1.2	0.384 9	0.386 9	0.388 8	0.390 7	0.392 5	0.394 4	0.396 2	0.398 0	0.399 7	0.401 5
1.3	0.403 2	0.404 9	0.406 6	0.408 2	0.409 9	0.411 5	0.413 1	0.414 7	0.416 2	0.417 7
1.4	0.419 2	0.420 7	0.422 2	0.423 6	0.425 1	0.426 5	0.427 9	0.429 2	0.430 6	0.431 9
1.5	0.433 2	0.434 5	0.435 7	0.437 0	0.438 2	0.439 4	0.440 6	0.441 8	0.442 9	0.444 1
1.6	0.445 2	0.446 3	0.447 4	0.448 4	0.449 5	0.450 5	0.451 5	0.452 5	0.453 5	0.454 5
1.7	0.445 4	0.456 4	0.457 3	0.458 2	0.459 1	0.459 9	0.460 8	0.461 6	0.462 5	0.463 3
1.8	0.464 1	0.464 9	0.465 6	0.466 4	0.467 1	0.467 8	0.468 6	0.469 3	0.469 9	0.470 6
1.9	0.471 3	0.471 9	0.472 6	0.473 2	0.473 8	0.474 4	0.475 0	0.475 6	0.476 1	0.476 7
2.0	0.477 2	0.477 8	0.478 3	0.478 8	0.479 3	0.479 8	0.480 3	0.480 8	0.481 2	0.481 7
2.1	0.482 1	0.482 6	0.483 0	0.483 4	0.483 8	0.484 2	0.484 6	0.485 0	0.485 4	0.485 7
2.2	0.486 1	0.486 4	0.486 8	0.487 1	0.487 5	0.487 8	0.488 1	0.488 4	0.488 7	0.489 0
2.3	0.489 3	0.489 6	0.489 8	0.490 1	0.490 4	0.490 6	0.490 9	0.491 1	0.491 3	0.491 6
2.4	0.491 8	0.492 0	0.492 2	0.492 5	0.492 7	0.492 9	0.493 1	0.493 2	0.493 4	0.493 6
2.5	0.493 8	0.494 0	0.494 1	0.494 3	0.494 5	0.494 6	0.494 8	0.494 9	0.495 1	0.495 2
2.6	0.495 3	0.495 5	0.495 6	0.495 7	0.495 9	0.496 0	0.496 1	0.496 2	0.496 3	0.496 4
2.7	0.496 5	0.496 6	0.496 7	0.496 8	0.496 9	0.497 0	0.497 1	0.497 2	0.497 3	0.497 4
2.8	0.497 4	0.497 5	0.497 6	0.497 7	0.497 7	0.497 8	0.497 9	0.497 9	0.498 0	0.498 1
2.9	0.498 1	0.498 2	0.498 2	0.498 3	0.498 4	0.498 4	0.498 5	0.498 5	0.498 6	0.498 6
3.0	0.498 7	0.498 7	0.498 7	0.498 8	0.498 8	0.498 9	0.498 9	0.498 9	0.499 0	0.499 0

注：本表给出该分布的右侧（即 Z≥0）面积。由于正态分布是围绕 Z=0 而对称分布的，故左侧面积与相应的右侧面积相同。例如，$P(-1.96\leqslant Z\leqslant 0)=0.475\ 0$。因此，$P(-1.96\leqslant Z\leqslant 1.96)=2(0.475\ 0)=0.95$。

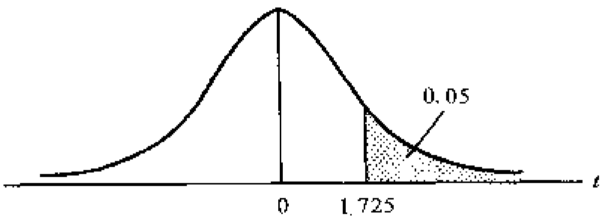
表 D. 2　　　t 分布的百分点

例

$\Pr(t>2.086)=0.025$

$\Pr(t>1.725)=0.05$　自由度＝20

$\Pr(|t|>1.725)=0.10$



df \ Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.05	0.01 0.02	0.005 0.010	0.001 0.002
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	1.638	2.353	3.182	4.541	5.841	10.214
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

注：每列顶头的较小概率指单侧面积；较大的概率则指双侧面积。

资料来源：取自 E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 12, Cambridge University Press, New York, 1966。生物统计学报编辑部与董事会准予复制。

表 D.3　　F 分布的上端百分点

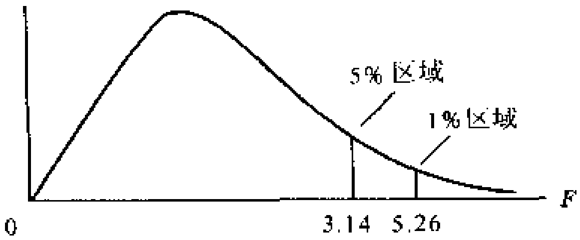
例

$\Pr(F>1.59)=0.25$

$\Pr(F>2.42)=0.10$ 对于自由度 $N_1=10$

$\Pr(F>3.14)=0.05$ 　和 $N_2=9$

$\Pr(F>5.26)=0.01$



分母自由度 N_2	分子自由度 N_1												
	Pr	1	2	3	4	5	6	7	8	9	10	11	12
1	0.25	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.36	9.41
	0.10	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
	0.05	161	200	216	225	230	234	237	239	241	242	243	244
2	0.25	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.39
	0.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
	0.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
	0.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	0.25	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.45
	0.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
	0.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
	0.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1
4	0.25	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	0.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
	0.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4
5	0.25	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89
	0.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27
	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
	0.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89
6	0.25	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.77
	0.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
	0.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	0.25	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.69	1.68
	0.10	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
	0.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
	0.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47
8	0.25	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.63	1.62
	0.10	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
	0.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
9	0.25	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
	0.10	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
	0.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11

资料来源：取自 E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for statisticians*, vol. 1, 3d ed., table 18, Cambridge University Press, New York, 1966。生物统计学报编辑部与董事会准予复制。

续前表

分子自由度 N_1													分母自由度 N_2
15	20	24	30	40	50	60	100	120	200	500	∞	Pr	
9.49	9.58	9.63	9.67	9.71	9.74	9.76	9.78	9.80	9.82	9.84	9.85	0.25	
61.2	61.7	62.0	62.3	62.5	62.7	62.8	63.0	63.1	63.2	63.3	63.3	0.10	1
246	248	249	250	251	252	252	253	253	254	254	254	0.05	
3.41	3.43	3.43	3.44	3.45	3.45	3.46	3.47	3.47	3.48	3.48	3.48	0.25	
9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.48	9.49	9.49	9.49	0.10	2
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	0.05	
99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	0.01	
2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	0.25	
5.20	5.18	5.18	5.17	5.16	5.15	5.15	5.14	5.14	5.14	5.14	5.13	0.10	3
8.70	8.66	8.64	8.62	8.59	8.58	8.57	8.55	8.55	8.54	8.53	8.53	0.05	
26.9	26.7	26.6	26.5	26.4	26.4	26.3	26.2	26.2	26.2	26.1	26.1	0.01	
2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	0.25	
3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.78	3.77	3.76	3.76	0.10	4
5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.66	5.65	5.64	5.63	0.05	
14.2	14.0	13.9	13.8	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	0.01	
1.89	1.88	1.88	1.88	1.88	1.88	1.87	1.87	1.87	1.87	1.87	1.87	0.25	
3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11	3.10	0.10	5
4.62	4.56	4.53	4.50	4.46	4.44	4.43	4.41	4.40	4.39	4.37	4.36	0.05	
9.72	9.55	9.47	9.38	9.29	9.24	9.20	9.13	9.11	9.08	9.04	9.02	0.01	
1.76	1.76	1.75	1.75	1.75	1.75	1.74	1.74	1.74	1.74	1.74	1.74	0.25	
2.87	2.84	2.82	2.80	2.78	2.77	2.76	2.75	2.74	2.73	2.73	2.72	0.10	6
3.94	3.87	3.84	3.81	3.77	3.75	3.74	3.71	3.70	3.69	3.68	3.67	0.05	
7.56	7.40	7.31	7.23	7.14	7.09	7.06	6.99	6.97	6.93	6.90	6.88	0.01	
1.68	1.67	1.67	1.66	1.66	1.66	1.65	1.65	1.65	1.65	1.65	1.65	0.25	
2.63	2.59	2.58	2.56	2.54	2.52	2.51	2.50	2.49	2.48	2.48	2.47	0.10	7
3.51	3.44	3.41	3.38	3.34	3.32	3.30	3.27	3.27	3.25	3.24	3.23	0.05	
6.31	6.16	6.07	5.99	5.91	5.86	5.82	5.75	5.74	5.70	5.67	5.65	0.01	
1.62	1.61	1.60	1.60	1.59	1.59	1.59	1.58	1.58	1.58	1.58	1.58	0.25	
2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.32	2.31	2.30	2.29	0.10	8
3.22	3.15	3.12	3.08	3.04	2.02	3.01	2.97	2.97	2.95	2.94	2.93	0.05	
5.52	5.36	5.28	5.20	5.12	5.07	5.03	4.96	4.95	4.91	4.88	4.86	0.01	
1.57	1.56	1.56	1.55	1.55	1.54	1.54	1.53	1.53	1.53	1.53	1.53	0.25	
2.34	2.30	2.28	2.25	2.23	2.22	2.21	2.19	2.18	2.17	2.17	2.16	0.10	9
3.01	2.94	2.90	2.86	2.83	2.80	2.79	2.76	2.75	2.73	2.72	2.71	0.05	
4.96	4.81	4.73	4.65	4.57	4.52	4.48	4.42	4.40	4.36	4.33	4.31	0.01	

