《计量经济学》

《Econometrics》

《经济计量学》

第一章　绪论

0.1 关于绪论

0.2 课程教学大纲

§1.1 计量经济学

§1.2 经典计量经济学模型的建模步骤

§1.3 计量经济学模型的应用

0 .1 关于绪论

- 绪论是课程的纲。
- 学好绪论，可以说学好了课程的一半。参观一个城市，先站在最高处俯瞰，然后走街串巷；了解一座建筑，先看模型，后走进每一个房间。各起一半作用。
- 绪论课的目的：了解课程的性质和在课程体系中的地位；了解课程完整的内容体系和将要讲授的内容；了解课程的重点和难点；了解课程的学习方法；介绍课程中不讲的但是必须了解的课程内容。
- 不必全懂，只需似懂非懂。

0.2 《计量经济学》教学大纲

1. 课程

   计量经济学

   课号：10114513

   学分：3

   课程性质：教育部规定核心课程

2. 教师

主讲教师：李子奈

办公地点：经管楼南512

电话：62789793，

E-mail: lizinai@mail.tsinghua.edu.cn

特聘教授：艾春荣，讲授部分内容，英语授课

助教：孙云，电话：62778492

E-mail: suny2.03@em.tsinghua.edu.cn

3. 课程说明

(1) 教学目的

经济学是一门科学，实证的方法，尤其是数量分析方法是经济学研究的基本方法论。通过该门课程教学，使学生掌握计量经济学的基本理论与方法，并能够建立实用的计量经济学应用模型。

(2) 先修课程

中级微观经济学、中级宏观经济学、经济统计学、微积分、线性代数、概率论与数理统计、应用数理统计

(3) 教材及参考书

《计量经济学》，李子奈，高等教育出版社，2000年7月

《Basic Econometrics》,Damodar N. Gujarrati，2001

《计量经济学—方法与应用》，李子奈，清华大学出版社，1992年

《高等计量经济学》，李子奈、叶阿忠，清华大学出版社，2000年

《计量经济学—理论、方法与模型》，唐国兴，复旦大学出版社，1988年

<Econometric Models,Techniques,and Applications>, Michael D. Intriligator, Prentice-Hall Inc.,1978(FirstEdition),1997(Second Edition)

(4) 课堂资料下载

内容：教材电子版、补充资料、课件、习题集、教学基本要求、教学大纲、复习要点等。

路径：院内：网上邻居→整个网络→**Mislab**

→Ntserver→Lizn

院外：**\\166.111.96.50→Lizn**

(5) 教学讨论区

学院主页→分类讨论区→计量经济学讨论区

进入帐号：**s004**，口令：**kc9476**

(6) 答疑时间

(7) 课程内容提纲及学时安排

（总课时：48学时，课内外周学时：3/6）

第一章	绪论	3学时
第二章	单方程计量经济学模型理论与方法	15学时
第三章	单方程计量经济学应用模型	9学时
第四章	联立方程计量经济学理论与方法	9学时
第五章	宏观计量经济模型	3学时
第六章	时间序列计量经济学模型	9学时

(8) 课程成绩

综合练习一：10分

综合练习二：10分

课堂表现：　　10分

期末考核：　　70分

4. 关于学习方法的说明

(1) 理论与应用并重。既要重视理论方法，也要重视应用模型和应用中实际问题的解决；

(2) 以教材中的经典理论方法为主，也要理解适当引入的、教材中没有的非经典理论方法；

(3) 对于理论方法，重点是思路而不是数学过程；

(4) 对于应用模型，重点不是每种模型本身，而是它们演变与发展的方法论；

(5) 必须十分重视综合练习；

(6) 必须掌握一种应用软件，注意课堂的软件应用演示，“师傅领进门，修行在个人”，多练。

§1.1 计量经济学

一、计量经济学

二、计量经济学模型

三、计量经济学的内容体系

四、计量经济学是一门经济学科

五、计量经济学在经济学科中的地位

一、计量经济学

△ 经济学的一个分支学科

○1926年挪威经济学家R.Frish提出Econometrics

○ 1930年成立世界计量经济学会

○ 1933年创刊《Econometrica》

○ 20世纪40、50年代的大发展和60年代的扩张

○ 20世纪70年代以来非经典（现代）计量经济学的发展

△ 定义

“用数学方法探讨经济学可以从好几个方面着手，但任何一个方面都不能和计量经济学混为一谈。计量经济学与经济统计学绝非一码事；它也不同于我们所说的一般经济理论，尽管经济理论大部分具有一定的数量特征；计量经济学也不应视为数学应用于经济学的同义语。经验表明，统计学、经济理论和数学这三者对于真正了解现代经济生活的数量关系来说，都是必要的，但本身并非是充分条件。三者结合起来，就是力量，这种结合便构成了计量经济学。”

△ 在经济学科中占据极重要的地位

克莱因（**R.Klein**）：“计量经济学已经在经济学科中居于最重要的地位”，“在大多数大学和学院中，计量经济学的讲授已经成为经济学课程表中最有权威的一部分”。

萨缪尔森（**P.Samuelson**）：“第二次大战后的经济学是计量经济学的时代”。

二、计量经济学模型

- △ 模型
- △ 数学模型
- △ 经济数学模型
- △ 计量经济学模型
- △ 经济理论分析（行为分析）→数理分析 →数量分析

三、计量经济学的内容体系

- △ 广义计量经济学和狭义计量经济学
- △ 初、中、高级计量经济学
- △ 理论计量经济学和应用计量经济学
- △ 经典计量经济学和非经典计量经济学
- △ 微观计量经济学和宏观计量经济学

△广义计量经济学和狭义计量经济学

- 广义计量经济学是利用经济理论、数学以及统计学定量研究经济现象的经济计量方法的统称，包括回归分析方法、投入产出分析方法、时间序列分析方法等。
- 狭义计量经济学，也就是我们通常所说的计量经济学，以揭示经济现象中的因果关系为目的，在数学上主要应用回归分析方法。
- 本课程中的计量经济学模型，就是狭义计量经济学意义上的经济数学模型。

△ 初、中、高级计量经济学

- 初级以计量经济学的数理统计学基础知识和经典的线性单方程模型理论与方法为主要内容；
- 中级以用矩阵描述的经典的线性单方程模型理论与方法、经典的线性联立方程模型理论与方法，以及传统的应用模型为主要内容；
- 高级以非经典的、现代的计量经济学模型理论、方法与应用为主要内容。
- 本课程定位于中级水平上，适当引入高级的内容。

△ 理论计量经济学和应用计量经济学

- 理论计量经济学是以介绍、研究计量经济学的理论与方法为主要内容，侧重于理论与方法的数学证明与推导，与数理统计联系极为密切。除了介绍计量经济模型的数学理论基础、普遍应用的计量经济模型的参数估计方法与检验方法外，还研究特殊模型的估计方法与检验方法，应用了广泛的数学知识。
- 应用计量经济学则以建立与应用计量经济学模型为主要内容，强调应用模型的经济学和经济统计学基础，侧重于建立与应用模型过程中实际问题的处理。
- 本课程是二者的结合。

△ 经典计量经济学和非经典计量经济学

- 经典计量经济学（Classical Econometrics）一般指20世纪70年代以前发展并广泛应用的计量经济学。

  R. Frish创立

  T. Haavelmo建立了它的概率论基础

  L. R. Klein成为其理论与应用的集大成者

- 经典计量经济学在理论方法方面特征是：

  (1) 模型类型—随机模型；

  (2) 模型导向—理论导向；

  (3) 模型结构—线性或者可以化为线性，因果分析，解释变量具有同等地位，模型具有明确的形式和参数；

  (4) 数据类型—以时间序列数据或者截面数据为样本，被解释变量为服从正态分布的连续随机变量；

  (5) 估计方法—仅利用样本信息，采用最小二乘方法或者最大似然方法估计模型。

- 经典计量经济学在应用方面的特征是：

  (1) 应用模型方法论基础——实证分析、经验分析、归纳；

  (2) 应用模型的功能——结构分析、政策评价、经济预测、理论检验与发展；

  (3) 应用模型的领域——传统的应用领域，例如生产、需求、消费、投资、货币需求，以及宏观经济等。

- 非经典计量经济学一般指20世纪70年代以来发展的计量经济学理论、方法及应用模型，也称为现代计量经济学。
- 非经典计量经济学主要包括：微观计量经济学、非参数计量经济学、时间序列计量经济学和动态计量经济学等。
- 非经典计量经济学的内容体系：模型类型非经典的计量经济学问题、模型导向非经典的计量经济学问题、模型结构非经典的计量经济学问题、数据类型非经典的计量经济学问题和估计方法非经典的计量经济学问题。

- 本课程以经典计量经济学为主，适当引入一些简单的、应用较多的现代计量经济学理论方法。理由：

  一方面，从理论方法角度，经典计量经济学理论方法是非经典计量经济学理论方法的基础；

  另一方面，从应用的角度，经典计量经济学模型仍然是目前应用最为普遍的计量经济学模型。

△ 微观计量经济学和宏观计量经济学

- 微观计量经济学 于2000年诺贝尔经济学奖公报中正式提出。
- 微观计量经济学的内容集中于“对个人和家庭的经济行为进行经验分析”；
- “微观计量经济学的原材料是微观数据”，微观数据表现为截面数据和平行（penal）数据。
- 赫克曼（J.Heckman）和麦克法登（D.McFadden） 对微观计量经济学作出原创性贡献。

- 微观计量经济学教科书和课程有：
  “Microeconometrics”

  “Advanced Microeconometrics”

  “Applied Microeconometrics”

  “Topics in Microeconometrics”

  “Methods in Microeconometrics”

- 微观计量经济学的主要内容包括：

  平行（penal）数据模型的理论方法

  离散选择模型的理论方法

  选择性样本模型的理论方法

- 宏观计量经济学名称由来已久，但是它的主要内容和研究方向发生了变化。
- 经典宏观计量经济学：利用计量经济学理论方法，建立宏观经济模型，对宏观经济进行分析、评价和预测。
- 现代宏观计量经济学的主要研究方向：单位根检验、协整理论以及动态计量经济学。

四、计量经济学是一门经济学科

△ 从计量经济学的定义看

△ 从计量经济学在西方国家经济学科中的地位看

△ 从计量经济学与数理统计学的区别看

△ 从建立与应用计量经济学模型的全过程看

△ 从诺贝尔经济学奖看

△诺贝尔经济学奖与计量经济学

- 53位获奖者中10位直接因为对计量经济学发展的贡献而获奖

  1969 R. Frish J. Tinbergen

  1973 W. Leotief

  1980 L. R. Klein

  1984 R. Stone

  1989 T. Haavelmo

  2000 J. J. Heckman D. L. McFadden

  2003 R. F. EngleC. W. J. Granger

- 近20位担任过世界计量经济学会会长
- 30余位左右在获奖成果中应用了计量经济学

- 获奖者名单

2003 Robert F. Engle, Clive W. J. Granger

2002 Daniel Kahneman, Vernon L. Smith

2001 George A. Akerlof, A. Michael Spence, Joseph E. Stiglitz

2000 James J Heckman, Daniel L McFadden

1999 Robert A. Mundell

1998 Amartya Sen

1997 Robert C. Merton, Myron S. Scholes

1996 James A. Mirrlees, William Vickrey

1995 Robert E. Lucas Jr.

1994 John C. Harsanyi, John F. Nash Jr., Reinhard Selten

1993 Robert W. Fogel, Douglass C. North

1992 Gary S. Becker

1991 Ronald H. Coase

1990 Harry M. Markowitz, Merton H. Miller, William F. Sharpe

1989 Trygve Haavelmo

1988 Maurice Allais

1987 Robert M. Solow

1986 James M. Buchanan Jr.

1985 Franco Modigliani

1984 Richard Stone

1983 Gerard Debreu

1982 George J. Stigler

1981 James Tobin

1980 Lawrence R. Klein

1979 Theodore W. Schultz, Sir Arthur Lewis

1978 Herbert A. Simon

1977 Bertil Ohlin, James E. Meade

1976 Milton Friedman

1975 Leonid Vitaliyevich Kantorovich

Tjalling C. Koopmans

1974 Gunnar Myrdal

Friedrich August von Hayek

1973 Wassily Leontief

1972 John R. Hicks, Kenneth J. Arrow

1971 Simon Kuznets

1970 Paul A. Samuelson

1969 Ragnar Frisch, Jan Tinbergen

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1969

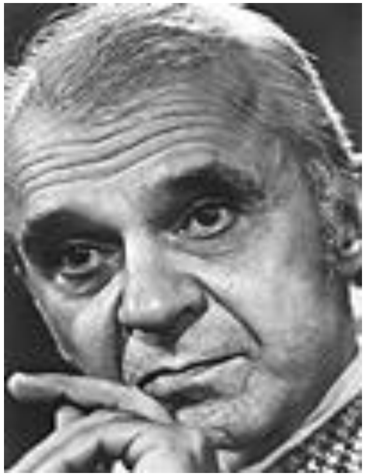
"for having developed and applied dynamic models for the analysis of economic processes"

Ragnar Frisch
Norway

Jan Tinbergen
the etherlands

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1973

"for the development of the input-output method and for its application to important economic problems"



Wassily Leontief
USA

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1980

"for the creation of econometric models and the application to the analysis of economic fluctuations and economic policies"

Lawrence R. Klein
USA

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1984

"for having made fundamental contributions to the development of systems of national accounts and hence greatly improved the basis for empirical economic analysis"

Richard Stone
Great Britain

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 1989

"for his clarification of the probability theory foundations of econometrics and his analyses of simultaneous economic structures"

Trygve Haavelmo
Norway

经典计量经济学		
	创立	Frisch
	建立第1个应用模型	Tinbergen
	建立概率论基础	Haavelmo
	发展数据基础	Stone
	发展应用模型	Klein
	建立投入产出模型	Leontief

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2000

"for his development of theory and methods for analyzing selective samples"

James J Heckman
USA

The Bank of Sweden Prize in Economic Sciences inMemory of Alfred Nobel 2000

"for his development of theory and methods for analyzing discrete choice"

Daniel L McFadden
USA

The Bank of Sweden Prize in Economic Sciences inMemory of Alfred Nobel 2003

"for methods of analyzing economic time series with common trends (cointegration)"

Clive W. J. Granger

UK

The Bank of Sweden Prize in Economic Sciences inMemory of Alfred Nobel 2003

"for methods of analyzing economic time series with time-varying volatility (ARCH)"

Robert F. Engle
USA

非经典计量经济学

微观计量：
选择性样本模型 ← Heckman

微观计量：
离散选择模型 ← McFadden

时间序列：
协整理论—现代宏观计量 ← Granger

时间序列：
ARCH—现代金融计量 ← Engle

五、计量经济学在经济学科中的地位

△ 从现代西方经济学的特征看

△ 从西方经济学的发展历史看

△ 从世界一流大学经济学课程表看

△ 从国际经济学刊物论文看

△ 从经济学的“世界先进水平”看

§1.2　建立计量经济学模型的步骤和要点

一、理论模型的设计

二、样本数据的收集

三、模型参数的估计

四、模型的检验

五、计量经济学模型成功的三要素

一、理论模型的建立

(1) 确定模型包含的变量

根据经济学理论和经济行为分析。

例如：同样是生产方程，电力工业和纺织工业应该选择不同的变量，为什么？

在时间序列数据样本下可以应用**Grange**统计检验等方法。

例如，消费和**GDP**之间的因果关系。

考虑数据的可得性。

注意因素和变量之间的联系与区别。

考虑入选变量之间的关系。

要求变量间互相独立。

(2) 确定模型的数学形式

利用经济学和数理经济学的成果

根据样本数据作出的变量关系图

选择可能的形式试模拟

(3) 拟定模型中待估计参数的理论期望值区间

符号、大小、 关系

例如：ln(人均食品需求量)=α+βln(人均收入)

+γln(食品价格) +δln(其它商品价格)+ε

其中α 、β、γ、δ的符号、大小、 关系

二、样本数据的收集

(1) 几类常用的样本数据

时间序列数据

截面数据

虚变量离散数据

联合应用

(2) 数据质量

完整性

准确性

可比性

一致性

三、模型参数的估计

(1) 各种模型参数估计方法

(2) 如何选择模型参数估计方法

(3) 关于应用软件的使用

课堂教学结合**Eviews**

能够熟练使用一种

四、模型的检验

(1) 经济意义检验

根据拟定的符号、大小、关系

例如：ln(人均食品需求量)=－2.0－0.5ln(人均收入)－4.5ln(食品价格) +0.8ln(其它商品价格)

ln(人均食品需求量)=－2.0+0.5ln(人均收入)－4.5ln(食品价格)+0.8ln(其它商品价格)

ln(人均食品需求量)=－2.0+0.5ln(人均收入)－0.8ln(食品价格) +0.8ln(其它商品价格)

(2) 统计检验

由数理统计理论决定

包括拟合优度检验

总体显著性检验

变量显著性检验

(3) 计量经济学检验

由计量经济学理论决定

包括异方差性检验

序列相关性检验

共线性检验

(4) 模型预测检验

由模型的应用要求决定

包括稳定性检验：扩大样本重新估计

预测性能检验：对样本外一点进行实际预测

五、计量经济学模型成功的三要素

- 理论
- 数据
- 方法

§1.3　计量经济学模型的应用

一、结构分析

二、经济预测

三、政策评价

四、理论检验与发展

一、结构分析

- 经济学中的结构分析是对经济现象中变量之间相互关系的研究。
- 结构分析所采用的主要方法是弹性分析、乘数分析与比较静力分析。
- 计量经济学模型的功能是揭示经济现象中变量之间的相互关系，即通过模型得到弹性、乘数等。
- 应用举例

二、经济预测

- 计量经济学模型作为一类经济数学模型，是从用于经济预测，特别是短期预测而发展起来的。
- 计量经济学模型是以模拟历史、从已经发生的经济活动中找出变化规律为主要技术手段。
- 对于非稳定发展的经济过程，对于缺乏规范行为理论的经济活动，计量经济学模型预测功能失效。
- 模型理论方法的发展以适应预测的需要。

三、政策评价

- 政策评价的重要性。
- 经济政策的不可试验性。
- 计量经济学模型的“经济政策实验室”功能。

四、理论检验与发展

- 实践是检验真理的唯一标准。
- 任何经济学理论，只有当它成功地解释了过去，才能为人们所接受。
- 计量经济学模型提供了一种检验经济理论的好方法。
- 对理论假设的检验可以发现和发展理论。

单方程计量经济学模型
理论与方法

Theory and Methodology of Single-Equation Econometric Model

第二章　经典单方程计量经济学模型：一元线性回归模型

- 回归分析概述
- 一元线性回归模型的参数估计
- 一元线性回归模型检验
- 一元线性回归模型预测
- 实例

§2.1　回归分析概述

一、变量间的关系及回归分析的基本概念

二、总体回归函数

三、随机扰动项

四、样本回归函数（SRF）

§2.1　回归分析概述

一、变量间的关系及回归分析的基本概念

1、变量间的关系

经济变量之间的关系，大体可分为两类：

（1）**确定性关系**或**函数关系**：研究的是确定现象非随机变量间的关系。

（2）**统计依赖**或**相关关系**：研究的是非确定现象随机变量间的关系。

例如：

函数关系：

$$\text{圆面积}=f\left(\pi,\text{半径}\right)=\pi\cdot\text{半径}^2$$

统计依赖关系/统计相关关系：

$$\text{农作物产量}=f\left(\text{气温, 降雨量, 阳光, 施肥量}\right)$$

<u>**对变量间统计依赖关系的考察主要是通过相关分析(correlation analysis)或回归分析(regression analysis)来完成的：**</u>

$$
\text{统计依赖关系}\left\{\begin{array}{l}
\text{线性相关}\left\{\begin{array}{l}\text{正相关}\\\text{不相关}\\\text{负相关}\end{array}\right\}\begin{array}{l}\text{相关系数：}\\-1\le\rho_{XY}\le 1\end{array}\\
\text{非线性相关}\left\{\begin{array}{l}\text{正相关}\\\text{不相关}\\\text{负相关}\end{array}\right.
\end{array}\right\}\left\{\begin{array}{l}\text{有因果关系}\longrightarrow\textbf{回归分析}\\\text{无因果关系}\longrightarrow\textbf{相关分析}\end{array}\right.
$$

▲注意：

①不线性相关并不意味着不相关；

②有相关关系并不意味着一定有因果关系；

③**回归分析**/**相关分析**研究一个变量对另一个（些）变量的统计依赖关系，但它们并不意味着一定有因果关系。

④**相关分析**对称地对待任何（两个）变量，两个变量都被看作是随机的。**回归分析**对变量的处理方法存在不对称性，即区分应变量（被解释变量）和自变量（解释变量）：前者是随机变量，后者不是。

2、回归分析的基本概念

回归分析(regression analysis)是研究一个变量关于另一个（些）变量的具体依赖关系的计算方法和理论。

其用意：在于通过后者的已知或设定值，去估计和（或）预测前者的（总体）均值。

这里：前一个变量被称为**被解释变量**（**Explained Variable**）或**应变量**（**Dependent Variable**），后一个（些）变量被称为**解释变量**（**Explanatory Variable**）或**自变量**（**Independent Variable**）。

回归分析构成计量经济学的方法论基础，其主要内容包括:

（1）根据样本观察值对经济计量模型参数进行估计，求得**回归方程；**

（2）对回归方程、参数估计值进行显著性检验；

（3）利用回归方程进行分析、评价及预测。

二、总体回归函数

由于变量间关系的随机性，**回归分析关心的是根据解释变量的已知或给定值，考察被解释变量的总体均值**，即当解释变量取某个确定值时，与之统计相关的被解释变量所有可能出现的对应值的平均值。

例2.1： 一个假想的社区有100户家庭组成，要研究该社区每月**家庭消费支出Y**与每月**家庭可支配收入X**的关系。

即如果知道了家庭的月收入，能否预测该社区家庭的平均月消费支出水平。

为达到此目的，将该100户家庭划分为组内收入差不多的10组，以分析每一收入组的家庭消费支出。

表 2.1.1　某社区家庭每月收入与消费支出统计表

	每月家庭可支配收入X（元）									
	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
每月家庭消费支出Y（元）	561	638	869	1023	1254	1408	1650	1969	2090	2299
	594	748	913	1100	1309	1452	1738	1991	2134	2321
	627	814	924	1144	1364	1551	1749	2046	2178	2530
	638	847	979	1155	1397	1595	1804	2068	2266	2629
		935	1012	1210	1408	1650	1848	2101	2354	2860
		968	1045	1243	1474	1672	1881	2189	2486	2871
			1078	1254	1496	1683	1925	2233	2552	
			1122	1298	1496	1716	1969	2244	2585	
			1155	1331	1562	1749	2013	2299	2640	
			1188	1364	1573	1771	2035	2310		
			1210	1408	1606	1804	2101			
				1430	1650	1870	2112			
				1485	1716	1947	2200			
						2002				
共计	2420	4950	11495	16445	19305	23870	25025	21450	21285	15510

分析：

（1）由于不确定因素的影响，对同一收入水平X，不同家庭的消费支出不完全相同；

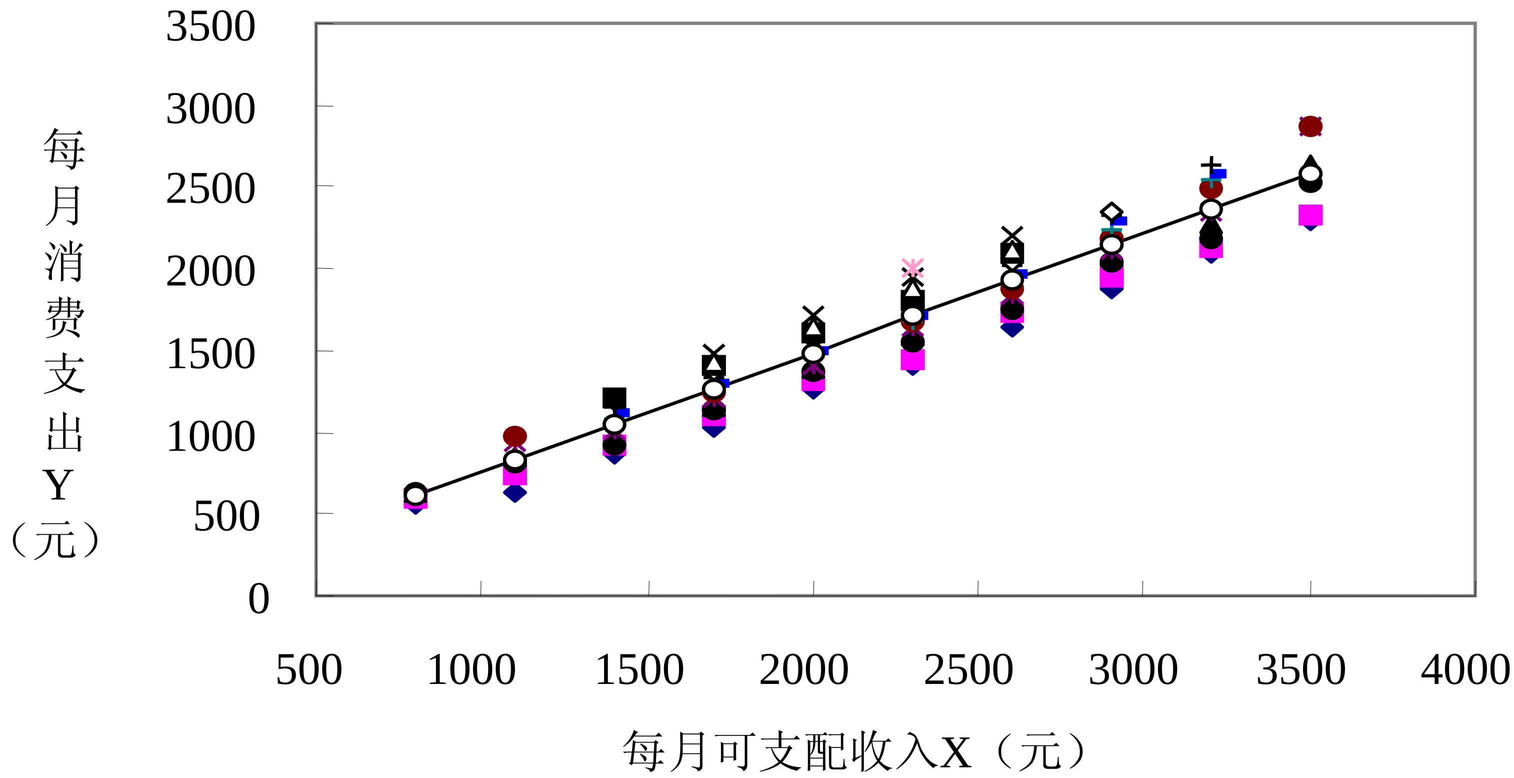
（2）但由于调查的完备性，给定收入水平X的消费支出Y的分布是确定的，即以X的给定值为条件的Y的**条件分布**（**Conditional distribution**）是已知的，如：　　$P(Y=561|X=800)=1/4$。

因此，给定收入X的值X_i，可得消费支出Y的**条件均值**（conditional mean）或**条件期望**（conditional expectation）：

$$E(Y|X=X_i)$$

该例中：$E(Y \mid X=800)=561$

描出散点图发现：随着收入的增加，消费“**平均地说**”也在增加，且Y的条件均值均落在一根正斜率的直线上。这条直线称为**总体回归线**。



- **概念:**

在给定解释变量X_i条件下被解释变量Y_i的期望轨迹称为**总体回归线**（population regression line），或更一般地称为**总体回归曲线**（population regression curve）。

相应的函数：

$$E(Y \mid X_i) = f(X_i)$$

称为（双变量）**总体回归函数**（population regression function, **PRF**）。

- **含义:**

回归函数（PRF）说明被解释变量Y的平均状态（总体条件期望）随解释变量X变化的规律。

- **函数形式:**

可以是线性或非线性的。

例2.1中，将居民消费支出看成是其可支配收入的线性函数时:

$$E(Y \mid X_i) = \beta_0 + \beta_1 X_i$$

为一**线性函数**。其中，β_0，β_1是未知参数，称为**回归系数**（regression coefficients）。。

三、随机扰动项

总体回归函数说明在给定的收入水平X_i下，该社区家庭平均的消费支出水平。

但对某一个别的家庭，其消费支出可能与该平均水平有偏差。

记 $$\mu_i = Y_i - E(Y \mid X_i)$$

称μ_i为观察值Y_i围绕它的期望值$E(Y|X_i)$的**离差**（**deviation**），是一个不可观测的随机变量，又称为**随机干扰项**（**stochastic disturbance**）或**随机误差项**（**stochastic error**）。

例2.1中，个别家庭的消费支出为：

$$Y_i = E(Y \mid X_i) + \mu_i = \beta_0 + \beta_1 X_i + \mu_i \qquad (*)$$

即，给定收入水平X_i ,个别家庭的支出可表示为两部分之和:

（1）该收入水平下所有家庭的平均消费支出$E(Y|X_i)$，称为系统性（**systematic**）或确定性（**deterministic**)部分。

（2）其他随机或非确定性（**nonsystematic**)部分μ_i。

（*）式称为**总体回归函数（方程）PRF的随机设定形式。表明被解释变量除了受解释变量的系统性影响外，还受其他因素的随机性影响。**

由于方程中引入了随机项，成为计量经济学模型，因此也称为**总体回归模型**。

随机误差项主要包括下列因素的影响：

1）在解释变量中被忽略的因素的影响；

2）变量观测值的观测误差的影响；

3）模型关系的设定误差的影响；

4）其它随机因素的影响。

产生并设计随机误差项的主要原因：

1）理论的含糊性；

2）数据的欠缺；

3）节省原则。

四、样本回归函数（SRF）

总体的信息往往无法掌握，现实的情况只能是在一次观测中得到总体的一个样本。

问题：能从一次抽样中获得总体的近似的信息吗？如果可以，如何从抽样中获得总体的近似信息？

例2.2：在例2.1的总体中有如下一个样本，

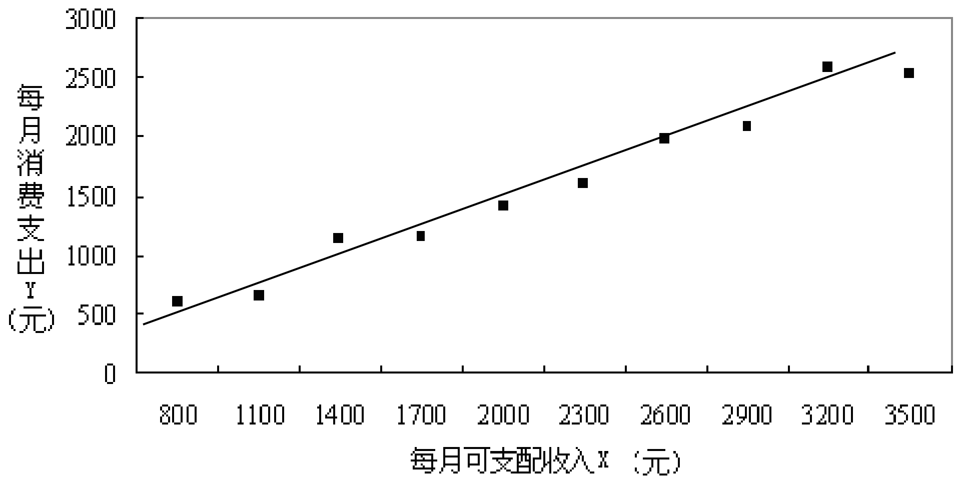
问：能否从该样本估计总体回归函数PRF？

表 2.1.3　家庭消费支出与可支配收入的一个随机样本

Y	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
X	594	638	1122	1155	1408	1595	1969	2078	2585	2530

回答：能

核样本的**散点图**（scatter diagram):



样本散点图近似于一条直线，画一条直线以尽好地拟合该散点图，由于样本取自总体，可以该线近似地代表总体回归线。该线称为**样本回归线**（**sample regression lines**）。

记样本回归线的函数形式为：

$$\hat{Y}_i = f(X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

称为**样本回归函数**（**sample regression function，SRF**）。

注意：

这里将**样本回归线**看成**总体回归线**的近似替代

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$\Downarrow$$

$$\begin{aligned} Y_i &= E(Y \mid X_i) + \mu_i \\ &= \beta_0 + \beta_1 X_i + \mu_i \end{aligned}$$

则 $\hat{Y}_i$ 为 $E(Y \mid X_i)$ 的估计量；

$\hat{\beta}_i$ 为 β_i 的估计量， $i = (0,1)$

样本回归函数的随机形式/样本回归模型：

同样地，样本回归函数也有如下的随机形式：

$$Y_i = \hat{Y}_i + \hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

式中，e_i 称为**（样本）残差**（或**剩余**）**项**（residual），代表了其他影响Y_i 的随机因素的集合，可看成是μ_i 的估计量$\hat{\mu}_i$ 。

由于方程中引入了随机项，成为计量经济模型，因此也称为**样本回归模型**（**sample regression model**）。

▼**回归分析的主要目的**：根据样本回归函数SRF，估计总体回归函数PRF。

即，根据 $$Y_i = \hat{Y}_i + e_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

估计 $$Y_i = E(Y \mid X_i) + \mu_i = \beta_0 + \beta_1 X_i + \mu_i$$

这就要求：

设计一"方法"构造SRF，以使SRF尽可能"接近"PRF，或者说使 $\hat{\beta}_i (i=0,1)$ 尽可能接近 $\beta_i (i=0,1)$。

注意：这里PRF可能永远无法知道。

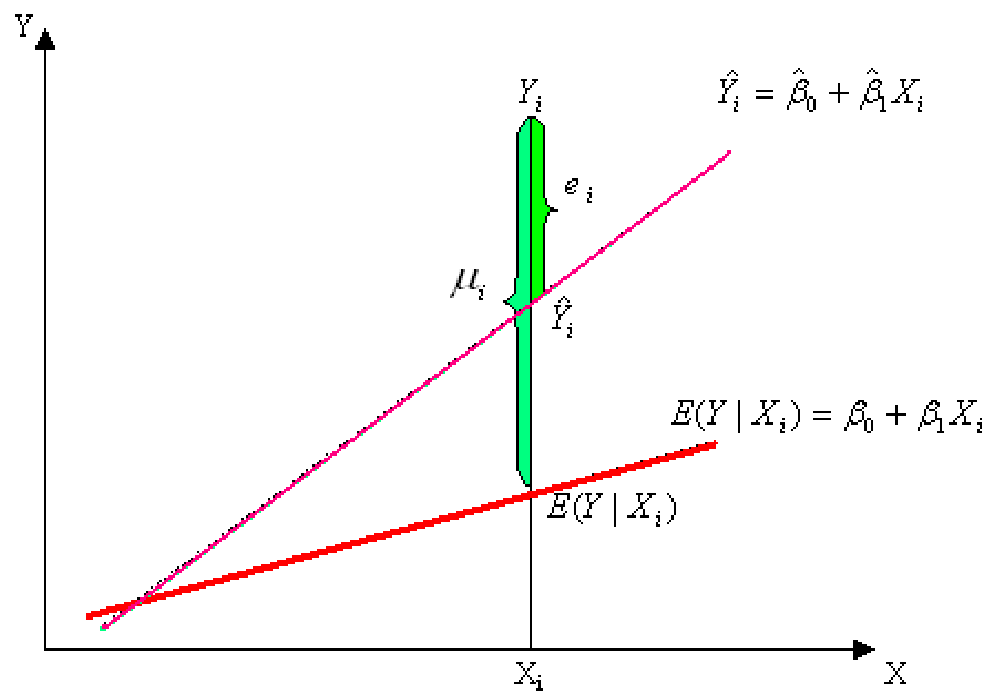


图 2.1.3 总体回归线与样本回归线的基本关系

§2.2　一元线性回归模型的参数估计

一、一元线性回归模型的基本假设

二、参数的普通最小二乘估计（OLS）

三、参数估计的最大或然法(ML)

四、最小二乘估计量的性质

五、参数估计量的概率分布及随机干扰项方差的估计

单方程计量经济学模型分为两大类：

线性模型和非线性模型

- 线性模型中，变量之间的关系呈线性关系
- 非线性模型中，变量之间的关系呈非线性关系

一元线性回归模型：只有一个解释变量

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \qquad i=1,2,\ldots,n$$

Y为被解释变量，X为解释变量，β_0与β_1为**待估参数**，μ为**随机干扰项**

回归分析的主要目的是要通过样本回归函数（模型）SRF尽可能准确地估计总体回归函数（模型）PRF。

估计方法有多种，其种最广泛使用的是**普通最小二乘法**（ordinary least squares, OLS）。

为保证参数估计量具有良好的性质，通常对模型提出若干基本假设。

注：实际这些假设与所采用的估计方法紧密相关。

一、线性回归模型的基本假设

假设1、解释变量X是确定性变量，不是随机变量；

假设2、随机误差项μ具有零均值、同方差和不序列相关性：

$$E(\mu_i)=0 \qquad i=1,2,\ldots,n$$

$$Var(\mu_i)=\sigma_\mu^2 \qquad i=1,2,\ldots,n$$

$$Cov(\mu_i,\mu_j)=0 \qquad i\neq j \quad i,j=1,2,\ldots,n$$

假设3、随机误差项μ与解释变量X之间不相关：

$$Cov(X_i,\mu_i)=0 \qquad i=1,2,\ldots,n$$

假设4、μ服从零均值、同方差、零协方差的正态分布

$$\mu_i \sim N(0,\sigma_\mu^2) \qquad i=1,2,\ldots,n$$

注意:

1、如果假设1、2满足，则假设3也满足;

2、如果假设4满足，则假设2也满足。

以上假设也称为线性回归模型的**经典假设**或**高斯（Gauss）假设**，满足该假设的线性回归模型，也称为**经典线性回归模型**（Classical Linear Regression Model, CLRM）。

另外，在进行模型回归时，还有两个暗含的假设：

假设5：随着样本容量的无限增加，解释变量X的样本方差趋于一有限常数。即

$$\sum (X_i - \bar{X})^2 / n \to Q, \qquad n \to \infty$$

假设6：回归模型是正确设定的

假设5旨在排除时间序列数据出现持续上升或下降的变量作为解释变量，因为这类数据不仅使大样本统计推断变得无效，而且往往产生所谓的**伪回归问题**（spurious regression problem）。

假设6也被称为模型没有**设定偏误**（specification error）

二、参数的普通最小二乘估计（OLS）

给定一组样本观测值（X_i, Y_i）（i=1,2,…n）要求样本回归函数尽可能好地拟合这组值.

普通最小二乘法（Ordinary least squares, OLS）给出的判断标准是：二者之差的平方和

$$Q=\sum_{1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{1}^{n}(Y_i-(\hat{\beta}_0+\hat{\beta}_1X_i))^2$$

最小。

即在给定样本观测值之下，选择出 $\hat{\beta}_0$、$\hat{\beta}_1$ 能使 Y_i 与 $\hat{Y}_i$ 之差的平方和最小。

根据微分运算，可推得用于估计$\hat{\beta}_0$ 、$\hat{\beta}_1$ 的下列方程组：

$$\begin{cases} \sum(\hat{\beta}_0+\hat{\beta}_1 X_i - Y_i)=0 \\ \sum(\hat{\beta}_0+\hat{\beta}_1 X_i - Y_i)X_i=0 \end{cases} \qquad (*)$$

或
$$\begin{cases} \Sigma Y_i = n\hat{\beta}_0+\hat{\beta}_1\Sigma X_i \\ \Sigma Y_i X_i = \hat{\beta}_0\Sigma X_i+\hat{\beta}_1\Sigma X_i^2 \end{cases}$$

解得：
$$\begin{cases} \hat{\beta}_0=\dfrac{\Sigma X_i^2\Sigma Y_i-\Sigma X_i\Sigma Y_i X_i}{n\Sigma X_i^2-(\Sigma X_i)^2} \\ \hat{\beta}_1=\dfrac{n\Sigma Y_i X_i-\Sigma Y_i\Sigma X_i}{n\Sigma X_i^2-(\Sigma X_i)^2} \end{cases}$$

方程组（*）称为**正规方程组**（**normal equations**）。

记 $$\sum x_i^2=\sum(X_i-\bar{X})^2=\sum X_i^2-\frac{1}{n}\left(\sum X_i\right)^2$$

$$\sum x_i y_i=\sum(X_i-\bar{X})(Y_i-\bar{Y})=\sum X_i Y_i-\frac{1}{n}\sum X_i\sum Y_i$$

上述参数估计量可以写成：
$$\begin{cases}\hat{\beta}_1=\dfrac{\Sigma x_i y_i}{\Sigma x_i^2}\\ \hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}\end{cases}$$

称为OLS估计量的**离差形式**（**deviation form**）。

由于参数的估计结果是通过最小二乘法得到的，故称为**普通最小二乘估计量**（**ordinary least squares estimators**）。

顺便指出，记 $\hat{y}_i = \hat{Y}_i - \bar{Y}$

则有

$$\hat{y}_i = (\hat{\beta}_0 + \hat{\beta}_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{X} + \bar{e})$$
$$= \hat{\beta}_1 (X_i - \bar{X}) - \tfrac{1}{n} \sum e_i$$

可得 $$\hat{y}_i = \hat{\beta}_1 x_i \qquad (**)$$

（**）式也称为**样本回归函数**的**离差形式**。

注意：

在计量经济学中，往往以小写字母表示对均值的离差。

三、参数估计的最大或然法(ML)

最大或然法 (**Maximum Likelihood**,简称**ML**)，也称**最大似然法**，是不同于最小二乘法的另一种参数估计方法，是从最大或然原理出发发展起来的其它估计方法的基础。

基本原理:

对于**最大或然法**，当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大。

在满足基本假设条件下，对一元线性回归模型：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

随机抽取n组样本观测值（X_i, Y_i）（i=1,2,…n）。

假如模型的参数估计量已经求得，为 $\hat{\beta}_0$ 、$\hat{\beta}_1$

那么Y_i服从如下的正态分布：

$$Y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 X_i, \sigma^2)$$

于是，Y的概率函数为

$$P(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2} \qquad \text{（i=1,2,…n）}$$

因为Y_i是相互独立的，所以的所有样本观测值的联合概率，也即**或然函数**(likelihood function)为：

$$L(\hat{\beta}_0,\hat{\beta}_1,\sigma^2)=P(Y_1,Y_2,\cdots,Y_n)$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}e^{-\frac{1}{2\sigma^2}\Sigma(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2}$$

将该或然函数极大化，即可求得到模型参数的极大或然估计量。

由于或然函数的极大化与或然函数的对数的极大化是等价的，所以，取对数或然函数如下：

$$
\begin{aligned}
L^* &= \ln(L) \\
&= -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2
\end{aligned}
$$

对 L^* 求极大值，等价于对 $\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$ 求极小值:

$$
\begin{cases}
\dfrac{\partial}{\partial\hat{\beta}_0}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0 \\
\dfrac{\partial}{\partial\hat{\beta}_1}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0
\end{cases}
$$

解得模型的参数估计量为：

$$\begin{cases} \hat{\beta}_0 = \dfrac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma Y_i X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \\ \hat{\beta}_1 = \dfrac{n\Sigma Y_i X_i - \Sigma Y_i \Sigma X_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \end{cases}$$

可见，在满足一系列基本假设的情况下，模型结构参数的**最大或然估计量**与**普通最小二乘估计量**是相同的。

例2.2.1：在上述家庭**可支配收入-消费支出**例中，对于所抽出的一组样本数，参数估计的计算可通过下面的表2.2.1进行。

表 2.2.1　参数估计的计算表

	X_i	Y_i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	X_i^2	Y_i^2
1	800	594	-1350	-973	1314090	1822500	947508	640000	352836
2	1100	638	-1050	-929	975870	1102500	863784	1210000	407044
3	1400	1122	-750	-445	334050	562500	198381	1960000	1258884
4	1700	1155	-450	-412	185580	202500	170074	2890000	1334025
5	2000	1408	-150	-159	23910	22500	25408	4000000	1982464
6	2300	1595	150	28	4140	22500	762	5290000	2544025
7	2600	1969	450	402	180720	202500	161283	6760000	3876961
8	2900	2078	750	511	382950	562500	260712	8410000	4318084
9	3200	2585	1050	1018	1068480	1102500	1035510	10240000	6682225
10	3500	2530	1350	963	1299510	1822500	926599	12250000	6400900
求和	21500	15674			5769300	7425000	4590020	53650000	29157448
平均	2150	1567							

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{5769300}{7425000} = 0.777$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_0 \overline{X} = 1567 - 0.777 \times 2150 = -103.172$$

因此，由该样本估计的回归方程为：

$$\hat{Y}_i = -103.172 + 0.777 X_i$$

四、最小二乘估计量的性质

当模型参数估计出后，需考虑参数估计值的精度，即是否能代表总体参数的真值，或者说需考察参数估计量的统计性质。

一个用于考察总体的估计量，可从如下几个方面考察其优劣性：

（1）**线性性**，即它是否是另一随机变量的线性函数；

（2）**无偏性**，即它的均值或期望值是否等于总体的真实值；

（3）**有效性**，即它是否在所有线性无偏估计量中具有最小方差。

这三个准则也称作估计量的**小样本性质。**

拥有这类性质的估计量称为**最佳线性无偏估计量**（**best liner unbiased estimator, BLUE**）。

当不满足小样本性质时，需进一步考察估计量的**大样本**或**渐近性质**：

（4）渐近无偏性，即样本容量趋于无穷大时，是否它的均值序列趋于总体真值；

（5）一致性，即样本容量趋于无穷大时，它是否依概率收敛于总体的真值；

（6）渐近有效性，即样本容量趋于无穷大时，是否它在所有的一致估计量中具有最小的渐近方差。

高斯—马尔可夫定理(Gauss-Markov theorem)

在给定经典线性回归的假定下，最小二乘估计量是具有最小方差的线性无偏估计量。

1、**线性性**，即估计量$\hat{\beta}_0$ 、$\hat{\beta}_1$ 是Y_i 的线性组合。

证： $$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} + \frac{\bar{Y}\sum x_i}{\sum x_i^2}$$

令 $k_i = \dfrac{x_i}{\sum x_i^2}$， 因 $\sum x_i = \sum (X_i - \bar{X}) = 0$， 故有

$$\hat{\beta}_1 = \sum \frac{x_i}{\sum x_i^2} Y_i = \sum k_i Y_i$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{1}{n}\sum Y_i - \sum k_i Y_i \bar{X} = \sum (\frac{1}{n} - \bar{X} k_i) Y_i = \sum w_i Y_i$$

2、无偏性，即估计量$\hat{\beta}_0$ 、$\hat{\beta}_1$ 的均值（期望）等于总体回归参数真值β_0与β_1

证： $\hat{\beta}_1 = \sum k_i Y_i = \sum k_i (\beta_0 + \beta_1 X_i + \mu_i) = \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i \mu_i$

易知 $\sum k_i = \dfrac{\sum x_i}{\sum x_i^2} = 0$ $\qquad \sum k_i X_i = 1$

故 $\hat{\beta}_1 = \beta_1 + \sum k_i \mu_i$

$$E(\hat{\beta}_1) = E(\beta_1 + \sum k_i \mu_i) = \beta_1 + \sum k_i E(\mu_i) = \beta_1$$

同样地，容易得出

$$E(\hat{\beta}_0) = E(\beta_0 + \sum w_i \mu_i) = E(\beta_0) + \sum w_i E(\mu_i) = \beta_0$$

3、有效性（最小方差性），即在所有线性无偏估计量中，最小二乘估计量$\hat{\beta}_0$ 、$\hat{\beta}_1$ 具有最小方差。

(1)先求$\hat{\beta}_0$与$\hat{\beta}_1$ 的方差

$$\operatorname{var}(\hat{\beta}_1) = \operatorname{var}(\sum k_i Y_i) = \sum k_i^2 \operatorname{var}(\beta_0 + \beta_1 X_i + \mu_i) = \sum k_i^2 \operatorname{var}(\mu_i)$$

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

$$\operatorname{var}(\hat{\beta}_0) = \operatorname{var}(\sum w_i Y_i) = \sum w_i^2 \operatorname{var}(\beta_0 + \beta_1 X_i + \mu_i) = \sum (1/n - \bar{X} k_i)^2 \sigma^2$$

$$= \sum \left[\left(\frac{1}{n} \right)^2 - 2\frac{1}{n} \bar{X} k_i + \bar{X}^2 k_i^2 \right] \sigma^2 = \left(\frac{1}{n} - \frac{2}{n} \bar{X} \sum k_i + \bar{X}^2 \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 \right) \sigma^2$$

$$= \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right) \sigma^2 = \frac{\sum x_i^2 + n\bar{X}^2}{n \sum x_i^2} \sigma^2 = \frac{\sum X_i^2}{n \sum x_i^2} \sigma^2$$

（2）证明最小方差性

假设$\hat{\beta}_1^*$是其他估计方法得到的关于β_1的线性无偏估计量：

$$\hat{\beta}_1^* = \sum c_i Y_i$$

其中，$c_i = k_i + d_i$，d_i为不全为零的常数

则容易证明

$$\operatorname{var}(\hat{\beta}_1^*) \geq \operatorname{var}(\hat{\beta}_1)$$

同理，可证明β_0的最小二乘估计量$\hat{\beta}_0$具有最的小方差

普通最小二乘估计量（ordinary least Squares Estimators）称为**最佳线性无偏估计量**（**best linear unbiased estimator, BLUE**）

由于最小二乘估计量拥有一个“好”的估计量所应具备的小样本特性，它自然也拥有大样本特性。

如考察$\hat{\beta}_1$的一致性

$$P\lim(\hat{\beta}_1) = P\lim(\beta_1 + \sum k_i \mu_i) = P\lim(\beta_1) + P\lim(\frac{\sum x_i \mu_i}{\sum x_i^2})$$

$$= \beta_1 + \frac{P\lim(\sum x_i \mu_i / n)}{P\lim(\sum x_i^2 / n)}$$

$$= \beta_1 + \frac{Cov(X, \mu)}{Q} = \beta_1 + \frac{0}{Q} = \beta_1$$

五、参数估计量的概率分布及随机干扰项方差的估计

1、参数估计量$\hat{\beta}_0$和$\hat{\beta}_1$的概率分布

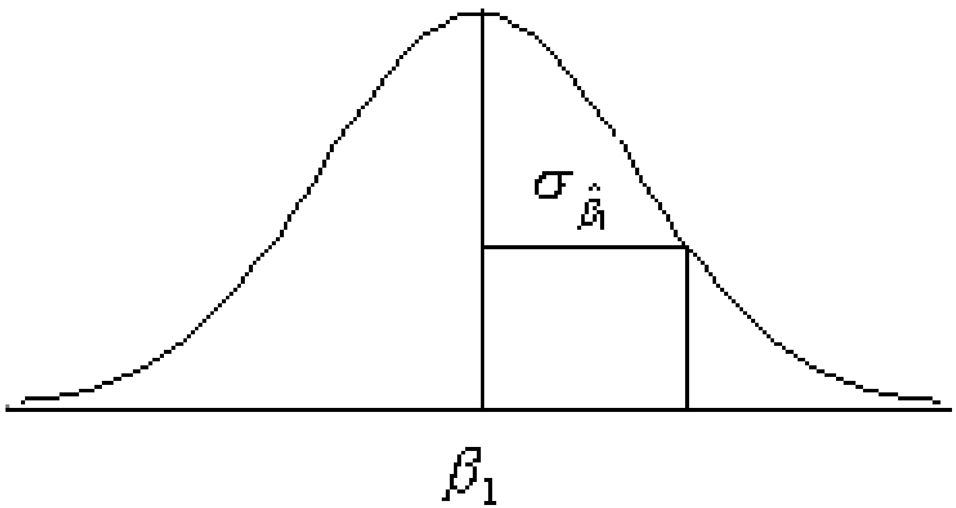
普通最小二乘估计量$\hat{\beta}_0$、$\hat{\beta}_1$分别是Y_i的线性组合，因此，$\hat{\beta}_0$和$\hat{\beta}_1$的概率分布取决于Y的分布特征

在μ是正态分布的假设下，Y是正态分布，则$\hat{\beta}_0$、$\hat{\beta}_1$也服从正态分布，因此

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum x_i^2}) \qquad \hat{\beta}_0 \sim N(\beta_0, \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2)$$

$\hat{\beta}_0$和$\hat{\beta}_1$的标准差

$$\sigma_{\hat{\beta}_1} = \sqrt{\sigma^2 / \sum x_i^2} \qquad \sigma_{\hat{\beta}_0} = \sqrt{\frac{\sigma^2 \sum X_i^2}{n \sum x_i^2}}$$



2、随机误差项μ的方差σ^2的估计

在估计的参数 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差表达式中，都含有随机扰动项 μ 的方差 σ^2。σ^2又称为**总体方差**。

由于σ^2实际上是未知的，因此 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差实际上无法计算，这就需要对其进行估计。

由于随机项$\boldsymbol{\mu}_i$不可观测，只能从$\boldsymbol{\mu}_i$的估计——残差$\boldsymbol{e}_i$出发，对总体方差进行估计。

可以证明，σ^2的**最小二乘估计量**为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2}$$

它是关于σ^2的无偏估计量。

在最大或然估计法中，

解或然方程

$$\frac{\partial}{\partial \sigma^2} L^* = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = 0$$

即可得到 σ^2 的最大或然估计量为：

$$\hat{\sigma}^2 = \frac{1}{n} \Sigma (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{\sum e_i^2}{n}$$

因此， σ^2的最大或然估计量不具无偏性，但却具有一致性。

在随机误差项μ的方差σ^2估计出后，参数$\hat{\beta}_0$和$\hat{\beta}_1$的**方差**和**标准差**的估计量分别是：

$\hat{\beta}_1$的样本方差：　$S_{\hat{\beta}_1}^2 = \hat{\sigma}^2 / \sum x_i^2$

$\hat{\beta}_1$的样本标准差：　$S_{\hat{\beta}_1} = \hat{\sigma} / \sqrt{\sum x_i^2}$

$\hat{\beta}_0$的样本方差：　$S_{\hat{\beta}_0}^2 = \hat{\sigma}^2 \sum X_i^2 / n \sum x_i^2$

$\hat{\beta}_0$的样本标准差：　$S_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\sum X_i^2 / n \sum x_i^2}$

§2.3　一元线性回归模型的统计检验

一、拟合优度检验

二、变量的显著性检验

三、参数的置信区间

- 回归分析是要通过样本所估计的参数来代替总体的真实参数，或者说是用样本回归线代替总体回归线。

- 尽管从统计性质上已知，如果有足够多的重复 抽样，参数的估计值的期望（均值）就等于其总体的参数真值，但在一次抽样中，估计值不一定就等于该真值。
- 那么，在一次抽样中，参数的估计值与真值的差异有多大，是否显著，这就需要进一步进行统计检验。
- 主要包括**拟合优度检验**、变量的**显著性检验**及参数的**区间估计**。

一、拟合优度检验

拟合优度检验：对样本回归直线与样本观测值之间拟合程度的检验。

度量拟合优度的指标：判定系数（可决系数）$\boldsymbol{R}^2$

问题：采用普通最小二乘估计方法，已经保证了模型最好地拟合了样本观测值，为什么还要检验拟合程度？

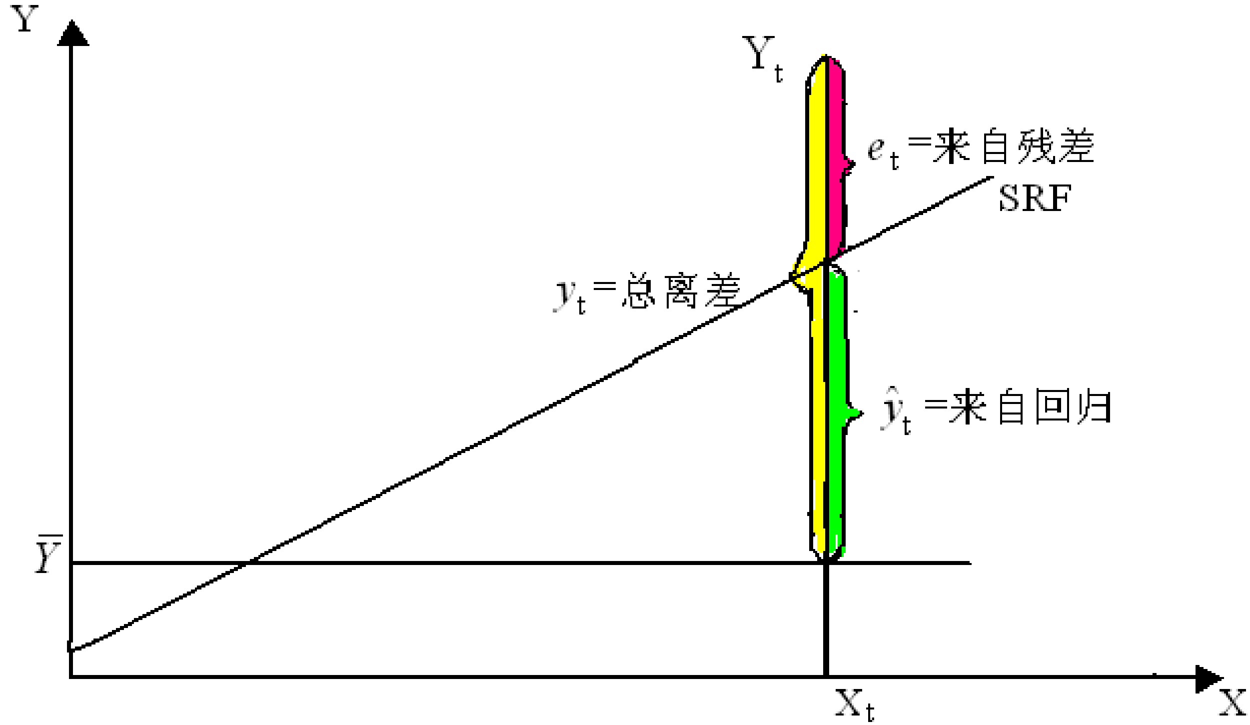
1、总离差平方和的分解

已知由一组样本观测值（**X_i,Y_i**），i=1,2…,n 得到如下样本回归直线

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

而**Y**的第*i*个观测值与样本均值的离差 $y_i = (Y_i - \bar{Y})$ 可分解为两部分之和

$$y_i = Y_i - \bar{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y}) = e_i + \hat{y}_i$$



$\hat{y}_i=(\hat{Y}_i-\bar{Y})$ 是样本回归拟合值与观测值的平均值之差，可认为是**由回归直线解释的部分**；

$e_i=(Y_i-\hat{Y}_i)$ 是实际观测值与回归拟合值之差，是**回归直线不能解释的部分**。

如果 $Y_i=\hat{Y}_i$ 即实际观测值落在样本回归“线”上，则**拟合最好**。可认为，“离差”全部来自回归线，而与“残差”无关。

对于所有样本点，则需考虑这些点与样本均值离差的平方和,可以证明：

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 + 2\sum \hat{y}_i e_i$$

$$= \sum \hat{y}_i^2 + \sum e_i^2$$

记 $TSS = \sum y_i^2 = \sum (Y_i - \bar{Y})^2$ 总体平方和（**Total Sum of Squares**）

$ESS = \sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2$ 回归平方和（**Explained Sum of Squares**）

$RSS = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ 残差平方和（**Residual Sum of Squares** ）

TSS=ESS+RSS

Y的观测值围绕其均值的***总离差*****(total variation)**可分解为两部分：**一部分来自回归线(ESS)，另一部分则来自随机势力(RSS)**。

在给定样本中，TSS不变，

如果实际观测点离样本回归线越近，则**ESS**在**TSS**中占的比重越大，因此

拟合优度：回归平方和ESS/Y的总离差TSS

2、可决系数R^2统计量

记 $$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

称 R^2 为（样本）**可决系数/判定系数**（**coefficient of determination)**。

可决系数的取值范围：[0，1]

R^2越接近1，说明实际观测点离样本线越近，拟合优度越高。

在实际计算可决系数时，在$\hat{\beta}_1$已经估计出后：

$$R^2=\hat{\beta}_1^2\left(\frac{\sum x_i^2}{\sum y_i^2}\right)$$

在例2.1.1的**收入-消费支出**例中，

$$R^2=\hat{\beta}_1^2\frac{\sum x_i^2}{\sum y_i^2}=\frac{(0.777)^2\times 7425000}{4590020}=0.9766$$

注：可决系数是一个非负的统计量。它也是随着抽样的不同而不同。为此，对可决系数的统计可靠性也应进行检验，这将在第3章中进行。

二、变量的显著性检验

回归分析是要判断**解释变量**X是否是**被解释变量**Y的一个显著性的影响因素。

在**一元线性模型**中，就是要判断X是否对Y具有显著的线性性影响。这就需要进行**变量的显著性检验。**

变量的显著性检验所应用的方法是数理统计学中的假设检验。

计量经计学中，主要是针对变量的参数真值是否为零来进行显著性检验的。

1、假设检验

- 所谓**假设检验**，就是事先对总体参数或总体分布形式作出一个假设，然后利用样本信息来判断原假设是否合理，即判断样本信息与原假设是否有显著差异，从而决定是否接受或否定原假设。
- **假设检验采用的逻辑推理方法是反证法。**

先假定原假设正确，然后根据样本信息，观察由此假设而导致的结果是否合理，从而判断是否接受原假设。

- **判断结果合理与否，是基于“小概率事件不易发生”这一原理的**

2、变量的显著性检验

对于一元线性回归方程中的 $\hat{\beta}_1$，已经知道它服从正态分布

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

由于真实的σ^2未知，在用它的无偏估计量 $\hat{\sigma}^2 = \sum e_i^2 / (n-2)$ 替代时，可构造如下统计量

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / \sum x_i^2}} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2)$$

检验步骤：

（1）对总体参数提出假设

H_0： $\beta_1=0$， H_1： $\beta_1\neq 0$

（2）以原假设H0构造t统计量，并由样本计算其值

$$t=\frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

（3）给定显著性水平α，查t分布表，得临界值 $t_{\alpha/2}(n-2)$

(4) 比较，判断

若 $|t|> t_{\alpha/2}(n-2)$，则拒绝H_0，接受H_1；

若 $|t|\leq t_{\alpha/2}(n-2)$，则拒绝$H_1$，接受$H_0$；

对于一元线性回归方程中的β_0，可构造如下t统计量进行显著性检验：

$$t=\frac{\hat{\beta}_0-\beta_0}{\sqrt{\hat{\sigma}^2\sum X_i^2/n\sum x_i^2}}=\frac{\hat{\beta}_0}{S_{\hat{\beta}_0}}\sim t(n-2)$$

在上述收入-消费支出例中，首先计算σ_2的估计值

$$\hat{\sigma}^2=\frac{\sum e_i^2}{n-2}=\frac{\sum y_i^2-\hat{\beta}_1^2\sum x_i^2}{n-2}=\frac{4590020-0.777^2\times 7425000}{10-2}=13402$$

于是$\hat{\beta}_1$和$\hat{\beta}_0$的标准差的估计值分别是：

$$S_{\hat{\beta}_1}=\sqrt{\hat{\sigma}^2/\sum x_i^2}=\sqrt{13402/7425000}=\sqrt{0.0018}=0.0425$$

$$S_{\hat{\beta}_0}=\sqrt{\hat{\sigma}^2\sum X_i^2/n\sum x_i^2}=\sqrt{13402\times 53650000/10\times 7425000}=98.41$$

t统计量的计算结果分别为：

$$t_1 = \hat{\beta}_1 \Big/ S_{\hat{\beta}_1} = 0.777/0.0425 = 18.29$$

$$t_0 = \hat{\beta}_0 \Big/ S_{\hat{\beta}_0} = -103.17/98.41 = -1.048$$

给定显著性水平α=0.05，查t分布表得临界值

$t_{0.05/2}(8)=2.306$

$|t_1|>2.306$，说明**家庭可支配收入在95%的置信度下显著，即是消费支出的主要解释变量；**

$|t_2|<2.306$, 表明在95%的置信度下，无法拒绝截距项为零的假设。

三、参数的置信区间

回归分析希望通过样本所估计出的参数 $\hat{\beta}_1$ 来代替总体的参数 β_1。

假设检验可以通过一次抽样的结果检验总体参数可能的假设值的范围（如是否为零），但它并没有指出在一次抽样中样本参数值到底离总体参数的真值有多"近"。

要判断样本参数的估计值在多大程度上可以"近似"地替代总体参数的真值，往往需要通过构造一个以样本参数的估计值为中心的"区间"，来考察它以多大的可能性（概率）包含着真实的参数值。这种方法就是参数检验的**置信区间估计**。

要判断估计的参数值$\hat{\beta}$离真实的参数值β有多“近”，可预先选择一个概率 α(0<α<1)，并求一个正数 δ，使得随机区间 $(\hat{\beta}-\delta, \hat{\beta}+\delta)$ 包含参数的真值的概率为1-α。即：

$$P(\hat{\beta}-\delta \le \beta \le \hat{\beta}+\delta) = 1-\alpha$$

如果存在这样一个区间，称之为**置信区间**（**confidence interval**）；1-α称为**置信系数**（**置信度**）（**confidence coefficient**），α称为**显著性水平**（**level of significance**）；置信区间的端点称为**置信限**（**confidence limit**）或**临界值**（**critical values**）。

一元线性模型中，β_i (i=1，2）的置信区间:

在变量的显著性检验中已经知道：

$$t = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \sim t(n-2)$$

意味着，如果给定置信度（1-α），从分布表中查得自由度为(n-2)的临界值，那么t值处在($-t_{\alpha/2}$， $t_{\alpha/2}$)的概率是(1-α)。表示为：

$$P(-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}) = 1-\alpha$$

即

$$P(-t_{\frac{\alpha}{2}} < \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} < t_{\frac{\alpha}{2}}) = 1-\alpha$$

$$P(\hat{\beta}_i - t_{\frac{\alpha}{2}} \times s_{\hat{\beta}_i} < \beta_i < \hat{\beta}_i + t_{\frac{\alpha}{2}} \times s_{\hat{\beta}_i}) = 1-\alpha$$

于是得到：(1-α)的置信度下，β_i的置信区间是

$$(\hat{\beta}_i - t_{\frac{\alpha}{2}} \times s_{\hat{\beta}_i}, \hat{\beta}_i + t_{\frac{\alpha}{2}} \times s_{\hat{\beta}_i})$$

在上述收入-消费支出例中，如果给定α =0.01，查表得：

$$t_{\frac{\alpha}{2}}(n-2) = t_{0.005}(8) = 3.355$$

由于 $s_{\hat{\beta}_1} = 0.042$ $S_{\hat{\beta}_0} = 98.41$

于是，β_1、β_0的置信区间分别为：

（0.6345,0.9195）

（-433.32,226.98）

由于置信区间一定程度地给出了样本参数估计值与总体参数真值的“接近”程度，因此置信区间越小越好。

要缩小置信区间，需

（1）增大样本容量n，因为在同样的置信水平下，n越大，t分布表中的临界值越小；同时，增大样本容量，还可使样本参数估计量的标准差减小；

（2）提高模型的拟合优度，因为样本参数估计量的标准差与残差平方和呈正比，模型拟合优度越高，残差平方和应越小。

§2.4 一元线性回归分析的应用：预测问题

一、$\hat{Y}_0$是条件均值$E(Y|X=X_0)$或个值Y_0的一个无偏估计

二、总体条件均值与个值预测值的置信区间

对于一元线性回归模型

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

给定样本以外的解释变量的观测值X_0，可以得到被解释变量的预测值$\hat{Y}_0$，可以此作为其**条件均值**$\mathbf{E(Y|X=X_0)}$或**个别值**$\mathbf{Y_0}$的一个近似估计。

注意：

严格地说，这只是被解释变量的预测值的估计值，而不是预测值。

原因：（1）参数估计量不确定；

（2）随机项的影响

一、$\hat{Y}_0$是条件均值$E(Y|X=X_0)$或个值Y_0的一个无偏估计

对总体回归函数$E(Y|X=X_0)=\beta_0+\beta_1X$，$X=X_0$时

$$E(Y|X=X_0)=\beta_0+\beta_1X_0$$

通过样本回归函数$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$，求得的拟合值为

$$\hat{Y}_0=\hat{\beta}_0+\hat{\beta}_1X_0$$

于是 $E(\hat{Y}_0)=E(\hat{\beta}_0+\hat{\beta}_1X_0)=E(\hat{\beta}_0)+X_0E(\hat{\beta}_1)=\beta_0+\beta_1X_0$

可见，$\hat{Y}_0$是条件均值$E(Y|X=X_0)$的无偏估计。

对总体回归模型$Y=\beta_0+\beta_1X+\mu$，当$X=X_0$时

$$Y_0 = \beta_0 + \beta_1 X_0 + \mu$$

于是

$$E(Y_0) = E(\beta_0 + \beta_1 X_0 + \mu) = \beta_0 + \beta_1 X_0 + E(\mu) = \beta_0 + \beta_1 X_0$$

而通过样本回归函数$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$，求得拟合值

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

的期望为

$$E(\hat{Y}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_0) = E(\hat{\beta}_0) + X_0 E(\hat{\beta}_1) = \beta_0 + \beta_1 X_0$$

$\hat{Y}_0$是个值Y_0的无偏估计。

二、总体条件均值与个值预测值的置信区间

1、总体均值预测值的置信区间

由于 $$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$$

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum x_i^2}) \qquad \hat{\beta}_0 \sim N(\beta_0, \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2)$$

于是 $$E(\hat{Y}_0) = E(\hat{\beta}_0) + X_0 E(\hat{\beta}_1) = \beta_0 + \beta_1 X_0$$

$$Var(\hat{Y}_0) = Var(\hat{\beta}_0) + 2X_0 Cov(\hat{\beta}_0, \hat{\beta}_1) + X_0^2 Var(\hat{\beta}_1)$$

可以证明 $$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \bar{X} / \sum x_i^2$$

因此

$$Var(\hat{Y}_0)=\frac{\sigma^2\sum X_i^2}{n\sum x_i^2}-\frac{2X_0\bar{X}\sigma^2}{\sum x_i^2}+\frac{X_0^2\sigma^2}{\sum x_i^2}$$

$$=\frac{\sigma^2}{\sum x_i^2}\left(\frac{\sum X_i^2-n\bar{X}^2}{n}+\bar{X}^2-2X_0\bar{X}+X_0^2\right)$$

$$=\frac{\sigma^2}{\sum x_i^2}\left(\frac{\sum x_i^2}{n}+(X_0-\bar{X})^2\right)=\sigma^2\left(\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right)$$

故

$$\hat{Y}_0\sim N\left(\beta_0+\beta_1X_0,\sigma^2\left(\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right)\right)$$

将未知的σ^2代以它的无偏估计量$\hat{\sigma}^2$，可构造**t统计量**

$$t=\frac{\hat{Y}_0-(\beta_0+\beta_1X_0)}{S_{\hat{Y}_0}}\sim t(n-2)$$

其中 $S_{\hat{Y}_0}=\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right)}$

于是，在1-α的置信度下，**总体均值E(Y|X$_0$)的置信区间为**

$$\hat{Y}_0-t_{\frac{\alpha}{2}}\times S_{\hat{Y}_0}<E(Y\mid X_0)<\hat{Y}_0+t_{\frac{\alpha}{2}}\times S_{\hat{Y}_0}$$

2、总体个值预测值的预测区间

由 $Y_0=\beta_0+\beta_1X_0+\mu$ 知：

$$Y_0 \sim N(\beta_0+\beta_1 X_0, \sigma^2)$$

于是 $$\hat{Y}_0 - Y_0 \sim N\left(0, \sigma^2\left(1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right)\right)$$

将未知的σ^2代以它的无偏估计量 $\hat{\sigma}^2$，可构造**t统计量**

$$t=\frac{\hat{Y}_0-Y_0}{S_{\hat{Y}_0-Y_0}} \sim t(n-2)$$

式中： $$S_{\hat{Y}_0-Y_0}=\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right)}$$

从而在1-α的置信度下， **Y_0的置信区间**为

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times S_{\hat{Y}_0-Y_0} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times S_{\hat{Y}_0-Y_0}$$

在上述**收入-消费支出**例中，得到的样本回归函数为

$$\hat{Y}_i = -103.172 + 0.777 X_i$$

则在 X_0=1000处， $\hat{\mathbf{Y}}_{\mathbf{0}}$ = –103.172+0.777×1000=673.84

而

$$Var(\hat{Y}_0) = 13402 \left[\frac{1}{10} + \frac{(1000 - 2150)^2}{7425000} \right] = 3727.29$$

$$S(\hat{Y}_0) = 61.05$$

因此，**总体均值**E(Y|X=1000)的95%的置信区间为：

673.84-2.306×61.05< E(Y|X=1000) <673.84+2.306×61.05

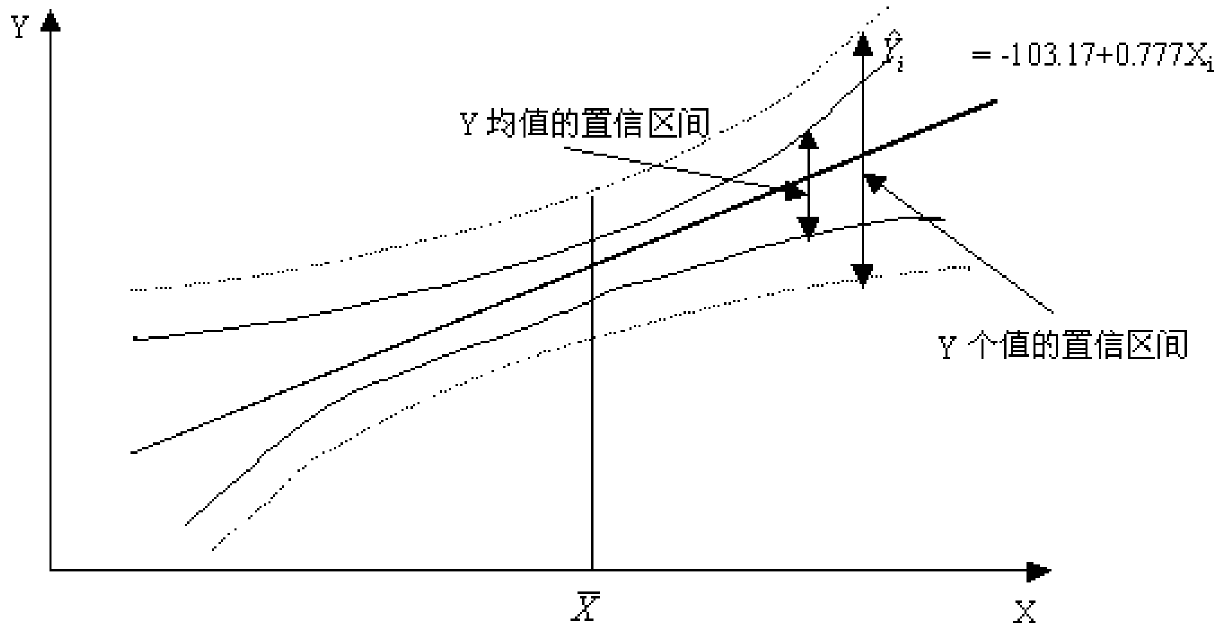
或 （533.05, 814.62）

同样地，对于Y在X=1000的**个体值**，其95%的置信区间为：

$$673.84 - 2.306 \times 61.05 < Y_{x=1000} < 673.84 + 2.306 \times 61.05$$

或 (372.03, 975.65)

- 总体回归函数的**置信带（域）**（confidence band）
- 个体的**置信带（域）**



对于Y的总体均值E(Y|X)与个体值的预测区间（置信区间）：

（1）样本容量n越大，预测精度越高，反之预测精度越低；

（2）样本容量一定时，置信带的宽度当在X均值处最小，其附近进行预测（插值预测）精度越大；X越远离其均值，置信带越宽，预测可信度下降。

§2.5 实例：时间序列问题

一、中国居民人均消费模型

二、时间序列问题

一、中国居民人均消费模型

例2.5.1 考察中国居民收入与消费支出的关系。

GDPP：*人均国内生产总值*（1990年不变价）

CONSP：*人均居民消费*（以居民消费价格指数（1990=100）缩减）。

表 2.5.1　中国居民人均消费支出与人均 GDP（元/人）

年份	人均居民消费 CONSP	人均GDP GDPP	年份	人均居民消费 CONSP	人均GDP GDPP
1978	395.8	675.1	1990	797.1	1602.3
1979	437.0	716.9	1991	861.4	1727.2
1980	464.1	763.7	1992	966.6	1949.8
1981	501.9	792.4	1993	1048.6	2187.9
1982	533.5	851.1	1994	1108.7	2436.1
1983	572.8	931.4	1995	1213.1	2663.7
1984	635.6	1059.2	1996	1322.8	2889.1
1985	716.0	1185.2	1997	1380.9	3111.9
1986	746.5	1269.6	1998	1460.6	3323.1
1987	788.3	1393.6	1999	1564.4	3529.3
1988	836.4	1527.0	2000	1690.8	3789.7
1989	779.7	1565.9			