分母自由度 N_2	Pr	分子自由度 N_1: 1	2	3	4	5	6	7	8	9	10	11	12
10	0.25	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.55	1.54
	0.10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
	0.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
11	0.25	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.52	1.51
	0.10	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
	0.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
	0.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40
12	0.25	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.50	1.49
	0.10	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
	0.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
	0.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16
13	0.25	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47
	0.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
	0.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
	0.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	0.25	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.46	1.45
	0.10	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.08	2.05
	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
	0.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	0.25	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.44
	0.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
	0.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
	0.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	0.25	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.44	1.43
	0.10	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
	0.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
17	0.25	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.42	1.41
	0.10	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
	0.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
	0.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46
18	0.25	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
	0.10	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.96	1.93
	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	0.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
19	0.25	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.40
	0.10	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.94	1.91
	0.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
	0.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	0.25	1.40	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.39
	0.10	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.92	1.89
	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23

续前表

分子自由度 N_1													分母自由度 N_2
15	20	24	30	40	50	60	100	120	200	500	∞	Pr	
1.53	1.52	1.52	1.51	1.51	1.50	1.50	1.49	1.49	1.49	1.48	1.48	0.25	
2.24	2.20	2.18	2.16	2.13	2.12	2.11	2.09	2.08	2.07	2.06	2.06	0.10	10
2.85	2.77	2.74	2.70	2.66	2.64	2.62	2.59	2.58	2.56	2.55	2.54	0.05	
4.56	4.41	4.33	4.25	4.17	4.12	4.08	4.01	4.00	3.96	3.93	3.91	0.01	
1.50	1.49	1.49	1.48	1.47	1.47	1.47	1.46	1.46	1.46	1.45	1.45	0.25	
2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	2.00	1.99	1.98	1.97	0.10	11
2.72	2.65	2.61	2.57	2.53	2.51	2.49	2.46	2.45	2.43	2.42	2.40	0.05	
4.25	4.10	4.02	3.94	3.86	3.81	3.78	3.71	3.69	3.66	3.62	3.60	0.01	
1.48	1.47	1.46	1.45	1.45	1.44	1.44	1.43	1.43	1.43	1.42	1.42	0.25	
2.10	2.06	2.04	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	0.10	12
2.62	2.54	2.51	2.47	2.43	2.40	2.38	2.35	2.34	2.32	2.31	2.30	0.05	
4.01	3.86	3.78	3.70	3.62	3.57	3.54	3.47	3.45	3.41	3.38	3.36	0.01	
1.46	1.45	1.44	1.43	1.42	1.42	1.42	1.41	1.41	1.40	1.40	1.40	0.25	
2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.88	1.86	1.85	1.85	0.10	13
2.53	2.46	2.42	2.38	2.34	2.31	2.30	2.26	2.25	2.23	2.22	2.21	0.05	
3.82	3.66	3.59	3.51	3.43	3.38	3.34	3.27	3.25	3.22	3.19	3.17	0.01	
1.44	1.43	1.42	1.41	1.41	1.40	1.40	1.39	1.39	1.39	1.38	1.38	0.25	
2.01	1.96	1.94	1.91	1.89	1.87	1.86	1.83	1.83	1.82	1.80	1.80	0.10	14
2.46	2.39	2.35	2.31	2.27	2.24	2.22	2.19	2.18	2.16	2.14	2.13	0.05	
3.66	3.51	3.43	3.35	3.27	3.22	3.18	3.11	3.09	3.06	3.03	3.00	0.01	
1.43	1.41	1.41	1.40	1.39	1.39	1.38	1.38	1.37	1.37	1.36	1.36	0.25	
1.97	1.92	1.90	1.87	1.85	1.83	1.82	1.79	1.79	1.77	1.76	1.76	0.10	15
2.40	2.33	2.29	2.25	2.20	2.18	2.16	2.12	2.11	2.10	2.08	2.07	0.05	
3.52	3.37	3.29	3.21	3.13	3.08	3.05	2.98	2.96	2.92	2.89	2.87	0.01	
1.41	1.40	1.39	1.38	1.37	1.37	1.36	1.36	1.35	1.35	1.34	1.34	0.25	
1.94	1.89	1.87	1.84	1.81	1.79	1.78	1.76	1.75	1.74	1.73	1.72	0.10	16
2.35	2.28	2.24	2.19	2.15	2.12	2.11	2.07	2.06	2.04	2.02	2.01	0.05	
3.41	3.26	3.18	3.10	3.02	2.97	2.93	2.86	2.84	2.81	2.78	2.75	0.01	
1.40	1.39	1.38	1.37	1.36	1.35	1.35	1.34	1.34	1.34	1.33	1.33	0.25	
1.91	1.86	1.84	1.81	1.78	1.76	1.75	1.73	1.72	1.71	1.69	1.69	0.10	17
2.31	2.23	2.19	2.15	2.10	2.08	2.06	2.02	2.01	1.99	1.97	1.96	0.05	
3.31	3.16	3.08	3.00	2.92	2.87	2.83	2.76	2.75	2.71	2.68	2.65	0.01	
1.39	1.38	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.32	1.32	1.32	0.25	
1.89	1.84	1.81	1.78	1.75	1.74	1.72	1.70	1.69	1.68	1.67	1.66	0.10	18
2.27	2.19	2.15	2.11	2.06	2.04	2.02	1.98	1.97	1.95	1.93	1.92	0.05	
3.23	3.08	3.00	2.92	2.84	2.78	2.75	2.68	2.66	2.62	2.59	2.57	0.01	
1.38	1.37	1.36	1.35	1.34	1.33	1.33	1.32	1.32	1.31	1.31	1.30	0.25	
1.86	1.81	1.79	1.76	1.73	1.71	1.70	1.67	1.67	1.65	1.64	1.63	0.10	19
2.23	2.16	2.11	2.07	2.03	2.00	1.98	1.94	1.93	1.91	1.89	1.88	0.05	
3.15	3.00	2.92	2.84	2.76	2.71	2.67	2.60	2.58	2.55	2.51	2.49	0.01	
1.37	1.36	1.35	1.34	1.33	1.33	1.32	1.31	1.31	1.30	1.30	1.29	0.25	
1.84	1.79	1.77	1.74	1.71	1.69	1.68	1.65	1.64	1.63	1.62	1.61	0.10	20
2.20	2.12	2.08	2.04	1.99	1.97	1.95	1.91	1.90	1.88	1.86	1.84	0.05	
3.09	2.94	2.86	2.78	2.69	2.64	2.61	2.54	2.52	2.48	2.44	2.42	0.01	

表 D.3 **F分布的上端百分点(续)**

分母自由度 N_2	分子自由度 N_1												
	Pr	1	2	3	4	5	6	7	8	9	10	11	12
22	0.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	0.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	0.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	0.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	0.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	0.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	0.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	0.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	0.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
	0.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	0.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	0.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	0.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	0.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	0.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	0.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	0.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	0.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	0.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	0.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	0.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	0.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	0.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	0.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	0.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	0.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	0.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	0.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	0.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	0.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	0.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	0.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	0.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	0.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	0.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	0.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

续前表

分子自由度 N_1													分母自由度 N_2
15	20	24	30	40	50	60	100	120	200	500	∞	Pr	
1.36	1.34	1.33	1.32	1.31	1.31	1.30	1.30	1.30	1.29	1.29	1.28	0.25	22
1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.61	1.60	1.59	1.58	1.57	0.10	
2.15	2.07	2.03	1.98	1.94	1.91	1.89	1.85	1.84	1.82	1.80	1.78	0.05	
2.98	2.83	2.75	2.67	2.58	2.53	2.50	2.42	2.40	2.36	2.33	2.31	0.01	
1.35	1.33	1.32	1.31	1.30	1.29	1.29	1.28	1.28	1.27	1.27	1.26	0.25	24
1.78	1.73	1.70	1.67	1.64	1.62	1.61	1.58	1.57	1.56	1.54	1.53	0.10	
2.11	2.03	1.98	1.94	1.89	1.86	1.84	1.80	1.79	1.77	1.75	1.73	0.05	
2.89	2.74	2.66	2.58	2.49	2.44	2.40	2.33	2.31	2.27	2.24	2.21	0.01	
1.34	1.32	1.31	1.30	1.29	1.28	1.28	1.26	1.26	1.26	1.25	1.25	0.25	26
1.76	1.71	1.68	1.65	1.61	1.59	1.58	1.55	1.54	1.53	1.51	1.50	0.10	
2.07	1.99	1.95	1.90	1.85	1.82	1.80	1.76	1.75	1.73	1.71	1.69	0.05	
2.81	2.66	2.58	2.50	2.42	2.36	2.33	2.25	2.23	2.19	2.16	2.13	0.01	
1.33	1.31	1.30	1.29	1.28	1.27	1.27	1.26	1.25	1.25	1.24	1.24	0.25	28
1.74	1.69	1.66	1.63	1.59	1.57	1.56	1.53	1.52	1.50	1.49	1.48	0.10	
2.04	1.96	1.91	1.87	1.82	1.79	1.77	1.73	1.71	1.69	1.67	1.65	0.05	
2.75	2.60	2.52	2.44	2.35	2.30	2.26	2.19	2.17	2.13	2.09	2.06	0.01	
1.32	1.30	1.29	1.28	1.27	1.26	1.26	1.25	1.24	1.24	1.23	1.23	0.25	30
1.72	1.67	1.64	1.61	1.57	1.55	1.54	1.51	1.50	1.48	1.47	1.46	0.10	
2.01	1.93	1.89	1.84	1.79	1.76	1.74	1.70	1.68	1.66	1.64	1.62	0.05	
2.70	2.55	2.47	2.39	2.30	2.25	2.21	2.13	2.11	2.07	2.03	2.01	0.01	
1.30	1.28	1.26	1.25	1.24	1.23	1.22	1.21	1.21	1.20	1.19	1.19	0.25	40
1.66	1.61	1.57	1.54	1.51	1.48	1.47	1.43	1.42	1.41	1.39	1.38	0.10	
1.92	1.84	1.79	1.74	1.69	1.66	1.64	1.59	1.58	1.55	1.53	1.51	0.05	
2.52	2.37	2.29	2.20	2.11	2.06	2.02	1.94	1.92	1.87	1.83	1.80	0.01	
1.27	1.25	1.24	1.22	1.21	1.20	1.19	1.17	1.17	1.16	1.15	1.15	0.25	60
1.60	1.54	1.51	1.48	1.44	1.41	1.40	1.36	1.35	1.33	1.31	1.29	0.10	
1.84	1.75	1.70	1.65	1.59	1.56	1.53	1.48	1.47	1.44	1.41	1.39	0.05	
2.35	2.20	2.12	2.03	1.94	1.88	1.84	1.75	1.73	1.68	1.63	1.60	0.01	
1.24	1.22	1.21	1.19	1.18	1.17	1.16	1.14	1.13	1.12	1.11	1.10	0.25	120
1.55	1.48	1.45	1.41	1.37	1.34	1.32	1.27	1.26	1.24	1.21	1.19	0.10	
1.75	1.66	1.61	1.55	1.50	1.46	1.43	1.37	1.35	1.32	1.28	1.25	0.05	
2.19	2.03	1.95	1.86	1.76	1.70	1.66	1.56	1.53	1.48	1.42	1.38	0.01	
1.23	1.21	1.20	1.18	1.16	1.14	1.12	1.11	1.10	1.09	1.08	1.06	0.25	200
1.52	1.46	1.42	1.38	1.34	1.31	1.28	1.24	1.22	1.20	1.17	1.14	0.10	
1.72	1.62	1.57	1.52	1.46	1.41	1.39	1.32	1.29	1.26	1.22	1.19	0.05	
2.13	1.97	1.89	1.79	1.69	1.63	1.58	1.48	1.44	1.39	1.33	1.28	0.01	
1.22	1.19	1.18	1.16	1.14	1.13	1.12	1.09	1.08	1.07	1.04	1.00	0.25	∞
1.49	1.42	1.38	1.34	1.30	1.26	1.24	1.18	1.17	1.13	1.08	1.00	0.10	
1.67	1.57	1.52	1.46	1.39	1.35	1.32	1.24	1.22	1.17	1.11	1.00	0.05	
2.04	1.88	1.79	1.70	1.59	1.52	1.47	1.36	1.32	1.25	1.15	1.00	0.01	

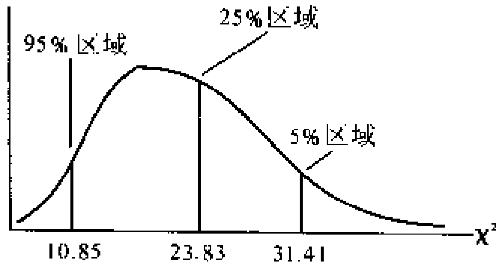
表 D.4 χ^2 分布的上端百分点

例

$\Pr(\chi^2>10.85)=0.95$

$\Pr(\chi^2>23.83)=0.25$ 对于 df=20

$\Pr(\chi^2>31.41)=0.05$



自由度 \ Pr	0.995	0.990	0.975	0.950	0.900
1	$392\,704\times10^{-10}$	$157\,088\times10^{-9}$	$982\,069\times10^{-9}$	$393\,214\times10^{-8}$	0.015 790 8
2	0.010 025 1	0.020 100 7	0.050 635 6	0.102 587	0.210 720
3	0.071 721 2	0.114 832	0.215 795	0.351 846	0.584 375
4	0.206 990	0.297 110	0.484 419	0.710 721	1.063 623
5	0.411 740	0.554 300	0.831 211	1.145 476	1.610 31
6	0.675 727	0.872 085	1.237 347	1.635 39	2.204 13
7	0.989 265	1.239 043	1.689 87	2.167 35	2.833 11
8	1.344 419	1.646 482	2.179 73	2.732 64	3.489 54
9	1.734 926	2.087 912	2.700 39	3.325 11	4.168 16
10	2.155 85	2.558 21	3.246 97	3.940 30	4.865 18
11	2.603 21	3.053 47	3.815 75	4.574 81	5.577 79
12	3.073 82	3.570 56	4.403 79	5.226 03	6.303 80
13	3.565 03	4.106 91	5.008 74	5.891 86	7.041 50
14	4.074 68	4.660 43	5.628 72	6.570 63	7.789 53
15	4.600 94	5.229 35	6.262 14	7.260 94	8.546 75
16	5.142 24	5.812 21	6.907 66	7.961 64	9.312 23
17	5.697 24	6.407 76	7.564 18	8.671 76	10.085 2
18	6.264 81	7.014 91	8.230 75	9.390 46	10.864 9
19	6.843 98	7.632 73	8.906 55	10.117 0	11.650 9
20	7.433 86	8.260 40	9.590 83	10.850 8	12.442 6
21	8.033 66	8.897 20	10.282 93	11.591 3	13.239 6
22	8.642 72	9.542 49	10.982 3	12.338 0	14.041 5
23	9.260 42	10.195 67	11.688 5	13.090 5	14.847 9
24	9.886 23	10.856 4	12.401 1	13.848 4	15.658 7
25	10.519 7	11.524 0	13.119 7	14.611 4	16.473 4
26	11.160 3	12.198 1	13.843 9	15.379 1	17.291 9
27	11.807 6	12.878 6	14.573 3	16.151 3	18.113 8
28	12.461 3	13.564 8	15.307 9	16.927 9	18.939 2
29	13.121 1	14.256 5	16.047 1	17.708 3	19.767 7
30	13.786 7	14.953 5	16.790 8	18.492 6	20.599 2
40	20.706 5	22.164 3	24.433 1	26.509 3	29.050 5
50	27.990 7	29.706 7	32.357 4	34.764 2	37.688 6
60	35.534 6	37.484 8	40.481 7	43.187 9	46.458 9
70	43.275 2	45.441 8	48.757 6	51.739 3	55.329 0
80	51.172 0	53.540 0	57.153 2	60.391 5	64.277 8
90	59.196 3	61.754 1	65.646 6	69.126 0	73.291 2
100*	67.327 6	70.064 8	74.221 9	77.929 5	83.358 1