该两组数据是1978~2000年的**时间序列数据**（time series data）；

前述**收入-消费支出例**中的数据是**截面数据**（cross-sectional data）。

1、建立模型

拟建立如下一元回归模型

$$CONSP = C + \beta\, GDPP + \mu$$

采用**Eviews软件**进行回归分析的结果见下表

表 **2.5.2** 中国居民人均消费支出对人均 **GDP** 的回归（**1978~2000**）

LS // Dependent Variable is CONSP

Sample: 1978 2000

Included observations: 23

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	201.1071	14.88514	13.51060	0.0000
GDPP1	0.386187	0.007222	53.47182	0.0000

R-squared	0.992709	Mean dependent var	905.3331
Adjusted R-squared	0.992362	S.D. dependent var	380.6428
S.E. of regression	33.26711	Akaike info criterion	7.092079
Sum squared resid	23240.71	Schwarz criterion	7.190818
Log likelihood	-112.1945	F-statistic	2859.235
Durbin-Watson stat	0.550288	Prob(F-statistic)	0.000000

一般可写出如下回归分析结果：

$$\widehat{CONSP} = 201.107 + 0.3862GDPP$$

(13.51)　(53.47)

R^2=0.9927　F=2859.23　DW=0.5503

2、模型检验

R^2=0.9927

T值：C：13.51，　GDPP：53.47

临界值: $t_{0.05/2}(21)=2.08$

斜率项：0<0.3862<1，符合绝对收入假说

3、预测

2001年：GDPP=4033.1（元）（90年不变价）

点估计：$CONSP_{2001}$=201.107 + 0.3862×4033.1 = 1758.7（元）

2001年实测的CONSP（1990年价）:1782.2元，

相对误差: -1.32%。

2001年人均居民消费的**预测区间**

人均GDP的**样本均值**与**样本方差**：

$$E(GDPP)=1823.5 \qquad Var(GDPP)=982.04^2=964410.4$$

在95%的置信度下，**$E(CONSP_{2001})$的预测区间**为：

$$1758.7 \pm 2.306 \times \sqrt{\frac{23240.71}{23-2} \times \left(\frac{1}{23}+\frac{(4033.1-1823.5)^2}{(23-1)\times 964410.4}\right)}$$

$=1758.7\pm40.13$

或：（1718.6,1798.8）

同样地，在95%的置信度下，**$CONSP_{2001}$的预测区间**为：

$$1758.7 \pm 2.306 \times \sqrt{\frac{23240.71}{23-2} \times \left(1+\frac{1}{23}+\frac{(4033.1-1823.5)^2}{(23-1)\times 964410.4}\right)}$$

$=1758.7\pm86.57$

或 （1672.1, 1845.3）

二、时间序列问题

上述实例表明，时间序列完全可以进行类似于截面数据的回归分析。

然而，在时间序列回归分析中，有两个需注意的问题：

第一，关于抽样分布的理解问题。

能把表2.5.1中的数据理解为是从某个总体中抽出的一个样本吗？

第二，关于“伪回归问题”（spurious regression problem）。

可决系数R^2，考察被解释变量Y的变化中可由解释变量X的变化“解释”的部分。

这里“解释”能否换为“引起”？

在现实经济问题中，对时间序列数据作回归，即使两个变量间没有任何的实际联系，也往往会得到较高的可决系数，尤其对于具有相同变化趋势（同时上升或下降）的变量，更是如此。

这种现象被称为“伪回归”或“虚假回归”。

第三章 经典单方程计量经济学模型：多元回归

- 多元线性回归模型
- 多元线性回归模型的参数估计
- 多元线性回归模型的统计检验
- 多元线性回归模型的预测
- 回归模型的其他形式
- 回归模型的参数约束

§ 3.1 多元线性回归模型

一、多元线性回归模型

二、多元线性回归模型的基本假定

一、多元线性回归模型

多元线性回归模型:表现在线性回归模型中的解释变量有多个。

一般表现形式:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \qquad i=1,2\ldots,n$$

其中:k为解释变量的数目,β_j称为**回归参数**(regression coefficient)。

习惯上:把**常数项**看成为一**虚变量**的系数,该虚变量的样本观测值始终取1。这样:

模型中解释变量的数目为(k+1)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

也被称为总体回归函数的随机表达形式。它 的非随机表达式为:

$$E(Y_i \mid X_{1i}, X_{2i}, \cdots X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

方程表示：各变量X值固定时Y的平均响应。

β_j也被称为偏回归系数，表示在其他解释变量保持不变的情况下，X_j每变化1个单位时，Y的均值E(Y)的变化;

或者说β_j给出了X_j的单位变化对Y均值的“直接”或“净”（不含其他变量）影响。

总体回归模型n个随机方程的矩阵表达式为

$$\mathbf{Y} = \mathbf{X}\beta + \mu$$

其中

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}_{n\times(k+1)}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}_{n\times 1}$$

样本回归函数：用来估计总体回归函数

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{ki}$$

其**随机表示式**：

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{ki} + e_i$$

e_i称为**残差**或**剩余项**(residuals)，可看成是总体回归函数中随机扰动项μ_i的近似替代。

样本回归函数的**矩阵表达**:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} \qquad \text{或} \qquad \mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}$$

其中：

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} \qquad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

二、多元线性回归模型的基本假定

假设1，解释变量是非随机的或固定的，且各X之间互不相关（无多重共线性）。

假设2，随机误差项具有零均值、同方差及不序列相关性

$$E(\mu_i) = 0$$

$$Var(\mu_i) = E(\mu_i^2) = \sigma^2 \qquad i \neq j \quad i, j = 1,2,\cdots,n$$

$$Cov(\mu_i, \mu_j) = E(\mu_i \mu_j) = 0$$

假设3，解释变量与随机项不相关

$$Cov(X_{ji}, \mu_i) = 0 \qquad j = 1,2\cdots,k$$

假设4，随机项满足正态分布

$$\mu_i \sim N(0, \sigma^2)$$

上述假设的矩阵符号表示式：

假设1，$n\times(k+1)$矩阵**X**是非随机的，且**X**的秩$\rho=k+1$，即**X**满秩。

假设2，
$$E(\mu)=E\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}=\begin{pmatrix}E(\mu_1)\\ \vdots\\ E(\mu_n)\end{pmatrix}=0$$

$$E(\mu\,\mu')=E\left(\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}\begin{pmatrix}\mu_1 & \cdots & \mu_n\end{pmatrix}\right)=E\begin{pmatrix}\mu_1^2 & \cdots & \mu_1\mu_n\\ \vdots & \ddots & \vdots\\ \mu_n\mu_1 & \cdots & \mu_n^2\end{pmatrix}$$

$$=\begin{pmatrix}\operatorname{var}(\mu_1) & \cdots & \operatorname{cov}(\mu_1,\mu_n)\\ \vdots & \ddots & \vdots\\ \operatorname{cov}(\mu_n,\mu_1) & \cdots & \operatorname{var}(\mu_n)\end{pmatrix}=\begin{pmatrix}\sigma^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma^2\end{pmatrix}=\sigma^2\mathbf{I}$$

假设3，E(**X'**$\boldsymbol{\mu}$)=0，即

$$E\begin{pmatrix}\sum\mu_i\\ \sum X_{1i}\mu_i\\ \vdots\\ \sum X_{Ki}\mu_i\end{pmatrix}=\begin{pmatrix}\sum E(\mu_i)\\ \sum X_{1i}E(\mu_i)\\ \vdots\\ \sum X_{Ki}E(\mu_i)\end{pmatrix}=0$$

假设4，向量$\boldsymbol{\mu}$有一多维正态分布，即

$$\mu \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

同一元回归一样，多元回归还具有如下两个重要假设：

假设5，样本容量趋于无穷时，各解释变量的方差趋于有界常数，即$\mathbf{n} \to \infty$时，

$$\frac{1}{n}\sum x_{ji}^2 = \frac{1}{n}\sum (X_{ji} - \bar{X}_j)^2 \to Q_j \qquad \text{或} \qquad \frac{1}{n}\mathbf{x}'\mathbf{x} \to \mathbf{Q}$$

其中：$\mathbf{Q}$为一非奇异固定矩阵，矩阵$\mathbf{x}$是由各解释变量的离差为元素组成的$n \times k$阶矩阵

$$\mathbf{x} = \begin{pmatrix} x_{11} & \cdots & x_{k1} \\ \vdots & \cdots & \vdots \\ x_{1n} & \cdots & x_{kn} \end{pmatrix}$$

假设6，回归模型的设定是正确的。

§3.2 多元线性回归模型的估计

估计目标：结构参数 $\hat{\beta}_j$ 及随机误差项的方差 $\hat{\sigma}^2$

估计方法：OLS、ML或者MM

一、普通最小二乘估计

*二、最大或然估计

*三、矩估计

四、参数估计量的性质

五、样本容量问题

六、估计实例

一、普通最小二乘估计

对于随机抽取的n组观测值 $(Y_i, X_{ji}), i=1,2,\cdots,n, j=0,1,2,\cdots k$

如果**样本函数**的参数估计值已经得到，则有：

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_{Ki} \qquad \text{i=1,2…n}$$

根据**最小二乘原理**，参数估计值应该是下列方程组的解

$$\begin{cases} \dfrac{\partial}{\partial \hat{\beta}_0} Q = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_1} Q = 0 \\ \dfrac{\partial}{\partial \hat{\beta}_2} Q = 0 \\ \quad \vdots \\ \dfrac{\partial}{\partial \hat{\beta}_k} Q = 0 \end{cases}$$

其中 $Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$

$$= \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}))^2$$

于是得到关于待估参数估计值的**正规方程组**：

$$\begin{cases}\Sigma(\hat{\beta}_0+\hat{\beta}_1 X_{1i}+\hat{\beta}_2 X_{2i}+\cdots+\hat{\beta}_k X_{ki})=\Sigma Y_i \\ \Sigma(\hat{\beta}_0+\hat{\beta}_1 X_{1i}+\hat{\beta}_2 X_{2i}+\cdots+\hat{\beta}_k X_{ki})X_{1i}=\Sigma Y_i X_{1i} \\ \Sigma(\hat{\beta}_0+\hat{\beta}_1 X_{1i}+\hat{\beta}_{2i} X_{2i}+\cdots+\hat{\beta}_k X_{ki})X_{2i}=\Sigma Y_i X_{2i} \\ \vdots \\ \Sigma(\hat{\beta}_0+\hat{\beta}_1 X_{1i}+\hat{\beta}_2 X_{2i}+\cdots+\hat{\beta}_k X_{ki})X_{ki}=\Sigma Y_i X_{ki}\end{cases}$$

解该（k+1）个方程组成的线性代数方程组，即可得到(k+1)个待估参数的估计值$\hat{\beta}_j, j=0,1,2,\cdots,k$ 。

正规方程组的矩阵形式

$$\begin{pmatrix} n & \sum X_{1i} & \cdots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \cdots & \sum X_{1i}X_{ki} \\ \cdots & \cdots & \cdots & \cdots \\ \sum X_{ki} & \sum X_{ki}X_{1i} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix}$$

即

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{Y}$$

由于**X’X**满秩，故有

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

将上述过程用矩阵表示如下：

寻找一组参数估计值 $\hat{\beta}$，使得残差平方和

$$Q=\sum_{i=1}^{n}e_i^2=\mathbf{e}'\mathbf{e}=(\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta})$$

最小。

即求解方程组： $\dfrac{\partial}{\partial\hat{\beta}}(\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta})=0$

$$\frac{\partial}{\partial\hat{\beta}}(\mathbf{Y}'\mathbf{Y}-\hat{\beta}'\mathbf{X}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\hat{\beta}+\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})=0$$

$$\frac{\partial}{\partial\hat{\beta}}(\mathbf{Y}'\mathbf{Y}-2\mathbf{Y}'\mathbf{X}\hat{\beta}+\hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta})=0$$

$$-\mathbf{X}'\mathbf{Y}+\mathbf{X}'\mathbf{X}\hat{\beta}=0$$

得到： $\mathbf{X}'\mathbf{Y}=\mathbf{X}'\mathbf{X}\hat{\beta}$

于是： $\hat{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

例3.2.1: 在例2.1.1的**家庭收入-消费支出**例中,

$$(\mathbf{X}'\mathbf{X})=\begin{pmatrix}1 & 1 & \cdots & 1\\ X_1 & X_2 & \cdots & X_n\end{pmatrix}\begin{pmatrix}1 & X_1\\ 1 & X_2\\ \cdots & \cdots\\ 1 & X_n\end{pmatrix}=\begin{pmatrix}n & \sum X_i\\ \sum X_i & \sum X_i^2\end{pmatrix}=\begin{pmatrix}10 & 21500\\ 21500 & 53650000\end{pmatrix}$$

$$\mathbf{X}'\mathbf{Y}=\begin{pmatrix}1 & 1 & \cdots & 1\\ X_1 & X_2 & \cdots & X_n\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \cdots\\ Y_n\end{pmatrix}=\begin{pmatrix}\sum Y_i\\ \sum X_iY_i\end{pmatrix}=\begin{pmatrix}15674\\ 39468400\end{pmatrix}$$

可求得

$$(\mathbf{X}'\mathbf{X})^{-1}=\begin{pmatrix}0.7226 & -0.0003\\ -0.0003 & 1.35E-07\end{pmatrix}$$

于是

$$\hat{\beta}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}0.7226 & -0.0003\\ -0.0003 & 1.35E-07\end{pmatrix}\begin{pmatrix}15674\\ 39648400\end{pmatrix}=\begin{pmatrix}-103.172\\ 0.7770\end{pmatrix}$$

◇正规方程组的另一种写法

对于**正规方程组**

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\hat{\beta}$$

将 $\mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}$ 代入得

$$\mathbf{X}'\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{e} = \mathbf{X}'\mathbf{X}\hat{\beta}$$

于是

$$\mathbf{X}'\mathbf{e} = \mathbf{0} \qquad (*)$$

或

$$\begin{cases} \sum e_i = 0 \\ \sum_i X_{ji} e_i = 0 \end{cases} \qquad (**)$$

(*)或（**）是多元线性回归模型**正规方程组**的另一种写法

样本回归函数的离差形式

$$y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki} + e_i \qquad \text{i=1,2…n}$$

其矩阵形式为

$$\mathbf{y} = \mathbf{x}\hat{\beta} + \mathbf{e}$$

其中：

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

在离差形式下，参数的最小二乘估计结果为

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{Y}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \cdots - \hat{\beta}_k \bar{X}_k$$

◇随机误差项μ的方差σ的无偏估计

可以证明，随机误差项μ的方差的无偏估计量为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$$

*二、最大或然估计

对于多元线性回归模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

易知 $Y_i \sim N(\mathbf{X_i \beta}, \sigma^2)$

Y的随机抽取的n组样本观测值的联合概率

$$L(\hat{\boldsymbol{\beta}}, \sigma^2) = P(Y_1, Y_2, \cdots, Y_n)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2} \Sigma (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}))^2}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})}$$

即为变量Y的**或然函数**

对数或然函数为

$$
\begin{aligned}
L^* &= Ln(L) \\
&= -nLn(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\hat{\beta})\ (\mathbf{Y}-\mathbf{X}\hat{\beta})
\end{aligned}
$$

对对数或然函数求极大值，也就是对

$$(\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta})$$

求极小值。

因此，参数的**最大或然估计**为

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

结果与参数的普通最小二乘估计相同

*三、矩估计（Moment Method, MM）

OLS估计是通过得到一个关于参数估计值的**正规方程组**

$$(\mathbf{X}'\mathbf{X})\hat{\beta} = \mathbf{X}'\mathbf{Y}$$

并对它进行求解而完成的。

该正规方程组 可以从另外一种思路来导：

$$\mathbf{Y} = \mathbf{X}\beta + \mu$$

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mu$$

$$\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{X}'\boldsymbol{\mu}$$

求期望： $E(\mathbf{X}'(\mathbf{Y} - \mathbf{X}\beta)) = \mathbf{0}$

$$E(\mathbf{X}'(\mathbf{Y}-\mathbf{X}\beta))=\mathbf{0}$$

称为原总体回归方程的一组**矩条件**，表明了原总体回归方程所具有的内在特征。

如果随机抽出原总体的一个样本，估计出的样本回归方程

$$\hat{\mathbf{Y}}=\mathbf{X}\hat{\beta}$$

能够近似代表总体回归方程的话，则应成立：

$$\frac{1}{n}\mathbf{X}'(\mathbf{Y}-\mathbf{X}\hat{\beta})=0$$

由此得到**正规方程组**

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{Y}$$

解此正规方程组即得参数的MM估计量。

易知MM估计量与OLS、ML估计量等价。

矩方法是工具变量方法(Instrumental Variables,IV)和广义矩估计方法(Generalized Moment Method, GMM)的基础

- 在矩方法中关键是利用了

$$\mathrm{E}(\boldsymbol{X'\mu})=\mathbf{0}$$

- 如果某个解释变量与随机项相关，只要能找到1个工具变量，仍然可以构成一组矩条件。这就是IV。

- 如果存在＞k+1个变量与随机项不相关，可以构成一组包含＞k+1方程的矩条件。这就是GMM。

四、参数估计量的性质

在满足基本假设的情况下，其结构参数β的普通最小二乘估计、最大或然估计及矩估计仍具有：

线性性、无偏性、有效性。

同时，随着样本容量增加，参数估计量具有：

渐近无偏性、渐近有效性、一致性。

1、线性性

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{CY}$$

其中, **C=(X'X)**$^{-1}$ **X'** 为一仅与固定的**X**有关的行向量

2、无偏性

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\
&= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\mu})) \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}E(\mathbf{X}'\boldsymbol{\mu}) \\
&= \boldsymbol{\beta}
\end{aligned}
$$

这里利用了假设：E(*X'μ*)=**0**

3、有效性（最小方差性）

参数估计量 $\hat{\beta}$ 的**方差-协方差矩阵**

$$
\begin{aligned}
Cov(\hat{\beta}) &= E(\hat{\beta}-E(\hat{\beta}))(\hat{\beta}-E(\hat{\beta}))' \\
&= E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'
\end{aligned}
$$

$$= E((\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}'\mu\mu'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mu\mu')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

$$= E(\mu\mu')(\mathbf{X}'\mathbf{X})^{-1}$$

$$= \sigma^2\mathbf{I}(\mathbf{X}'\mathbf{X})^{-1}$$

$$= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

其中利用了

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mu)$$

$$= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu$$

和 $$E(\mu\mu') = \sigma^2\mathbf{I}$$

根据高斯—马尔可夫定理，$Cov(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ 在所有无偏估计量的方差中是最小的。

五、样本容量问题

1. 最小样本容量

所谓“**最小样本容量**”，即从最小二乘原理和最大或然原理出发，欲得到参数估计量，不管其质量如何，所要求的样本容量的下限。

样本最小容量必须不少于模型中解释变量的数目（包括常数项），即

$$n \geq k+1$$

因为，无多重共线性要求：秩$(\mathbf{X})=k+1$

2、满足基本要求的样本容量

从统计检验的角度：

n>30 时，Z检验才能应用；

n-k≥8时， t分布较为稳定

一般经验认为：

当n≥30或者至少n≥3(k+1)时，才能说满足模型估计的基本要求。

模型的良好性质只有在大样本下才能得到理论上的证明

六、多元线性回归模型的参数估计实例

例3.2.2 在例2.5.1中，已建立了**中国居民人均消费**一元线性模型。这里我们再考虑建立多元线性模型。

解释变量： 人均GDP：GDPP

前期消费：CONSP(-1)

估计区间： 1979~2000年

Eviews软件估计结果

LS // Dependent Variable is CONS

Sample(adjusted): 1979 2000

Included observations: 22 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	120.7000	36.51036	3.305912	0.0037
GDPP	0.221327	0.060969	3.630145	0.0018
CONSP(-1)	0.451507	0.170308	2.651125	0.0158

R-squared	0.995403	Mean dependent var	928.4946
Adjusted R-squared	0.994920	S.D. dependent var	372.6424
S.E. of regression	26.56078	Akaike info criterion	6.684995
Sum squared resid	13404.02	Schwarz criterion	6.833774
Log likelihood	-101.7516	F-statistic	2057.271
Durbin-Watson stat	1.278500	Prob(F-statistic)	0.000000

随机误差项的方差的估计值为$\hat{\sigma}^2 = 13404.02/(22-3) = 705.47$

§3.3 多元线性回归模型的统计检验

一、拟合优度检验

二、方程的显著性检验(F检验)

三、变量的显著性检验（t检验）

四、参数的置信区间

一、拟合优度检验

1、可决系数与调整的可决系数

总离差平方和的分解

记 $TSS=\sum(Y_i-\bar{Y})^2$ 总离差平方和

$ESS=\sum(\hat{Y}_i-\bar{Y})^2$ 回归平方和

$RSS=\sum(Y_i-\hat{Y}_i)^2$ 剩余平方和

则
$$\begin{aligned}TSS&=\Sigma(Y_i-\bar{Y})^2\\&=\Sigma((Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y}))^2\\&=\Sigma(Y_i-\hat{Y}_i)^2+2\Sigma(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})+\Sigma(\hat{Y}_i-\bar{Y})^2\end{aligned}$$

由于 $\sum (Y_i - \hat{Y})(\hat{Y}_i - \bar{Y}) = \sum e_i (\hat{Y}_i - \bar{Y})$

$$= \hat{\beta}_0 \sum e_i + \hat{\beta}_1 \sum e_i X_{1i} + \cdots + \hat{\beta}_k \sum e_i X_{ki} + \bar{Y} \sum e_i$$

$=0$

所以有：

$$TSS = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2 = RSS + ESS$$

注意：一个有趣的现象

$$\left(Y_i - \bar{Y}\right) = \left(Y_i - \hat{Y}_i\right) + \left(\hat{Y}_i - \bar{Y}\right)$$

$$\left(Y_i - \bar{Y}\right)^2 \neq \left(Y_i - \hat{Y}_i\right)^2 + \left(\hat{Y}_i - \bar{Y}\right)^2$$

$$\sum \left(Y_i - \bar{Y}\right)^2 = \sum \left(Y_i - \hat{Y}_i\right)^2 + \sum \left(\hat{Y}_i - \bar{Y}\right)^2$$

可决系数

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

该统计量越接近于1，模型的拟合优度越高。

问题：

在应用过程中发现，如果在模型中增加一个解释变量，R^2往往增大（Why?）

这就给人一个错觉：***要使得模型拟合得好，只要增加解释变量即可***。

但是，现实情况往往是，由增加解释变量个数引起的R^2的增大与拟合好坏无关，**R^2需调整。**

调整的可决系数（adjusted coefficient of determination）

在样本容量一定的情况下，增加解释变量必定使得自由度减少，所以调整的思路是:**将残差平方和与总离差平方和分别除以各自的自由度，以剔除变量个数对拟合优度的影响:**

$$\bar{R}^2 = 1 - \frac{RSS/(n-k-1)}{TSS/(n-1)}$$

其中：n-k-1为残差平方和的自由度，n-1为总体平方和的自由度。

$\bar{R}^2$与R^2之间存在如下关系：

$$\bar{R}^2 = 1-(1-\bar{R}^2)\frac{n-1}{n-k-1}$$

在中国居民消费支出的二元模型例中，$\bar{R}^2$ =0.9954

在中国居民消费支出的一元模型例中，$\bar{R}^2$ =0.9927

问题：$\bar{R}^2$ 多大才算通过拟合优度检验？

*2、赤池信息准则和施瓦茨准则

为了比较所含解释变量个数不同的多元回归模型的拟合优度，常用的标准还有：

赤池信息准则（Akaike information criterion, AIC）

$$AIC = \ln \frac{\mathbf{e}'\mathbf{e}}{n} + \frac{2(k+1)}{n}$$

施瓦茨准则（Schwarz criterion，SC）

$$AC = \ln \frac{\mathbf{e}'\mathbf{e}}{n} + \frac{k}{n} \ln n$$

这两准则均要求仅当所增加的解释变量能够减少AIC值或AC值时才在原模型中增加该解释变量。

Eviews的估计结果显示:

中国居民消费一元例中:

AIC=6.68　　　AC=6.83

中国居民消费二元例中:

AIC=7.09　　　AC=7.19

从这点看，可以说前期人均居民消费CONSP(-1)应包括在模型中。

二、方程的显著性检验(F检验)

方程的显著性检验，旨在对模型中被解释变量与解释变量之间的线性关系在总体上是否显著成立作出推断。

1、方程显著性的F检验

即检验模型

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+\mu_i \qquad i=1,2,\ldots,n$$

中的参数β_j是否显著不为0。

可提出如下原假设与备择假设：

H_0： $\beta_0=\beta_1=\beta_2=\ldots=\beta_k=0$

H_1： β_j不全为0

F检验的思想来自于总离差平方和的分解式：

TSS=ESS+RSS

由于回归平方和 $ESS=\sum \hat{y}_i^2$ 是解释变量 X 的联合体对被解释变量 Y 的线性作用的结果，考虑比值

$$ESS/RSS=\sum \hat{y}_i^2 \Big/ \sum e_i^2$$

如果这个比值较大，则X的联合体对Y的解释程度高，可认为总体存在线性关系，反之总体上可能不存在线性关系。

因此，可通过该比值的大小对总体线性关系进行推断。

根据数理统计学中的知识，在原假设H_0成立的条件下，统计量

$$F=\frac{ESS/k}{RSS/(n-k-1)}$$

服从自由度为$(k, n-k-1)$的F分布

给定显著性水平α，可得到临界值$F_\alpha(k,n-k-1)$，由样本求出统计量**F**的数值，通过

$F > F_\alpha(k,n-k-1)$ 或 $F \le F_\alpha(k,n-k-1)$

来拒绝或接受原假设H_0，以判定原方程**总体上**的线性关系是否显著成立。

对于中国居民人均消费支出的例子:

一元模型：F=285.92

二元模型：F=2057.3

给定显著性水平α =0.05，查分布表，得到临界值:

一元例：$F_{\alpha}(1,21)=4.32$

二元例：$F_{\alpha}(2,19)=3.52$

显然有 $F > F_{\alpha}(k,n-k-1)$

即二个模型的线性关系在95%的水平下显著成立。

2、关于拟合优度检验与方程显著性检验关系的讨论

由 $\bar{R}^2 = 1 - \dfrac{RSS/(n-k-1)}{TSS/(n-1)}$ 与 $F = \dfrac{ESS/k}{RSS/(n-k-1)}$

可推出：$\bar{R}^2 = 1 - \dfrac{n-1}{n-k-1+kF}$

或 $F = \dfrac{\bar{R}^2/k}{(1-\bar{R}^2)/(n-k-1)}$

F 与 $\bar{R}^2$ 同向变化：当 $\bar{R}^2$ ＝ 0 时， F ＝ 0 ；

$\bar{R}^2$ 越大， F 值也越大；

当 $\bar{R}^2$ ＝ 1 时， F 为无穷大。

因此，F检验是所估计回归的总显著性的一个度量，也是 $\bar{R}^2$ 的一个显著性检验。亦即

检验 $H_0: \beta_1 = 0, \beta_2 = 0, \cdots, \beta_k = 0$ 等价于检验 $\bar{R}^2 = 0$

回答前面的问题：$\bar{R}^2$ 多大才算通过拟合优度检验

- 在中国居民人均收入-消费一元模型中，

F>4.32 → $\bar{R}^2$>0.131 → 模型在95%的水平下显著成立

- 在中国居民人均收入-消费二元模型中，

F>3.52 → $\bar{R}^2$>0.194 → 模型在95%的水平下显著成立

三、变量的显著性检验（t检验）

方程的**总体线性**关系显著≠**每个解释变量**对被解释变量的影响都是显著的

因此，必须对每个解释变量进行显著性检验，以决定是否作为解释变量被保留在模型中。

这一检验是由对变量的 t 检验完成的。

1、t统计量

由于

$$Cov(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

以c_{ii}表示矩阵$(X'X)^{-1}$主对角线上的第i个元素，于是参数估计量的方差为：

$$Var(\hat{\beta}_i) = \sigma^2 c_{ii}$$

其中σ^2为随机误差项的方差，在实际计算时，用它的估计量代替：

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$$

易知 $\hat{\beta}$ 服从如下正态分布

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 c_{ii})$$

因此，可构造如下t统计量

$$t = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}} \frac{\hat{\beta}_i - \beta_i}{\sqrt{c_{ii} \dfrac{\mathbf{e}'\mathbf{e}}{n-k-1}}} \sim t(n-k-1)$$

2、t检验

设计原假设与备择假设：

H_0： $\beta_i=0$ （i=1,2…k）

H_1： $\beta_i\neq 0$

给定显著性水平α，可得到临界值$t_{\alpha/2}(n-k-1)$，由样本求出统计量**t**的数值，通过

$$|t| > t_{\alpha/2}(n-k-1) \quad 或 \quad |t| \leq t_{\alpha/2}(n-k-1)$$

来拒绝或接受原假设H_0，从而**判定对应的解释变量是否应包括在模型中。**

注意：一元线性回归中，t检验与F检验一致

一方面，t检验与F检验都是对相同的原假设 **H_0：$\beta_1=0$** 进行检验；

另一方面，两个统计量之间有如下关系：

$$F=\frac{\sum \hat{y}_i^2}{\sum e_i^2/(n-2)}=\frac{\hat{\beta}_1^2\sum x_i^2}{\sum e_i^2/(n-2)}=\frac{\hat{\beta}_1^2}{\sum e_i^2/(n-2)\sum x_i^2}$$

$$=\left(\frac{\hat{\beta}_1}{\sqrt{\sum e_i^2/(n-2)\sum x_i^2}}\right)^2=\left(\hat{\beta}_1\bigg/\sqrt{\frac{\sum e_i^2}{n-2}\cdot\frac{1}{\sum x_i^2}}\right)^2=t^2$$

在中国居民人均收入-消费支出二元模型例中，由应用软件计算出参数的t值:

$$|t_0| = 3.306 \qquad |t_1| = 3.630 \qquad |t_2| = 2.651$$

给定显著性水平α=0.05，查得相应临界值: $t_{0.025}(19)$ =2.093。

可见，**计算的所有t值都大于该临界值**，所以拒绝原假设。即:

包括常数项在内的3个解释变量都在95%的水平下显著，都通过了变量显著性检验。

四、参数的置信区间

参数的置信区间用来考察：**在一次抽样中所估计的参数值离参数的真实值有多“近”。**

在变量的显著性检验中已经知道：

$$t=\frac{\hat{\beta}_i-\beta_i}{S_{\hat{\beta}_i}}\frac{\hat{\beta}_i-\beta_i}{\sqrt{c_{ii}\frac{\mathbf{e}'\mathbf{e}}{n-k-1}}}\sim t(n-k-1)$$

容易推出：在(1-α)的置信水平下β_i的置信区间是

$$(\hat{\beta}_i-t_{\frac{\alpha}{2}}\times s_{\hat{\beta}_i},\hat{\beta}_i+t_{\frac{\alpha}{2}}\times s_{\hat{\beta}_i})$$

其中，$t_{\alpha/2}$为显著性水平为α 、自由度为n-k-1的临界值。

在中国居民人均收入-消费支出二元模型例中，

给定α=0.05，查表得临界值：$t_{0.025}(19)=2.093$

从回归计算中已得到：

$$\hat{\beta}_0 = 120.70 \qquad s_{\hat{\beta}_0} = 36.51$$

$$\hat{\beta}_1 = 0.2213 \qquad s_{\hat{\beta}_1} = 0.061$$

$$\hat{\beta}_2 = 0.4515 \qquad s_{\hat{\beta}_2} = 0.170$$

计算得参数的置信区间：

$$\beta_0: \quad (44.284, 197.116)$$

$$\beta_1: \quad (0.0937, 0.3489)$$

$$\beta_2: \quad (0.0951, 0.8080)$$

如何才能缩小置信区间？

- 增大样本容量n，因为在同样的样本容量下，n越大，t分布表中的临界值越小，同时，增大样本容量，还可使样本参数估计量的标准差减小；

- 提高模型的拟合优度，因为样本参数估计量的标准差与残差平方和呈正比，模型优度越高，残差平方和应越小。

- 提高样本观测值的分散度，一般情况下，样本观测值越分散，$(X'X)^{-1}$的分母的$|X'X|$的值越大，致使区间缩小。

§3.4 多元线性回归模型的预测

一、$E(Y_0)$的置信区间

二、Y_0的置信区间

对于模型

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$$

给定样本以外的解释变量的观测值 $\mathbf{X}_0=(1,X_{10},X_{20},\ldots,X_{k0})$，可以得到被解释变量的预测值：

$$\hat{Y}_0 = \mathbf{X}_0\hat{\beta}$$

它可以是总体均值 $E(Y_0)$ 或个值 Y_0 的预测。

但严格地说，这只是被解释变量的预测值的估计值，而不是预测值。

为了进行科学预测，还需求出预测值的置信区间，包括 $E(Y_0)$ 和 Y_0 的置信区间。

一、E(Y_0)的置信区间

易知

$$E(\hat{Y}_0) = E(\mathbf{X_0}\hat{\beta}) = \mathbf{X_0}E(\hat{\beta}) = \mathbf{X_0}\beta = E(Y_0)$$

$$Var(\hat{Y}_0) = E(\mathbf{X_0}\hat{\beta} - \mathbf{X_0}\beta)^2 = E(\mathbf{X_0}(\hat{\beta} - \beta)\mathbf{X_0}(\hat{\beta} - \beta))$$

由于$\mathbf{X_0}(\hat{\beta} - \beta)$为标量，因此

$$\begin{aligned} Var(\hat{Y}_0) &= E(\mathbf{X_0}(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\mathbf{X}_0') \\ &= \mathbf{X_0}E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\mathbf{X}_0' \\ &= \sigma^2\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0' \end{aligned}$$

容易证明

$$\hat{Y}_0 \sim N(\mathbf{X_0}\beta, \sigma^2 \mathbf{X_0}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0')$$

取随机扰动项的样本估计量$\hat{\sigma}^2$，构造如下t统计量

$$\frac{\hat{Y}_0 - E(Y_0)}{\hat{\sigma}\sqrt{\mathbf{X_0}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X_0}'}} \sim t(n-k-1)$$

于是，得到(1−α)的置信水平下E(Y_0)的**置信区间**：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'} < E(Y_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'}$$

其中，$t_{\alpha/2}$为(1−α)的置信水平下的**临界值**。

二、Y_0的置信区间

如果已经知道实际的预测值Y_0，那么预测误差为：

$$e_0 = Y_0 - \hat{Y}_0$$

容易证明

$$\begin{aligned} E(e_0) &= E(\mathbf{X}_0\beta + \mu_0 - \mathbf{X}_0\hat{\beta}) \\ &= E(\mu_0 - \mathbf{X}_0(\hat{\beta} - \beta)) \\ &= E(\mu_0 - \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Var(e_0) &= E(e_0^2) \\ &= E(\mu_0 - \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mu)^2 \\ &= \sigma^2(1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0') \end{aligned}$$

e_0服从正态分布，即

$$e_0 \sim N(0, \sigma^2(1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'))$$

取随机扰动项的样本估计量 $\hat{\sigma}^2$，可得e_0的方差的估计量

$$\hat{\sigma}_{e_0}^2 = \hat{\sigma}^2(1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0')$$

构造t统计量

$$t = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma}_{e_0}} \sim t(n-k-1)$$

可得给定(1−α)的置信水平下Y_0的**置信区间**：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'}$$

中国居民人均收入-消费支出二元模型例中：

2001年人均GDP：4033.1元，

于是人均居民消费的预测值为

$\hat{Y}_{2001}$=120.7+0.2213×4033.1+0.4515×1690.8=1776.8（元）

实测值（90年价）=1782.2元，**相对误差**：-0.31%

预测的置信区间：

在95%的置信度下，临界值$t_{\alpha/2}(19)=2.093$，$\hat{\sigma}^2=705.5$，

$$(\mathbf{X}'\mathbf{X})^{-1}=\begin{pmatrix} 1.88952 & 0.00285 & -0.00828 \\ 0.00285 & 0.00001 & -0.00001 \\ -0.00828 & -0.00001 & 0.00004 \end{pmatrix}$$

$$\mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 = 0.3938$$

于是$E(\hat{Y}_{2001})$的95%的置信区间为:

$$1776.8 \pm 2.093 \times \sqrt{705.5} \times \sqrt{0.3938}$$

或 （1741.8，1811.7）

同样，易得$\hat{Y}_{2001}$的95%的置信区间为

$$1776.8 \pm 2.093 \times \sqrt{705.5} \times \sqrt{1.3938}$$

或 （1711.1, 1842.4）

§3.5 回归模型的其他函数形式

一、模型的类型与变换

二、非线性回归实例

在实际经济活动中，经济变量的关系是复杂的，直接表现为线性关系的情况并不多见。

如著名的**恩格尔曲线**(Engle curves)表现为**幂函数曲线**形式、宏观经济学中的**菲利普斯曲线**（Pillips cuves）表现为**双曲线**形式等。

但是，大部分非线性关系又可以通过一些简单的数学处理，使之化为数学上的线性关系，从而可以运用线性回归的方法进行计量经济学方面的处理。

一、模型的类型与变换

1、倒数模型、多项式模型与变量的直接置换法

例如，描述税收与税率关系的**拉弗曲线**：抛物线

$$s = a + b\,r + c\,r^2 \qquad c<0$$

s：税收；　r：税率

设$X_1 = r$，$X_2 = r^2$，则原方程变换为

$$s = a + b\,X_1 + c\,X_2 \qquad c<0$$

2、幂函数模型、指数函数模型与对数变换法

例如，Cobb-Dauglas生产函数：幂函数

$$Q = AK^{\alpha}L^{\beta}$$

Q：产出量，K：投入的资本；L：投入的劳动

方程两边取对数：

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L$$

3、复杂函数模型与级数展开法

例如，常替代弹性CES生产函数

$$Q = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} e^{\mu} \qquad (\delta_1+\delta_2=1)$$

Q:产出量，K：资本投入，L：劳动投入

ρ：替代参数，δ_1、δ_2：分配参数

方程两边取对数后，得到：

$$LnQ = LnA - \frac{1}{\rho} Ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \mu$$

将式中$\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})$在$\rho=0$处展开台劳级数,取关于ρ的线性项，即得到一个线性近似式。

如取0阶、1阶、2阶项，可得

$$\ln Y = \ln A + \delta_1 m \ln K + \delta_2 m \ln L - \frac{1}{2} \rho\, m \delta_1 \delta_2 \left(\ln\left(\frac{K}{L}\right)\right)^2$$