续前表

0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
0.101 530 8	0.454 937	1.323 30	2.705 54	3.841 46	5.023 89	6.634 90	7.879 44
0.575 364	1.386 29	2.772 59	4.605 17	5.991 47	7.377 76	9.210 34	10.596 6
1.212 534	2.365 97	4.108 35	6.251 39	7.814 73	9.348 40	11.344 9	12.838 1
1.922 55	3.356 70	5.385 27	7.779 44	9.487 73	11.143 3	13.276 7	14.860 2
2.674 60	4.351 46	6.625 68	9.236 35	11.070 5	12.832 5	15.086 3	16.749 6
3.454 60	5.348 12	7.840 80	10.644 6	12.591 6	14.449 4	16.811 9	18.547 6
4.254 85	6.345 81	9.037 15	12.017 0	14.067 1	16.012 8	18.475 3	20.277 7
5.070 64	7.344 12	10.218 8	13.361 6	15.507 3	17.534 6	20.090 2	21.955 0
5.898 83	8.342 83	11.388 7	14.683 7	16.919 0	19.022 8	21.666 0	23.589 3
6.737 20	9.341 82	12.548 9	15.987 1	18.307 0	20.483 1	23.209 3	25.188 2
7.584 12	10.341 0	13.700 7	17.275 0	19.675 1	21.920 0	24.725 0	26.756 9
8.438 42	11.340 3	14.845 4	18.549 4	21.026 1	23.336 7	26.217 0	28.299 5
9.299 06	12.339 8	15.983 9	19.811 9	22.362 1	24.735 6	27.688 3	29.819 4
10.165 3	13.339 3	17.117 0	21.064 2	23.684 8	26.119 0	29.141 3	31.319 3
11.036 5	14.338 9	18.245 1	22.307 2	24.995 8	27.488 4	30.577 9	32.801 3
11.912 2	15.338 5	19.368 8	23.541 8	26.296 2	28.845 4	31.999 9	34.267 2
12.791 9	16.338 1	20.488 7	24.769 0	27.587 1	30.191 0	33.408 7	35.718 5
13.675 3	17.337 9	21.604 9	25.989 4	28.869 3	31.526 4	34.805 3	37.156 4
14.562 0	18.337 6	22.717 8	27.203 6	30.143 5	32.852 3	36.190 8	38.582 2
15.451 8	19.337 4	23.827 7	28.412 0	31.410 4	34.169 6	37.566 2	39.996 8
16.344 4	20.337 2	24.934 8	29.615 1	32.670 5	35.478 9	38.932 1	41.401 0
17.239 6	21.337 0	26.039 3	30.813 3	33.924 4	36.780 7	40.289 4	42.795 6
18.137 3	22.336 9	27.141 3	32.006 9	35.172 5	38.075 7	41.638 4	44.181 3
19.037 2	23.336 7	28.241 2	33.196 3	36.415 1	39.364 1	42.979 8	45.558 5
19.939 3	24.336 6	29.338 9	34.381 6	37.652 5	40.646 5	44.314 1	46.927 8
20.843 4	25.336 4	30.434 5	35.563 1	38.885 2	41.923 2	45.641 7	48.289 9
21.749 4	26.336 3	31.528 4	36.741 2	40.113 3	43.194 4	46.963 0	49.644 9
22.657 2	27.336 3	32.620 5	37.915 9	41.337 2	44.460 7	48.278 2	50.993 3
23.566 6	28.336 2	33.710 9	39.087 5	42.556 9	45.722 2	49.587 9	52.335 6
24.477 6	29.336 0	34.799 8	40.256 0	43.772 9	46.979 2	50.892 2	53.672 0
33.660 3	39.335 4	45.616 0	51.805 0	55.758 5	59.341 7	63.690 7	66.765 9
42.942 1	49.334 9	56.333 6	63.167 1	67.504 8	71.420 2	76.153 9	79.490 0
52.293 8	59.334 7	66.981 4	74.397 0	79.081 9	83.297 6	88.379 4	91.951 7
61.698 3	69.334 4	77.576 6	85.527 1	90.531 2	95.023 1	100.425	104.215
71.144 5	79.334 3	88.130 3	96.578 2	101.879	106.629	112.329	116.321
80.624 7	89.334 2	98.649 9	107.565	113.145	118.136	124.116	128.299
90.133 2	99.334 1	109.141	118.498	124.342	129.561	135.807	140.169

* 对于大于100的自由度，表达式 $\sqrt{2\chi^2}-\sqrt{(2k-1)}=Z$ 遵循标准化正态分布，其中 k 代表自由度。

资料来源：摘自 E. S. Pearson and H. O. Hartley, eds., *Biometrika Tables for Statisticians*, vol. 1, 3d ed., table 8, Cambridge University Press, New York, 1966。生物统计学报编辑部与董事会授权复制。

表 D. 5a　　德宾-沃森 d 统计量：在 0.05 显著性水平上 d_L 和 d_U 的显著点

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.368	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005			—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.734
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309	0.650	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.682	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.795	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.198
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.907	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.837	1.369	1.873	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862	1.594	1.877
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

续前表

n	$k'=1$ d_L	$k'=1$ d_U	$k'=2$ d_L	$k'=2$ d_U	$k'=3$ d_L	$k'=3$ d_U	$k'=4$ d_L	$k'=4$ d_U	$k'=5$ d_L	$k'=5$ d_U	$k'=6$ d_L	$k'=6$ d_U	$k'=7$ d_L	$k'=7$ d_U	$k'=8$ d_L	$k'=8$ d_U	$k'=9$ d_L	$k'=9$ d_U	$k'=10$ d_L	$k'=10$ d_U
16	0.098	3.503	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
17	0.138	3.378	0.087	3.557	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
18	0.177	3.265	0.123	3.441	0.078	3.603	—	—	—	—	—	—	—	—	—	—	—	—	—	—
19	0.220	3.159	0.160	3.335	0.111	3.496	0.070	3.642	—	—	—	—	—	—	—	—	—	—	—	—
20	0.263	3.063	0.200	3.234	0.145	3.395	0.100	3.542	0.063	3.676	—	—	—	—	—	—	—	—	—	—
21	0.307	2.976	0.240	3.141	0.182	3.300	0.132	3.448	0.091	3.583	0.058	3.705	—	—	—	—	—	—	—	—
22	0.349	2.897	0.281	3.057	0.220	3.211	0.166	3.358	0.120	3.495	0.083	3.619	0.052	3.731	—	—	—	—	—	—
23	0.391	2.826	0.322	2.979	0.259	3.128	0.202	3.272	0.153	3.409	0.110	3.535	0.076	3.650	0.048	3.753	—	—	—	—
24	0.431	2.761	0.362	2.908	0.297	3.053	0.239	3.193	0.186	3.327	0.141	3.454	0.101	3.572	0.070	3.678	0.044	3.773	—	—
25	0.470	2.702	0.400	2.844	0.335	2.983	0.275	3.119	0.221	3.251	0.172	3.376	0.130	3.494	0.094	3.604	0.065	3.702	0.041	3.790
26	0.508	2.649	0.438	2.784	0.373	2.919	0.312	3.051	0.256	3.179	0.205	3.303	0.160	3.420	0.120	3.531	0.087	3.632	0.060	3.724
27	0.544	2.600	0.475	2.730	0.409	2.859	0.348	2.987	0.291	3.112	0.238	3.233	0.191	3.349	0.149	3.460	0.112	3.563	0.081	3.658
28	0.578	2.555	0.510	2.680	0.445	2.805	0.383	2.928	0.325	3.050	0.271	3.168	0.222	3.283	0.178	3.392	0.138	3.495	0.104	3.592
29	0.612	2.515	0.544	2.634	0.479	2.755	0.418	2.874	0.359	2.992	0.305	3.107	0.254	3.219	0.208	3.327	0.166	3.431	0.129	3.528
30	0.643	2.477	0.577	2.592	0.512	2.708	0.451	2.823	0.392	2.937	0.337	3.050	0.286	3.160	0.238	3.266	0.195	3.368	0.156	3.465
31	0.647	2.443	0.608	2.553	0.545	2.665	0.484	2.776	0.425	2.887	0.370	2.996	0.317	3.103	0.269	3.208	0.224	3.309	0.183	3.406
32	0.703	2.411	0.638	2.517	0.576	2.625	0.515	2.733	0.457	2.840	0.401	2.946	0.349	3.050	0.299	3.153	0.253	3.252	0.211	3.348
33	0.731	2.382	0.668	2.484	0.606	2.588	0.546	2.692	0.488	2.796	0.432	2.899	0.379	3.000	0.329	3.100	0.283	3.198	0.239	3.293
34	0.758	2.355	0.695	2.454	0.634	2.554	0.575	2.654	0.518	2.754	0.462	2.854	0.409	2.954	0.359	3.051	0.312	3.147	0.267	3.240
35	0.783	2.330	0.722	2.425	0.662	2.521	0.604	2.619	0.547	2.716	0.492	2.813	0.439	2.910	0.388	3.005	0.340	3.099	0.295	3.190
36	0.808	2.306	0.748	2.398	0.689	2.492	0.631	2.586	0.575	2.680	0.520	2.774	0.467	2.868	0.417	2.961	0.369	3.053	0.323	3.142
37	0.831	2.285	0.772	2.374	0.714	2.464	0.657	2.555	0.602	2.646	0.548	2.738	0.495	2.829	0.445	2.920	0.397	3.009	0.351	3.097
38	0.854	2.265	0.796	2.351	0.739	2.438	0.683	2.526	0.628	2.614	0.575	2.703	0.522	2.792	0.472	2.880	0.424	2.968	0.378	3.054
39	0.875	2.246	0.819	2.329	0.763	2.413	0.707	2.499	0.653	2.585	0.600	2.671	0.549	2.757	0.499	2.843	0.451	2.929	0.404	3.013
40	0.896	2.228	0.840	2.309	0.785	2.391	0.731	2.473	0.678	2.557	0.626	2.641	0.575	2.724	0.525	2.808	0.477	2.892	0.430	2.974
45	0.988	2.156	0.938	2.225	0.887	2.296	0.838	2.367	0.788	2.439	0.740	2.512	0.692	2.586	0.644	2.659	0.598	2.733	0.553	2.807
50	1.064	2.103	1.019	2.163	0.973	2.225	0.927	2.287	0.882	2.350	0.836	2.414	0.792	2.479	0.747	2.544	0.703	2.610	0.660	2.675
55	1.129	2.062	1.087	2.116	1.045	2.170	1.003	2.225	0.961	2.281	0.919	2.338	0.877	2.396	0.836	2.454	0.795	2.512	0.754	2.571
60	1.184	2.031	1.145	2.079	1.106	2.127	1.068	2.177	1.029	2.227	0.990	2.278	0.951	2.330	0.913	2.382	0.874	2.434	0.836	2.487
65	1.231	2.006	1.195	2.049	1.160	2.093	1.124	2.138	1.088	2.183	1.052	2.229	1.016	2.276	0.980	2.323	0.944	2.371	0.908	2.419
70	1.272	1.986	1.239	2.026	1.206	2.066	1.172	2.106	1.139	2.148	1.105	2.189	1.072	2.232	1.038	2.275	1.005	2.318	0.971	2.362
75	1.308	1.970	1.277	2.006	1.247	2.043	1.215	2.080	1.184	2.118	1.153	2.156	1.121	2.195	1.090	2.235	1.058	2.275	1.027	2.315
80	1.340	1.957	1.311	1.991	1.283	2.024	1.253	2.059	1.224	2.093	1.195	2.129	1.165	2.165	1.136	2.201	1.106	2.238	1.076	2.275
85	1.369	1.946	1.342	1.977	1.315	2.009	1.287	2.040	1.260	2.073	1.232	2.105	1.205	2.139	1.177	2.172	1.149	2.206	1.121	2.241
90	1.395	1.937	1.369	1.966	1.344	1.995	1.318	2.025	1.292	2.055	1.266	2.085	1.240	2.116	1.213	2.148	1.187	2.179	1.160	2.211
95	1.418	1.929	1.394	1.956	1.370	1.984	1.345	2.012	1.321	2.040	1.296	2.068	1.271	2.097	1.247	2.126	1.222	2.156	1.197	2.186
100	1.439	1.923	1.416	1.948	1.393	1.974	1.371	2.000	1.347	2.026	1.324	2.053	1.301	2.080	1.277	2.108	1.253	2.135	1.229	2.164
150	1.579	1.892	1.564	1.908	1.550	1.924	1.535	1.940	1.519	1.956	1.504	1.972	1.489	1.989	1.474	2.006	1.458	2.023	1.443	2.040
200	1.654	1.885	1.643	1.896	1.632	1.908	1.621	1.919	1.610	1.931	1.599	1.943	1.588	1.955	1.576	1.967	1.565	1.979	1.554	1.991

注：n＝观测个数，k'＝不包含常数项的解释变量个数。

例，若 $n=40$ 和 $k'=4$，则 $d_L=1.285$ 和 $d_U=1.721$。如果所计算的 d 值小于 1.285，即表明有正的一阶序列相关；如果它大于 1.721，则表明无正的一阶序列相关迹象，但若 d 落在这两个上下限之间，则表明尚无迹象足以判明是否有正的一阶序列相关。

资料来源：本表是原始德宾-沃森表的扩充。先复制自 N. E. Savin and K. J. White, "The Durbin-Watson Test for Serial Correlation With Extreme Small Samples or Many Regressors," *Econometrica*, vol. 45, November 1977, pp. 1989—1996。其后校正于 R. W. Farebrother, *Econometrica*, vol. 48, September 1980, p. 1554。计量经济学会允许复制。

表 D.5b　　德宾-沃森 d 统计量：在 0.01 显著性水平上 d_L 和 d_U 的显著点

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.390	1.142				—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.435	1.036	0.294	1.676	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.497	1.003	0.345	1.489	0.229	2.102	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.554	0.998	0.408	1.389	0.279	1.875	0.183	2.433	—	—	—	—	—	—	—	—	—	—	—	—
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.150	2.690	—	—	—	—	—	—	—	—	—	—
11	0.653	1.010	0.519	1.297	0.396	1.640	0.286	2.030	0.193	2.453	0.124	2.892	—	—	—	—	—	—	—	—
12	0.697	1.023	0.569	1.274	0.449	1.575	0.339	1.913	0.244	2.280	0.164	2.665	0.105	3.053	—	—	—	—	—	—
13	0.738	1.038	0.616	1.261	0.499	1.526	0.391	1.826	0.294	2.150	0.211	2.490	0.140	2.838	0.090	3.182	—	—	—	—
14	0.776	1.054	0.660	1.254	0.547	1.490	0.441	1.757	0.343	2.049	0.257	2.354	0.183	2.667	0.122	2.981	0.078	3.287	—	—
15	0.811	1.070	0.700	1.252	0.591	1.464	0.488	1.704	0.391	1.967	0.303	2.244	0.226	2.530	0.161	2.817	0.107	3.101	0.068	3.374
16	0.844	1.086	0.737	1.252	0.633	1.446	0.532	1.663	0.437	1.900	0.349	2.153	0.269	2.416	0.200	2.681	0.142	2.944	0.094	3.201
17	0.874	1.102	0.772	1.255	0.672	1.432	0.574	1.630	0.480	1.847	0.393	2.078	0.313	2.319	0.241	2.566	0.179	2.811	0.127	3.053
18	0.902	1.118	0.805	1.259	0.708	1.422	0.613	1.604	0.522	1.803	0.435	2.015	0.355	2.238	0.282	2.467	0.216	2.697	0.160	2.925
19	0.928	1.132	0.835	1.265	0.742	1.415	0.650	1.584	0.561	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.255	2.597	0.196	2.813
20	0.952	1.147	0.863	1.271	0.773	1.411	0.685	1.567	0.598	1.737	0.515	1.918	0.436	2.110	0.362	2.308	0.294	2.510	0.232	2.714
21	0.975	1.161	0.890	1.277	0.803	1.408	0.718	1.554	0.633	1.712	0.552	1.881	0.474	2.059	0.400	2.244	0.331	2.434	0.268	2.625
22	0.997	1.174	0.914	1.284	0.831	1.407	0.748	1.543	0.667	1.691	0.587	1.849	0.510	2.015	0.437	2.188	0.368	2.367	0.304	2.548
23	1.018	1.187	0.938	1.291	0.858	1.407	0.777	1.534	0.698	1.673	0.620	1.821	0.545	1.977	0.473	2.140	0.404	2.308	0.340	2.479
24	1.037	1.199	0.960	1.298	0.882	1.407	0.805	1.528	0.728	1.658	0.652	1.797	0.578	1.944	0.507	2.097	0.439	2.255	0.375	2.417
25	1.055	1.211	0.981	1.305	0.906	1.409	0.831	1.523	0.756	1.645	0.682	1.776	0.610	1.915	0.540	2.059	0.473	2.209	0.409	2.362
26	1.072	1.222	1.001	1.312	0.928	1.411	0.855	1.518	0.783	1.635	0.711	1.759	0.640	1.889	0.572	2.026	0.505	2.168	0.441	2.313
27	1.089	1.233	1.019	1.319	0.949	1.413	0.878	1.515	0.808	1.626	0.738	1.743	0.669	1.867	0.602	1.997	0.536	2.131	0.473	2.269
28	1.104	1.244	1.037	1.325	0.969	1.415	0.900	1.513	0.832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.566	2.098	0.504	2.229
29	1.119	1.254	1.054	1.332	0.988	1.418	0.921	1.512	0.855	1.611	0.788	1.718	0.723	1.830	0.658	1.947	0.595	2.068	0.533	2.193
30	1.133	1.263	1.070	1.339	1.006	1.421	0.941	1.511	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.925	0.622	2.041	0.562	2.160
31	1.147	1.273	1.085	1.345	1.023	1.425	0.960	1.510	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017	0.589	2.131
32	1.160	1.282	1.100	1.352	1.040	1.428	0.979	1.510	0.917	1.597	0.856	1.690	0.794	1.788	0.734	1.889	0.674	1.995	0.615	2.104
33	1.172	1.291	1.114	1.358	1.055	1.432	0.996	1.510	0.936	1.594	0.876	1.683	0.816	1.776	0.757	1.874	0.698	1.975	0.641	2.080
33	1.172	1.291	1.114	1.358	1.055	1.432	0.996	1.510	0.936	1.594	0.876	1.683	0.816	1.776	0.757	1.874	0.698	1.975	0.641	2.080
34	1.184	1.299	1.128	1.364	1.070	1.435	1.012	1.511	0.954	1.591	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.957	0.665	2.057
35	1.195	1.307	1.140	1.370	1.085	1.439	1.028	1.512	0.971	1.589	0.914	1.671	0.857	1.757	0.800	1.847	0.744	1.940	0.689	2.037
36	1.206	1.315	1.153	1.376	1.098	1.442	1.043	1.513	0.988	1.588	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.925	0.711	2.18
37	1.217	1.323	1.165	1.382	1.112	1.446	1.058	1.514	1.004	1.586	0.950	1.662	0.895	1.742	0.841	1.825	0.787	1.911	0.733	2.001
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.515	1.019	1.585	0.566	1.658	0.913	1.735	0.860	1.816	0.807	1.899	0.754	1.985
39	1.237	1.337	1.187	1.393	1.137	1.453	1.085	1.517	1.034	1.584	0.982	1.655	0.930	1.729	0.878	1.807	0.826	1.887	0.774	1.970
40	1.246	1.344	1.198	1.398	1.148	1.457	1.098	1.518	1.048	1.584	0.997	1.652	0.946	1.724	0.895	1.799	0.844	1.876	0.749	1.956
45	1.288	1.376	1.245	1.423	1.201	1.474	1.156	1.528	1.111	1.584	1.065	1.643	1.019	1.704	0.974	1.768	0.927	1.834	0.881	1.902
50	1.324	1.403	1.285	1.446	1.245	1.491	1.205	1.538	1.164	1.587	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.805	0.955	1.864
55	1.356	1.427	1.320	1.466	1.284	1.506	1.247	1.548	1.209	1.592	1.072	1.638	1.134	1.685	1.095	1.734	1.057	1.785	1.018	1.837
60	1.383	1.449	1.350	1.484	1.317	1.520	1.283	1.558	1.249	1.598	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771	1.072	1.817
65	1.407	1.468	1.377	1.500	1.346	1.534	1.315	1.568	1.283	1.604	1.251	1.642	1.218	1.680	1.186	1.720	1.153	1.761	1.120	1.802
70	1.429	1.485	1.400	1.515	1.372	1.546	1.343	1.578	1.313	1.611	1.283	1.645	1.253	1.680	1.223	1.716	1.192	1.754	1.162	1.792
75	1.448	1.501	1.422	1.529	1.395	1.557	1.368	1.587	1.340	1.617	1.313	1.649	1.284	1.682	1.256	1.714	1.227	1.748	1.199	1.783
80	1.466	1.515	1.441	1.541	1.416	1.568	1.390	1.595	1.364	1.624	1.338	1.653	1.312	1.683	1.285	1.714	1.259	1.745	1.232	1.777
85	1.482	1.528	1.458	1.553	1.435	1.578	1.411	1.603	1.386	1.630	1.362	1.657	1.337	1.685	1.312	1.714	1.287	1.743	1.262	1.773
90	1.496	1.540	1.474	1.563	1.452	1.587	1.429	1.611	1.406	1.636	1.382	1.661	1.360	1.687	1.336	1.714	1.312	1.741	1.288	1.769
95	1.510	1.552	1.489	1.573	1.468	1.596	1.446	1.618	1.425	1.642	1.403	1.666	1.381	1.690	1.358	1.715	1.336	1.741	1.313	1.767
100	1.522	1.562	1.503	1.583	1.482	1.604	1.462	1.625	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.357	1.741	1.335	1.765
150	1.611	1.637	1.598	1.651	1.584	1.665	1.571	1.679	1.557	1.693	1.543	1.708	1.530	1.722	1.515	1.737	1.501	1.752	1.486	1.767
200	1.664	1.684	1.653	1.693	1.643	1.704	1.633	1.715	1.623	1.725	1.613	1.735	1.603	1.746	1.592	1.757	1.582	1.768	1.571	1.779