并非所有的函数形式都可以线性化

无法线性化模型的一般形式为:

$$Y = f(X_1, X_2, \cdots, X_k) + \mu$$

其中，$f(x_1,x_2,\ldots,X_k)$为非线性函数。如：

$$Q = AK^{\alpha}L^{\beta} + \mu$$

二、非线性回归实例

例3.5.1 建立中国城镇居民食品消费需求函数模型。

根据需求理论，居民对食品的消费需求函数大致为

$$Q = f(X, P_1, P_0) \qquad (*)$$

Q:居民对食品的需求量，X：消费者的消费支出总额

P_1：食品价格指数，P_0：居民消费价格总指数。

零阶齐次性，当所有商品和消费者货币支出总额按同一比例变动时，需求量保持不变

$$Q = f(X/P_0, P_1/P_0) \qquad (**)$$

为了进行比较，将同时估计（*）式与（）式。**

首先，确定具体的函数形式

根据**恩格尔定律**，居民对食品的消费支出与居民的总支出间呈**幂函数**的变化关系：

$$Q = AX^{\beta_1} P_1^{\beta_2} P_0^{\beta_3}$$

对数变换：

$$\ln(Q) = \beta_0 + \beta_1 \ln X + \beta_2 \ln P_1 + \beta_3 \ln P_0 + \mu \qquad (***)$$

考虑到**零阶齐次性**时

$$\ln(Q) = \beta_0 + \beta_1 \ln(X / P_0) + \beta_2 \ln(P_1 / P_0) + \mu \qquad (****)$$

(****)式也可看成是对（***）式施加如下约束而得

$$\beta_1 + \beta_2 + \beta_3 = 0$$

因此，**对（****）式进行回归，就意味着原需求函数满足零阶齐次性条件。**

表 3.5.1　中国城镇居民消费支出（元）及价格指数

	X (当年价)	X1 (当年价)	GP (上年=100)	FP (上年=100)	XC (1990年价)	Q (1990年价)	P0 (1990=100)	P1 (1990=100)
1981	456.8	420.4	102.5	102.7	646.1	318.3	70.7	132.1
1982	471.0	432.1	102.0	102.1	659.1	325.0	71.5	132.9
1983	505.9	464.0	102.0	103.7	672.2	337.0	75.3	137.7
1984	559.4	514.3	102.7	104.0	690.4	350.5	81.0	146.7
1985	673.2	351.4	111.9	116.5	772.6	408.4	87.1	86.1
1986	799.0	418.9	107.0	107.2	826.6	437.8	96.7	95.7
1987	884.4	472.9	108.8	112.0	899.4	490.3	98.3	96.5
1988	1104.0	567.0	120.7	125.2	1085.5	613.8	101.7	92.4
1989	1211.0	660.0	116.3	114.4	1262.5	702.2	95.9	94.0
1990	1278.9	693.8	101.3	98.8	1278.9	693.8	100.0	100.0
1991	1453.8	782.5	105.1	105.4	1344.1	731.3	108.2	107.0
1992	1671.7	884.8	108.6	110.7	1459.7	809.5	114.5	109.3
1993	2110.8	1058.2	116.1	116.5	1694.7	943.1	124.6	112.2
1994	2851.3	1422.5	125.0	134.2	2118.4	1265.6	134.6	112.4
1995	3537.6	1766.0	116.8	123.6	2474.3	1564.3	143.0	112.9
1996	3919.5	1904.7	108.8	107.9	2692.0	1687.9	145.6	112.8
1997	4185.6	1942.6	103.1	100.1	2775.5	1689.6	150.8	115.0
1998	4331.6	1926.9	99.4	96.9	2758.9	1637.2	157.0	117.7
1999	4615.9	1932.1	98.7	95.7	2723.0	1566.8	169.5	123.3
2000	4998.0	1958.3	100.8	97.6	2744.8	1529.2	182.1	128.1
2001	5309.0	2014.0	100.7	100.7	2764.0	1539.9	192.1	130.8

X：人均消费

X1：人均食品消费

GP：居民消费价格指数

FP：居民食品消费价格指数

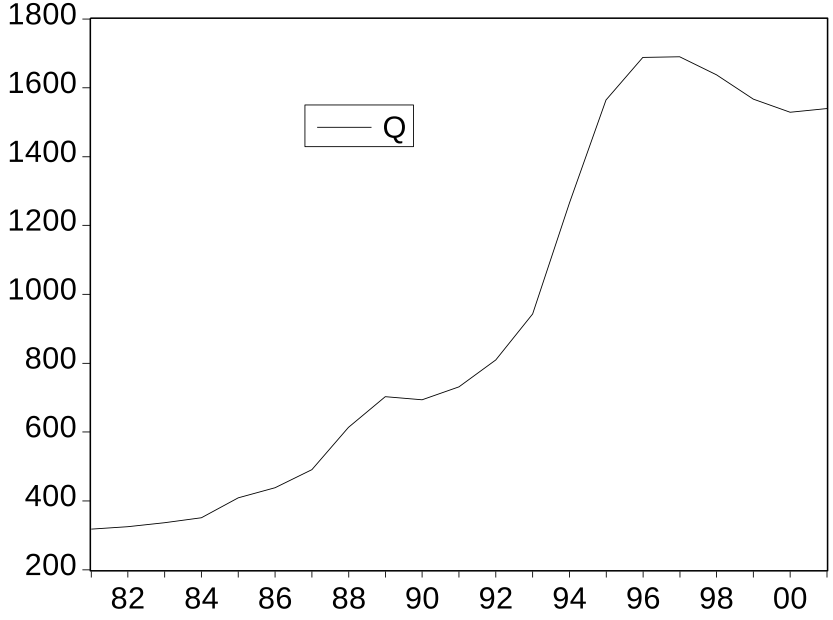
XC：人均消费（90年价）

Q：人均食品消费（90年价）

P0：居民消费价格缩减指数（1990=100）

P：居民食品消费价格缩减指数（1990=100

中国城镇居民人均食品消费



特征：

消费行为在1981~1995年间表现出较强的一致性

1995年之后呈现出另外一种变动特征。

建立1981~1994年中国城镇居民对食品的消费需求模型:

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0)$$

(9.03) (25.35) (-2.28) (-7.34)

$R^2=0.9987$ $\bar{R}^2=0.9983$ DW=1.50 F=2583.28

各变量的弹性和 $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$ =0.05，比较接近于零，但不为零。

按零阶齐次性表达式回归：

$$\ln(\hat{Q}) = 3.83 + 1.07\ln(X / P_0) - 0.09\ln(P_1 / P_0)$$

（75.86）(52.66)　　　(-3.62)

$R^2=0.9986$　　$\bar{R}^2=0.9984$　　DW=1.51　　F=4166.3

为了比较，改写该式为：

$$\begin{aligned}\ln\hat{Q} &= 3.83 + 1.07(\ln X - \ln P_0) - 0.09(\ln P_1 - \ln P_0) \\ &= 3.83 + 1.07\ln X - 0.09\ln P_1 - 0.98\ln P_0\end{aligned}$$

发现与

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0)$$

接近。

意味着：所建立的食品需求函数满足零阶齐次性特征

§3.6 受约束回归

在建立回归模型时，有时根据经济理论需对模型中变量的参数施加一定的约束条件。

如： 0阶齐次性 条件的消费需求函数

1阶齐次性 条件的C-D生产函数

模型施加约束条件后进行回归，称为受约束回归（**restricted regression**）；

不加任何约束的回归称为无约束回归（**unrestricted regression**）。

受约束回归

一、模型参数的线性约束

二、对回归模型增加或减少解释变量

三、参数的稳定性

*四、非线性约束

一、模型参数的线性约束

对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \qquad (*)$$

施加约束

$$\beta_1+\beta_2=1 \qquad \beta_{k-1}=\beta_k$$

得

$$Y=\beta_0+\beta_1X_1+(1-\beta_1)X_2+\cdots+\beta_{k-1}X_{k-1}+\beta_{k-1}X_k+\mu^*$$

或 $$Y^*=\beta_0+\beta_1X_1^*+\beta_3X_3+\cdots+\beta_{k-1}X_{k-1}^*+\mu^* \qquad (**)$$

如果对（**）式回归得出 $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_3,\cdots,\hat{\beta}_{k-1}$

则由约束条件可得： $\hat{\beta}_2=1-\hat{\beta}_1 \qquad \hat{\beta}_k=\hat{\beta}_{k-1}$

然而，对所考查的具体问题能否施加约束？需进一步进行相应的检验。常用的检验有：

F检验、x^2检验与t检验，

主要介绍F检验

在同一样本下，记无约束样本回归模型为

$$\mathbf{Y} = \mathbf{X}\hat{\beta} + \mathbf{e}$$

受约束样本回归模型为

$$\mathbf{Y} = \mathbf{X}\hat{\beta}_* + \mathbf{e}_*$$

于是

$$\mathbf{e}_* = \mathbf{Y} - \mathbf{X}\hat{\beta}_* = \mathbf{X}\hat{\beta} + \mathbf{e} - \mathbf{X}\hat{\beta}_* = \mathbf{e} - \mathbf{X}(\hat{\beta}_* - \hat{\beta})$$

受约束样本回归模型的**残差平方和**$\mathbf{RSS_R}$

$$\mathbf{e}_*'\mathbf{e}_* = \mathbf{e}'\mathbf{e} + (\hat{\beta}_* - \hat{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta}_* - \hat{\beta})$$

于是

$$\mathbf{e}_*'\mathbf{e}_* \geq \mathbf{e}'\mathbf{e} \qquad (*)$$

e'e为无约束样本回归模型的**残差平方**和$\mathbf{RSS_U}$

受约束与无约束模型都有**相同的TSS**

由（*）式 $\quad \mathbf{RSS_R} \geq \mathbf{RSS_U}$

从而 $\quad \mathbf{ESS_R} \leq \mathbf{ESS_U}$

这意味着，通常情况下，对模型施加约束条件会降低模型的解释能力。

但是，如果约束条件为真，则受约束回归模型与无约束回归模型具有相同的解释能力，RSS_R 与 RSS_U的差异变小。

可用RSS_R - RSS_U的大小来检验约束的真实性

根据数理统计学的知识：

$$RSS_U / \sigma^2 \sim \chi^2(n - k_U - 1)$$

$$RSS_R / \sigma^2 \sim \chi^2(n - k_R - 1)$$

$$(RSS_R - RSS_U) / \sigma^2 \sim \chi^2(k_U - k_R)$$

于是：

$$\boxed{F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{RSS_U /(n - k_U - 1)} \sim F(k_U - k_R, n - k_U - 1)}$$

讨论：

如果约束条件无效， **RSS_R** 与 **RSS_U**的差异较大，计算的F值也较大。

于是，可用计算的F统计量的值与所给定的显著性水平下的临界值作比较，对约束条件的真实性进行检验。

注意，$\boldsymbol{k_U - k_R}$恰为约束条件的个数。

例3.6.1 中国城镇居民对食品的人均消费需求实例中，对零阶齐次性检验:

无约束回归: RSS_U=0.00324，　　k_U=3

受约束回归: RSS_R=0.00332，　　K_R=2

样本容量n=14，　　约束条件个数$k_U - k_R$=3-2=1

$$F=\frac{(0.003315-0.003240)/1}{0.003240/10}=0.231$$

取α=5%，查得临界值$F_{0.05}(1,10)$=4.96

判断：不能拒绝中国城镇居民对食品的人均消费需求函数具有零阶齐次特性这一假设。

这里的F检验适合所有关于参数线性约束的检验

如：多元回归中对方程总体线性性的F检验：

H_0： $\beta_j=0$　　　　　　j=1,2,…,k

这里：受约束回归模型为

$$Y=\beta_0+\mu_*$$

$$F=\frac{(RSS_R-RSS_U)/(k_U-k_R)}{RSS_U/(n-k_U-1)}=\frac{(TSS-ESS_R-RSS_U)/k}{RSS_U/(n-k-1)}$$

$$=\frac{(TSS-RSS_U)/k}{RSS_U/(n-k-1)}=\frac{ESS_U/k}{RSS_U/(n-k-1)}$$

这里，运用了$ESS_R=0$。

二、对回归模型增加或减少解释变量

考虑如下两个回归模型

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\mu \qquad (*)$$

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\beta_{k+1}X_{k+1}+\cdots\beta_{k+q}X_{k+q}+\mu \qquad (**)$$

(*)式可看成是（**）式的**受约束回归**：

H0: $\beta_{k+1}=\beta_{k+2}=\cdots=\beta_{k+q}=0$

相应的 F 统计量为：

$$F=\frac{(RSS_R-RSS_U)/q}{RSS_U/(n-(k+q+1))}$$

$$=\frac{(ESS_U-ESS_R)/q}{RSS_U/(n-(k+q+1))}\sim F(q,n-(k+q+1))$$

讨论：

如果约束条件为真，即额外的变量X_{k+1}，…，X_{k+q}对Y没有解释能力，则F统计量较小；

否则，约束条件为假，意味着额外的变量对Y有较强的解释能力，则F统计量较大。

因此，可通过F的**计算值**与**临界值**的比较，来判断额外变量是否应包括在模型中。

F统计量的另一个等价式

$$F=\frac{(R_U^2-R_R^2)/q}{(1-R_U^2)/(n-(k+q+1))}$$

三、参数的稳定性

建立模型时往往希望模型的参数是稳定的，即所谓的**结构不变**，这将提高模型的预测与分析功能。**如何检验？**

1、邹氏参数稳定性检验

假设**需要建立的模型**为

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu$$

在两个连续的时间序列（1,2,…，n_1）与（n_1+1,…，n_1+n_2）中，相应的模型分别为：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu_1$$

$$Y = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_k X_k + \mu_2$$

合并两个时间序列为(1,2,…， n_1 ， n_1+1,…， n_1+n_2)，则可写出如下无约束回归模型

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad (*)$$

如果α=β，表示没有发生结构变化，因此可针对如下假设进行检验：

H_0: α=β

(*)式施加上述约束后变换为受约束回归模型

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \beta + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad (**)$$

因此，检验的F统计量为：

$$F=\frac{(RSS_R-RSS_U)/k}{RSS_U/[n_1+n_2-2(k+1)]}\sim F[k,n_1+n_2-2(k+1)]$$

记RSS$_1$与RSS$_2$为在两时间段上分别回归后所得的残差平方和，容易验证，

$$RSS_U=RSS_1+RSS_2$$

于是

$$F=\frac{[RSS_R-(RSS_1+RSS_2)]/k}{(RSS_1+RSS_2)/[n_1+n_2-2(k+1)]}\sim F[k,n_1+n_2-2(k+1)]$$

参数稳定性的检验步骤：

（1）分别以两连续时间序列作为两个样本进行回归，得到相应的残差平方：RSS_1与RSS_2

（2）将两序列并为一个大样本后进行回归，得到大样本下的残差平方和RSS_R

（3）计算F统计量的值，与临界值比较：

若F值大于临界值，则拒绝原假设，认为发生了结构变化，参数是非稳定的。

该检验也被称为邹氏参数稳定性检验（**Chow test for parameter stability**）。

2、邹氏预测检验

上述参数稳定性检验要求$n_2>k$。

如果出现$n_2<k$ ，则往往进行如下的邹氏预测检验（**Chow test for predictive failure**）。

邹氏预测检验的基本思想:

先用前一时间段n_1个样本估计原模型，再用估计出的参数进行后一时间段n_2个样本的预测。

如果预测误差较大，则说明参数发生了变化，否则说明参数是稳定的。

分别以β、α 表示第一与第二时间段的参数，则

$$\begin{cases} \mathbf{Y}_1 = \mathbf{X}_1\beta + \mu_1 \\ \mathbf{Y}_2 = \mathbf{X}_2\alpha + \mu_2 = \mathbf{X}_2\beta + \mathbf{X}_2(\alpha - \beta) + \mu_2 = \mathbf{X}_2\beta + \gamma + \mu_2 \end{cases} \quad (*)$$

其中， $\gamma = \mathbf{X}_2(\alpha - \beta)$

如果γ =**0**，则 α = β，表明参数在估计期与预测期相同

(*)的矩阵式：

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{I}_{n_2} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (**)$$

可见，用前n_1个样本估计可得前k个参数β的估计，而γ不外是用后n_2个样本测算的预测误差$\mathbf{X}_2(\alpha - \beta)$

如果参数没有发生变化，则$\gamma=0$，矩阵式简化为

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \beta + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad (***)$$

（***）式与（**）式 $\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{I}_{\mathbf{n}_2} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

分别可看成受约束与无约束回归模型，于是有如下F检验：

$$F = \frac{(RSS_R - RSS_U)/(k_U - k_R)}{RSS_U/(n - k_U - 1)} = \frac{(RSS_R - RSS_1)/n_2}{RSS_1/(n_1 - k - 1)}$$

这里：$K_U - K_R = n_2$

$RSS_U = RSS_1$

邹氏预测检验步骤：

第一步，在两时间段的合成大样本下做OLS回归，得受约束模型的残差平方和RSS_R ；

第二步，对前一时间段的n_1个子样做OLS回归，得残差平方和RSS_1 ；

第三步，计算检验的F统计量，做出判断：

给定显著性水平α，查F分布表，得临界值$F_{\alpha}(n_2, n_1-k-1)$

如果 $F>F(n_2, n_1-k-1)$ ，则拒绝原假设，认为预测期发生了结构变化。

例3.6.2 中国城镇居民食品人均消费需求的邹氏检验。

1、参数稳定性检验

1981~1994:

$$\ln(\hat{Q}) = 3.63 + 1.05\ln(X) - 0.08\ln(P_1) - 0.92\ln(P_0) \qquad RSS_1=0.003240$$

1995~2001:

$$\ln Q = 13.78 + 0.55\ln X - 3.06\ln P_1 + 0.71\ln P_0$$

(9.96) (7.14) (-5.13) (1.81)

$R^2=0.9946$ $\bar{R}^2=0.9893$ DW=2.80 F=185.37 $RSS_2=0.000058$

1981~2001:

$$\ln Q = 5.00 + 1.21\ln X - 0.14\ln P_1 - 1.39\ln P_0$$

(14.83) (27.26) (-3.24) (-11.17)

$R^2=0.9982$ $\bar{R}^2=0.9979$ DW=0.93 F=3228.0 $RSS_R=0.013789$

$$F=\frac{[0.013789-(0.003240+0.0000580)]/4}{(0.003240+0.000058)/(21-8)}=10.34$$

给定$\alpha=5\%$，查表得临界值$F_{0.05}(4, 13)=3.18$

判断：F值>临界值，拒绝参数稳定的原假设，表明中国城镇居民食品人均消费需求在1994年前后发生了显著变化。

2、邹氏预测检验

$$F=\frac{(0.013789-0.003240)/7}{0.003240/(14-3-1)}=4.65$$

给定$\alpha=5\%$，查表得临界值$F_{0.05}(7, 10)=3.18$

判断： F值>临界值，拒绝参数稳定的原假设

*四、非线性约束

也可对模型参数施加**非线性约束**, 如对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu$$

施加非线性约束$\beta_1\beta_2=1$, 得到**受约束回归模型**:

$$Y=\beta_0+\beta_1X_1+\frac{1}{\beta_1}X_2+\cdots+\beta_kX_k+\mu^*$$

该模型必需采用**非线性最小二乘法**（**nonlinear least squares**）进行估计。

非线性约束检验是建立在**最大似然原理**基础上的, 有**最大似然比检验**、**沃尔德检验**与**拉格朗日乘数检验**.

1、最大似然比检验 (likelihood ratio test, LR)

估计:无约束回归模型与受约束回归模型,

方法:最大似然法,

检验:两个似然函数的值的差异是否“足够”大。

记$L(\boldsymbol{\beta}, \sigma^2)$为一似然函数:

无约束回归: Max: $L(\hat{\beta}, \hat{\sigma}^2)$

受约束回归: Max: $L(\tilde{\beta}, \tilde{\sigma}^2)$ 约束: $g(\boldsymbol{\beta})=0$

或求极值: $\Phi = L(\beta, \sigma^2) - \lambda' g(\beta)$

$g(\boldsymbol{\beta})$:以各约束条件为元素的列向量,

λ': 以相应拉格朗日乘数为元素的行向量

受约束的函数值不会超过无约束的函数值，但如果**约束条件为真**，则两个函数值就非常“**接近**”由此，定义**似然比**（**likelihood ratio**）：

$$L\left(\tilde{\beta},\ \tilde{\sigma}^2\right)/L\left(\hat{\beta},\ \hat{\sigma}^2\right)$$

如果比值很小，**说明**两似然函数值差距较大，则应**拒绝**约束条件为真的假设；

如果比值接近于 1 ，**说明**两似然函数值很接近，应**接受**约束条件为真的假设。

具体检验时，由于大样本下：

$$LR = -2[\ln L(\tilde{\beta},\tilde{\sigma}^2) - \ln L(\hat{\beta},\hat{\sigma}^2)] \sim \chi^2(h)$$

h是约束条件的个数。因此：

通过LR统计量的χ^2分布特性来进行判断。

在中国城镇居民人均食品消费需求例中，对零阶齐次性的检验：

受约束回归模型：$\ln L(\tilde{\beta},\tilde{\sigma}^2)=38.57$

无约束回归模型：$\ln L(\hat{\beta},\hat{\sigma}^2)=38.73$

$$LR=-2(38.57-38.73)=0.32$$

给出$\alpha=5\%$、查得临界值$\chi^2_{0.05}(1)=3.84$，

判断： $LR<\chi^2_{0.05}(1)$，不拒绝原约束的假设，

表明：中国城镇居民对食品的人均消费需求函数满足零阶齐次性条件。

2、沃尔德检验（Wald test, W）

沃尔德检验中，只须估计无约束模型。如对

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu$$

要检验约束$\beta_1+\beta_2=1$，只须对该模型进行回归，并判断$\widehat{\beta}_1+\widehat{\beta}_2$与1的差距是否足够大。

在所有古典假设都成立的条件下，容易证明

$$\hat{\beta}_1 + \hat{\beta}_2 \sim N(\beta_1 + \beta_2, \sigma^2_{\hat{\beta}_1+\hat{\beta}_2})$$

因此，在$\beta_1+\beta_2=1$的约束条件下

$$z = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{\sigma_{\hat{\beta}_1+\hat{\beta}_2}} \sim N(0,1)$$

$\sigma^2_{\hat{\beta}_1+\hat{\beta}_2}$是$\widehat{\beta}_1+\widehat{\beta}_2$的方差，可记为$\sigma^2_{\hat{\beta}_1+\hat{\beta}_2} = \sigma^2 f(X)$

以σ^2的极大似然估计量 $\tilde{\sigma}^2 = \mathbf{e}'\mathbf{e}/n$ 代入 $\sigma^2_{\hat{\beta}_1+\hat{\beta}_2}$

记 $\tilde{\sigma}^2_{\hat{\beta}_1+\hat{\beta}_2} = \tilde{\sigma}^2 f(\mathbf{X})$ 可建立**沃尔德统计量:**

$$W = \frac{(\hat{\beta}_1 + \hat{\beta}_2 - 1)^2}{\tilde{\sigma}^2_{\hat{\beta}_1+\hat{\beta}_2}} \sim \chi^2(1)$$

如果有h个约束条件，可得到h个统计量$z_1, z_2, \ldots, z_h$

约束条件为真时，可建立大样本下的服从自由度为h的渐近χ^2分布统计量

$$W = \mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z} \sim \chi^2(h)$$

其中，**Z**为以z_i为元素的列向量，**C**是**Z**的方差-协方差矩阵。

因此，**W从总体上测量了无约束回归不满足约束条件的程度。**

对非线性约束，沃尔德统计量W的算法描述要复杂得多。

3、拉格朗日乘数检验

拉格朗日乘数检验则只需估计受约束模型.

受约束回归是求最大似然法的极值问题：

$$\Phi = L(\beta, \sigma^2) - \lambda' g(\beta)$$

λ'是拉格朗日乘数行向量，衡量各约束条件对最大似然函数值的影响程度。

如果某一约束为真，则该约束条件对最大似然函数值的影响很小，于是，相应的拉格朗日乘数的值应接近于零。

因此，拉格朗日乘数检验就是检验某些拉格朗日乘数的值是否“足够大”，如果“足够大”，则拒绝约束条件为真的假设。

拉格朗日统计量LM本身是一个关于拉格朗日乘数的复杂的函数，在各约束条件为真的情况下，服从一自由度恰为约束条件个数的渐近χ^2分布。

同样地，如果为线性约束，LM服从一精确的χ^2分布：

$$LM = nR^2 \qquad (*)$$

n为样本容量，R^2为如下被称为**辅助回归**（auxiliary regression）的可决系数：

$$\hat{e}_R = \hat{\delta}_0 + \hat{\delta}_1 X_1 + \hat{\delta}_2 X_2 + \cdots + \hat{\delta}_k X_k$$

这里，$\hat{e}_R$为受约束回归模型的残差序列

如果约束是非线性的，辅助回归方程的估计比较复杂，但仍可按（*）式计算LM统计量的值。

最后，一般地有：LM≤LR≤W

第四章 经典单方程计量经济学模型：放宽基本假定的模型

基本假定违背：不满足基本假定的情况。主要 包括：

（1）随机误差项序列存在**异方差**性；

（2）随机误差项序列存在**序列相关**性；

（3）解释变量之间存在**多重共线**性；

（4）解释变量是随机变量且与随机误差项相关（**随机解释变量**）；

此外：

（5）模型设定有偏误

（6）解释变量的方差不随样本容量的增而收敛

计量经济检验：对模型基本假定的检验

本章主要学习：前4类

§4.1 异方差性

一、异方差的概念

二、异方差的类型

三、实际经济问题中的异方差性

四、异方差性的后果

五、异方差性的检验

六、异方差的修正

七、案例

一、异方差的概念

对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

如果出现

$$Var(\mu_i) = \sigma_i^2$$

即对于不同的样本点，随机误差项的方差不再是常数，而互不相同，则认为出现了**异方差性**(**Heteroskedasticity**)。

二、异方差的类型

同方差性假定：σ_i^2 = 常数 $\neq f(X_i)$

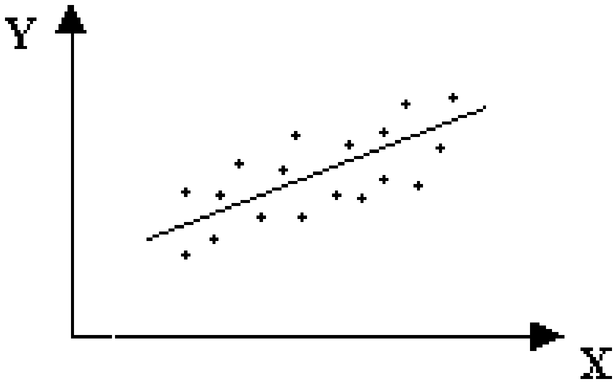
异方差时：$\sigma_i^2 = f(X_i)$

异方差一般可归结为三种类型：

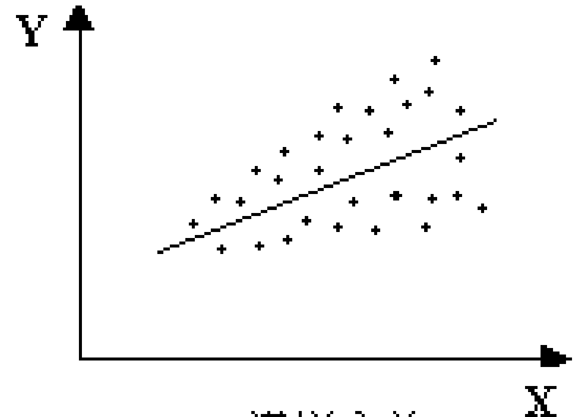
(1)单调递增型：σ_i^2随X的增大而增大

(2)单调递减型：σ_i^2随X的增大而减小

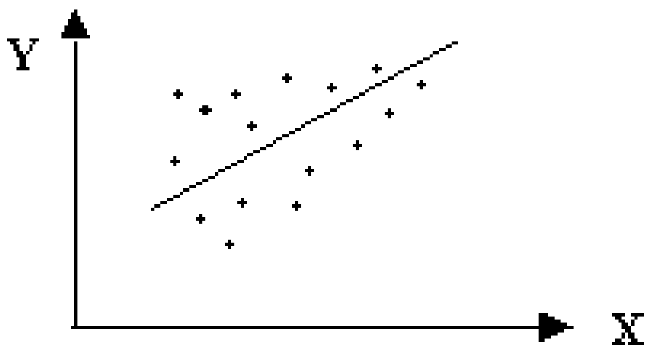
(3)复　杂　型：σ_i^2与X的变化呈复杂形式



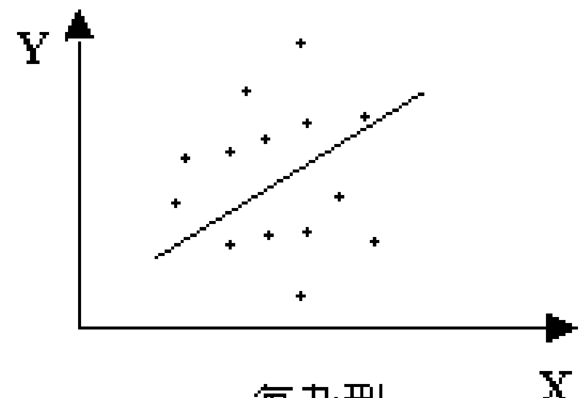
同方差



递增方差



递减方差



复杂型

三、实际经济问题中的异方差性

例4.1.1：截面资料下研究居民家庭的储蓄行为

$$Y_i=\beta_0+\beta_1 X_i+\mu_i$$

Y_i:第i个家庭的储蓄额 X_i:第i个家庭的可支配收入

高收入家庭：储蓄的差异较大

低收入家庭：储蓄则更有规律性，差异较小

μ_i的方差呈现单调递增型变化

例4.1,2，以绝对收入假设为理论假设、以截面数据为样本建立居民消费函数：

$$C_i=\beta_0+\beta_1 Y_i+\mu_I$$

将居民按照收入等距离分成n组，取组平均数为样本观测值。

一般情况下，居民收入服从正态分布：中等收入组人数多，两端收入组人数少。而人数多的组平均数的误差小，人数少的组平均数的误差大。

所以样本观测值的**观测误差**随着解释变量观测值的不同而不同，往往引起异方差性。

例4.1.3，以某一行业的企业为样本建立企业生产函数模型

$$Y_i=A_i^{\beta 1}K_i^{\beta 2}L_i^{\beta 3}e^{\mu i}$$

被解释变量：产出量Y

解释变量：资本K、劳动L、技术A，

那么：每个企业所处的外部环境对产出量的影响被包含在随机误差项中。

每个企业所处的外部环境对产出量的影响程度不同，造成了随机误差项的异方差性。

这时，随机误差项的方差并不随某一个解释变量观测值的变化而呈规律性变化，呈现复杂型。

四、异方差性的后果

计量经济学模型一旦出现异方差性，如果仍采用OLS估计模型参数，会产生下列不良后果：

1、参数估计量非有效

OLS估计量仍然具有无偏性，但不具有有效性

因为在有效性证明中利用了

$$E(\mu\mu')=\sigma^2 \mathbf{I}$$

而且，在大样本情况下，尽管参数估计量具有一致性，但仍然不具有渐近有效性。

2、变量的显著性检验失去意义

变量的显著性检验中，构造了t统计量

$$t = \hat{\beta}_1 / S_{\hat{\beta}_1}$$

它是建立在σ^2不变而正确估计了参数方差$S_{\hat{\beta}_1}$的基础之上的。

如果出现了异方差性，估计的$S_{\hat{\beta}_1}$出现偏误（偏大或偏小），t检验失去意义。

其他检验也是如此。

3、模型的预测失效

一方面，由于上述后果，使得模型不具有良好的统计性质；

另一方面，在预测值的置信区间中也包含有参数方差的估计量 $S_{\hat{\beta}_i}$。

所以，当模型出现异方差性时，参数OLS估计值的变异程度增大，从而造成对Y的预测误差变大，降低预测精度，预测功能失效。

五、异方差性的检验

- **检验思路：**

由于异方差性就是相对于不同的解释变量观测值，随机误差项具有不同的方差。那么：

检验异方差性，也就是检验随机误差项的方差与解释变量观测值之间的相关性及其相关的“形式”。

问题在于用什么来表示随机误差项的方差

一般的处理方法：

首先采用 OLS 法估计模型，以求得随机误差项的估计量（注意，该估计量是不严格的），我们称之为“**近似估计量**”，用$\tilde{e}_i$表示。于是有

$$Var(\mu_i) = E(\mu_i^2) \approx \tilde{e}_i^2$$

$$\tilde{e}_i = y_i - (\hat{y}_i)_{0ls}$$

即用$\tilde{e}_i^2$来表示随机误差项的方差。

几种异方差的检验方法：

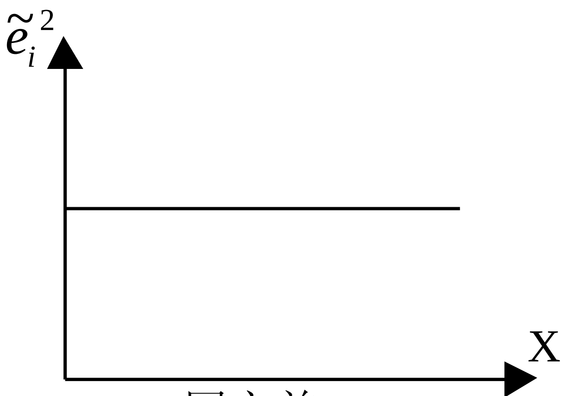
1、图示法

（1）用X-Y的散点图进行判断

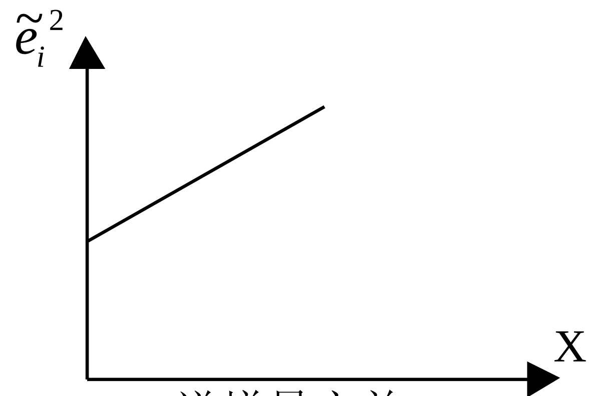
看是否存在明显的散点扩大、缩小或复杂型趋势（即不在一个固定的带型域中）

(2) X-$\tilde{e}_i^2$的散点图进行判断

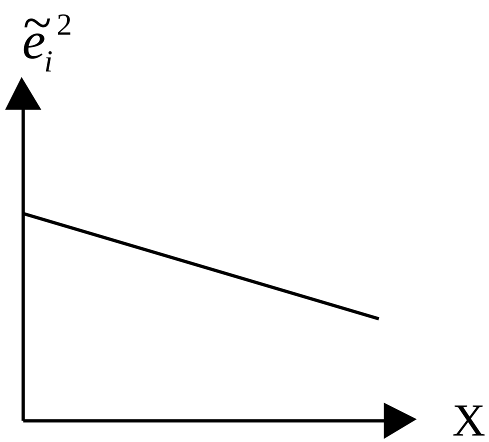
看是否形成一斜率为零的直线



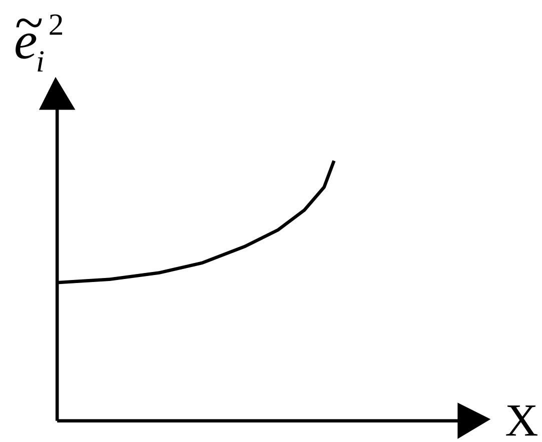
同方差



递增异方差



递减异方差



复杂型异方差

2、帕克(Park)检验与戈里瑟(Gleiser)检验

基本思想：

偿试建立方程：

$$\tilde{e}_i^2 = f(X_{ji}) + \varepsilon_i \quad 或 \quad |\tilde{e}_i| = f(X_{ji}) + \varepsilon_i$$

选择关于变量X的不同的函数形式，对方程进行估计并进行显著性检验，如果存在某一种函数形式，使得方程显著成立，则说明原模型存在异方差性。

如：帕克检验常用的函数形式：

$$f(X_{ji}) = \sigma^2 X_{ji}^{\alpha} e^{\varepsilon_i} \quad 或 \quad \ln(\tilde{e}_i^2) = \ln\sigma^2 + \alpha \ln X_{ji} + \varepsilon_i$$

若α在统计上是显著的，表明存在异方差性。

3、戈德菲尔德-匡特(Goldfeld-Quandt)检验

G-Q检验以F检验为基础，适用于样本容量较大、异方差递增或递减的情况。

G-Q检验的思想：

先将样本一分为二，对子样①和子样②分别作回归，然后利用两个子样的残差平方和之比构造统计量进行异方差检验。

由于该统计量服从F分布，因此假如存在递增的异方差，则F远大于1；反之就会等于1（同方差）、或小于1（递减方差）。

G-Q检验的步骤：

①将n对样本观察值(X_i,Y_i)按观察值X_i的大小排队

②将序列中间的c=n/4个观察值除去，并将剩下的观察值划分为较小与较大的相同的两个子样本，每个子样样本容量均为(n-c)/2

③对每个子样分别进行OLS回归，并计算各自的残差平方和

分别用 $\sum \tilde{e}_{1i}^2$ 与 $\sum \tilde{e}_{2i}^2$ 表示较小与较大的残差平方和(自由度均为 $\frac{n-c}{2}-k-1$)；

④在同方差性假定下，构造如下满足F分布的统计量

$$F=\frac{\sum\tilde{e}_{2i}^{2}\Big/\left(\frac{n-c}{2}-k-1\right)}{\sum\tilde{e}_{1i}^{2}\Big/\left(\frac{n-c}{2}-k-1\right)}\sim F\left(\frac{n-c}{2}-k-1,\frac{n-c}{2}-k-1\right)$$

⑤给定显著性水平α，确定临界值$F_{\alpha}(v_1,v_2)$，若$F> F_{\alpha}(v_1,v_2)$，则拒绝同方差性假设，表明存在异方差。

当然，还可根据两个残差平方和对应的子样的顺序判断是递增型异方差还是递减异型方差。

3、怀特（White）检验

怀特检验不需要排序，且适合任何形式的异方差

怀特检验的基本思想与步骤（以二元为例）：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$$

先对该模型作OLS回归，得到 $\tilde{e}_i^2$

然后做如下辅助回归

$$\tilde{e}_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + \varepsilon_i \qquad (*)$$

可以证明，在同方差假设下：

$$nR^2 \stackrel{\bullet}{\sim} \chi^2(h)$$

R^2为(*)的可决系数，h为(*)式解释变量的个数，

$\stackrel{\bullet}{\sim}$ 表示渐近服从某分布。

注意：

辅助回归仍是检验与解释变量可能的组合的显著性，因此，辅助回归方程中还可引入解释变量的更高次方。

如果存在异方差性，则表明确与解释变量的某种组合有显著的相关性，这时往往显示出有较高的可决系数以及某一参数的t检验值较大。

当然，在多元回归中，由于辅助回归方程中可能有太多解释变量，从而使自由度减少，有时可去掉交叉项。

六、异方差的修正

模型检验出存在异方差性，可用**加权最小二乘法**（**Weighted Least Squares, WLS**）进行估计。

加权最小二乘法的基本思想：

加权最小二乘法是对原模型加权，使之变成一个新的不存在异方差性的模型，然后采用OLS估计其参数。

$$\sum W_i e_i^2 = \sum W_i [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)]^2$$

在采用OLS方法时:

对较小的残差平方e_i^2赋予较大的权数，

对较大的残差平方e_i^2赋予较小的权数。

例如，如果对一多元模型，经检验知：

$$Var(\mu_i)=E(\mu_i)^2=\sigma_i^2=f(X_{ji})\sigma^2$$

可以用$\sqrt{f(X_j)}$去除该模型，得

$$\frac{1}{\sqrt{f(X_{ji})}}Y_i=\beta_0\frac{1}{\sqrt{f(X_{ji})}}+\beta_1\frac{1}{\sqrt{f(X_{ji})}}X_{1i}+\beta_2\frac{1}{\sqrt{f(X_{ji})}}X_{2i}+\cdots$$

$$+\beta_k\frac{1}{\sqrt{f(X_{ji})}}X_{ki}+\frac{1}{\sqrt{f(X_{ji})}}\mu_i$$

新模型中，存在

$$Var(\frac{1}{\sqrt{f(X_{ji})}}\mu_i)=E(\frac{1}{\sqrt{f(X_{ji})}}\mu_i)^2=\frac{1}{f(X_{ji})}E(\mu_i)^2=\sigma^2$$

即满足同方差性，可用OLS法估计。

一般情况下:

对于模型

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\mu}$$

存在

$$E(\mu)=0$$

$$Cov(\mu)=E(\mu\,\mu')=\sigma^2\mathbf{W}$$

$$\boldsymbol{W}=\begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}$$

即存在异方差性。

W是一对称正定矩阵，存在一可逆矩阵**D**使得

W=**DD**’

用$\mathbf{D}^{-1}$左乘　　Y=**X**β+μ

两边，得到一个新的模型：

$$\mathbf{D}^{-1}\mathbf{Y} = \mathbf{D}^{-1}\mathbf{X}\beta + \mathbf{D}^{-1}\mu$$

$$\mathbf{Y}_* = \mathbf{X}_*\beta + \mu_*$$

该模型具有同方差性。因为

$$E(\mu_*\mu_*{}') = E(\mathbf{D}^{-1}\mu\mu'\mathbf{D}^{-1}{}') = \mathbf{D}^{-1}E(\mu\mu')\mathbf{D}^{-1}{}'$$

$$= \mathbf{D}^{-1}\sigma^2\Omega\mathbf{D}^{-1}{}' = \mathbf{D}^{-1}\sigma^2\mathbf{DD}'\mathbf{D}'^{-1} = \sigma^2\mathbf{I}$$

用OLS法估计新模型，记参数估计量为 $\hat{\beta}_*$ ，则

$$\hat{\beta}_* = (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{Y}_*$$

$$= (\mathbf{X}'\mathbf{D}^{-1'}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^{-1'}\mathbf{D}^{-1}\mathbf{Y}$$

$$= (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{Y}$$

这就是原模型　$\mathrm{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\mu}$

的**加权最小二乘估计量**，是无偏、有效的估计量。

这里权矩阵为D^{-1}，它来自于原模型残差项$\boldsymbol{\mu}$的方差-协方差矩阵$\sigma^2\mathbf{W}$ 。

如何得到$\sigma^2\mathbf{W}$？

从前面的推导过程看，它来自于原模型残差项μ的方差-协方差矩阵。因此

仍对原模型进行OLS估计，得到随机误差项的近似估计量$\check{e}_i$，以此构成权矩阵的估计量，即

$$\sigma^2\hat{\mathbf{W}} = \begin{bmatrix} \tilde{e}_1^2 & & \\ & \ddots & \\ & & \tilde{e}_n^2 \end{bmatrix}$$

这时可直接以 $\mathbf{D}^{-1} = diag\{1/|\tilde{e}_1|, 1/|\tilde{e}_2|, \cdots, 1/|\tilde{e}_n|\}$

作为权矩阵。

注意：

在实际操作中人们通常采用如下的经验方法：

不对原模型进行异方差性检验，而是直接选择加权最小二乘法，尤其是采用截面数据作样本时。

如果确实存在异方差，则被有效地消除了；

如果不存在异方差性，则加权最小二乘法等价于普通最小二乘法

七、案例--中国农村居民人均消费函数

例4.1.4 中国农村居民人均消费支出主要由人均纯收入来决定。

农村人均纯收入包括(1)从事农业经营的收入，(2)包括从事其他产业的经营性收入(3)工资性收入、(4)财产收入(4)转移支付收入。

考察**从事农业经营的收入(X_1)**和**其他收入(X_2)**对中国**农村居民消费支出(Y)**增长的影响:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \mu$$

表 4.1.1　中国 2001 年各地区农村居民家庭人均纯收入与消费支出相关数据（单位：元）

地区	人均消费支出 Y	从事农业经营的收入 X_1	其他收入 X_2	地区	人均消费支出 Y	从事农业经营的收入 X_1	其他收入 X_2
北　京	3552.1	579.1	4446.4	湖　北	2703.36	1242.9	2526.9
天　津	2050.9	1314.6	2633.1	湖　南	1550.62	1068.8	875.6
河　北	1429.8	928.8	1674.8	广　东	1357.43	1386.7	839.8
山　西	1221.6	609.8	1346.2	广　西	1475.16	883.2	1088.0
内蒙古	1554.6	1492.8	480.5	海　南	1497.52	919.3	1067.7
辽　宁	1786.3	1254.3	1303.6	重　庆	1098.39	764.0	647.8
吉　林	1661.7	1634.6	547.6	四　川	1336.25	889.4	644.3
黑龙江	1604.5	1684.1	596.2	贵　州	1123.71	589.6	814.4
上　海	4753.2	652.5	5218.4	云　南	1331.03	614.8	876.0
江　苏	2374.7	1177.6	2607.2	西　藏	1127.37	621.6	887.0
浙　江	3479.2	985.8	3596.6	陕　西	1330.45	803.8	753.5
安　徽	1412.4	1013.1	1006.9	甘　肃	1388.79	859.6	963.4
福　建	2503.1	1053.0	2327.7	青　海	1350.23	1300.1	410.3
江　西	1720.0	1027.8	1203.8	宁　夏	2703.36	1242.9	2526.9
山　东	1905.0	1293.0	1511.6	新　疆	1550.62	1068.8	875.6
河　南	1375.6	1083.8	1014.1				

普通最小二乘法的估计结果：

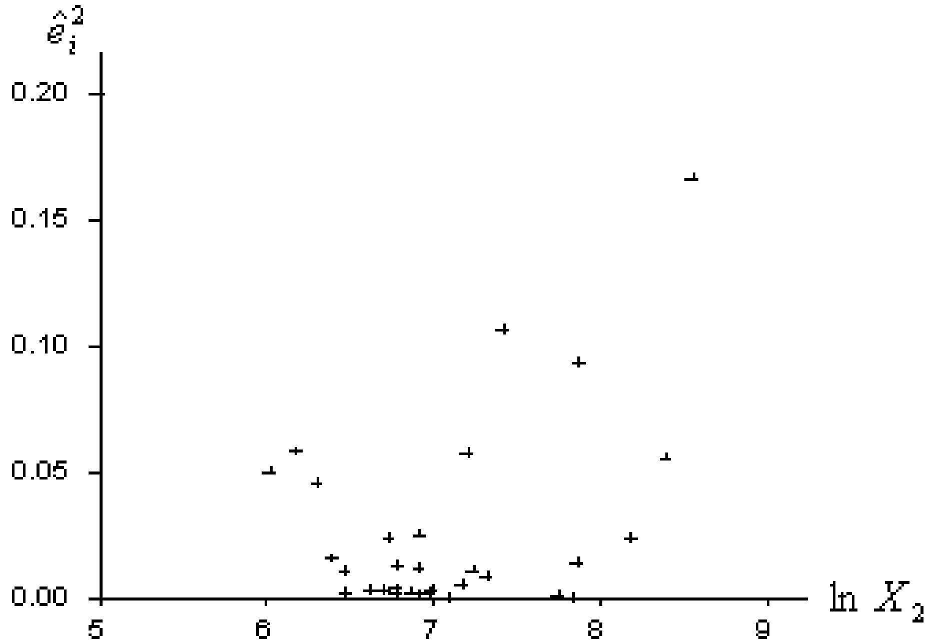
$$\ln \hat{Y} = 1.655 + 0.3166 \ln X_1 + 0.5084 \ln X_2$$

(1.87)　　(3.02)　　（10.04）

R^2 =0.7831　　$\bar{R}^2$ =0.7676　　DW=1.89　F=50.53　　RSS=0.8232

异方差检验

OLS回归的残差平方项 $\hat{e}_i^2$ 与 $\ln X_2$ 的散点图



进一步的统计检验

(1)G-Q检验

将原始数据按X_2排成升序，去掉中间的7个数据，得两个容量为12的子样本。

对两个子样本分别作OLS回归，求各自的残差平方和RSS_1和RSS_2：

子样本1：
$$\ln \hat{Y} = 4.061 + 0.343 \ln X_1 + 0.119 \ln X_2$$
$$(3.18) \quad (4.13) \qquad (0.94)$$

$R^2=0.7068$，　$RSS_1=0.0648$

子样本2：
$$\ln \hat{Y} = 0.791 + 0.138 \ln X_1 + 0.776 \ln X_2$$
$$(0.43) \quad (0.73) \qquad (6.53)$$

$R^2=0.8339$，　$RSS_2=0.2729$

计算F统计量：

$$F = RSS_2/RSS_1 = 0.2792/0.0648 = 4.31$$

查表

给定α=5%，查得临界值 $F_{0.05}(9,9)=2.97$

判断

$F > F_{0.05}(9,9)$

否定两组子样方差相同的假设，从而该总体随机项**存在递增异方差性**。

（2）怀特检验

作辅助回归:

$$\hat{e}^2 = -0.17 + 0.102\ln X_1 + 0.015(\ln X_1)^2 - 0.055\ln X_2 + 0.026(\ln X_2)^2$$

（-0.04）(0.10)　(0.21)　(-0.12)　(1.47)

$$-0.043\ln X_1 \ln X_2$$

(-1.11)

R^2 =0.4638

似乎没有哪个参数的t检验是显著的 。但

nR^2 =31*0.4638=14.38

α=5%下，临界值 $\chi^2_{0.05}(5)$=11.07，**拒绝**同方差性

去掉交叉项后的辅助回归结果

$$\hat{e}^2 = 3.842 - 0.570\ln X_1 + 0.042(\ln X_1)^2 - 0.539\ln X_2 + 0.039(\ln X_2)^2$$

(1.36) (-0.64) (064) (-2.76) (2.90)

$R^2 = 0.4374$

X_2项与X_2的平方项的参数的t检验是显著的，且

$$n R^2 = 31 \times 0.4374 = 13.56$$

α=5%下，临界值 $\chi^2_{0.05}(4)=9.49$

拒绝同方差的原假设

原模型的加权最小二乘回归

对原模型进行OLS估计，得到随机误差项的近似估计量$\check{e}_i$，以此构成权矩阵$\sigma^2 \mathbf{W}$的估计量；
再以$1/|\check{e}_i|$为权重进行WLS估计，得

$$\ln \hat{Y} = 1.497 + 0.319 \ln X_1 + 0.527 \ln X_2$$

(5.12)　(5.94)　（28.94）

R^2=0.9999　$\bar{R}^2$=0.9999　DW=2.49　F=924432　RSS=0.0706

各项统计检验指标全面改善

§4.2 序列相关性

Serial Correlation

§4.2 序列相关性

一、序列相关性概念
二、实际经济问题中的序列相关性
三、序列相关性的后果
四、序列相关性的检验
五、具有序列相关性模型的估计
六、案例

一、序列相关性概念

对于模型

$$Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_k X_{ki}+\mu_i \qquad i=1,2,\ldots,n$$

随机项互不相关的基本假设表现为

$$\mathrm{Cov}(\mu_i,\mu_j)=0 \qquad i\neq j,\ i,j=1,2,\ldots,n$$

如果对于不同的样本点，随机误差项之间不再是不相关的，而是存在某种相关性，则认为出现了*序列相关性*。

在其他假设仍成立的条件下，**序列相关**即意味着

$$E(\mu_i \mu_j) \neq 0$$

或

$$Cov(\boldsymbol{\mu}) = E(\boldsymbol{\mu}\boldsymbol{\mu}') = \begin{pmatrix} \sigma^2 & \cdots & E(\mu_1\mu_n) \\ \vdots & \ddots & \vdots \\ E(\mu_n\mu_1) & \cdots & \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma^2 \end{pmatrix}$$

$$= \sigma^2 \boldsymbol{\Omega} \neq \sigma^2 \mathbf{I}$$

如果仅存在

$$E(\mu_i\ \mu_{i+1})\neq 0 \qquad i=1,2,\ldots,n$$

称为**一阶列相关**，或**自相关**（autocorrelation）

自相关往往可写成如下形式：

$$\mu_i=\rho\mu_{i-1}+\varepsilon_i \qquad -1<\rho<1$$

其中：ρ被称为**自协方差系数**（**coefficient of autocovariance**）或**一阶自相关系数**（**first-order coefficient of autocorrelation**）

ε_i是满足以下标准的OLS假定的随机干扰项：

$$E(\varepsilon_i)=0\ , \qquad \mathrm{var}(\varepsilon_i)=\sigma^2\ , \qquad \mathrm{cov}(\varepsilon_i,\varepsilon_{i-s})=0 \qquad s\neq 0$$

由于序列相关性经常出现在以时间序列为样本的模型中，因此，本节将用下标*t*代表*i*。

二、实际经济问题中的序列相关性

1、经济变量固有的惯性

大多数经济时间数据都有一个明显的特点：**惯性**，表现在时间序列不同时间的前后关联上。

例如，绝对收入假设下**居民总消费函数模型**：

$$C_t=\beta_0+\beta_1 Y_t+\mu_t \qquad t=1,2,\ldots,n$$

由于**消费习惯**的影响被包含在随机误差项中，则可能出现序列相关性（往往是正相关 ）。

2、模型设定的偏误

所谓模型**设定偏误**（**Specification error**）是指所设定的模型“不正确”。主要表现在模型中丢掉了重要的解释变量或模型函数形式有偏误。

例如，本来应该估计的模型为

$$Y_t=\beta_0+\beta_1X_{1t}+\beta_2X_{2t}+\beta_3X_{3t}+\mu_t$$

但在模型设定中做了下述回归：

$$Y_t=\beta_0+\beta_1X_{1t}+\beta_1X_{2t}+v_t$$

因此，$v_t=\beta_3X_{3t}+\mu_t$，如果X_3确实影响Y，则出现序列相关。

又如：如果真实的边际成本回归模型应为：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \mu_t$$

其中：Y=边际成本，X=产出，

但建模时设立了如下模型：

$$Y_t = \beta_0 + \beta_1 X_t + v_t$$

因此，由于$v_t = \beta_2 X_t^2 + \mu_{t,}$，包含了产出的平方对随机项的系统性影响，随机项也呈现序列相关性。

3、数据的“编造”

在实际经济问题中，有些数据是通过已知数据生成的。

因此，新生成的数据与原数据间就有了内在的联系，表现出序列相关性。

例如：**季度数据**来自月度数据的简单平均，这种平均的计算减弱了每月数据的波动性，从而使随机干扰项出现序列相关。

还有就是两个时间点之间的“**内插**”技术往往导致随机项的序列相关性。

二、序列相关性的后果

计量经济学模型一旦出现序列相关性，如果仍采用OLS法估计模型参数，会产生下列不良后果：

1、参数估计量非有效

因为，在有效性证明中利用了

$$E(\mathbf{NN}')=\sigma^2\mathbf{I}$$

即同方差性和互相独立性条件。

而且，在大样本情况下，**参数估计量虽然具有一致性，但仍然不具有渐近有效性。**

2、变量的显著性检验失去意义

在变量的显著性检验中，统计量是建立在参数方差正确估计基础之上的，这只有当随机误差项具有同方差性和互相独立性时才能成立。

如果存在序列相关，估计的参数方差 $s_{\hat{\beta}_i}$ 出现偏误（偏大或偏小），t检验就失去意义。

其他检验也是如此。

3、模型的预测失效

区间预测与参数估计量的方差有关，在方差有偏误的情况下，使得预测估计不准确，预测精度降低。

所以，**当模型出现序列相关性时，它的预测功能失效。**

三、序列相关性的检验

三、序列相关性的检验

基本思路:

序列相关性检验方法有多种，但基本思路相同：

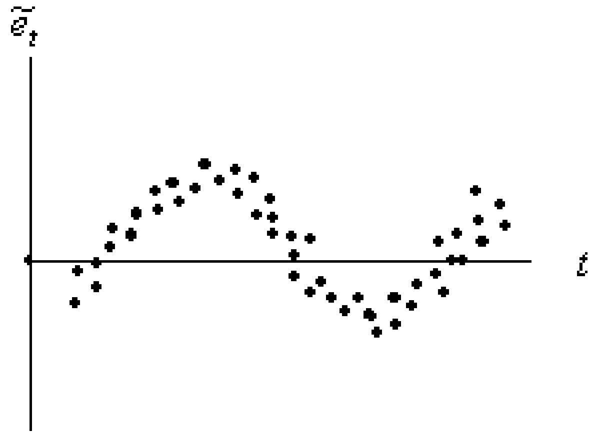
首先，采用 OLS 法估计模型，以求得随机误差项的“**近似估计量**”，用$\tilde{e}_i$表示：

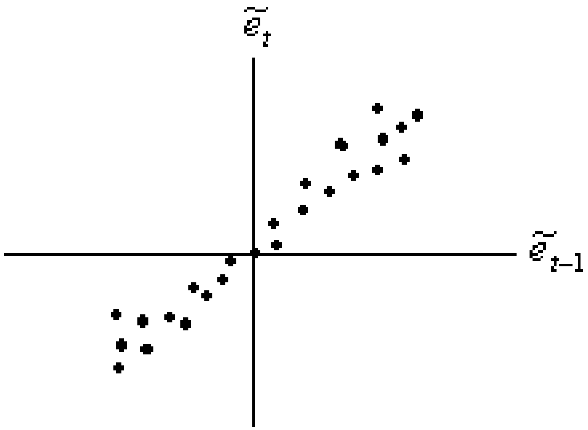
$$\tilde{e}_i = Y_i - (\hat{Y}_i)_{0ls}$$

然后，通过分析这些“**近似估计量**”之间的相关性，以判断随机误差项是否具有序列相关性。

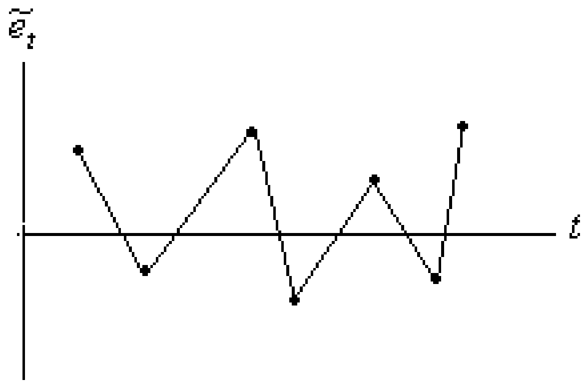
1、图示法

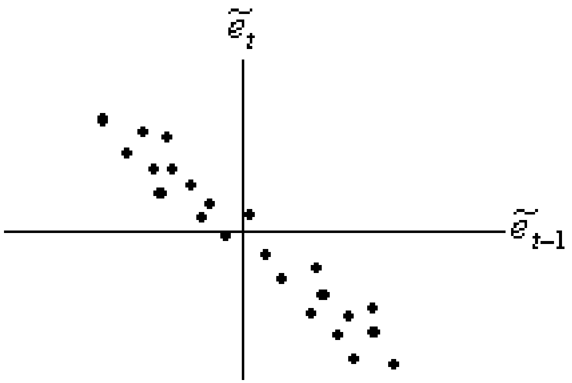
用$\tilde{e}_i$的变化图形来判断μ_i的序列相关性：





正序列相关（正自相关）





负序列相关（负自相关）

2、回归检验法

以$\tilde{e}_t$为被解释变量，以各种可能的相关量，诸如以$\tilde{e}_{t-1}$、$\tilde{e}_{t-2}$、$\tilde{e}_t^2$等为解释变量，建立各种方程：

$$\tilde{e}_t = \rho\tilde{e}_{t-1} + \varepsilon_t$$

$$\tilde{e}_t = \rho_1\tilde{e}_{t-1} + \rho_2\tilde{e}_{t-2} + \varepsilon_t$$

……

如果存在某一种函数形式，使得方程显著成立，则说明原模型存在序列相关性。

回归检验法的**优点**是：（1）能够确定序列相关的形式，（2）适用于任何类型序列相关性问题的检验。

3、杜宾-瓦森（Durbin-Watson）检验法

D-W检验是杜宾（J.Durbin）和瓦森(G.S. Watson)**于1951年提出的一种检验序列自相关的方法，该方法的假定条件是**：

（1）解释变量X非随机；

（2）随机误差项μ_i为一阶自回归形式：

$$\mu_i=\rho\mu_{i-1}+\varepsilon_i$$

（3）回归模型中不应含有滞后应变量作为解释变量，即不应出现下列形式：

$$Y_i=\beta_0+\beta_1X_{1i}+\ldots\beta_kX_{ki}+\gamma Y_{i-1}+\mu_i$$

（4）回归含有截距项

D.W. 统计量:

杜宾和瓦森针对原**假设**：H_0: $\rho=0$， 即不存在一阶自回归，构如下造统计量:

$$D.W. = \frac{\sum_{t=2}^{n}(\tilde{e}_t - \tilde{e}_{t-1})^2}{\sum_{t=1}^{n}\tilde{e}_t^{\,2}}$$

该统计量的分布与出现在给定样本中的X值有复杂的关系，因此其**精确的分布很难得到**。

但是，**他们**成功地导出了临界值的下限d_L和上限d_U，且这些上下限只与样本的容量n和解释变量的个数k有关，而与解释变量X的取值无关。

D.W检验步骤:

（1）计算DW值

（2）给定α，由n和k的大小查DW分布表，得临界值d_L和d_U

（3）比较、判断

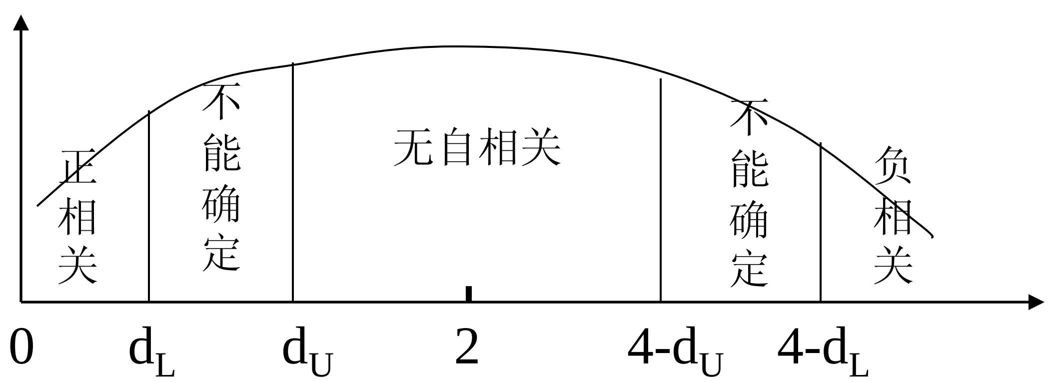
若 $0 < D.W. < d_L$ 存在正自相关

$d_L < D.W. < d_U$ 不能确定

$d_U < D.W. < 4 - d_U$ 无自相关

$4 - d_U < D.W. < 4 - d_L$ 不能确定

$4 - d_L < D.W. < 4$ 存在负自相关



当D.W.值在2左右时，模型不存在一阶自相关。

证明：

展开D.W.统计量：

$$D.W. = \frac{\sum_{t=2}^{n} \tilde{e}_t^2 + \sum_{t=2}^{n} \tilde{e}_{t-1}^2 - 2\sum_{t=2}^{n} \tilde{e}_t \tilde{e}_{t-1}}{\sum_{t=1}^{n} \tilde{e}_t^2} \qquad (*)$$

当n较大时，$\sum_{t=2}^{n} \tilde{e}_t^2$，$\sum_{t=2}^{n} \tilde{e}_{t-1}^2$，$\sum_{t=1}^{n} \tilde{e}_t^2$大致相等，则(*)可以简化为：

$$D.W. \approx 2\left(1 - \frac{\sum_{t=2}^{n} \tilde{e}_t \tilde{e}_{t-1}}{\sum_{t=1}^{n} \tilde{e}_t^2}\right) \approx 2(1-\rho)$$

$$D.W. \approx 2\left(1-\frac{\sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1}}{\sum_{t=1}^{n}\tilde{e}_t^{\,2}}\right) \approx 2(1-\rho)$$

这里，

$$\sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1} \Big/ \sum_{t=1}^{n}\tilde{e}_t^{\,2} \approx \sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1} \Big/ \sum_{t=2}^{n}\tilde{e}_t^{\,2} = \rho$$

为一阶自回归模型

$$\mu_i=\rho\mu_{i-1}+\varepsilon_i$$

的参数估计。

如果存在完全一阶正相关，即**ρ=1**，则 **D.W.≈ 0**

完全一阶负相关，即**ρ= -1**, 则 **D.W.≈ 4**

完全不相关，即**ρ=0**，则 **D.W.≈2**

4、拉格朗日乘数（Lagrange multiplier）检验

拉格朗日乘数检验克服了DW检验的缺陷，适合于高阶序列相关以及模型中存在滞后被解释变量的情形。

它是由布劳殊（Breusch）与戈弗雷（Godfrey）于1978年提出的，也被称为**GB检验**。

对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

如果怀疑随机扰动项存在**p阶序列相关**：

$$\mu_t = \rho_1 \mu_{t-1} + \rho_2 \mu_{t-2} \cdots + \rho_p \mu_{t-p} + \varepsilon_t$$

GB检验可用来检验如下受约束回归方程

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \rho_1 \mu_{t-1} + \cdots + \rho_p \mu_{t-p} + \varepsilon_t$$

约束条件为：

H_0: $\rho_1 = \rho_2 = \ldots = \rho_p = 0$

约束条件H_0为真时，大样本下

$$LM = (n-p)R^2 \sim \chi^2(p)$$

其中，n为样本容量，R^2为如下辅助回归的可决系数：

$$\tilde{e}_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + \tilde{e}_1 \mu_{t-1} + \cdots + \rho_p \tilde{e}_{t-p} + \varepsilon_t$$

给定α，查临界值$\chi_\alpha^2(p)$，与LM值比较，做出判断，

实际检验中，可从1阶、2阶、…逐次向更高阶检验。

四、序列相关的补救

如果模型被检验证明存在序列相关性，则需要发展新的方法估计模型。

最常用的方法是**广义最小二乘法**（**GLS: Generalized least squares**）和**广义差分法** **(Generalized Difference)**。

1、广义最小二乘法

对于模型

$$\mathbf{Y=X\beta+\mu}$$

如果存在序列相关，同时存在异方差，即有

$$\text{Cov}(\mu,\mu') = E(\mu\mu') = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \sigma^2 \mathbf{\Omega}$$

Ω是一对称正定矩阵，存在一可逆矩阵**D**，使得

$$\mathbf{\Omega=DD'}$$

变换原模型：

$$\mathbf{D}^{-1}\mathbf{Y}=\mathbf{D}^{-1}\mathbf{X}\,\boldsymbol{\beta}+\mathbf{D}^{-1}\boldsymbol{\mu}$$

即

$$\mathbf{Y}_*=\mathbf{X}_*\boldsymbol{\beta}+\boldsymbol{\mu}_* \qquad (*)$$

该模型具有同方差性和随机误差项互相独立性：

$$E(\mu_*\mu_*') = E(\mathbf{D}^{-1}\mu\,\mu'\mathbf{D}^{-1'}) = \mathbf{D}^{-1}E(\mu\,\mu')\mathbf{D}^{-1'}$$

$$= \mathbf{D}^{-1}\sigma^2\boldsymbol{\Omega}\,\mathbf{D}^{-1'} = \mathbf{D}^{-1}\sigma^2\mathbf{D}\mathbf{D}'\mathbf{D}'^{-1}$$

$$= \sigma^2\mathbf{I}$$

(*)式的OLS估计：

$$\hat{\beta}_* = (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{Y}_*$$

$$= (\mathbf{X}'\mathbf{D}^{-1'}\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^{-1'}\mathbf{D}^{-1}\mathbf{Y}$$

$$= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}$$

这就是原模型的广义最小二乘估计量(GLS estimators)，是无偏的、有效的估计量。

如何得到矩阵Ω？

对**Ω**的形式进行特殊设定后，才可得到其估计值。

如设定随机扰动项为一阶序列相关形式

$$\mu_i=\rho\mu_{i-1}+\varepsilon_i$$

则

$$Cov(\mu,\mu')=\frac{\sigma_\varepsilon^2}{1-\rho^2}\begin{pmatrix}1 & \rho & \cdots & \rho^{n-1}\\ \rho & 1 & \cdots & \rho^{n-2}\\ \cdots & \cdots & \cdots & \cdots\\ \rho^{n-1} & \rho^{n-2} & \cdots & 1\end{pmatrix}=\sigma^2\mathbf{\Omega}$$

$$\mathbf{D}^{-1}=\begin{pmatrix}\sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0\\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0\\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1\end{pmatrix}$$

2、广义差分法

广义差分法是将原模型变换为满足OLS法的差分模型，再进行OLS估计。

如果原模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

存在

$$\mu_t = \rho_1 \mu_{t-1} + \rho_2 \mu_{t-2} + \cdots + \rho_l \mu_{t-l} + \varepsilon_t$$

可以将原模型变换为:

$$Y_t - \rho_1 Y_{t-1} - \cdots - \rho_l Y_{t-l} = \beta_0 (1 - \rho_1 - \cdots - \rho_l) + \beta_1 (X_{1t} - \rho_1 X_{1t-1} - \cdots - \rho_l X_{1t-l})$$
$$+ \cdots + \beta_k (X_{kt} - \rho_1 X_{kt-1} - \cdots - \rho_l X_{kt-l}) + \varepsilon_t$$

该模型为**广义差分模型**，不存在序列相关问题。可进行**OLS**估计。

注意：

- **广义差分法**就是上述**广义最小二乘法**，但是却损失了部分样本观测值。

如：一阶序列相关的情况下，广义差分是估计

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + \cdots + \beta_k(X_{kt} - \rho X_{kt-1}) + \varepsilon_t$$

$$t = 2,3,\cdots,n$$

这相当于

$$\mathbf{D}^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}$$

去掉第一行后左乘原模型**Y=Xβ+ μ** 。即运用了**GLS**法，但第一次观测值被排除了。

3、随机误差项相关系数的估计

应用广义最小二乘法或广义差分法，必须已知随机误差项的相关系数$\rho_1, \rho_2, \ldots, \rho_L$。

实际上，人们并不知道它们的具体数值，所以必须首先对它们进行估计。

常用的估计方法有：

- 科克伦-奥科特（Cochrane-Orcutt）迭代法。
- 杜宾（durbin）两步法

（1）科克伦-奥科特迭代法。

以一元线性模型为例：

首先，采用OLS法估计原模型

$$Y_i=\beta_0+\beta_1X_i+\mu_i$$

得到的μ的“近似估计值”，并以之作为观测值使用OLS法估计下式

$$\mu_i=\rho_1\mu_{i-1}+\rho_2\mu_{i-2}+\dots\rho_L\mu_{i-L}+\varepsilon_i$$

得到$\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_l$，作为随机误差项的相关系数$\rho_1,\rho_2,\cdots,\rho_l$的**第一次估计值**。

其次，将$\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_L$代入广义差分模型

$$Y_i - \rho_1 Y_{i-1} - \cdots - \rho_l Y_{i-l} = \beta_0(1 - \hat{\rho}_1 - \cdots - \hat{\rho}_l) + \beta_1(X_i - \hat{\rho}_1 X_{i-1} - \cdots - \hat{\rho}_l X_{i-l}) + \varepsilon_i$$

$$i = 1 + l, 2 + l, \cdots, n$$

进行OLS估计，得到$\hat{\hat{\beta}}_0$，$\hat{\hat{\beta}}_1$

再次，将$\hat{\hat{\beta}}_0, \hat{\hat{\beta}}_1$代回原模型　$Y_i=\beta_0+\beta_1 X_i+\mu_i$

求出μ_i新的“近拟估计值”，并以之作为样本观测值，再次估计

$$\mu_i=\rho_1\mu_{i-1}+\rho_2\mu_{i-2}+\ldots\rho_L\mu_{i-L}+\varepsilon_i$$

得到$\rho_1, \rho_2, \ldots, \rho_L$的第二次估计值$\hat{\hat{\rho}}_1, \hat{\hat{\rho}}_2, \ldots, \hat{\hat{\rho}}_L$

类似地，可进行第三次、第四次迭代。

关于迭代的次数，可根据具体的问题来定。

一般是事先给出一个精度，当相邻两次$\rho_1, \rho_2, \ldots, \rho_L$的估计值之差小于这一精度时，迭代终止。

实践中，有时只要迭代两次，就可得到较满意的结果。两次迭代过程也被称为**科克伦-奥科特两步法**。

（2）杜宾（durbin）两步法

该方法仍是先估计$\rho_1,\rho_2,\ldots,\rho_l$，再对差分模型进行估计

第一步，变换差分模型为下列形式

$$Y_i = \rho_1 Y_{i-1} + \cdots + \rho_l Y_{i-l} + \beta_0(1-\hat{\rho}_1 - \cdots - \hat{\rho}_l) + \beta_1(X_i - \hat{\rho}_1 X_{i-1} - \cdots - \hat{\rho}_l X_{i-l}) + \varepsilon_i$$

$$i = 1+l, 2+l, \cdots, n$$

进行OLS估计，得各Y_j（$j=i-1, i-2, \ldots, i-l$)前的系数$\rho_1,\rho_2, \ldots, \rho_l$的估计值$\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_L$

第二步，将估计的$\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_l$代入差分模型

$$Y_i-\rho_1 Y_{i-1}-\cdots-\rho_l Y_{i-l}=\beta_0(1-\rho_1-\cdots-\rho_l)+\beta_1(X_i-\rho_1 X_{i-1}-\cdots-\rho_l X_{i-l})+\varepsilon_i$$

$$i=1+l,2+l,\cdots,n$$

采用 OLS 法估计，得到参数$\beta_0(1-\hat{\rho}_1-\cdots-\hat{\rho}_l),\beta_1$的估计量，记为$\hat{\beta}_0^*$，$\hat{\beta}_1^*$。

于是：

$$\hat{\beta}_0=\hat{\beta}_0^*\Big/(1-\hat{\rho}_1-\cdots-\hat{\rho}_l)\,,\qquad \hat{\beta}_1=\hat{\beta}_1^*$$

•应用软件中的广义差分法

在**Eview**/**TSP**软件包下，广义差分采用了科克伦-奥科特（**Cochrane-Orcutt**）迭代法估计ρ。

在解释变量中引入**AR(1)**、**AR(2)**、**…**，即可得到参数和ρ_1、ρ_2、**…**的估计值。

其中**AR(m)**表示随机误差项的**m**阶自回归。在估计过程中自动完成了ρ_1、ρ_2、**…**的迭代。

注意：

- 如果能够找到一种方法，求得**Ω**或各序列相关系数ρ_j的估计量，使得GLS能够实现，则称为**可行的广义最小二乘法**（**FGLS, Feasible Generalized Least Squares**）。
- **FGLS**估计量，也称为**可行的广义最小二乘估计量**（**feasible general least squares estimators**）
- 可行的广义最小二乘估计量不再是无偏的，但却是一致的，而且在科克伦-奥科特迭代法下，估计量也具有渐近有效性。
- 前面提出的方法，就是**FGLS**

4、虚假序列相关问题

由于随机项的序列相关往往是在模型设定中遗漏了重要的解释变量或对模型的函数形式设定有误，这种情形可称为**虚假序列相关(false autocorrelation)**，应在模型设定中排除。

避免产生虚假序列相关性的措施是在开始时建立一个“一般”的模型，然后逐渐剔除确实不显著的变量。

五、案例：中国商品进口模型

经济理论指出，**商品进口**主要由进口国的**经济发展水平**，以及**商品进口价格指数**与**国内价格指数**对比因素决定的。

由于无法取得中国商品进口价格指数，我们主要研究中国商品进口与国内生产总值的关系。（下表）。

表 4.2.1　1978~2001 年中国商品进口与国内生产总值

	国内生产总值 GDP（亿元）	商品进口 M（亿美元）		国内生产总值 GDP（亿元）	商品进口 M（亿美元）
1978	3624.1	108.9	1990	18547.9	533.5
1979	4038.2	156.7	1991	21617.8	637.9
1980	4517.8	200.2	1992	26638.1	805.9
1981	4862.4	220.2	1993	34634.4	1039.6
1982	5294.7	192.9	1994	46759.4	1156.1
1983	5934.5	213.9	1995	58478.1	1320.8
1984	7171.0	274.1	1996	67884.6	1388.3
1985	8964.4	422.5	1997	74462.6	1423.7
1986	10202.2	429.1	1998	78345.2	1402.4
1987	11962.5	432.1	1999	82067.46	1657
1988	14928.3	552.7	2000	89442.2	2250.9
1989	16909.2	591.4	2001	95933.3	2436.1

资料来源：《中国统计年鉴》（1995、2000、2002）。

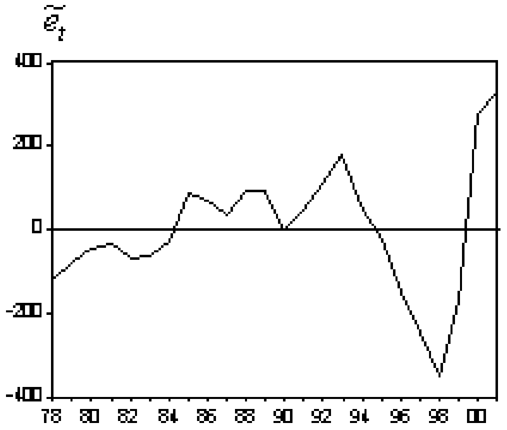
1. 通过OLS法建立如下中国商品进口方程:

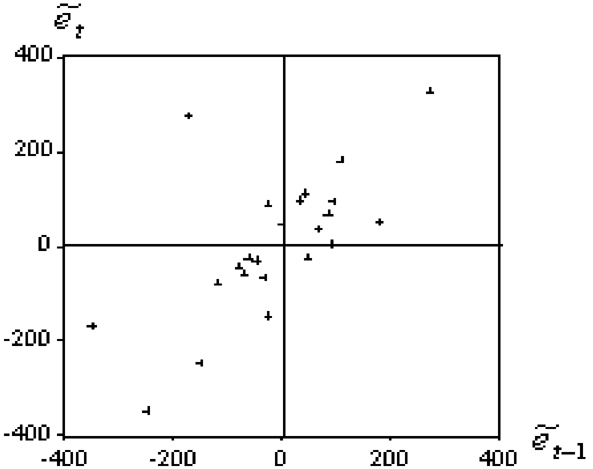
$$\hat{M}_t = 152.91 + 0.02GDP_t$$

（2.32）　（20.12）

$R^2=0.948$　$\bar{R}^2=0.946$　$SE=154.9$　$DW=0.628$

2. 进行序列相关性检验。





- **DW检验**

取α=5%，由于n=24，k=2(包含常数项)，查表得：

d_l=1.27，　　d_u=1.45

由于　　DW=0.628< d_l ，故：存在正自相关。

- **拉格朗日乘数检验**

2阶滞后：

$$\tilde{e}_t = 6.593 - 0.0003GDP_t + 1.094\tilde{e}_{t-1} - 0.786\tilde{e}_{t-2}$$

（0.23）（-0.50）　(6.23)　（-3.69）

R^2=0. 6614

于是，LM=22×0.6614=14.55

取α=5%，χ^2分布的临界值$\chi^2_{0.05}(2)$=5.991

$LM > \chi^2_{0.05}(2)$　　故：存在正自相关

3阶滞后：

$$\tilde{e}_t = 6.692 - 0.0003GDP + 1.108\tilde{e}_{t-1} - 0.819\tilde{e}_{t-2} + 0.032\tilde{e}_{t-3}$$

(0.22) (-0.497) (4.541) （-1.842） （0.087）

R^2=0. 6615

于是，LM=21×0.6614=13.89

取α=5%，χ^2分布的临界值$\chi^2_{0.05}(3)=7.815$

$$LM > \chi^2_{0.05}(3)$$

表明：存在正自相关；但$\check{e}_{t-3}$的参数不显著，说明不存在3阶序列相关性。

3、运用广义差分法进行自相关的处理

（1）采用杜宾两步法估计ρ

第一步，估计模型

$$M_t = \beta_0^* + \rho_1 M_{t-1} + \rho_2 M_{t-2} + \beta_1^* GDP_t + \beta_2^* GDP_{t-1} + \beta_3^* GDP_{t-2} + \varepsilon_t$$

$$\hat{M}_t = 78.09 + 0.938M_{t-1} - 0.469M_{t-2} + 0.055GDP_t - 0.096GDP_{t-1} + 0.054GDP_{t-2}$$

（1.76） (6.64) (-1.76) (5.88) (-5.19) (5.30)

$R^2 = 0.9913$, $\bar{R}^2 = 0.9886$, D.W.=2.31

第二步，作差分变换：

$$M_t^* = M_t - (0.938M_{t-1} - 0.469M_{t-2})$$

$$GDP_t^* = GDP_t - (0.938GDP_{t-1} - 0.469GDP_{t-2})$$

则M*关于GDP*的OLS估计结果为：

$$\hat{M}_t^* = 86.18 + 0.020 GDP_t^*$$

（2.76） (16.46)

R^2 =0.9313 $\overline{R}^2$ =0.9279 D.W.=1.583

取α=5%，DW>d_u=1.43 (样本容量24-2=22)

表明：已不存在自相关

为了与OLS估计结果对比，计算$\hat{\beta}_0$：

$$\hat{\beta}_0 = \hat{\beta}_0^* /(1 - \hat{\rho}_1 - \hat{\rho}_2) = 86.18 /(1 - 0.938 + 0.469) = 162.30$$

于是原模型为：

$$\hat{M}_t = 162.30 + 0.020 GDP_t$$

与OLS估计结果的差别只在截距项：$\hat{M}_t = 152.91 + 0.02 GDP_t$

（2）采用科克伦-奥科特迭代法估计ρ

在Eviews软包下，2阶广义差分的结果为：

$$\hat{M}_t = 169.32 + 0.020 GDP_t + 1.108 AR[1] - 0.801 AR[2]$$

(3.81)　(18.45)　(6.11)　(-3.61)

$R^2=0.982$　$\bar{R}^2=0.979$　D.W.=1.85

取$\alpha=5\%$ ，$DW>d_u=1.66$(样本容量:22)

表明:广义差分模型已不存在序列相关性。

可以验证：仅采用1阶广义差分，变换后的模型仍存在1阶自相关性；

采用3阶广义差分，变换后的模型不再有自相关性，但AR[3]的系数的t值不显著。

§4.3 多重共线性

Multi-Collinearity

§4.3 多重共线性

- 一、多重共线性的概念
- 二、实际经济问题中的多重共线性
- 三、多重共线性的后果
- 四、多重共线性的检验
- 五、克服多重共线性的方法
- 六、案例
- *七、分部回归与多重共线性

一、多重共线性的概念

对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \mu_i$$

$$i = 1, 2, \ldots, n$$

其基本假设之一是解释变量是互相独立的。

如果某两个或多个解释变量之间出现了相关性，则称为多重共线性(Multicollinearity)。

如果存在

$$c_1X_{1i}+c_2X_{2i}+\ldots+c_kX_{ki}=0 \qquad i=1,2,\ldots,n$$

其中：c_i不全为0，**则称为解释变量间存在完全共线性**（**perfect multicollinearity**）。

如果存在

$$c_1X_{1i}+c_2X_{2i}+\ldots+c_kX_{ki}+v_i=0 \qquad i=1,2,\ldots,n$$

其中c_i不全为0，v_i为随机误差项，则称为 **近似共线性**（**approximate multicollinearity**）或**交互相关** **(intercorrelated)**。

在矩阵表示的线性回归模型

Y=Xβ+μ

中，**完全共线性指：秩(X)<k+1**，即

$$X=\begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{pmatrix}$$

中，至少有一列向量可由其他列向量（不包括第一列）线性表出。

如：$X_2=\lambda X_1$，则X_2对Y的作用可由X_1代替。

注意：

完全共线性的情况并不多见，一般出现的是在一定程度上的共线性，即近似共线性。

二、实际经济问题中的多重共线性

一般地，产生多重共线性的主要原因有以下三个方面：

（1）经济变量相关的共同趋势

时间序列样本：经济繁荣时期，各基本经济变量（收入、消费、投资、价格）都趋于增长；衰退时期，又同时趋于下降。

横截面数据：生产函数中，资本投入与劳动力投入往往出现高度相关情况，大企业二者都大，小企业都小。

（2）滞后变量的引入

在经济计量模型中，往往需要引入滞后经济变量来反映真实的经济关系。

例如，消费=f(当期收入, 前期收入）

显然，两期收入间有较强的线性相关性。

（3）样本资料的限制

由于完全符合理论模型所要求的样本数据较难收集，特定样本可能存在某种程度的多重共线性。

一般经验：

时间序列数据样本：简单线性模型，往往存在多重共线性。

截面数据样本：问题不那么严重，但多重共线性仍然是存在的。

二、多重共线性的后果

1、完全共线性下参数估计量不存在

$$\mathbf{Y} = \mathbf{X}\beta + \mu$$

的OLS估计量为：

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

如果存在完全共线性，则$(\mathbf{X}'\mathbf{X})^{-1}$不存在，无法得到参数的估计量。

例：对离差形式的二元回归模型

$$y = \beta_1 x_1 + \beta_2 x_2 + \mu$$

如果两个解释变量完全相关，如$x_2 = \lambda x_1$，则

$$y = (\beta_1 + \lambda\beta_2) x_1 + \mu$$

这时，只能确定综合参数$\beta_1+\lambda\beta_2$的估计值：

$$\widehat{\beta_1 + \lambda\beta_2} = \sum x_{1i} y_i \Big/ \sum x_{1i}^2$$

2、近似共线性下OLS估计量非有效

近似共线性下，可以得到OLS参数估计量，但参数估计量方差的表达式为

$$Cov(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

由于$|X'X| \approx 0$，引起$(X'X)^{-1}$主对角线元素较大，使参数估计值的方差增大，**OLS参数估计量非有效。**

仍以二元线性模型 $y=\beta_1 x_1+\beta_2 x_2+\mu$ 为例:

$$\operatorname{var}(\hat{\beta}_1)=\sigma^2(X'X)_{11}^{-1}=\frac{\sigma^2\sum x_{2i}^2}{\sum x_{1i}^2\sum x_{2i}^2-(\sum x_{1i}x_{2i})^2}=\frac{\sigma^2/\sum x_{1i}^2}{1-(\sum x_{1i}x_{2i})^2\Big/\sum x_{1i}^2\sum x_{2i}^2}$$

$$=\frac{\sigma^2}{\sum x_{1i}^2}\cdot\frac{1}{1-r^2}$$

$\dfrac{(\sum x_{1i}x_{2i})^2}{\sum x_{1i}^2\sum x_{2i}^2}$ 恰为X_1与X_2的线性相关系数的平方r^2

由于 $r^2\leq 1$，故 $1/(1-r^2)\geq 1$

当完全不共线时, $r^2=0$ $\quad \operatorname{var}(\hat{\beta}_1)=\sigma^2/\sum x_{1i}^2$

当近似共线时, $0<r^2<1$ $\quad \operatorname{var}(\hat{\beta}_1)=\dfrac{\sigma^2}{\sum x_{1i}^2}\cdot\dfrac{1}{1-r^2}>\dfrac{\sigma^2}{\sum x_{1i}^2}$

多重共线性使参数估计值的方差增大，$1/(1-r^2)$为**方差膨胀因子(Variance Inflation Factor, VIF)**

表 4.3.1 方差膨胀因子表

相关系数平方	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99	0.999
方差膨胀因子	1	2	5	10	20	25	33	50	100	1000

当完全共线时， $r^2=1$， $\quad \operatorname{var}(\hat{\beta}_1)=\infty$

3、参数估计量经济含义不合理

如果模型中两个解释变量具有线性相关性，例如 $X_2 = \lambda X_1$，

这时，X_1和X_2前的参数β_1、β_2并不反映各自与被解释变量之间的结构关系，而是反映它们对被解释变量的共同影响。

β_1、β_2已经失去了应有的经济含义，于是经常表现出**似乎反常的现象**：例如β_1本来应该是正的，结果恰是负的。

4、变量的显著性检验失去意义

存在多重共线性时

↓

参数估计值的方差与标准差变大

↓

容易使通过样本计算的t值小于临界值，
误导作出参数为0的推断

↓

可能将重要的解释变量排除在模型之外

5、模型的预测功能失效

变大的方差容易使区间预测的“区间”变大，使预测失去意义。

注意：

除非是完全共线性，多重共线性并不意味着任何基本假设的违背；

因此，即使出现较高程度的多重共线性，OLS估计量仍具有线性性等良好的统计性质。

问题在于，即使OLS法仍是最好的估计方法，它却不是“完美的”，尤其是在统计推断上无法给出真正有用的信息。

三、多重共线性的检验

多重共线性表现为解释变量之间具有相关关系，所以用于多重共线性的检验方法主要是统计方法：如判定系数检验法、逐步回归检验法等。

多重共线性检验的任务是：

（1）检验多重共线性是否存在；

（2）估计多重共线性的范围，即判断哪些变量之间存在共线性。

1、检验多重共线性是否存在

(1) 对两个解释变量的模型，采用简单相关系数法

求出X_1与X_2的简单相关系数r，若|r|接近1，则说明两变量存在较强的多重共线性。

(2) 对多个解释变量的模型，采用综合统计检验法

若 在OLS法下：R^2与F值较大，但t检验值较小，说明各解释变量对Y的联合线性作用显著，但各解释变量间存在共线性而使得它们对Y的独立作用不能分辨，故t检验不显著。

2、判明存在多重共线性的范围

如果存在多重共线性，需进一步确定究竟由哪些变量引起。

(1) 判定系数检验法

使模型中每一个解释变量分别以其余解释变量为解释变量进行回归，并计算相应的拟合优度。

如果某一种回归

$$X_{ji}=\alpha_1 X_{1i}+\alpha_2 X_{2i}+\ldots\alpha_L X_{Li}$$

的**判定系数**较大，说明X_j与其他X间存在**共线性**。

具体可进一步对上述回归方程作F检验：

构造如下F统计量

$$F_j=\frac{R_{j.}^2/(k-2)}{(1-R_{j.}^2)/(n-k+1)}\sim F(k-2,n-k+1)$$

式中：$R_{j\cdot}^2$为第j个解释变量对其他解释变量的回归方程的决定系数，

若存在较强的共线性，则$R_{j\cdot}^2$较大且接近于1，这时（$1-R_{j\cdot}^2$）较小，从而F_j的值较大。

因此，给定显著性水平α，计算F值，并与相应的临界值比较，来判定是否存在相关性。

另一等价的检验是:

在模型中排除某一个解释变量X_j，估计模型;

如果拟合优度与包含X_j时十分接近，则说明X_j与其它解释变量之间存在共线性。

(2)逐步回归法

以Y为被解释变量，逐个引入解释变量，构成回归模型，进行模型估计。

根据拟合优度的变化决定新引入的变量是否独立。

如果拟合优度变化显著，则说明新引入的变量是一个独立解释变量；

如果拟合优度变化很不显著，则说明新引入的变量与其它变量之间存在共线性关系。

四、克服多重共线性的方法

如果模型被检验证明存在多重共线性，则需要发展新的方法估计模型，最常用的方法有三类。

1、第一类方法：排除引起共线性的变量

找出引起多重共线性的解释变量，将它排除出去。

以**逐步回归法**得到最广泛的应用。

- **注意：**

这时，剩余解释变量参数的经济含义和数值都发生了变化。

2、第二类方法：差分法

时间序列数据、线性模型：将原模型变换为差分模型:

$$\Delta Y_i=\beta_1 \Delta X_{1i}+\beta_2 \Delta X_{2i}+\ldots+\beta_k \Delta X_{ki}+\Delta \mu_i$$

可以有效地消除原模型中的多重共线性。

一般讲，增量之间的线性关系远比总量之间的线性关系弱得多。

例如：

表 4.3.2　中国 GDP 与居民消费 C 的总量与增量数据（亿元）

年份	C	Y	C/Y	△C	△Y	△C/△Y
1978	1759.1	3605.6	0.488			
1979	2005.4	4074.0	0.492	246.3	468.4	0.526
1980	2317.1	4551.3	0.509	311.7	477.3	0.653
1981	2604.1	4901.4	0.531	287.0	350.1	0.820
1982	2867.9	5489.2	0.522	263.8	587.8	0.449
1983	3182.5	6076.3	0.524	314.6	587.1	0.536
1984	3674.5	7164.4	0.513	492.0	1088.1	0.452
1985	4589.0	8792.1	0.522	914.5	1627.7	0.562
1986	5175.0	10132.8	0.511	586.0	1340.7	0.437
1987	5961.2	11784.7	0.506	786.2	1651.9	0.476
1988	7633.1	14704.0	0.519	1671.9	2919.3	0.573
1989	8523.5	16466.0	0.518	890.4	1762.0	0.505
1990	9113.2	18319.5	0.497	589.7	1853.5	0.318
1991	10315.9	21280.4	0.485	1202.7	2960.9	0.406
1992	12459.8	25863.7	0.482	2143.9	4583.3	0.468
1993	15682.4	34500.7	0.455	3222.6	8637.0	0.373
1994	20809.8	46690.7	0.446	5127.4	12190.0	0.421
1995	26944.5	58510.5	0.461	6134.7	11819.8	0.519
1996	32152.3	68330.4	0.471	5207.8	9819.9	0.530
1997	34854.6	74894.2	0.465	2702.3	6563.8	0.412
1998	36921.1	79003.3	0.467	2066.5	4109.1	0.503
1999	39334.4	82673.1	0.476	2413.3	3669.8	0.658
2000	42911.9	89112.5	0.482	3577.5	6439.4	0.556

由表中的比值可以直观地看到，**增量的线性关系弱于总量之间的线性关系**。

进一步分析：

Y与C(-1)之间的判定系数为0.9988，

△Y与△C(-1)之间的判定系数为0.9567

3、第三类方法：减小参数估计量的方差

多重共线性的主要后果是参数估计量具有较大的方差，所以

采取适当方法减小参数估计量的方差，虽然没有消除模型中的多重共线性，但确能消除多重共线性造成的后果。

例如：

①增加样本容量，**可使参数估计量的方差减小。**

*②岭回归法（Ridge Regression）

70年代发展的岭回归法，**以引入偏误为代价减小参数估计量的方差**，受到人们的重视。

具体方法是：引入矩阵**D**，使参数估计量为

$$\hat{\beta} = (\mathbf{X}'\mathbf{X} + \mathbf{D})^{-1}\mathbf{X}'\mathbf{Y} \qquad (*)$$

其中矩阵**D**一般选择为主对角阵，即

D=aI

a为大于0的常数。

显然，与未含D的参数B的估计量相比，(*)式的估计量有较小的方差。

六、案例——中国粮食生产函数

根据理论和经验分析，影响粮食生产（Y）的主要因素有：

农业化肥施用量（X_1）；粮食播种面积(X_2)

成灾面积(X_3); 农业机械总动力(X_4);

农业劳动力(X_5)

已知中国粮食生产的相关数据，建立中国粮食生产函数：

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\beta_4 X_4+\beta_4 X_5+\mu$$

表 4.3.3　中国粮食生产与相关投入资料

年份	粮食产量 Y (万吨)	农业化肥施用量 X_1 （万公斤）	粮食播种面积 X_2 （千公顷）	受灾面积 X_3 （公顷）	农业机械总动力 X_4 （万千瓦）	农业劳动 X_5 （万人）
1983	38728	1659.8	114047	16209.3	18022	31645.1
1984	40731	1739.8	112884	15264.0	19497	31685.0
1985	37911	1775.8	108845	22705.3	20913	30351.5
1986	39151	1930.6	110933	23656.0	22950	30467.0
1987	40208	1999.3	111268	20392.7	24836	30870.0
1988	39408	2141.5	110123	23944.7	26575	31455.7
1989	40755	2357.1	112205	24448.7	28067	32440.5
1990	44624	2590.3	113466	17819.3	28708	33330.4
1991	43529	2806.1	112314	27814.0	29389	34186.3
1992	44264	2930.2	110560	25894.7	30308	34037.0
1993	45649	3151.9	110509	23133.0	31817	33258.2
1994	44510	3317.9	109544	31383.0	33802	32690.3
1995	46662	3593.7	110060	22267.0	36118	32334.5
1996	50454	3827.9	112548	21233.0	38547	32260.4
1997	49417	3980.7	112912	30309.0	42016	32434.9
1998	51230	4083.7	113787	25181.0	45208	32626.4
1999	50839	4124.3	113161	26731.0	48996	32911.8
2000	46218	4146.4	108463	34374.0	52574	32797.5

1、用OLS法估计上述模型：

$$\hat{Y} = -12816.44 + 6.213X_1 + 0.421X_2 - 0.166X_3 - 0.098X_4 - 0.028X_5$$

(-0.91) (8.39) (3.32) (-2.81) (-1.45) (-0.14)

$R^2=0.9828$ $\bar{R}^2=0.9756$ $F=137.11$ $DW=1.81$

R^2接近于1；

给定$\alpha=5\%$，得F临界值 $F_{0.05}(5,12)=3.11$

$F=638.4 > 15.19$，

故认上述粮食生产的总体线性关系显著成立。

但X_4、X_5的参数未通过t检验，且符号不正确，故**解释变量间可能存在多重共线性**。

2、检验简单相关系数

列出X_1，X_2，X_3，X_4，X_5的相关系数矩阵：

	X1	X2	X3	X4	X5
X1	1.00	0.01	0.64	0.96	0.55
X2	0.01	1.00	-0.45	-0.04	0.18
X3	0.64	-0.45	1.00	0.69	0.36
X4	0.96	-0.04	0.69	1.00	0.45
X5	0.55	0.18	0.36	0.45	1.00

- **发现：** X_1与X_4间存在高度相关性。

3、找出最简单的回归形式

分别作Y与X_1，X_2，X_4，X_5间的回归：

$$\hat{Y} = 30867.64 + 4.576 X_1$$
(25.58)　　(11.49)

R^2=0.8919　F=132.1　DW=1.56

$$\hat{Y} = -33821.18 + 0.699 X_2$$
(-0.49)　　(1.14)

R^2=0.075　F=1.30　DW=0.12

$$\hat{Y} = 31919.0 + 0.380 X_4$$
(17.45)　　(6.68)

R^2=0.7527　F=48.7　DW=1.11

$$\hat{Y} = -28259.19 + 2.240 X_5$$
(-1.04)　　(2.66)

R^2=0.3064　F=7.07　DW=0.36

- 可见，应选第1个式子为初始的回归模型。

4、逐步回归

将其他解释变量分别导入上述初始回归模型，寻找最佳回归方程。

	C	X1	X2	X3	X4	X5	$\bar{R}^2$	DW
Y=f(X1)	30868	4.23					0.8852	1.56
t 值	25.58	11.49						
Y=f(X1,X2)	-43871	4.65	0.67				0.9558	2.01
t 值	-3.02	18.47	5.16					
Y=f(X1,X2,X3)	-11978	5.26	0.41	-0.19			0.9752	1.53
t 值	0.85	19.6	3.35	-3.57				
Y=f(X1,X2,X3,X4)	-13056	6.17	0.42	-0.17	-0.09		0.9775	1.80
t 值	-0.97	9.61	3.57	-3.09	-1.55			
Y=f(X1,X3,X4,X5)	-12690	5.22	0.40	-0.20		0.07	0.9798	1.55
t 值	-0.87	17.85	3.02	-3.47		0.37		

5、结论

回归方程以Y=f（X_1，X_2，X_3）为最优：

$$Y = -11978 + 5.26X_1 + 0.41X_2 - 0.19X_3$$

*七、分部回归与多重共线性

1、分部回归法(Partitioned Regression)

对于模型

$$\mathbf{Y} = \mathbf{X}\mathrm{B} + \mathrm{N}$$

将解释变量分为两部分，对应的参数也分为两部分：

$$\mathbf{Y} = \mathbf{X}_1\mathrm{B}_1 + \mathbf{X}_2\mathrm{B}_2 + \mathrm{N}$$

在满足解释变量与随机误差项不相关的情况下，可以写出关于参数估计量的方程组：

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{Y} \\ \mathbf{X}_2'\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 \\ \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathrm{B}}_1 \\ \hat{\mathrm{B}}_2 \end{bmatrix}$$

$$\hat{B}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{B}_2$$
$$= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'(\mathbf{Y} - \mathbf{X}_2\hat{B}_2)$$

如果存在 $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$

则有 $\hat{B}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$

这就是仅以X_1作为解释变量时的参数估计量

同样有 $\hat{B}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y}$

这就是仅以X_2作为解释变量时的参数估计量。

2、由分部回归法导出

- 如果一个多元线性模型的解释变量之间完全正交，可以将该多元模型分为多个一元模型、二元模型、...进行估计，参数估计结果不变；
- 实际模型由于存在或轻或重的共线性，如果将它们分为多个一元模型、二元模型、...进行估计，参数估计结果将发生变化；

- 当模型存在共线性，将某个共线性变量去掉，剩余变量的参数估计结果将发生变化，而且经济含义有发生变化；

- 严格地说，实际模型由于总存在一定程度的共线性，所以每个参数估计量并不真正反映对应变量与被解释变量之间的结构关系。

第五章 经典单方程计量经济学模型：专门问题

§5.1　虚拟变量

§5.2　滞后变量

§5.3　设定误差

§5.4　建模理论

§5.1 虚拟变量模型

一、虚拟变量的基本含义

二、虚拟变量的引入

三、虚拟变量的设置原则

一、虚拟变量的基本含义

- 许多经济变量是**可以定量度量**的，**如：** 商品需求量、价格、收入、产量等
- 但也有一些影响经济变量的因素**无法定量度量**，**如：** 职业、性别对收入的影响，战争、自然灾害对GDP的影响，季节对某些产品（如冷饮）销售的影响等等。
- 为了在模型中能够反映这些因素的影响，并提高模型的精度，需要将它们“量化”，

这种“量化”通常是通过引入“虚拟变量”来完成的。根据这些因素的属性类型，构造只取“0”或“1”的人工变量，通常称为**虚拟变量**（**dummy variables**），记为D。

- 例如，**反映文程度的虚拟变量可取为：**

$$D=\begin{cases}1, & \text{本科学历}\\[2ex] 0, & \text{非本科学历}\end{cases}$$

一般地，在虚拟变量的设置中：

- **基础类型、肯定类型取值为1；**
- **比较类型，否定类型取值为0。**

概念：

同时含有一般解释变量与虚拟变量的模型称为虚拟变量模型或者方差分析（analysis-of variance: ANOVA）模型。

一个以性别为虚拟变量考察企业职工薪金的模型：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \mu_i$$

其中：Y_i为企业职工的薪金，X_i为工龄，

D_i=1，若是男性，D_i=0，若是女性。

二、虚拟变量的引入

- 虚拟变量做为解释变量引入模型有两种基本方式：**加法方式**和**乘法方式**。

1、加法方式

上述企业职工薪金模型中性别虚拟变量的引入采取了加法方式。

在该模型中，如果仍假定$E(\mu_i)=0$，则

企业女职工的平均薪金为：

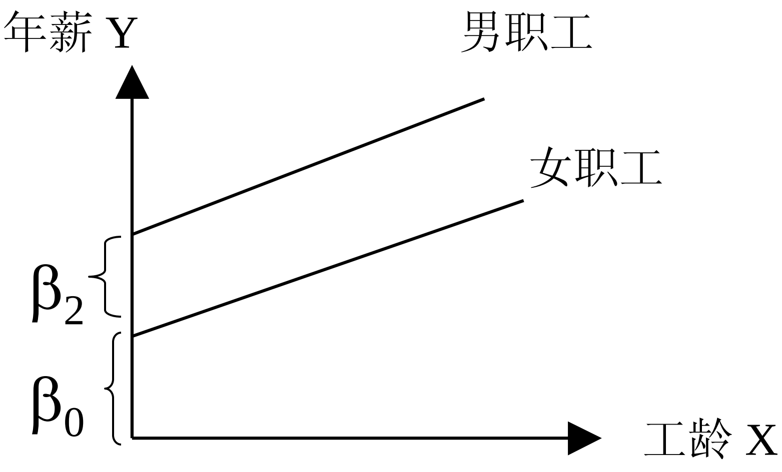
$$E(Y_i \mid X_i, D_i = 0) = \beta_0 + \beta_1 X_i$$

企业男职工的平均薪金为：

$$E(Y_i \mid X_i, D_i = 1) = (\beta_0 + \beta_2) + \beta_1 X_i$$

几何意义：

- 假定$\beta_2>0$，则两个函数有相同的斜率，但有不同的截距。意即，男女职工平均薪金对教龄的变化率是一样的，但两者的平均薪金水平相差β_2。
- 可以通过传统的回归检验，对β_2的统计显著性进行检验，以判断企业男女职工的平均薪金水平是否有显著差异。



又例：在横截面数据基础上，考虑个人保健支出对个人收入和教育水平的回归。

教育水平考虑三个层次：高中以下，

高中，

大学及其以上

这时需要引入两个虚拟变量：

$$D_1=\begin{cases}1 & \text{高中}\\ 0 & \text{其他}\end{cases}\qquad D_2=\begin{cases}1 & \text{大学及其以上}\\ 0 & \text{其他}\end{cases}$$

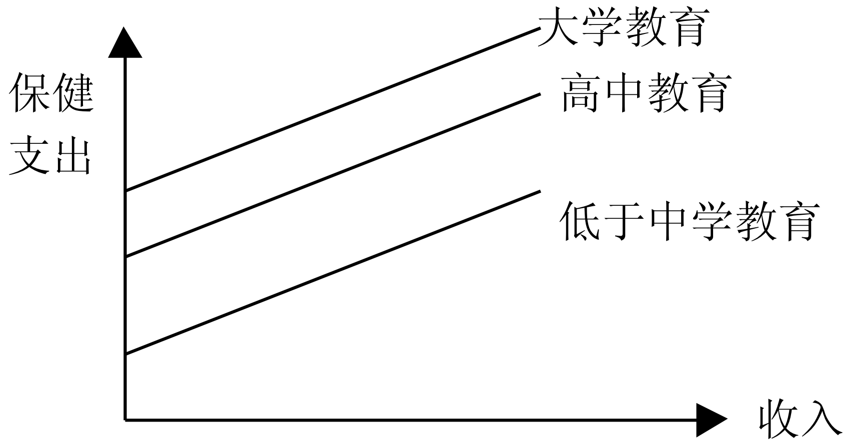
模型可设定如下：

$$Y_i=\beta_0+\beta_1X_i+\beta_2D_1+\beta_3D_2+\mu_i$$

在$E(\mu_i)=0$的初始假定下，高中以下、高中、大学及其以上教育水平下个人保健支出的函数：

- 高中以下：$E(Y_i \mid X_i, D_1=0, D_2=0)=\beta_0+\beta_1 X_i$
- 高中：$E(Y_i \mid X_i, D_1=1, D_2=0)=(\beta_0+\beta_2)+\beta_1 X_i$
- 大学及其以上：$E(Y_i \mid X_i, D_1=0, D_2=1)=(\beta_0+\beta_3)+\beta_1 X_i$

假定$\beta_3>\beta_2$，其几何意义：



- **还可将多个虚拟变量引入模型中以考察多种“定性”因素的影响。**

如在上述职工薪金的例中，再引入代表学历的虚拟变量D_2：

$$D_2 = \begin{cases} 1 & \text{本科及以上学历} \\ 0 & \text{本科以下学历} \end{cases}$$

职工薪金的回归模型可设计为：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \mu_i$$

于是，不同性别、不同学历职工的平均薪金分别为：

- 女职工本科以下学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 X_i$$

- 男职工本科以下学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 1, D_2 = 0) = (\beta_0 + \beta_2) + \beta_1 X_i$$

- 女职工本科以上学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 0, D_2 = 1) = (\beta_0 + \beta_3) + \beta_1 X_i$$

- 男职工本科以上学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 1, D_2 = 1) = (\beta_0 + \beta_2 + \beta_3) + \beta_1 X_i$$

2、乘法方式

- 加法方式引入虚拟变量，考察：**截距的不同**，
- 许多情况下：往往是斜率就有变化，**或斜率、截距同时发生变化**。
- **斜率的变化可通过以乘法的方式引入虚拟变量来测度。**

例： 根据消费理论，消费水平C主要取决于收入水平Y，但在一个较长的时期，人们的消费倾向会发生变化，尤其是在自然灾害、战争等反常年份，消费倾向往往出现变化。这种消费倾向的变化可通过在收入的系数中引入虚拟变量来考察。

如，设 $D_t = \begin{cases} 1 & 正常年份 \\ 0 & 反常年份 \end{cases}$ 消费模型可建立如下：

$$C_t = \beta_0 + \beta_1 X_t + \beta_2 D_t X_t + \mu_t$$

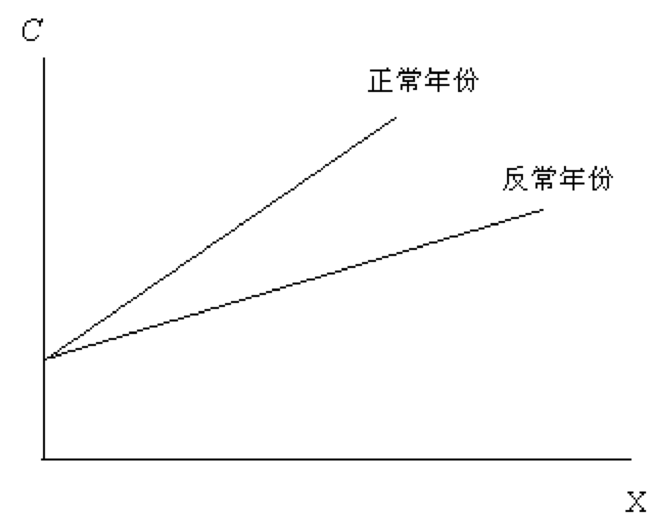
- 这里，虚拟变量D以与X相乘的方式引入了模型中，从而可用来考察消费倾向的变化。
- 假定$E(\mu_i)= 0$，上述模型所表示的函数可化为：

正常年份：

$$E(C_t \mid X_t, D_t = 1) = \beta_0 + (\beta_1 + \beta_2) X_t$$

反常年份：

$$E(C_t \mid X_t, D_t = 0) = \beta_0 + \beta_1 X_t$$



当截距与斜率发生变化时，则需要同时引入加法与乘法形式的虚拟变量。

- **例5.1.1，** 考察1990年前后的中国居民的总储蓄-收入关系是否已发生变化。

  表5.1.1中给出了中国1979~2001年以城乡储蓄存款余额代表的居民储蓄以及以GNP代表的居民收入的数据。

表 5.1.1　　1979~2001 年中国居民储蓄与收入数据（亿元）

90年前	储蓄	GNP	90年后	储蓄	GNP
1979	281	4038.2	1991	9107	21662.5
1980	399.5	4517.8	1992	11545.4	26651.9
1981	523.7	4860.3	1993	14762.4	34560.5
1982	675.4	5301.8	1994	21518.8	46670.0
1983	892.5	5957.4	1995	29662.3	57494.9
1984	1214.7	7206.7	1996	38520.8	66850.5
1985	1622.6	8989.1	1997	46279.8	73142.7
1986	2237.6	10201.4	1998	53407.5	76967.2
1987	3073.3	11954.5	1999	59621.8	80579.4
1988	3801.5	14922.3	2000	64332.4	88228.1
1989	5146.9	16917.8	2001	73762.4	94346.4
1990	7034.2	18598.4			

以Y为储蓄，X为收入，可令：

- 1990年前： $Y_i=\alpha_1+\alpha_2X_i+\mu_{1i}$ $i=1,2\ldots,n_1$
- 1990年后： $Y_i=\beta_1+\beta_2X_i+\mu_{2i}$ $i=1,2\ldots,n_2$

则有可能出现下述四种情况中的一种：

(1) $\alpha_1=\beta_1$，且$\alpha_2=\beta_2$，即两个回归相同，称为**重合回归**（Coincident Regressions）；

(2) $\alpha_1\neq\beta_1$,但$\alpha_2=\beta_2$，即两个回归的差异仅在其截距，称为**平行回归**（Parallel Regressions）；

(3) $\alpha_1=\beta_1$，但$\alpha_2\neq\beta_2$，即两个回归的差异仅在其斜率，称为**汇合回归**(Concurrent Regressions);

(4) $\alpha_1\neq\beta_1$，且$\alpha_2\neq\beta_2$，即两个回归完全不同，称为**相异回归**（Dissimilar Regressions）。

可以运用**邹氏结构变化的检验**。这一问题也可通过引入乘法形式的虚拟变量来解决。

将n_1与n_2次观察值合并，并用以估计以下回归：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_3 D_i + \beta_4 (D_i X_i) + \mu_i$$

D_i为引入的虚拟变量：
$$D_i = \begin{cases} 1 & 90\text{年前} \\ 0 & 90\text{年后} \end{cases}$$

于是有：

$$E(Y_i \mid D_i = 0, X_i) = \beta_0 + \beta_1 X_i$$

$$E(Y_i \mid D_i = 1, X_i) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) X_i$$

可分别表示1990年后期与前期的储蓄函数。

在统计检验中，如果$\beta_4=0$的假设被拒绝，则说明两个时期中储蓄函数的斜率不同。

- 具体的回归结果为：

$$\hat{Y}_i = -15452 + 0.8881X_i + 13802.3D_i - 0.4765D_iX_i$$

(-6.11)　　(22.89)　　(4.33)　　(-2.55)

$\bar{R}^2$ =0.9836

由β_3与β_4的t检验可知：参数显著地不等于0，强烈示出两个时期的回归是相异的，

储蓄函数分别为：

1990年前：　$\hat{Y}_i = -1649.7 + 0.4116X_i$

1990年后：　$\hat{Y}_i = -15452 + 0.8881X_i$

3、临界指标的虚拟变量的引入

在经济发生转折时期，可通过建立临界指标的虚拟变量模型来反映。

例如，进口消费品数量Y主要取决于国民收入X的多少，中国在改革开放前后，Y对X的回归关系明显不同。

这时，可以t*=1979年为转折期，以1979年的国民收入X_t*为临界值，设如下虚拟变量：

$$D_t = \begin{cases} 1 & t \geq t^* \\ 0 & t < t^* \end{cases}$$

则进口消费品的回归模型可建立如下：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 (X_t - X_t^*) D_t + \mu_t$$

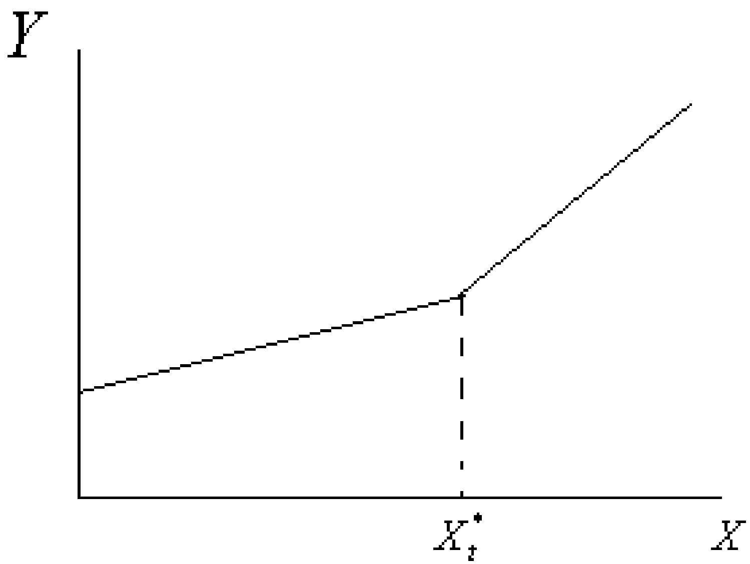
OLS法得到该模型的回归方程为

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t + \hat{\beta}_2 (X_t - X_t^*) D_t$$