续前表

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
16	0.060	3.446	—		—	—	—	—	—	—	—	—	—	—			—	—	—	—
17	0.084	3.286	0.053	3.506	—	—	—	—	—	—	—	—		—	—	—	—	—	—	—
18	0.113	3.146	0.075	3.358	0.047	3.357	—		—	—	—	—	—	—	—				—	—
19	0.145	3.023	0.102	3.227	0.067	3.420	0.043	3.601	—	—	—	—			—	—	—	—	—	
20	0.178	2.914	0.131	3.109	0.092	3.297	0.061	3.474	0.038	3.639	—			—	—	—	—	—	—	
21	0.212	2.817	0.162	3.004	0.119	3.185	0.084	3.358	0.055	3.521	0.035	3.671		—	—	—	—	—		
22	0.246	2.729	0.194	2.909	0.148	3.084	0.109	3.252	0.077	3.412	0.050	3.562	0.032	3.700	—	—	—		—	—
23	0.281	2.651	0.227	2.822	0.178	2.991	0.136	3.155	0.100	3.311	0.070	3.459	0.046	3.597	0.029	3.725		—	—	—
24	0.315	2.580	0.260	2.744	0.209	2.906	0.165	3.065	0.125	3.218	0.092	3.363	0.065	3.501	0.043	3.629	0.027	3.747	—	
25	0.348	2.517	0.292	2.674	0.240	2.829	0.194	2.982	0.152	3.131	0.116	3.274	0.085	3.410	0.060	3.538	0.039	3.657	0.025	3.766
26	0.381	2.460	0.324	2.610	0.272	2.758	0.224	2.906	0.180	3.050	0.141	3.191	0.107	3.325	0.079	3.452	0.055	3.572	0.036	3.682
27	0.413	2.409	0.356	2.552	0.303	2.694	0.253	2.836	0.208	2.976	0.167	3.113	0.131	3.245	0.100	3.371	0.073	3.490	0.051	3.602
28	0.444	2.363	0.387	2.499	0.333	2.635	0.283	2.772	0.237	2.907	0.194	3.040	0.156	3.169	0.122	3.294	0.093	3.412	0.068	3.524
29	0.474	2.321	0.417	2.451	0.363	2.582	0.313	2.713	0.266	2.843	0.222	2.972	0.182	3.098	0.146	3.220	0.114	3.338	0.087	3.450
30	0.503	2.283	0.447	2.407	0.393	2.533	0.342	2.659	0.294	2.785	0.249	2.909	0.208	3.032	0.171	3.152	0.137	3.267	0.107	3.379
31	0.531	2.248	0.475	2.367	0.422	2.487	0.371	2.609	0.322	2.730	0.277	2.851	0.234	2.970	0.196	3.087	0.160	3.201	0.128	3.311
32	0.558	2.216	0.503	2.330	0.450	2.446	0.399	2.563	0.350	2.680	0.304	2.797	0.261	2.912	0.221	3.026	0.184	3.137	0.151	3.246
33	0.585	2.187	0.530	2.296	0.477	2.408	0.426	2.520	0.377	2.633	0.331	2.746	0.287	2.858	0.246	2.969	0.209	3.078	0.174	3.184
34	0.610	2.160	0.556	2.266	0.503	2.373	0.452	2.481	0.404	2.590	0.357	2.699	0.313	2.808	0.272	2.915	0.233	3.022	0.197	3.126
35	0.634	2.136	0.581	2.237	0.529	2.340	0.478	2.444	0.430	2.550	0.383	2.655	0.339	2.761	0.297	2.865	0.257	2.969	0.221	3.071
36	0.658	2.113	0.605	2.210	0.554	2.310	0.504	2.410	0.455	2.512	0.409	2.614	0.364	2.717	0.322	2.818	0.282	2.919	0.244	3.019
37	0.680	2.092	0.628	2.186	0.578	2.282	0.528	2.379	0.480	2.477	0.434	2.576	0.389	2.675	0.347	2.774	0.306	2.872	0.268	2.969
38	0.702	2.073	0.651	2.164	0.601	2.256	0.552	2.350	0.504	2.445	0.458	2.540	0.614	2.637	0.371	2.733	0.330	2.828	0.291	2.923
39	0.723	2.055	0.673	2.143	0.623	2.232	0.575	2.323	0.528	2.414	0.482	2.507	0.438	2.600	0.395	2.694	0.354	2.787	0.315	2.879
40	0.744	2.039	0.694	2.123	0.645	2.210	0.597	2.297	0.551	2.386	0.505	2.476	0.461	2.566	0.418	2.657	0.377	2.748	0.338	2.838
45	0.835	1.972	0.790	2.044	0.744	2.118	0.700	2.193	0.655	2.269	0.612	2.346	0.570	2.424	0.528	2.503	0.488	2.582	0.448	2.661
50	0.913	1.925	0.871	1.987	0.829	2.051	0.787	2.116	2.746	2.182	0.705	2.250	0.665	2.318	0.625	2.387	0.586	2.456	0.548	2.526
55	0.979	1.891	0.940	1.945	0.902	2.002	0.863	2.059	0.825	2.117	0.786	2.176	0.748	2.237	0.711	2.298	0.674	2.359	0.637	2.421
60	1.037	1.865	1.001	1.914	0.965	1.964	0.929	2.015	0.893	2.067	0.857	2.120	0.822	2.173	0.786	2.227	0.751	2.283	0.716	2.338
65	1.087	1.845	1.053	1.889	1.020	1.934	0.986	1.980	0.953	2.027	0.919	2.075	0.886	2.123	0.852	2.172	0.819	2.221	0.786	2.272
70	1.131	1.831	1.099	1.870	1.068	1.911	1.037	1.953	1.005	1.995	0.974	2.038	0.943	2.082	0.911	2.127	0.880	2.172	0.849	2.217
75	1.170	1.819	1.141	1.856	1.111	1.893	1.082	1.931	1.052	1.970	1.023	2.009	0.993	2.049	0.964	2.090	0.934	2.131	0.905	2.172
80	1.205	1.810	1.177	1.844	1.150	1.878	1.122	1.913	1.094	1.949	1.066	1.984	1.039	2.022	1.011	2.059	0.983	2.097	0.955	2.135
85	1.236	1.803	1.210	1.834	1.184	1.866	1.158	1.898	1.132	1.931	1.106	1.965	1.080	1.999	1.053	2.033	1.027	2.068	1.000	2.104
90	1.264	1.798	1.240	1.827	1.215	1.856	1.191	1.886	1.166	1.917	1.141	1.948	1.116	1.979	1.091	2.012	1.066	2.044	1.041	2.077
95	1.290	1.793	1.267	1.821	1.244	1.848	1.221	1.876	1.197	1.905	1.174	1.934	1.150	1.963	1.126	1.993	1.102	2.023	1.079	2.054
100	1.314	1.790	1.292	1.816	1.270	1.841	1.248	1.868	1.225	1.895	1.203	1.922	1.181	1.949	1.158	1.977	1.136	2.006	1.113	2.034
150	1.473	1.783	1.458	1.799	1.444	1.814	1.429	1.830	1.414	1.847	1.400	1.863	1.385	1.880	1.370	1.897	1.355	1.913	1.340	1.931
200	1.561	1.791	1.550	1.801	1.539	1.813	1.528	1.824	1.518	1.836	1.507	1.847	1.495	1.860	1.484	1.871	1.474	1.883	1.462	1.896

注：n＝观测个数

k'＝不包含常数项的解释变量个数。

资料来源：Savin and White，前引文献，经计量经济学会允许转载。

表 D. 6A　　　　游程检验中的游程临界值

N_1 \ N_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	30	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

注：表 D. 6A 和表 D. 6B 对 N_1（正号）和 N_2（负号）的不同值给出游程 n 的临界值。对于一个样本的游程检验，任何等于或小于表 D. 6A 所给的 n 值，或等于或大于表 D. 6B 所给的 n 值，都是在 0. 05 水平上显著的。

资料来源：Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill Book Company, New York, 1956, table F, pp. 252—253. 该表由 Siegel 改编自原始来源：Frieda S. Swed and C. Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," *Annals of Mathematical Statistics*, vol. 14, 1943。McGraw-Hill Book Company 与 *Annals of Mathematical Statistics* 允许转载。

表 D6. B

游程检验中的游程临界值

N_1	N_2: 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4				9	9														
5			9	10	10	11	11												
6			9	10	11	12	12	13	13	13	13								
7				11	12	13	13	14	14	14	14	15	15	15					
8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
16							17	18	19	20	21	21	22	23	23	24	25	25	25
17							17	18	19	20	21	22	23	23	24	25	25	26	26
18							17	18	19	20	21	22	23	24	25	25	26	26	27
19							17	18	20	21	22	23	23	24	25	26	26	27	27
20							17	18	20	21	22	23	24	25	25	26	27	27	28

例，在一个由 20 个"＋"号（＝N_1）和 10 个"－"号（＝N_2）组成的 30 个观测值序列中，0.05 显著性水平的游程临界值是 9 和 20。如表 D. 6A 和表 D. 6B 分别所示。因此在一项应用中，如果发现游程数等于或小于 9 或者等于或大于 20，就可（在 0.05 显著性水平上）拒绝所观测的序列是随机的这一假设。

表 D.7　　单位检验的1%和5%临界迪基-富勒 $t(=\tau)$ 和 F 值

样本容量	t_{nc}^{*}		t_{c}^{*}		t_{ct}^{*}		F^{**}		$F^{\dagger}$	
	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
25	−2.66	−1.95	−3.75	−3.00	−4.38	−3.60	10.61	7.24	8.21	5.68
50	−2.62	−1.95	−3.58	−2.93	−4.15	−3.50	9.31	6.73	7.02	5.13
100	−2.60	−1.95	−3.51	−2.89	−4.04	−3.45	8.73	6.49	6.50	4.88
250	−2.58	−1.95	−3.46	−2.88	−3.99	−3.43	8.43	6.34	6.22	4.75
500	−2.58	−1.95	−3.44	−2.87	−3.98	−3.42	8.34	6.30	6.15	4.71
∞	−2.58	−1.95	−3.43	−2.86	−3.96	−3.41	8.27	6.25	6.09	4.68

* 下脚标 nc、c 和 ct 分别表示在回归(21.9.5)中无常数项、有常数项和同时含有常数项和趋势项。

** (21.9.5)中常数项和 δ 项同时为零的联合假设的临界 F 值。

† (21.9.5)中常数项、趋势项和 δ 项同时为零的联合假设的临界 F 值。

资料来源：选自 W. A. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, New York, 1976, p. 373（τ 检验部分）和 D. A. Dickey and W. A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, vol. 49, 1981, p. 1063。

附录 E　互联网上的经济数据*

976 经济统计简报：这是一个优秀的数据来源，包括产出、收入、就业、失业、工资、生产和商业活动、价格与货币、信贷与证券市场、国际统计等方面的数据。

http://www.whitehouse.gov/fsbr/esbr.htm

美联储米色书：简要介绍 12 个联邦储备区当前的经济状况。

http://www.bog.frb.fed.us/fomc/bb/current

政府信息共享资源：提供区域经济方面的信息；1990 年人口和住房状况普查数据；1992 年经济普查数据；1982、1987 和 1992 年农业普查数据；1991—1995 年美国进出口数据；1990 年就业机会平等方面的数据。

http://govinfo.keer.orst.edu

国民经济研究局(**NBER**)主页：这个受同行高度评价的私人研究机构搜集了资产价格、劳动就业、生产、货币、供给、商业周期指标等方面的大量数据，而且建立了许多与其他网站的链接。

http://www.nber.org

977 综列研究：提供了对美国个人和家庭代表样本纵向调查的数据，这些数据自 1968 年开始搜集。

http://www.umich.edu/~psid

经济学家的网上资源：包含经济活动信息和数据方面十分宽泛的数据，并建立了与许多网站的链接。对学术性和非学术性的经济学家而言都是十分有价值的资源。

http://econwpa.wwstl.edu/EconFAO/EconFaq.html

联邦网上定位器：提供联邦政府几乎所有部门的信息，并建立了国际链接。

http://www.law.vill.edu/Fed-Agency/fedwebloc.html

经济学网上资源(WebEC)：经济现象和数据最综合的图书馆。

http://wuecon.wustl.edu/~adnetec/WebEc/WebEc.html

美国股票交易数据网：约 700 家在第二大证券市场上挂牌的公司的信息。

http://www.amex.com

经济分析局(BEA)主页：经济分析局是美国商业部的一个机构，出版《当代商业调查》(*Survey of Current Business*)，这个网站是各种经济活动数据的优秀来源。

http://www.bea.doc.gov/

商业周期指标：你在这里可以找到约 256 个经济时间序列的数据。

http://www.globalexposure.com/bci.html

CIA 出版物：你将找到每年定期出版的《世界各国浏览》(*World Fact Book*)和《国际统计手册》(*Hondbook of International Statistics*)。

http://www.odic.gov/cia/publications/pubs.html

能源信息管理(DOE)：每个能源细类的经济信息和数据。

http://www.eia.doe.gov/

联邦储备银行经济数据库：圣路易斯联邦储备银行发布历史上的一些经济与社会数据，包括利率、货币和商业指标、汇率等。

http://www.stls.frb.org/fred/fred.html

978 国际贸易事物：提供与许多贸易统计、跨国项目等网站的链接。

http://www.ita.doc.gov/

美国统计数据库：国家贸易数据搜集处提供了国际贸易数据和鼓励出口信息等方面最全面的来源。还包括几个国家人口、政治和社会经济状况方面的大量数据。

http://www.stat-usa.gov/BEN/databases.html

网上统计资料来源/经济学：从各联邦机构、经济指标、联邦储备委员会、消费价格数据等整理出来的优秀统计资源，并与其他统计资源相链接。

http://www.lib.umich.edu/libhome/Documents.centers/stecom.html

劳工统计局：其主页涉及到就业、失业和工资等各方面的数据，并建立

与其他统计网站的链接。

http://www.stats.bls.gov:80/

美国人口普查局主页：提供了关于收入、就业、收入分配和贫穷等社会、人口和经济数据的基本来源。

http://www.census.gov/

一般社会调查数据：从 1972 年开始每年对美国家庭进行个人采访所得到的调查数据。35 000 人以上对约 2 500 个不同的问题做出了回答，从而形成大量数据。

http://www.icpsr.umich.edu/GSS/

贫穷状况研究所：由无党派、非营利的大学为基础的研究中心通过一系列有关贫穷状况和社会不平等状况的问题搜集到的数据。

http://www.ssc.wisc.edu/irp/

社会保障管理：提供社会保障事物方面大量数据的官方网站。

http://www.sa.gov/

参考文献

初级（Introductory）

979 Frank, C. R., Jr.: *Statistics and Econometrics*, Holt, Rinehart and Winston, New York, 1971.

Goldberger Arthur S.: Introductory Econometrics, Harvard University Press, 1998.

Gujarati, Damodar N.: *Essentials of Econometrics*, McGraw-Hill, New York, 1999.

Hill, Carter, William Griffiths, and George Judge: *Undergraduate Econometrics*, John Wiley & Sons, New York, 2001.

Hu, Teh-Wei: *Econometrics: An Introductory Analysis*, University Park Press, Baltimore, 1973.

Katz, David A.: *Econometric Theory and Applications*, Prentice Hall, Englewood Cliffs, N. J., 1982.

Klein, Lawrence R.: *An Introduction to Econometrics*, Prentice Hall, Englewood Cliffs, N. J., 1962.

Koop, Gary: *Analysis of Economic Data*, John Wiley & Sons, New York, 2000.

Walters, A. A.: *An Introduction to Econometrics*, Macmillan, London, 1968.

中级 (Intermediate)

Aigner, D. J.: *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N. J., 1971.

Dhrymes, Phoebus J: *Introductory Econometrics*, Springer-Verlag, New York, 1978.

Dielman, Terry E.: *Applied Regression Analysis for Business and Economics*, PWS-Kent, Boston, 1991.

980 Draper, N. R., and H. Smith: *Applied Regression Analysis*, 3d ed., John Wiley & Sons, New York, 1998.

Dutta, M., *Econometric Methods*, South-Western Publishing Company, Cincinnati, 1975.

Goldberger A. S.: *Topics in Regression Analysis*, Macmillan, New York, 1968.