则两时期进口消费品函数分别为：

当t<t*=1979年， $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_t$

当t≥t*=1979年， $\hat{Y}_t = (\hat{\beta}_0 - \hat{\beta}_2 X_i^*) + (\hat{\beta}_1 + \hat{\beta}_2) X_t$



三、虚拟变量的设置原则

虚拟变量的个数须按以下原则确定：

每一定性变量所需的虚拟变量个数要比该定性变量的类别数少1，即如果有m个定性变量，只在模型中引入m-1个虚拟变量。

例。已知冷饮的销售量Y除受k种定量变量X_k的影响外，还受春、夏、秋、冬四季变化的影响，要考察该四季的影响，只需引入三个虚拟变量即可：

$$D_{1t}=\begin{cases}1 & \text{春季}\\ 0 & \text{其他}\end{cases}\qquad D_{2t}=\begin{cases}1 & \text{夏季}\\ 0 & \text{其他}\end{cases}\qquad D_{3t}=\begin{cases}1 & \text{秋季}\\ 0 & \text{其他}\end{cases}$$

则冷饮销售量的模型为：

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots \beta_k X_{kt} + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \mu_t$$

- 在上述模型中，若再引入第四个虚拟变量

$$D_{4t} = \begin{cases} 1 & \text{冬季} \\ 0 & \text{其他} \end{cases}$$

则冷饮销售模型变量为：

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots \beta_k X_{kt} + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \mu_t$$

其矩阵形式为：

$$\mathbf{Y} = (\mathbf{X}, \mathbf{D}) \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + \mu$$

如果只取六个观测值，其中春季与夏季取了两次，秋、冬各取到一次观测值，则式中的：

$$(\mathbf{X},\mathbf{D})=\begin{pmatrix}1 & X_{11} & \cdots & X_{k1} & 1 & 0 & 0 & 0\\ 1 & X_{12} & \cdots & X_{k2} & 0 & 1 & 0 & 0\\ 1 & X_{13} & \cdots & X_{k3} & 0 & 0 & 1 & 0\\ 1 & X_{14} & \cdots & X_{k4} & 0 & 0 & 0 & 1\\ 1 & X_{15} & \cdots & X_{k5} & 0 & 1 & 0 & 0\\ 1 & X_{16} & \cdots & X_{k6} & 1 & 0 & 0 & 0\end{pmatrix}\qquad \boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_k\end{pmatrix}\qquad \boldsymbol{\alpha}=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\\ \alpha_4\end{pmatrix}$$

显然，**(X,D)**中的第1列可表示成后4列的线性组合，从而**(X,D)**不满秩，参数无法唯一求出。

这就是所谓的“虚拟变量陷阱”，应避免。

§5.2 滞后变量模型

一、滞后变量模型

二、分布滞后模型的参数估计

三、自回归模型的参数估计

四、格兰杰因果关系检验

一、滞后变量模型

在经济运行过程中，广泛存在时间滞后效应。某些经济变量不仅受到同期各种因素的影响，而且也受到过去某些时期的各种因素甚至自身的过去值的影响。

通常把这种过去时期的，具有滞后作用的变量叫做滞后变量（Lagged Variable），含有滞后变量的模型称为**滞后变量模型**。

滞后变量模型考虑了时间因素的作用，使静态分析的问题有可能成为动态分析。**含有滞后解释变量的模型，又称动态模型（Dynamical Model）**。

1、滞后效应与与产生滞后效应的原因

因变量受到自身或另一解释变量的前几期值影响的现象称为**滞后效应**。

表示前几期值的变量称为**滞后变量**。

如：**消费函数**

通常认为，本期的消费除了受本期的收入影响之外，还受前1期，或前2期收入的影响：

$$C_t=\beta_0+\beta_1 Y_t+\beta_2 Y_{t-1}+\beta_3 Y_{t-2}+\mu_t$$

Y_{t-1}，Y_{t-2}为**滞后变量**。

- **产生滞后效应的原因**

1、心理因素：人们的心理定势，行为方式滞后于经济形势的变化，如中彩票的人不可能很快改变其生活方式。

2、技术原因：如当年的产出在某种程度上依赖于过去若干期内投资形成的固定资产。

3、制度原因：如定期存款到期才能提取，造成了它对社会购买力的影响具有滞后性。

2、滞后变量模型

以滞后变量作为解释变量，就得到**滞后变量模型**。它的一般形式为：

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_q Y_{t-q} + \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + \alpha_s X_{t-s} + \mu_t$$

q，s：滞后时间间隔

自回归分布滞后模型（autoregressive distributed lag model, ADL）：既含有Y对自身滞后变量的回归，还包括着X分布在不同时期的滞后变量

有限自回归分布滞后模型：滞后期长度有限

无限自回归分布滞后模型：滞后期无限，

（1）分布滞后模型（distributed-lag model）

分布滞后模型：模型中没有滞后被解释变量，仅有解释变量X的当期值及其若干期的滞后值：

$$Y_t = \alpha + \sum_{i=0}^{s} \beta_i X_{t-i} + \mu_t$$

β_0：**短期**(short-run)或**即期乘数**(impact multiplier)，表示本期X变化一单位对Y平均值的影响程度。

β_i（i=1, 2..., s）：**动态乘数**或**延迟系数**，表示各滞后期X的变动对Y平均值影响的大小。

$\sum_{i=0}^{s}\beta_i$ 称为**长期（long-run）**或**均衡乘数（total distributed-lag multiplier）**，表示X变动一个单位，由于滞后效应而形成的对Y平均值总影响的大小。

如果各期的X值保持不变，则X与Y间的长期或均衡关系即为

$$E(Y)=\alpha+(\sum_{i=0}^{s}\beta_i)X$$

2、自回归模型（autoregressive model）

自回归模型：模型中的解释变量仅包含X的当期值与被解释变量Y的一个或多个滞后值

$$Y_t = \alpha_0 + \alpha_1 X_t + \sum_{i=1}^{q} \beta_i Y_{t-i} + \mu_t$$

而

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \mu_t$$

称为**一阶自回归模型（first-order autoregressive model）**。

二、分布滞后模型的参数估计

1、分布滞后模型估计的困难

无限期的分布滞后模型，由于样本观测值的有限性，使得无法直接对其进行估计。

有限期的分布滞后模型，OLS会遇到如下问题：

1、没有先验准则确定滞后期长度；

2、如果滞后期较长，将缺乏足够的自由度进行估计和检验；

3、同名变量滞后值之间可能存在高度线性相关，即模型存在高度的多重共线性。

2、分布滞后模型的修正估计方法

人们提出了一系列的修正估计方法，但并不很完善。

各种方法的基本思想大致相同：都是**通过对各滞后变量加权，组成线性合成变量而有目的地减少滞后变量的数目，以缓解多重共线性，保证自由度。**

(1)经验加权法

根据实际问题的特点、实际经验给各滞后变量指定权数，滞后变量按权数线性组合，构成新的变量。权数据的类型有：

•**递减型：**

即认为**权数是递减的**，X的近期值对Y的影响较远期值大。

如消费函数中，收入的近期值对消费的影响作用显然大于远期值的影响。

例如：**滞后期为 3**的一组权数可取值如下：

1/2， 1/4， 1/6， 1/8

则新的线性组合变量为：

$$W_{1t}=\frac{1}{2}X_{t}+\frac{1}{4}X_{t-1}+\frac{1}{6}X_{t-2}+\frac{1}{8}X_{t-3}$$

- 矩型：

即认为权数是相等的，X的逐期滞后值对值Y的影响相同。

如滞后期为3，指定相等权数为1/4，则新的线性组合变量为：

$$W_{2t}=\frac{1}{4}X_{t}+\frac{1}{4}X_{t-1}+\frac{1}{4}X_{t-2}+\frac{1}{4}X_{t-3}$$

- **倒V型**

权数先递增后递减呈倒“V”型。

例如：在一个较长建设周期的投资中，历年投资X为产出Y的影响，往往在周期期中投资对本期产出贡献最大。

如滞后期为4，权数可取为

1/6， 1/4， 1/2， 1/3， 1/5

则新变量为

$$W_{3t}=\frac{1}{6}X_t+\frac{1}{4}X_{t-1}+\frac{1}{2}X_{t-2}+\frac{1}{3}X_{t-3}+\frac{1}{5}X_{t-4}$$

例5.2.1 对一个分布滞后模型：

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \mu_t$$

给定递减权数：1/2， 1/4， 1/6， 1/8

令 $$W_{1t} = \frac{1}{2} X_t + \frac{1}{4} X_{t-1} + \frac{1}{6} X_{t-2} + \frac{1}{8} X_{t-3}$$

原模型变为：$Y_t = \alpha_0 + \alpha_1 W_{1t} + \mu_t$

该模型可用OLS法估计。假如参数估计结果为

$\hat{\alpha}_0 = 0.5$ $\hat{\alpha}_1 = 0.8$

则原模型的估计结果为：

$$\hat{Y}_t = 0.5 + \frac{0.8}{2} X_t + \frac{0.8}{4} X_{t-1} + \frac{0.8}{6} X_{t-2} + \frac{0.8}{8} X_{t-3} = 0.5 + 0.4 X_t + 0.2 X_{t-1} + 0.133 X_{t-2} + 0.1 X_{t-3}$$

经验权数法的优点是：简单易行

缺点是：设置权数的随意性较大

通常的做法是：

多选几组权数，分别估计出几个模型，然后根据常用的统计检验（Ｒ方检验，Ｆ检验，ｔ检验，Ｄ－Ｗ检验），从中选择最佳估计式。

（2）阿尔蒙（Almon）多项式法

主要思想：针对有限滞后期模型，通过阿尔蒙变换，定义新变量，以减少解释变量个数，然后用OLS法估计参数。

主要步骤为：

第一步，阿尔蒙变换

对于分布滞后模型

$$Y_t = \alpha + \sum_{i=0}^{s} \beta_i X_{t-i} + \mu_t$$

假定其回归系数β_i可用一个关于滞后期i的适当阶数的多项式来表示，即：

$$\beta_i = \sum_{k=1}^{m} \alpha_k (i+1)^k \qquad i=0,1,\ldots,s$$

其中，m<s-1。阿尔蒙变换要求先验地确定适当阶数k，例如取k=2，得

$$\beta_i = \sum_{k=1}^{2} \alpha_k (i+1)^k = \alpha_1 (i+1) + \alpha_2 (i+1)^2 \qquad (*)$$

将(*)代入分布滞后模型 $Y_t = \alpha + \sum_{i=0}^{s} \beta_i X_{t-i} + \mu_t$ 得

$$Y_t = \alpha + \sum_{i=0}^{s} \left(\sum_{k=1}^{2} \alpha_k (i+1)^k\right) X_{t-i} + \mu_t$$

$$= \alpha + \alpha_1 \sum_{i=0}^{s} (i+1) X_{t-i} + \alpha_2 \sum_{i=0}^{s} (i+1)^2 X_{t-2} + \mu_t$$

定义新变量

$$W_{1t}=\sum_{i=0}^{s}(i+1)X_{t-i} \qquad W_{2t}=\sum_{i=0}^{s}(i+1)^2X_{t-i}$$

将原模型转换为：

$$Y_t=\alpha+\alpha_1W_{1t}+\alpha_2W_{2t}+\mu_t$$

第二步，模型的OLS估计

对变换后的模型进行OLS估计，得

$$\hat{\alpha},\hat{\alpha}_1,\hat{\alpha}_2$$

再计算出：

$$\beta_i=\sum_{k=1}^{2}\alpha_k(i+1)^k=\alpha_1(i+1)+\alpha_2(i+1)^2$$

求出滞后分布模型参数的估计值：

$$\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_s$$

由于$m+1<s$，可以认为原模型存在的自由度不足和多重共线性问题已得到改善。

需注意的是，在实际估计中，阿尔蒙多项式的阶数m一般取2或3，不超过4，否则达不到减少变量个数的目的。

例5.2.2 表5.2.1给出了中国**电力基本建设投资X**与**发电量Y**的相关资料，拟建立一多项式分布滞后模型来考察两者的关系。

表5.2.1 中国电力工业基本建设投资与发电量

年度	基本建设投资X（亿元）	发电量（亿千瓦时）	年度	基本建设投资X（亿元）	发电量（亿千瓦时）
1975	30.65	1958	1986	161.6	4495
1976	39.98	2031	1987	210.88	4973
1977	34.72	2234	1988	249.73	5452
1978	50.91	2566	1989	267.85	5848
1979	50.99	2820	1990	334.55	6212
1980	48.14	3006	1991	377.75	6775
1981	40.14	3093	1992	489.69	7539
1982	46.23	3277	1993	675.13	8395
1983	57.46	3514	1994	1033.42	9218
1984	76.99	3770	1995	1124.15	10070
1985	107.86	4107			

由于无法预见知电力行业基本建设投资对发电量影响的时滞期，需取不同的滞后期试算。

经过试算发现，在2阶阿尔蒙多项式变换下，滞后期数取到第6期，估计结果的经济意义比较合理。2阶阿尔蒙多项式估计结果如下：

$$\hat{Y}_t = 3319.5 + 3.061W_{0t} + 0.101W_{1t} - 0.271W_{2t}$$

（13.62）（1.86）（0.15）（−0.67）

$\overline{R^2}=0.9405$　　F=74.81　　DW=0.42

求得的分布滞后模型参数估计值为

$\hat{\beta}_0=0.323$，$\hat{\beta}_1=1.777$，$\hat{\beta}_2=2.690$，$\hat{\beta}_3=3.061$，$\hat{\beta}_4=2.891$，$\hat{\beta}_5=2.180$，$\hat{\beta}_6=0.927$

最后得到分布滞后模型估计式为:

$$\hat{Y}_t = 3319.5 + 0.323X_t + 1.777X_{t-1} + 2.690X_{t-2} + 3.061X_{t-3}$$

（13.62） (0.19) (2.14) (1.88) (1.86)

$$+ 2.891X_{t-4} + 2.180X_{t-5} + 0.927X_{t-6}$$

(1.96) (1.10) (0.24)

为了比较，下面给出直接对滞后6期的模型进行OLS估计的结果:

$$\hat{Y}_t = 3361.9 + 8.424X_t - 11.43X_{t-1} + 15.14X_{t-2} + 4.71X_{t-3}$$

（12.43） (1.80) (-1.89) (1.21) (0.36)

$$-14.70X_{t-4} + 26.94X_{t-5} - 25.42X_{t-6}$$

(-0.93) (1.09) (-1.12)

$\overline{R}^2$ =0.9770 F=42.54 DW=1.03

（3）科伊克（Koyck）方法

科伊克方法是将无限分布滞后模型转换为自回归模型，然后进行估计。

对于无限分布滞后模型：

$$Y_t = \alpha + \sum_{i=0}^{\infty} \beta_i X_{t-i} + \mu_t$$

科伊克变换假设β_i随滞后期i按几何级数衰减：

$$\beta_i = \beta_0 \lambda^i$$

其中，$0<\lambda<1$，称为分布滞后衰减率，$1-\lambda$称为**调整速率**（Speed of adjustment）。

科伊克变换的具体做法：

将科伊克假定$\beta_i=\beta_0\lambda^i$代入无限分布滞后模型，得

$$Y_t = \alpha + \beta_0 \sum_{i=0}^{\infty} \lambda^i X_{t-i} + \mu_t \qquad (*)$$

滞后一期并乘以λ ，得

$$\lambda Y_{t-1} = \lambda\alpha + \beta_0 \sum_{i=1}^{\infty} \lambda^i X_{t-i} + \lambda\mu_{t-1} \qquad (**)$$

将（*）减去（**）得科伊克变换模型：

$$Y_t - \lambda Y_{t-1} = (1-\lambda)\alpha + \beta_0 X_t + \mu_t - \lambda\mu_{t-1}$$

整理得科伊克模型的一般形式：

$$Y_t = a + bX_t + cY_{t-1} + v_t$$

其中：$a=(1-\lambda)\alpha$ ，$b=\beta_0$ ，$c=\lambda$ ，$v_t=\mu_t-\lambda\mu_{t-1}$

科伊克模型的特点：

（1）以一个滞后因变量Y_{t-1}代替了大量的滞后解释变量X_{t-i}，最大限度地节省了自由度，解决了滞后期长度s难以确定的问题；

（2）由于滞后一期的因变量Y_{t-1}与X_t的线性相关程度可以肯定小于X的各期滞后值之间的相关程度，从而缓解了多重共线性。

但科伊克变换也同时产生了两个新问题：

（1）模型存在随机项和v_t的一阶自相关性；

（2）滞后被解释变量Y_{t-1}与随机项v_t不独立。

这些新问题需要进一步解决。

三、自回归模型的参数估计

1、自回归模型的构造

- 一个无限期分布滞后模型可以通过科伊克变换转化为自回归模型。
- 事实上，**许多滞后变量模型都可以转化为自回归模型**，自回归模型是经济生活中更常见的模型。
- 以**适应预期模型**以及**局部调整模型**为例进行说明。

（1）自适应预期（Adaptive expectation）模型

在某些实际问题中，因变量Y_t并不取决于解释变量的当前实际值X_t，而取决于X_t的“预期水平”或“长期均衡水平”X_t^e。

例如，家庭本期消费水平，取决于本期收入的预期值；

市场上某种商品供求量，决定于本期该商品价格的均衡值。

因此，自适应预期模型最初表现形式是

$$Y_t = \beta_0 + \beta_1 X_t^e + \mu_t$$

由于预期变量是不可实际观测的，往往作如下**自适应预期假定：**

$$X_t^e - X_{t-1}^e = r(X_t - X_{t-1}^e)$$

其中：**r**为**预期系数**（coefficient of expectation），0≤**r** ≤1。

该式的经济含义为："经济行为者将根据过去的经验修改他们的预期"，即本期预期值的形成是一个逐步调整过程，**本期预期值的增量是本期实际值与前一期预期值之差的一部分**，其比例为**r** 。

这个假定还可写成：

$$X_t^e = rX_t + (1-r)X_{t-1}^e$$

将 $X_t^e = rX_t + (1-r)X_{t-1}^e$ 代入

$$Y_t = \beta_0 + \beta_1 X_t^e + \mu_t \qquad (*)$$

得

$$Y_t = \beta_0 + \beta_1[rX_t + (1-r)X_{t-1}^e] + \mu_t$$

将（*）式滞后一期并乘以(1-r), 得

$$(1-r)Y_{t-1} = \beta_0(1-r) + \beta_1(1-r)X_{t-1}^e + (1-r)\mu_{t-1} \qquad (**)$$

以(*)减去（**），整理得

$$Y_t = \beta_0 r + \beta_1 rX_t + (1-r)Y_{t-1} + v_t$$

其中 $v_t = \mu_t - (1-r)\mu_{t-1}$

可见自适应预期模型转化为自回归模型。

（2）局部调整(Partial Adjustment)模型

- 局部调整模型主要是用来研究物资储备问题的。
- **例如**，企业为了保证生产和销售，必须保持一定的原材料储备。对应于一定的产量或销售量X_t，存在着预期的最佳库存Y_t^e。
- 局部调整模型的最初形式为

$$Y_t^e = \beta_0 + \beta_1 X_t + \mu_t \qquad (9.3.7)$$

Y_t^e不可观测。由于生产条件的波动，生产管理方面的原因，库存储备Y_t的实际变化量只是预期变化的一部分。

储备按预定水平逐步进行调整，故有如下局部调整假设：

$$Y_t - Y_{t-1} = \delta(Y_t^e - Y_{t-1})$$

或：

$$Y_t = \delta Y_t^e + (1-\delta)Y_{t-1} \qquad (*)$$

其中，δ为调整系数，0≤ δ ≤1

将(*)式代入 $Y_t^e = \beta_0 + \beta_1 X_t + \mu_t$ 得

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\mu_t$$

可见，局部调整模型转化为自回归模型

2、自回归模型的参数估计

对于自回归模型

$$Y_t = \alpha_0 + \alpha_1 X_t + \sum_{i=1}^{q} \beta_i Y_{t-i} + \mu_t$$

估计时的主要问题：滞后被解释变量的存在可能导致它与随机扰动项相关，以及随机扰动项出现序列相关性。

考伊克模型：

$$Y_t = (1-\lambda)\alpha + \beta_0 X_t + \lambda Y_{t-1} + v_t$$

$$v_t = \mu_t - \lambda \mu_{t-1}$$

自适应预期模型：

$$Y_t = \beta_0 r + \beta_1 r X_t + (1-r) Y_{t-1} + v_t$$

$$v_t = \mu_t - (1-r)\mu_{t-1}$$

显然存在： $\mathrm{cov}(Y_{t-1}, v_t) \neq 0 \qquad \mathrm{cov}(v_t, v_{t-1}) \neq 0$

局部调整模型：

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\mu_t$$

存在：滞后被解释变量Y_{t-1}与随机扰动项$\delta\mu_t$的异期相关性。

因此，对自回归模型的估计主要需视滞后被解释变量与随机扰动项的不同关系进行估计。

以一阶自回归模型为例说明:

（1）工具变量法

对于一阶自回归模型

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \mu_t$$

若Y_{t-1}与μ_t同期相关，则OLS估计是有偏的，并且不是一致估计。

因此，对上述模型，通常采用工具变量法，即寻找一个新的经济变量Z_t，用来代替Y_{t-1}。

参数估计量具有一致性。

在实际估计中，一般用**X**的若干滞后的线性组合作为$\mathbf{Y}_{t-1}$的工具变量：

$$\hat{Y}_{t-1} = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_s X_{t-s}$$

由于原模型已假设随机扰动项μ_t与解释变量X及其滞后项不存在相关性，因此上述工具变量与μ_t不再线性相关。

一个更简单的情形是直接用X_{t-1}作为Y_{t-1}的工具变量。

（2）普通最小二乘法

若滞后被解释变量Y_{t-1}与随机扰动项μ_t同期无关（如局部调整模型），可直接使用OLS法进行估计，得到一致估计量。

注意：

上述工具变量法只解决了解释变量与μ_t相关对参数估计所造成的影响，但没有解决μ_t的自相关问题。

事实上，对于自回归模型，μ_t项的自相关问题始终存在，对于此问题，至今没有完全有效的解决方法。唯一可做的，就是尽可能地建立“正确”的模型，以使序列相关性的程度减轻。

例5.2.3 建立中国长期货币流通量需求模型

经验表明：中国改革开放以来，对**货币需求量(Y)**的影响因素，主要有资金运用中的**贷款额(X)**以及反映价格变化的**居民消费者价格指数(P)**。

长期货币流通量模型可设定为

$$Y_t^e = \beta_0 + \beta_1 X_t + \beta_2 P_t + \mu_t \qquad (*)$$

由于长期货币流通需求量不可观测，作局部调整：

$$Y_t - Y_{t-1} = \delta(Y_t^e - Y_{t-1}) \qquad (**)$$

将（*）式代入（**）得**短期货币流通量需求模型**：

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + \delta\beta_2 P_t + (1-\delta)Y_{t-1} + \delta\mu_t$$

表5.2.2　中国货币流通量、贷款额、居民消费价格指数历史数据

单位：亿元，上年=100

年度	货币流通量 Y	民民消费价格指数 P	贷款额 X	年度	货币流通量 Y	民民消费价格指数 P	贷款额 X
1978	212.0	100.7	1850	1990	2644.4	101.3	17680.7
1979	267.7	101.9	2039.6	1991	3177.8	105.1	21337.8
1980	346.2	107.5	2414.3	1992	4336.0	108.6	26322.9
1981	396.3	102.5	2860.2	1993	5864.7	116.1	32943.1
1982	439.1	102	3180.6	1994	7288.6	125	39976
1983	529.8	102	3589.9	1995	7885.3	116.8	50544.1
1984	792.1	102.7	4766.1	1996	8802.0	108.8	61156.6
1985	987.8	111.9	5905.6	1997	10177.6	103.1	74914.1
1986	1218.4	107	7590.8	1998	11204.2	99.4	86524.1
1987	1454.5	108.8	9032.5	1999	13455.5	98.7	93734.3
1988	2134.0	120.7	10551.3	2000	14652.7	100.8	99371.1
1989	2344.0	116.3	14360.1				

对局部调整模型

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + \delta\beta_2 P_t + (1-\delta)Y_{t-1} + \delta\mu_t$$

运用OLS法估计结果如下

$$Y_t = -3700.4 + 0.0714 X_t + 36.10 P_t + 0.5638 Y_{t-1}$$

（−2.93）（2.86）　　（3.10）　　（2.87）

R^2=0.9959　　$\overline{R^2}$=0.9953　　F=1467.96，　　D.W.=1.733

由$(1-\hat{\delta})=0.5638$，得$\hat{\delta}=0.4362$

最后得到长期货币流通需求模型的估计式：

$$Y_t^e = -8483.3 + 0.1637 X_t + 82.75 P_t$$

注意：

尽管D. W. =1. 733，但不能据此判断自回归模型不存在自相关(Why?)。

但　　LM=0. 7855，

α=5%下，临界值$\chi^2(1)=3.84$，

判断：模型已不存在一阶自相关。

如果直接对下式作OLS回归

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 P_t + \mu_t$$

得

$$Y_t = -5611.66 + 0.1427X_t + 54.19P_t$$

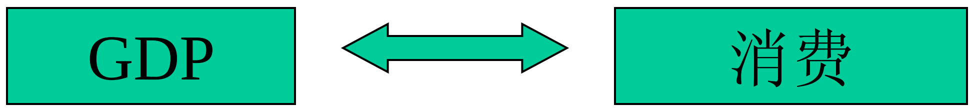
（−4. 81）　（58. 79）　（5. 05）

$R^2=0.9943$　$\bar{R}^2=0.9937$　F=1735. 36，　D. W. =1. 204

可见该模型随机扰动项具有序列相关性，

四、格兰杰因果关系检验

- 自回归分布滞后模型旨在揭示：某变量的变化受其自身及其他变量过去行为的影响。
- 然而，许多经济变量有着相互的影响关系



问题： 当两个变量在时间上有先导——滞后关系时，能否从统计上考察这种关系是单向的还是双向的？

即：主要是一个变量过去的行为在影响另一个变量的当前行为呢？还是双方的过去行为在相互影响着对方的当前行为？

格兰杰因果关系检验（Granger test of causality）

对两变量Y与X，格兰杰因果关系检验要求估计：

$$Y_t = \sum_{i=1}^{m} \alpha_i X_{t-i} + \sum_{i=1}^{m} \beta_i Y_{t-i} + \mu_{1t} \qquad (*)$$

$$X_t = \sum_{i=1}^{m} \lambda_i Y_{t-i} + \sum_{i=1}^{m} \delta_i X_{t-i} + \mu_{2t} \qquad (**)$$

可能存在有四种检验结果：

（1）**X对Y有单向影响**，表现为（*）式X各滞后项前的参数整体为零，而Y各滞后项前的参数整体不为零；

（2）**Y对X有单向影响**，表现为（**）式Y各滞后项前的参数整体为零，而X各滞后项前的参数整体不为零；

（3）**Y与X间存在双向影响**，表现为Y与X各滞后项前的参数整体不为零；

（4）**Y与X间不存在影响**，表现为Y与X各滞后项前的参数整体为零。

格兰杰检验是通过受约束的F检验完成的。如：

针对 $$Y_t = \sum_{i=1}^{m} \alpha_i X_{t-i} + \sum_{i=1}^{m} \beta_i Y_{t-i} + \mu_{1t}$$

中X滞后项前的参数整体为零的假设（X不是Y的格兰杰原因）

分别做包含与不包含X滞后项的回归，记前者与后者的残差平方和分别为RSS_U、RSS_R；再计算F统计量：

$$F = \frac{(RSS_R - RSS_U)/m}{RSS_U/(n-k)}$$

k为无约束回归模型的待估参数的个数。

如果：$F>F_{\alpha}(m,n-k)$ ，则拒绝原假设，认为**X是Y的格兰杰原因**。

注意：

格兰杰因果关系检验对于滞后期长度的选择有时很敏感。不同的滞后期可能会得到完全不同的检验结果。

因此，***一般而言***，常进行不同滞后期长度的检验，以检验模型中随机误差项不存在序列相关的滞后期长度来选取滞后期。

例5.2.4 检验1978~2000年间中国当年价GDP与居民消费CONS的因果关系。

表 5.2.3 中国 GDP 与消费支出（亿元）

年份	人均居民消费 CONSP	人均GDP GDPP	年份	人均居民消费 CONSP	人均GDP GDPP
1978	1759.1	3605.6	1990	9113.2	18319.5
1979	2005.4	4074.0	1991	10315.9	21280.4
1980	2317.1	4551.3	1992	12459.8	25863.7
1981	2604.1	4901.4	1993	15682.4	34500.7
1982	2867.9	5489.2	1994	20809.8	46690.7
1983	3182.5	6076.3	1995	26944.5	58510.5
1984	3674.5	7164.4	1996	32152.3	68330.4
1985	4589	8792.1	1997	34854.6	74894.2
1986	5175	10132.8	1998	36921.1	79003.3
1987	5961.2	11784.7	1999	39334.4	82673.1
1988	7633.1	14704.0	2000	42911.9	89112.5
1989	8523.5	16466.0			

取两阶滞后，Eviews给出的估计结果为：

Pairwise Granger Causality Tests

Sample: 1978 2000

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
GDP does not Granger Cause CONS	21	4.29749	0.03208
CONS does not Granger Cause GDP		1.82325	0.19350

判断：α=5%，临界值$F_{0.05}(2,17)=3.59$

拒绝“GDP不是CONS的格兰杰原因”的假设，不拒绝“CONS不是GDP的格兰杰原因”的假设。

因此，从2阶滞后的情况看，GDP的增长是居民消费增长的原因，而不是相反。

但在2阶滞后时，检验的模型存在1阶自相关性。

表 5.2.4　　格兰杰因果关系检验

滞后长度	格兰杰因果性	F 值	P 值	LM 值	AIC 值	结论
2	$GDP \xrightarrow{\times} CONS$	4.297	0.032	0.009	16.08	拒绝
	$CONS \xrightarrow{\times} GDP$	1.823	0.194	0.008	17.86	不拒绝
3	$GDP \xrightarrow{\times} CONS$	10.219	0.001	0.010	15.14	拒绝
	$CONS \xrightarrow{\times} GDP$	4.096	0.691	0.191	17.14	不拒绝
4	$GDP \xrightarrow{\times} CONS$	19.643	10E-04	0.110	14.70	拒绝
	$CONS \xrightarrow{\times} GDP$	5.247	0.015	0.027	16.42	拒绝
5	$GDP \xrightarrow{\times} CONS$	10.321	0.004	0.464	14.72	拒绝
	$CONS \xrightarrow{\times} GDP$	5.085	0.028	0.874	16.30	拒绝
6	$GDP \xrightarrow{\times} CONS$	4.705	0.078	0.022	14.99	不拒绝
	$CONS \xrightarrow{\times} GDP$	7.773	0.034	1.000	16.05	拒绝

分析:

随着滞后阶数的增加，拒绝“GDP是居民消费CONS的原因”的概率变大，而拒绝“居民消费CONS是GDP的原因”的概率变小。

如果同时考虑检验模型的序列相关性以及赤池信息准则，发现: **滞后4阶或5阶的检验模型不具有1阶自相关性，而且也拥有较小的AIC值**，这时判断结果是:**GDP与CONS有双向的格兰杰因果关系，即相互影响**。

§5.3 模型设定偏误问题

一、模型设定偏误的类型

二、模型设定偏误的后果

三、模型设定偏误的检验

一、模型设定偏误的类型

- 模型设定偏误主要有两大类:

(1) **关于解释变量选取的偏误**，主要包括**漏选相关变量**和**多选无关变量**，

(2) **关于模型函数形式选取的偏误**。

1、相关变量的遗漏（omitting relevant variables）

- 例如，如果“正确”的模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

而我们将模型设定为

$$Y = \alpha_0 + \alpha_1 X_1 + v$$

即设定模型时漏掉了一个相关的解释变量。

这类错误称为**遗漏相关变量**。

- **动态设定偏误**（dynamic mis-specification）:遗漏相关变量表现为对Y或X滞后项的遗漏 。

2、无关变量的误选
(including irrevelant variables)

- **例如**，如果

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu$$

仍为“真”，但我们将模型设定为

$$Y=\alpha_0+\alpha_1X_1+\alpha_2X_2+\alpha_3X_3+\mu$$

即设定模型时，多选了一个无关解释变量。

3、错误的函数形式 (wrong functional form)

- 例如，如果“真实”的回归函数为

$$Y = AX_1^{\beta_1} X_2^{\beta_2} e^{\mu}$$

但却将模型设定为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v$$

二、模型设定偏误的后果

- 当模型设定出现偏误时，模型估计结果也会与“实际”有偏差。这种**偏差的性质与程度与模型设定偏误的类型密切相关**。

1、遗漏相关变量偏误

采用遗漏相关变量的模型进行估计而带来的偏误称为**遗漏相关变量偏误**（omitting relevant variable bias）。

设正确的模型为

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\mu$$

却对

$$Y=\alpha_0+\alpha_1 X_1+v$$

进行回归，得

$$\hat{\alpha}_1=\frac{\sum x_{1i} y_i}{\sum x_{1i}^2}$$

将正确模型 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu$ 的离差形式

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \mu_i - \bar{\mu}$$

代入 $\hat{\alpha}_1 = \dfrac{\sum x_{1i} y_i}{\sum x_{1i}^2}$ 得

$$\hat{\alpha}_1 = \frac{\sum x_{1i} y_i}{\sum x_{1i}^2} = \frac{\sum x_{1i}(\beta_1 x_{1i} + \beta_2 x_{2i} + \mu_i - \bar{\mu})}{\sum x_{1i}^2}$$

$$= \beta_1 + \beta_2 \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2} + \frac{\sum x_{1i}(\mu_i - \bar{\mu})}{\sum x_{1i}^2}$$

(1)如果漏掉的X_2与X_1相关，则上式中的第二项在小样本下求期望与大样本下求概率极限都不会为零，从而使得**OLS估计量在小样本下有偏，在大样本下非一致**。

(2)如果X_2与X_1不相关，则α_1的估计满足无偏性与一致性；但这时α_0的估计却是有偏的。

(3) 随机扰动项μ的方差估计$\hat{\sigma}^2$也是有偏的。

(4)$\hat{\alpha}_1$的方差是真实估计量$\hat{\beta}_1$的方差的有偏估计。

由 $Y=\alpha_0+\alpha_1X_1+v$ 得

$$Var(\hat{\alpha}_1)=\frac{\sigma^2}{\sum x_{1i}^2}$$

由 $Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu$ 得

$$Var(\hat{\beta}_1)=\sigma^2\frac{\sum x_{2i}^2}{\sum x_{1i}^2\sum x_{2i}^2-(\sum x_{1i}x_{2i})^2}=\frac{\sigma^2}{\sum x_{1i}^2(1-r_{x_1x_2}^2)}$$

如果X_2与X_1相关，显然有 $Var(\hat{\alpha}_1)\neq Var(\hat{\beta}_1)$

如果X_2与X_1不相关，也有 $Var(\hat{\alpha}_1)\neq Var(\hat{\beta}_1)$ Why?