Griffiths, William E., R. Carter Hill and George G. Judge: *Learning and Practicing Econometrics* John Wiley & Sons, New York, 1993.

Huang, D. S.: *Regression and Econometric Methods*, John Wiley & Sons, New York, 1970.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee: *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, New York, 1982.

Kelejian, H. A., and W. E. Oates: *Introduction to Econometrics: Principles and Applications*, 2d ed., Harper & Row, New York, 1981.

Koutsoyiannis, A.: *Theory of Econometrics*, Harper & Row, New York, 1973.

Maddala G. S.: *Introduction to Econometrics*, John Wiley & Sons, 3d ed., New York, 2001.

Mark, Stewart B., and Kenneth F. Wallis: *Introductory Econometrics*, 2d ed., John Wiley & Sons, New York, 1981. A Halsted Press Book.

Murphy, James L.: *Introductory Econometrics*, Richard D. Irwin, Homewood, Ill., 1973.

Netter, J., and W. Wasserman: *Applied Linear Statistical Models*, Richard D. Irwin, Homewood, Ill., 1974.

Pindyck, R. S., and D. L. Rubinfeld: *Econometric Models and Econometric Forecasts*, 4d ed., McGraw-Hill, New York, 1990.

Sprent Peter: *Models in Regression and Related Topics*, Methuen, London, 1969.

Tintner, Gerhard: *Econometrics*, John Wiley & Sons (science ed.), New York, 1965.

Valavanis, Stefan: *Econometrics: An Introduction to Maximum-Likelihood Methods*, McGraw-Hill, New York, 1959.

Verbeek, Marno: *A Guide to Modern Econometrics*, John Wiley & Sons, New York, 2000.

Wonnacott, R. J., and T. H. Wonnacott: *Econometrics*, 2d ed., John Wiley & Sons, New York, 1979.

Wooldridge, Jeffrey M.: *Introductory Econometrics*, South-Western College Publishing, 2000.

高级 (Advanced)

Chow, Gregory C.: *Econometric Methods*, McGraw-Hill, New York, 1983.

Christ, C. F.: *Econometric Models and Methods*, John Wiley & Sons, New York, 1966.

Davidson, James: *Econometric Theory*, Blackwell Publishers, Oxford, U. K., 2000.

981 Dhrymes, P. J.: *Econometrics: Statistical Foundations and Applications*, Harper & Row, New York, 1970.

Fomby, Thomas B., Carter R. Hill, and Stanley R. Johnson: *Advanced Econometric Methods*, Springer-Verlag, New York, 1984.

Goldberger, A. S.: *A Course in Econometrics*, Harvard University Press, Cambridge, Mass., 1991.

Goldberger, A. S.: *Econometric Theory*, John Wiley & Sons, New York, 1964.

Greene, William H.: *Econometric Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, N. J., 2000.

Harvey A C.: *The Econometric Analysis of Time Series*, 2d ed., MIT Press, Cambridge, Mass., 1990.

Hayashi, Fumio: *Econometrics*, Princeton University Press, Princeton, N. J., 2000.

Johnston, J.: *Econometric Methods*, 3d ed., McGraw-Hill, New York, 1984.

Judge, George G., Carter R. Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, *Theory and Practice of Econometrics*, John Wiley & Sons, New York, 1980.

Klein, Lawrence R.: *A Textbook of Econometrics*, 2d ed., Prentice Hall, Englewood Cliffs, N. J., 1974.

Kmenta, Jan: *Elements of Econometrics*, 2d ed., Macmillan, New York, 1986.

Madansky, A.: *Foundations of Econometrics*, North-Holland, Amsterdam, 1976.

Maddala, G. S.: *Econometrics*, McGraw-Hill, New York, 1977.

Malinvaud, E.: *Statistical Methods of Econometrics*, 2d ed., North-Holland, Amsterdam, 1976.

Mittelhammer, Ron C., George G. Judge, and Douglas J. Miller: *Econometric Foundations*, Cambridge University Press, New York, 2000.

Theil, Henry: *Principles of Econometrics*, John Wiley & Sons, New York, 1971.

专论 (Specialized)

Belsley, David A., Edwin Kuh, and Roy E. Welsh: *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, New York, 1980.

Dhrymes, P. J.: *Distributed Lags: Problems of Estimation and Formulation*, Holden-Day, San Francisco, 1971.

Diebold, Francis X.: *Elements of Forecasting*, 2d ed., South-Western Publishers, 2001.

Goldfeld, S. M., and R. E. Quandt: *Nonlinear Methods of Econometrics*, North-Holland, Amsterdam, 1972.

Gourieroux, Christian: *Econometrics of Qualitative Dependent Variables*, Cambridge University Press, New York, 2000.

Graybill, F. A.: *An Introduction to Linear Statistical Models*, vol. 1, McGraw-Hill New York, 1961.

982 Hamilton, James D.: *Time Series Analysis*, Princeton University Press, Princeton, 1994.

Madda, G. S., and Kim In-Moo: *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press, New York, 1988.

Mills, T. C.: *Time Series Techniques for Economists*, Cambridge University Press, 1990.

Rao, C. R.: *Linear Statistical Inference and Its Applications*, 2d ed., John Wiley & Sons, New York, 1975.

Zellner, A.: *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York, 1971.

应用 (Applied)

Berndt, Ernst R.: *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley, 1991.

Bridge, J.I. : *Applied Econometrics*, North-Holland, Amsterdam, 1971.

Charemza, Wojciech W., and Derek F. Deadman: *New Directions in Econometric Practice: General to Specific Modelling, Cointegration and Vector Autoregression*, 2d ed., Edward Elgar Publisher, New York, 1997.

Cramer, J.S.: *Empirical Econometrics*, North-Holland, Amsterdam, 1969.

Desai, Meghnad : *Applied Econometrics*, McGraw-Hill, New York, 1976.

Kennedy, Peter : *A Guide to Econometrics*, 4th ed., MIT Press, Cambridge, Mass., 1998.

Leser, C.E.V. : *Econometric Techniques and Problems*, 2d ed., Hafner, London, 1974.

Mills, T.C. : *The Econometric Modelling of Financial Time Series*, Cambridge University Press, 1993.

Mukherjee, Chandan, Howard White, and Marc Wuyts: *Econometrics and Data Analysis for Developing Courntries*, Routledge, New York, 1998.

Patterson, Kerry: *An Introduction to Applied Econometrics: A Time Series Approach*, St. Martin's Press, New York, 2000.

Rao, Potluri, and Roger LeRoy Miller: *Applied Econometrics*, Wadsworth, Belmont, Calif., 1971.

(注：与本书所讨论各专题有关的经典论文，除上引文献外，还可以参考各章后面的扩充文献。)

人 名 索 引*

* 数字表示原书页码;数字 + n 表示注码在该页的注释。——译者注

D

E

F

M

N

O

T

U

V

W

X

Y

Z

标 题 索 引*

* 数字表示原书页码；数字＋n 表示注码在该页的注释。——译者注

D

E

F

N

Q

R

T

W

Z

译 后 记

经过近一年时间的整理和翻译，终于把古扎拉蒂教授的《计量经济学基础》第四版呈现在中文读者的面前。与第三版相比，第四版在综列数据模型、非线性回归模型、虚拟回归模型等方面都有很大的突破，而且对时间序列模型、计量经济建模等方面也有大幅度改进。作者重版的根本目的有两个，一是为了能在这本入门性的教材中反映出计量经济研究的新进展，特别是20世纪后半期快速发展的微观计量经济理论；二是为了使学生更容易地将计量经济理论与计量经济实践相结合。

在中国人民大学出版社的大力推动下，国内引进的计量经济学经典教材日益丰富和全面，中级计量经济学除这本《计量经济学基础》外，还有《计量经济学导论：现代观点》一书；高级计量经济学教材有即将出版的《计量经济分析》等。这些教材不仅有利于国内读者迅速接触国外教学现状，还为他们将来从事经济学研究奠定了良好的基础。最近，佛罗里达大学的艾春荣教授又寄来 Wooldridge 的《横截面和综列数据的计量经济分析》（*Econometric Analysis of Cross Section and Panel Data*）一书，也可以作为研究生学习高级计量经济学的教材。鉴于国内计量经济学教科书日益丰富和完善，我们特建议从第四版起把书名改称为《计量经济学基础》，以区别不同层次及同层次的其他教材。

本书仍是在华中科技大学经济学院的林少宫教授的组织下完成的。林少宫教授用一年的时间独立翻译了该书的第三版，现在又颇费周折地把第四版整理出来，而不是另外翻译一本高级教材，足见第四版有一定的价值。参加翻译的人员安排如下：华中科技大学公共管理学院的孙春霞博士负责第6～8章和第18～20章，经济学院的李卫兵博士负责第14和17章，王滨博士负责第15和16章，其他部分由费剑平完成。全部书稿经林老审校，费剑平汇总并整理而成。

感谢华中科技大学经济学院张培刚教授、徐长生教授、方齐云教授、张卫东教授的大力支持，数学系于寅教授的鼓励与帮助，梁晶教授的理解与关怀，以及责任编辑岳平的辛勤劳动。尽管如此，仍可能存在许多不足，请读者和专家不吝赐教。（027－87542678 或 jpingfei@yaho.com.cn。）

费剑平

华中科技大学

2004.12.3