2、包含无关变量偏误

采用包含无关解释变量的模型进行估计带来的偏误，称为**包含无关变量偏误**（including irrelevant variable bias）。

设 $$Y=\alpha_0+\alpha_1 X_1+v \qquad (*)$$

为正确模型，但却估计了

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\mu \qquad (**)$$

如果$\beta_2=0$，则(**)与(*)相同，因此，可将(**)式视为以$\beta_2=0$为约束的(*)式的特殊形式。

由于所有的经典假设都满足，因此对

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu \qquad (**)$$

式进行OLS估计，可得到无偏且一致的估计量。

注意： 由于$\beta_2=0$，因此，$E(\hat{\beta}_2)=0$。

但是，OLS估计量却不具有最小方差性。

$Y=\alpha_0+\alpha_1X_1+v$ 中X_1的方差: $$Var(\hat{\alpha}_1)=\frac{\sigma^2}{\sum x_{1i}^2}$$

$Y=\beta_0+\beta_1X_1+\beta_2X_2+\mu$ 中X_1的方差: $$Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_{1i}^2(1-r_{x_1x_2}^2)}$$

当X_1与X_2完全线性无关时: $Var(\hat{\alpha}_1)=Var(\hat{\beta}_1)$

否则: $Var(\hat{\beta}_1)>Var(\hat{\alpha}_1)$

3、错误函数形式的偏误

当选取了错误函数形式并对其进行估计时，带来的偏误称**错误函数形式偏误**（wrong functional form bias）。

容易判断，这种**偏误是全方位的**。

例如，如果“真实”的回归函数为

$$Y = AX_1^{\beta_1} X_2^{\beta_2} e^{\mu}$$

却估计线性式

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v$$

显然，两者的参数具有完全不同的经济含义，且估计结果一般也是不相同的。

三、模型设定偏误的检验

1、检验是否含有无关变量

可用t 检验与F检验完成。

检验的基本思想: 如果模型中误选了无关变量，则其系数的真值应为零。因此，只须对无关变量系数的显著性进行检验。

t检验：检验某1个变量是否应包括在模型中；

F检验：检验若干个变量是否应同时包括在模型中

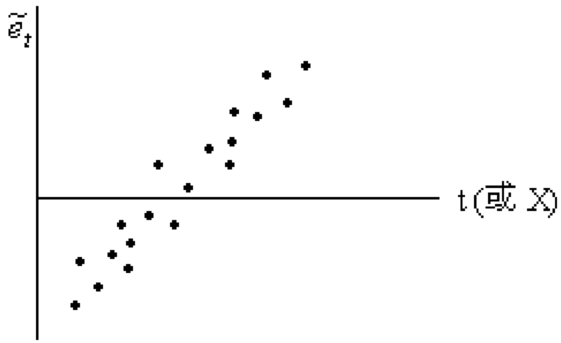
2、检验是否有相关变量的遗漏或函数形式设定偏误

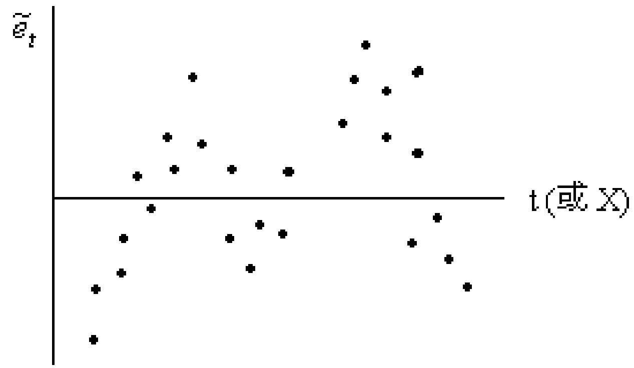
（1）残差图示法

对所设定的模型进行OLS回归，得到估计的残差序列$\tilde{e}_t$；

做出$\tilde{e}_t$与时间t或某解释变量X的散点图，考察$\tilde{e}_t$是否有规律地在变动，以判断是否遗漏了重要的解释变量或选取了错误的函数形式。

- 残差序列变化图





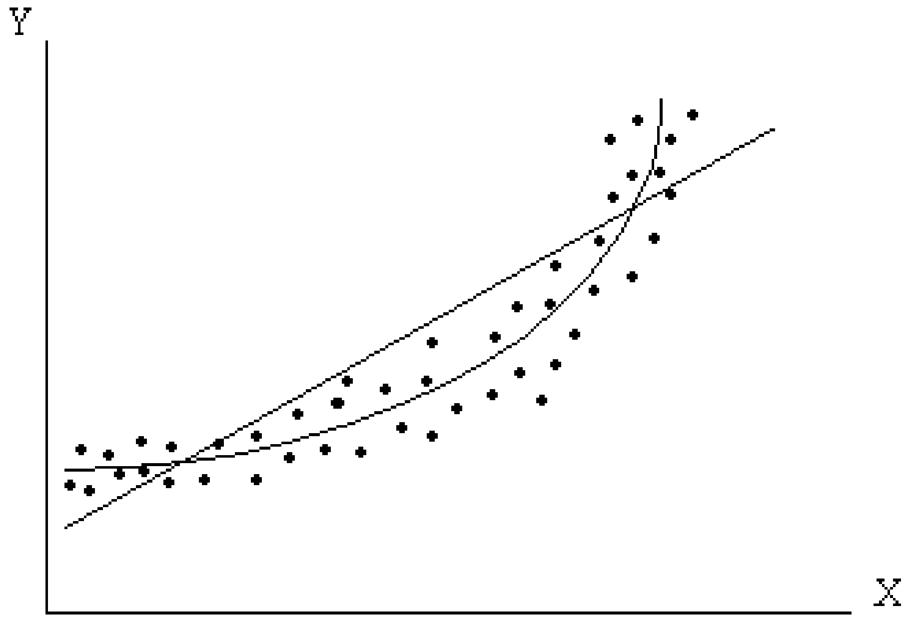
（a）趋势变化：

模型设定时可能遗漏了一随着时间的推移而持续上升的变量

（b）循环变化：

模型设定时可能遗漏了一随着时间的推移而呈现循环变化的变量

- 模型函数形式设定偏误时残差序列呈现正负交替变化



图示：一元回归模型中，真实模型呈幂函数形式，但却选取了线性函数进行回归。

（2）一般性设定偏误检验

但更准确更常用的判定方法是拉姆齐(Ramsey)于1969年提出的所谓**RESET 检验**（regression error specification test）。

基本思想：

如果事先知道遗漏了哪个变量，只需将此变量引入模型，估计并检验其参数是否显著不为零即可；

问题是不知道遗漏了哪个变量，需寻找一个替代变量Z，来进行上述检验。

RESET检验中，采用所设定模型中被解释变量Y的估计值$\hat{Y}$的若干次幂来充当该“替代”变量。

例如，先估计 $Y=\alpha_0+\alpha_1X_1+v$ 得

$$\hat{Y}=\hat{\alpha}_0+\hat{\alpha}_1X_1$$

再用通过残差项$\tilde{e}_t$与估计的$\hat{Y}$的图形判断引入$\hat{Y}$的若干次幂充当“替代”变量。

如$\tilde{e}_t$与Y的图形呈现曲线形变化时，回归模型可选为：

$$Y=\beta_0+\beta_1X_1+\gamma_1\hat{Y}^2+\gamma_2\hat{Y}^3+\mu$$

再根据第三章第五节介绍的**增加解释变量的F检验**来判断是否增加这些“替代”变量。

若仅增加一个“替代”变量，也可通过**t检验**来判断。

RESET检验也可用来检验函数形式设定偏误的问题。

例如，在一元回归中，假设真实的函数形式是非线性的，用泰勒定理将其近似地表示为多项式：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \cdots + \mu \qquad (*)$$

因此，如果设定了线性模型，就意味着遗漏了相关变量X_1^2、 X_1^3，等等。

因此，在一元回归中，可通过检验(*)式中的各高次幂参数的显著性来判断是否将非线性模型误设成了线性模型。

对多元回归，非线性函数可能是关于若干个或全部解释变量的非线性，这时可**按遗漏变量的程序进行检验**。

例如，估计 $Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\mu$

但却怀疑真实的函数形式是非线性的。

这时，只需以估计出的$\hat{Y}$的若干次幂为“替代”变量，进行类似于如下模型的估计

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\gamma_1 \hat{Y}^2+\gamma_2 \hat{Y}^3+\mu$$

再判断各“替代”变量的参数是否显著地不为零即可。

例5.3.1：在§4.3商品进口的例中,估计了中国商品进口M与GDP的关系，并发现具有强烈的一阶自相关性。

然而，由于仅用GDP来解释商品进口的变化，明显地遗漏了诸如商品进口价格、汇率等其他影响因素。因此，序列相关性的主要原因可能就是建模时遗漏了重要的相关变量造成的。

下面进行RESET检验。

用原回归模型估计出商品进口序列

$$\hat{M}_t = 152.91 + 0.020 GDP_t$$

R^2=0.9484

在原回归模型中加入$\hat{M}_t^2$、$\hat{M}_t^3$后重新进行估计，得：

$$\tilde{M}_t = -3.860 + 0.072GDP - 0.0028\hat{M}_t^2 + 8.59E-07\hat{M}_t^3$$

（-0.085）（8.274）（-6.457）（6.692）

$R^2=0.9842$

$$F = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-(k+q+1))} = \frac{(0.984-0.948)/2}{(1-0.984)/(24-4)} = 22.5$$

在$\alpha=5\%$下，查得临界值$F_{0.05}(2, 20)=3.49$

判断：拒绝原模型与引入新变量的模型可决系数无显著差异的假设，表明原模型确实存在遗漏相关变量的设定偏误。

*（3）同期相关性的豪斯蔓（Hausman）检验

由于在遗漏相关变量的情况下，往往导致解释变量与随机扰动项出现同期相关性，从而使得OLS估计量有偏且非一致。

因此，对模型遗漏相关变量的检验可以用模型是否出现解释变量与随机扰动项同期相关性的检验来替代。这就是**豪斯蔓检验（1978）的主要思想。**

当解释变量与随机扰动项同期相关时，通过工具变量法可得到参数的一致估计量。

而当解释变量与随机扰动项同期无关时， OLS估计量就可得到参数的一致估计量。

因此，**只须检验IV估计量与OLS估计量是否有显著差异来检验解释变量与随机扰动项是否同期无关。**

对一元线性回归模型

$$Y=\beta_0+\beta_1 X+\mu$$

所检验的假设是 H_0：X与μ无同期相关。

设一元样本回归模型为

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$$

以Z为工具变量，则IV估计量为：

$$\tilde{\beta} = \frac{\sum z_i y_i}{\sum z_i x_i}$$

$$= \frac{\sum z_i(\hat{\beta}_1 x_i + e_i)}{\sum z_i x_i} = \hat{\beta}_1 + \frac{\sum z_i e_i}{\sum z_i x_i} \qquad (*)$$

其中，$\hat{\beta}_1$为OLS估计量。

(*)式表明，IV估计量与OLS估计量无差异当且仅当$\sum z_i e_i=0$，即工具变量与OLS估计的残差项无关。

检验时，求Y关于X与Z的OLS回归式：

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\gamma} Z_i$$

如果$\hat{\gamma}$显著地异于零，就表明工具变量Z与$Y=\beta_0+\beta_1X+\mu$式OLS估计的残差相关，因此，拒绝原假设，说明X与μ同期相关。

在实际检验中，豪斯蔓检验主要针对多元回归进行，而且也不是直接对工具变量回归，而是对以各工具变量为自变量、分别以各解释变量为因变量进行回归。

如对二元回归模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i \qquad (*)$$

如果选取了若干个工具变量$Z_1, Z_2, \ldots, Z_m$，分别以X_1与X_2为因变量关于所有工具变量做回归，求出数据序列$\hat{X}_1$与$\hat{X}_2$，再估计下面的方程：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma_1 \hat{X}_{1i} + \gamma_2 \hat{X}_{2i}$$

通过**增加解释变量的F检验**，检验联合假设：
H_0：$\gamma_1=\gamma_2=0$ 。

拒绝原假设，就意味着（*）式中的解释变量与随机扰动项相关。

（4）线性模型与双对数线性模型的选择

无法通过判定系数的大小来辅助决策，因为在两类模型中被解释变量是不同的。

为了在两类模型中比较，可用Box-Cox变换:

第一步，计算Y的样本几何均值。

$$\tilde{Y} = (Y_1 Y_2 \cdots Y_n)^{1/n} = \exp(\frac{1}{n}\sum \ln Y_i)$$

第二步，用得到的样本几何均值去除原被解释变量Y，得到被解释变量的新序列Y*。

$$Y_i^* = Y_i / \tilde{Y}$$

第三步，用Y*替代Y，分别估计双对数线性模型与线性模型。并通过比较它们的残差平方和是否有显著差异来进行判断。

Zarembka（1968）提出的检验统计量为：

$$\frac{1}{2} n \ln\left(\frac{RSS_2}{RSS_1}\right)$$

其中，RSS_1与RSS_2分别为对应的较大的残差平方和与较小的残差平方和，n为样本容量。

可以证明：该统计量在两个回归的残差平方和无差异的假设下服从自由度为1 的χ^2分布。

因此，拒绝原假设时，就应选择RSS_2的模型。

例5.3.2 在§4.3中国商品进口的例中，

采用线性模型：R^2**=0.948**;

采用双对数线性模型: R^2**=0.973**，

但不能就此简单地判断双对数线性模型优于线性模型。下面进行Box-Cox变换。

计算原商品进口样本的几何平均值为：

$$\tilde{M} = \exp(\tfrac{1}{n}\sum \ln(M_t)) = 583.12$$

计算出新的商品进口序列：

$$M_t^* = M_t / .\tilde{M}$$

以M_t*替代M_t，分别进行双对数线性模型与线性模型的回归，得:

$$\ln(\hat{M}_t^*) = -1.3565 + 0.7836 \ln GDP_t$$ $RSS_1=0.5044$

$$\hat{M}_t^* = 0.2622 + 0.000035 GDP_t$$ $RSS_2=1.5536$

于是，$$\frac{1}{2} n \ln(\frac{RSS_2}{RSS_1}) = \frac{1}{2} \times 24 \ln(1.1249) = 13.49$$

在$\alpha=5\%$下，查得临界值$\chi^2_{0.05}(1)=3.841$

判断：拒绝原假设，表明**双对数线性模型确实“优于”线性模型。**

§5.4 从传统建模理论到约化建模理论

一、传统建模理论与数据开采问题

二、“从一般到简单”——约化建模理论

三、非嵌套假设检验

四、约化模型的准则

20世纪70年代中叶以来，计量经济学建模方法与建模理论得到了迅速发展。出现了利莫尔（Leamer）的**贝叶斯建模方法**，西姆斯（Sims）的**向量自回归建模型法**、亨德瑞（Hendry）的**约化建模理论**以及第10章将要学习的**协整建模理论**。这些现代建模理论是在对传统建模理论的不断质疑与修正中发展起来的，

亨德瑞的约化建模理论，吸收了向量自回归建模法与协整理论的部分内容，提出了**"从一般到简单"的建模思想**，在现代计量经济建模理论方面有着较大影响。

一、传统建模理论与数据开采问题

传统计量经济学的主导建模理论是“**结构模型方法论**”：

以先验给定的经济理论为建立模型的出发点，

以模型参数的估计为重心，

以参数估计值与其理论预期值相一致为判断标准，

是一个“**从简单到复杂**”的建模过程（simple-to-general approach）：

对不同变量及其数据的偿试与筛选过程.

这种传统的建模方法却有着某些固有的缺陷。其中备受质疑的是这种建模过程的所谓“数据开采”（Data minimg）问题。

数据开采:对不同变量及其数据的偿试与筛选

这一过程对最终选择的变量的t检验产生较大影响

当在众多备选变量中选择变量进入模型时，其中t检验的真实的显著性水平已不再是事先给出的名义显著性水平。

显著性水平意味着将一个无关变量作为相关变量选入模型而犯错误的概率。

罗维尔（Lovell）给出了一个从c个备选变量中选取k个变量进入模型时，真实显著性水平α^*与名义显著性水平α的关系：

$$\alpha^*=1-(1-\alpha)^{c/k}$$

如：　给定$\alpha=5\%$，如果有2个相互独立且与被解释变量无关的备选变量，误选一个进入模型的概率就成了 $1-(1-0.05)^2=0.0975$

传统建模方法的另一问题是它的“随意性”。

其结果是：对同一研究对象，使用同一数据，但不同的建模者往往得出不同的最终模型。

二、“从一般到简单”——约化建模型理论

该理论认为：在模型的最初设定上，就设立一个“一般”的模型，它包括了所有先验经济理论与假设中所应包括的全部变量，各种可能的“简单”模型都被“嵌套”（nested）在这个“一般”的模型之中。然后在模型的估计过程中逐渐剔除不显著的变量，最后得到一个较“简单”的最终模型。

这就是所谓的“**从一般到简单**”（general-to-specific）的建模理论。

特点：

(1)约化建模理论提出了一个对不同先验假设的更为系统的检验程序；

(2) 初始模型就是一个包括所有可能变量的"一般"模型，也就避免了过度的"数据开采"问题；

(3)由于初始模型的"一般"性，所有研究者的"起点"都有是相同的，因此，在相同的约化程序下，最后得到的最终模型也应该是相同的。

“从一般到简单”的建模理论例

例3.5.1曾建立了一个中国城镇居民食品消费模型：

$$Q=f(X,P_1,P_0)$$

然而，有理由认为X、P_1、P_0的变化可能会经过一段时期才会对Q起作用，因为消费者固有的消费习惯是不易改变的。于是，可建立如下更“一般”的模型：

$$\begin{aligned}\ln Q_t &= \alpha_0 + \alpha_1 \ln Q_{t-1} + \beta_1 \ln X_t + \beta_2 \ln X_{t-1} \\ &\quad + \gamma_1 \ln P_{1t} + \gamma_2 \ln P_{1t-1} + \delta_1 \ln P_{0t} + \delta_1 \ln P_{0t-1} + \mu_t\end{aligned}$$

在估计该模型之前，并不知道食品消费需求是怎样决定的，但可以考察几种可能的情况：

如，(1)对食品的消费需求是一个“静态”行为，只有当期的因素发生作用：

$$\ln Q_t = \alpha_0 + \beta_1 \ln X_t + \gamma_1 \ln P_{1t} + \delta_1 \ln P_{0t} + \mu_t \qquad (*)$$

也可以认为，(2)由于食品是必需品，P_1的变化并不对Q产生影响，但仍受P_0与X变动的影响，然而后者的影响却有着一期的滞后：

$$\ln Q_t = \alpha_0 + \beta_1 \ln X_t + \beta_2 \ln X_{t-1} + \delta_1 \ln P_{0t} + \delta_2 \ln P_{0t-1} + \mu_t \qquad (**)$$

可以看出，(*)、(**)都是**原一般模型**的特例，即都可通过对**原一般模型**施加约束得到。

如果一个模型可通过对“一般”模型施加约束得到，则称该模型“**嵌套**”在一般模型之中。

$$\ln Q_t = \alpha_0 + \alpha_1 \ln Q_{t-1} + \beta_1 \ln X_t + \beta_2 \ln X_{t-1} + \gamma_1 \ln P_{1t} + \gamma_2 \ln P_{1t-1} + \delta_1 \ln P_{0t} + \delta_1 \ln P_{0t-1} + \mu_t$$

$$\ln Q_t = \alpha_0 + \beta_1 \ln X_t + \beta_2 \ln X_{t-1} + \delta_1 \ln P_{0t} + \delta_2 \ln P_{0t-1} + \mu_t$$

约束：$\alpha_1=\gamma_1=\gamma_2=0$

$$\ln Q_t = \alpha_0 + \beta_1 \ln X_t + \gamma_1 \ln P_{1t} + \delta_1 \ln P_{0t} + \mu_t$$

约束：$\alpha_1=\beta_2=\gamma_2=\delta_2=0$

$$\ln Q_t = \alpha_0 + \beta_1 \ln(X_t / P_{0t}) + \gamma_1 \ln(P_{1t} / P_{0t}) + \mu_t$$

约束：$\beta_1+\gamma_1+\delta_1=0$

一般地，一个“一般模型”具有如下两个重要特性：

第一，与所考察问题相关的不同的先验理论与假设都“嵌套”在该一般模型中；

第二，能较好地拟合数据，并能满足模型设定偏误的各种检验。

该两条性质是相互关联的。例如，如果某一重要理论被忽略，则相关的变量也就被排除在该“一般”模型之外，从而使得该模型不能通过模型设定偏误的多种检验。

一个“一般”的模型是能够进行诸如遗漏相关变量、多选无关变量以及误设函数形式的多种设定偏误检验的。

- **从一般到简单的约化建模过程**

一旦建立了一个“一般”模型，就可对其进行**约化**（simplification research），寻找可能的简单模型。

这往往是通过检验“嵌套”于其中的各种简单模型进行的。主要包括（1）各种“约束”检验与（2）设定偏误检验，等。

一般模型的约化过程，是一个自上而下（top-down）逐级化简的建模过程。只有当观测数据不支持约束条件时，才退回到上一级，检验其他可能的约束，或者得到最终模型。

三、非嵌套假设检验

“从一般到简单”的建模程序面临的主要问题在于无法在两个没有嵌套关系的模型间进行选择。

这时，可能通过通常的拟合优度检验、池赤信息准则来帮助决策，更主要的检验是**非嵌套假设检验**。

假设要检验下面两个非嵌套模型：

H_0: $Y=\beta_0+\beta_1X+\beta_2Z+\mu$

H_1: $Y=\alpha_0+\alpha_1X+\alpha_2W+\varepsilon$

上述两模型之间没有嵌套关系，无法进行约束检验。

同时，H_0与H_1不是对立假设，拒绝假设H_0未必意味着接受假设H_1。因此，通常的假设检验程序无法直接使用。

为此，一种称为**包容性F检验**（encompassing F tests）被提了出来。这种检验是人为地构造一个“一般”模型：

$$Y = \gamma_0 + \gamma_1 X + \gamma_2 Z + \gamma_3 W + \mu \qquad (*)$$

于是，可针对一般模型(*)分别检验H_0与H_1 。

- 包容性F检验主要存在以下问题:

(1) **人为构造的一般模型没有实际的经济意义，** 尤其在H_0与H_1分别反映两种对立的经济理论的情况下更是如此；

(2) **有可能出现同时接受或拒绝H_0与H_1的现象；**

(3) **当Z与W高度相关时，往往导致既不能拒绝H_0，也不能拒绝H_1，** 因为在一般模型中去掉任何一个变量，都不会使拟合优度下降很多。

另一个解决办法是建立如下的一般模型：

$$Y=(1-\mu)(\beta_0+\beta_1 X+\beta_2 Z)+\mu(\alpha_0+\alpha_1 X+\alpha_2 W)+\varepsilon$$

如果μ=0，则为模型H_0，

如果μ=1，则为模型H_1。

因此，可通过检验施加的约束μ=0是否为真来判断H_0是否为正选模型。

问题是由该模型无法直接估计出μ的值。戴维森（Davidson）和麦金农（Mackinnon）建议通过下面步骤估计μ：

第一步，对模型H_1进行OLS估计，得到$\hat{Y}$：

$$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X + \hat{\alpha}_2 W$$

第二步，用估计的代替“一般模型”中的

$\alpha_0 + \alpha_1 X + \alpha_2 W$，并进行OLS估计：

$$Y = (1-\mu)(\beta_0 + \beta_1 X + \beta_2 Z) + \mu\hat{Y} + \varepsilon$$

戴维森和麦金农证明：**在大样本下，H_0为真时，μ的OLS估计量的t统计量服从标准正态分布：**

t~N(0, 1)。

因此，如果μ的t统计量的绝对值大于给定显著性水平下的临界值，就拒绝模型H_0。

如果要检验模型H_1是否为真，仍可通过上面两个步骤进行，但需先对H_0进行OLS估计，得到Ŷ，以它为另一解释变量估计如下模型：

$$Y=(1-\mu)(\alpha_0+\alpha_1 X+\alpha_2 W)+\mu\hat{Y}+\varepsilon$$

如果μ显著地异于0，则拒绝模型H_1为真的假设。

该非嵌套假设检验也被称为**J检验**（J test），因为需将两非嵌套模型联合起来进行参数的联合估计（joint estimation）。

注意：(1)拒绝H_0(或H_1)不意味着接受H_1(或H_0)；(2)J检验仍然存在同时接受或拒绝H_0与H_1的现象。

四、约化模型的准则

从一般到简单的建模过程，同样存在着数据开采问题。

一个“一般”模型经过k步约化后得到最终的简化模型，可以证明，每一步中的名义显著性水平α与最终模型中各种检验的实际显著性水平α^*间有如下关系：

$$\alpha^*=1-(1-\alpha)^k$$

然而，与“从简单到复杂”这一传统建模方法相比，“从一般到简单”的建模过程能够展现模型建立的全过程；

同时建模过程的程式化（systematic manner）也避免了过度的“数据开采”问题。

由于一定程度的数据开采不可避免，“从一般到简单”建模理论倡导更加关注模型的样本外预测（out-of-sample forecast）。

“从一般到简单”的建模方法，初始模型就可能包括了所有的相关变量，没有必要再进行遗漏相关变量的设定偏误检验。

“从一般到简单”的建模过程本身就是一项十分艰巨复杂的工作。各约化步骤往往是需要反复进行的，约化步骤的顺序也需要灵活按排。

而且，从实践上看，由于各种因素的影响，所建立的最终的简化模型不一定就是最“理想”的模型。亨德瑞给出了一个**约化模型的基本准则：**

第一，模型必须具有数据一致(data-coherent)性，即模型能够正确地解释已有的数据。约化过程中需不断进行设定偏误检验。

第二，模型必须与经济理论相一致（consistent with economic theory）。

第三，解释变量必须是弱外生的(exogenous)，

即解释变量应与随机扰动项不同期相关。

第四，模型具有恒定的参数(constant parameters)

第五，模型具有包容性，即模型应包容相竞争的对手模型。

第六，模型具有简洁性(parsimonious)，即在具有相同解释能力的情况下，一个拥有较少解释变量的模型优于拥有较多解释变量的模型。

例5.4.1 在§3.5的例3.5.1中，曾以传统的建模方法建立了1981~1994年间的中国城镇居民食品消费需求模型。

这里再以“从一般到简单”这一建模理论来做进一步的考察。

初始的一般模型设定为

$$\ln Q_t = \alpha_0 + \alpha_1 \ln Q_{t-1} + \beta_1 \ln X_t + \beta_2 \ln X_{t-1} + \gamma_1 \ln P_{1t} + \gamma_2 \ln P_{1t-1} + \delta_1 \ln P_{0t} + \delta_1 \ln P_{0t-1} + \mu_t$$

用小写字母代表变量的自然对数，则该一般模型的估计结果为:

$$\hat{q}_t = 3.297 + 0.042 q_{t-1} + 1.2620 x_t - 0.285 x_{t-1}$$

（1.41） (0.09) (8.24) (-0.57)

$$- 0.029 p_{1t} - 0.015 p_{1t-1} - 1.227 p_{0t} + 0.397 p_{0t-1}$$

(-0.65) (-0.24) (-6.03) (0.85)

$\overline{R}^2=0.999$ RSS=0.001166 DW=2.49

LM(1)=3.128 (p=0.077) RESET=8.515 (p=0.058)

CHOW=1.631 (p=0.305)

给定5%的显著性水平，可以判断，尽管若干个变量的t检验不显著，但**总体上看**，不存在模型的相关变量遗漏与函数形式的设定偏误问题，而且参数也具有稳定性。因此，**以它作为初始的一般模型是合适的。**

进一步考察模型的约化问题：

首先，检验模型

$$\ln Q_t = \alpha_0 + \beta_1 \ln X_t + \gamma_1 \ln P_{1t} + \delta_1 \ln P_{0t} + \mu_t$$

$$\hat{q}_t = 3.634 + 1.055 x_t - 0.080 p_{1t} - 0.925 p_{0t}$$

(9.03)　(25.35)　(-2.28)　(-7.35)

$\bar{R}^2=0.999$　RSS=0.00324　DW=1.50

LM(1)=0.856（p=0.355）　RESET=3.952（p=0.064）

CHOW=4.651 (p=0.015)

该模型是由“一般模型”去掉滞后变量得到，相当于对滞后变量施加了零约束，由受约束的F检验得检验值F=2.188，相伴概率p=0.207，可见：**在5%的显著性水平下，可接受该约束。**

但是，存在着结构变化，而且RSS有明显增大。

如果忽略存在结构变化这一特征，则上面模型能够作为一个可接受的模型，并可进一步检验

$$\ln Q_t = \alpha_0 + \beta_1 \ln(X_t / P_{0t}) + \gamma_1 \ln(P_{1t} / P_{0t}) + \mu_t$$

$$\hat{q}_t = 3.825 + 1.073(x_t - p_{0t}) - 0.091(p_{1t} - p_{0t})$$

（75.86）（52.66）（-3.62）

$\overline{R}^2=0.999$　RSS=0.003315　DW=1.51

LM(1)=0.382（p=0.536）　RESET=4.369（p=0.0047）

CHOW=9.856 (p=0.001)

取α=5%，RESET检验表明可能存在遗漏相关变量的设定偏误，这时RSS的值也有所增大，而且CHOW检验也表明存在明显的结构变化。

第六章　联立方程计量经济模型理论方法

Theory and Methodology of Simultaneous-Equations Econometrics Model

教学基本要求

- 本章是课程的重点内容之一。通过教学，要求学生达到：
- 了解（最低要求）：线性联立方程计量经济学模型的基本概念，线性联立方程模型的矩阵表示，有关模型识别的概念和实用的识别方法，几种主要的单方程估计方法（间接最小二乘法、工具变量法、两阶段最小二乘法）的原理与应用。

- 掌握（较高要求）：运用矩阵描述、推导和证明与间接最小二乘法、工具变量法和两阶段最小二乘法有关的过程和结论；为什么在实践中经常采用普通最小二乘法估计线性联立方程计量经济学模型；联立方程计量经济学模型系统检验的理论与方法。
- 应用（对应用能力的要求）：应用所学知识，在本章结束前独立完成一个综合练习，建立一个3-5个方程的中国宏观经济模型，自己建立理论模型，自己收集样本数据，采用几种方法应用计量经济学软件包进行模型的估计，对结果进行分析，最后提交一篇报告。

§6.1 问题的提出

一、经济研究中的联立方程计量经济学问题

二、计量经济学方法中的联立方程问题

一、经济研究中的联立方程计量经济学问题

1. 研究对象

- 经济系统，而不是单个经济活动

  “系统”的相对性

- 相互依存、互为因果，而不是单向因果关系
- 必须用一组方程才能描述清楚

2. 一个简单的宏观经济系统

- 由国内生产总值Y、居民消费总额C、投资总额I和政府消费额G等变量构成简单的宏观经济系统。
- 将政府消费额G由系统外部给定，其他内生。

$$
\begin{cases}
C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\
I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\
Y_t = C_t + I_t + G_t
\end{cases}
$$

- 在消费方程和投资方程中，国内生产总值决定居民消费总额和投资总额；
- 在国内生产总值方程中，它又由居民消费总额和投资总额所决定。

二、计量经济学方法中的联立方程问题

1. 随机解释变量问题

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 解释变量中出现随机变量，而且与误差项相关。
- 为什么？

2. 损失变量信息问题

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 如果用单方程模型的方法估计某一个方程，将损失变量信息。
- 为什么？

3. 损失方程之间的相关性信息问题

$$
\begin{cases}
C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\
I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\
Y_t = C_t + I_t + G_t
\end{cases}
$$

- 联立方程模型系统中每个随机方程之间往往存在某种相关性。
- 表现于不同方程随机误差项之间。
- 如果用单方程模型的方法估计某一个方程，将损失不同方程之间相关性信息。

4. 结论

- 必须发展新的估计方法估计联立方程计量经济学模型，以尽可能避免出现这些问题。
- 这就从计量经济学理论方法上提出了联立方程问题。

§6.2联立方程计量经济学模型的若干基本概念

- 变量
- 结构式模型
- 简化式模型
- 参数关系体系

一、变量

1. 内生变量 （Endogenous Variables）

- 对联立方程模型系统而言，已经不能用被解释变量与解释变量来划分变量，而将变量分为内生变量和外生变量两大类。
- 内生变量是具有某种概率分布的随机变量，它的参数是联立方程系统估计的元素。
- 内生变量是由模型系统决定的，同时也对模型系统产生影响。
- 内生变量一般都是经济变量。

- 一般情况下，内生变量与随机项相关，即

$$
\begin{aligned}
Cov(Y_i, \mu_i) &= E((Y_i - E(Y_i))(\mu_i - E(\mu_i))) \\
&= E((Y_i - E(Y_i))\mu_i) \\
&= E(Y_i\mu_i) - E(Y_i)E(\mu_i) \\
&= E(Y_i\mu_i) \\
&\neq 0
\end{aligned}
$$

- 在联立方程模型中，内生变量既作为被解释变量，又可以在不同的方程中作为解释变量。

2. 外生变量 (Exogenous Variables)

- 外生变量一般是确定性变量，或者是具有临界概率分布的随机变量，其参数不是模型系统研究的元素。
- 外生变量影响系统，但本身不受系统的影响。
- 外生变量一般是经济变量、条件变量、政策变量、虚变量。
- 一般情况下，外生变量与随机项不相关。

3. 先决变量（Predetermined Variables）

- 外生变量与滞后内生变量(Lagged Endogenous Variables)统称为先决变量。
- 滞后内生变量是联立方程计量经济学模型中重要的不可缺少的一部分变量，用以反映经济系统的动态性与连续性。
- 先决变量只能作为解释变量。

二、结构式模型
Structural Model

1.定义

- 根据经济理论和行为规律建立的描述经济变量之间直接结构关系的计量经济学方程系统称为结构式模型。
- 结构式模型中的每一个方程都是结构方程（ Structural Equations ）。
- 各个结构方程的参数被称为结构参数（ Structural Parameters or Coefficients ） 。
- 将一个内生变量表示为其它内生变量、先决变量和随机误差项的函数形式，被称为结构方程的正规形式。

2. 结构方程的方程类型

- 随机方程
  - 行为方程
  - 技术方程
  - 制度方程
  - 统计方程
- 恒等方程
  - 定义方程
  - 平衡方程
  - 经验方程

3. 完备的结构式模型

- 具有g个内生变量、k个先决变量、g个结构方程的模型被称为完备的结构式模型。
- 在完备的结构式模型中，独立的结构方程的数目等于内生变量的数目，每个内生变量都分别由一个方程来描述。

4. 完备的结构式模型的矩阵表示

- 习惯上用Y表示内生变量，X表示先决变量，μ表示随机项，β表示内生变量的结构参数，γ表示先决变量的结构参数，如果模型中有常数项，可以看成为一个外生的虚变量，它的观测值始终取**1**。

$$\mathrm{B}\boldsymbol{Y} + \Gamma\boldsymbol{X} = \mathrm{N}$$

$$(\mathrm{B}\Gamma)\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix} = \mathrm{N}$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{pmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & & & \\ y_{g1} & y_{g2} & \cdots & y_{gn} \end{bmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix}$$

$$\mathrm{N} = \begin{pmatrix} \mathrm{N}_1 \\ \mathrm{N}_2 \\ \vdots \\ \mathrm{N}_g \end{pmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & & & \\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gn} \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1g} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g} \\ \vdots & & & \\ \beta_{g1} & \beta_{g2} & \cdots & \beta_{gg} \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \vdots & & & \\ \gamma_{k1} & \gamma_{k2} & \cdots & \gamma_{kk} \end{bmatrix}$$

5. 简单宏观经济模型的矩阵表示

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

$$\mathbf{Y} = \begin{bmatrix} C_t \\ I_t \\ Y_t \end{bmatrix} = \begin{bmatrix} C_1 & C_2 & \cdots & C_n \\ I_1 & I_2 & \cdots & I_n \\ Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 \\ Y_{t-1} \\ G_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ Y_0 & Y_1 & \cdots & Y_{n-1} \\ G_1 & G_2 & \cdots & G_n \end{bmatrix}$$

$$\mathrm{N} = \begin{bmatrix} \mathrm{N}_1 \\ \mathrm{N}_2 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$(\mathrm{B}\Gamma) = \begin{bmatrix} 1 & 0 & -\alpha_1 & -\alpha_0 & 0 & 0 \\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0 \\ -1 & -1 & 1 & 0 & 0 & -1 \end{bmatrix}$$

三、简化式模型

Reduced-Form Model

1.定义

- 用所有先决变量作为每个内生变量的解释变量，所形成的模型称为简化式模型。
- 简化式模型并不反映经济系统中变量之间的直接关系，并不是经济系统的客观描述。
- 由于简化式模型中作为解释变量的变量中没有内生变量，可以采用普通最小二乘法估计每个方程的参数，所以它在联立方程模型研究中具有重要的作用。
- 简化式模型中每个方程称为简化式方程(Reduced-Form Equations)，方程的参数称为简化式参数(Reduced-Form Coefficients)。

2. 简化式模型的矩阵形式

$$\mathbf{Y} = \mathbf{\Pi X} + \mathbf{E}$$

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1k} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2k} \\ \vdots & & & \\ \pi_{g1} & \pi_{g2} & \cdots & \pi_{gk} \end{bmatrix} \qquad \mathrm{E} = \begin{bmatrix} \mathrm{E}_1 \\ \mathrm{E}_2 \\ \vdots \\ \mathrm{E}_{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1n} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2n} \\ \vdots & & & \\ \varepsilon_{g1} & \varepsilon_{g2} & \cdots & \varepsilon_{gn} \end{bmatrix}$$

3. 简单宏观经济模型的简化式模型

$$\begin{cases} C_t = \pi_{10} + \pi_{11} Y_{t-1} + \pi_{12} G_t + \varepsilon_t \\ I_t = \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} G_t + \varepsilon_t \\ Y_t = \pi_{30} + \pi_{31} Y_{t-1} + \pi_{32} G_t + \varepsilon_t \end{cases}$$

四、参数关系体系

1. 定义

$$\mathrm{B}\boldsymbol{Y} + \Gamma\boldsymbol{X} = \mathrm{N}$$

$$\mathrm{B}\mathbf{Y} = -\Gamma\mathbf{X} + \mathrm{N}$$

$$\mathbf{Y} = -\mathrm{B}^{-1}\Gamma\mathbf{X} + \mathrm{B}^{-1}\mathrm{N}$$

$$\mathbf{Y} = \Pi\mathbf{X} + \mathrm{E}$$

$$\Pi = -\mathrm{B}^{-1}\Gamma$$

- 该式描述了简化式参数与结构式参数之间的关系，称为参数关系体系。

2.作用

- 利用参数关系体系，首先估计简化式参数，然后可以计算得到结构式参数。
- 从参数关系体系还可以看出，简化式参数反映了先决变量对内生变量的直接与间接影响之和，这是简化式模型的另一个重要作用。

例如，在上述模型中存在如下关系：

$$\pi_{21} = \frac{\beta_2 - \alpha_1\beta_2}{1-\alpha_1-\beta_1} = \beta_2 + \frac{\beta_1\beta_2}{1-\alpha_1-\beta_1}$$

Π_{21}反映Y_{t-1}对I_t的直接与间接影响之和； 而其中的β_2正是结构方程中Y_{t-1}对I_t的结构参数，显然，它只反映Y_{t-1}对I_t的直接影响。

- 在这里，β_2是Y_{t-1}对I_t的部分乘数，Π_{21}反映Y_{t-1}对I_t的完全乘数。
- 注意：简化式参数与结构式参数之间的区别与联系。

§6.3联立方程计量经济学模型的识别

The Identification Problem

一、识别的概念

二、从定义出发识别模型

三、结构式识别条件

四、简化式识别条件

五、实际应用中的经验方法

一、识别的概念

1. 为什么要对模型进行识别？

- 从一个例子看

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + + \mu_{2t} \\ Y_t = C_t + I_t \end{cases}$$

- 消费方程是包含C、Y和常数项的直接线性方程。
- 投资方程和国内生产总值方程的某种线性组合（消去I）所构成的新方程也是包含C、Y和常数项的直接线性方程。

- 如果利用C、Y的样本观测值并进行参数估计后，很难判断得到的是消费方程的参数估计量还是新组合方程的参数估计量。
- 只能认为原模型中的消费方程是不可估计的。
- 这种情况被称为不可识别。
- 只有可以识别的方程才是可以估计的。

2. 识别的定义

- 3种定义：

"如果联立方程模型中某个结构方程不具有确定的统计形式，则称该方程为不可识别。"

"如果联立方程模型中某些方程的线性组合可以构成与某一个方程相同的统计形式，则称该方程为不可识别。"

"根据参数关系体系，在已知简化式参数估计值时，如果不能得到联立方程模型中某个结构方程的确定的结构参数估计值，则称该方程为不可识别。"

- 以是否具有确定的统计形式作为识别的基本定义。
- 什么是“统计形式”？
- 什么是“具有确定的统计形式”？

3. 模型的识别

- 上述识别的定义是针对结构方程而言的。
- 模型中每个需要估计其参数的随机方程都存在识别问题。
- 如果一个模型中的所有随机方程都是可以识别的，则认为该联立方程模型系统是可以识别的。反过来，如果一个模型系统中存在一个不可识别的随机方程，则认为该联立方程模型系统是不可以识别的。
- 恒等方程由于不存在参数估计问题，所以也不存在识别问题。但是，在判断随机方程的识别性问题时，应该将恒等方程考虑在内。

4. 恰好识别(Just Identification)与过度识别(Overidentification)

- 如果某一个随机方程具有一组参数估计量，称其为恰好识别；
- 如果某一个随机方程具有多组参数估计量，称其为过度识别。

二、从定义出发识别模型

1. 例题1

$$
\begin{cases}
C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\
I_t = \beta_0 + \beta_1 Y_t + + \mu_{2t} \\
Y_t = C_t + I_t
\end{cases}
$$

- 第2与第3个方程的线性组合得到的新方程具有与消费方程相同的统计形式，所以消费方程也是不可识别的。

- 第1与第3个方程的线性组合得到的新方程具有与投资方程相同的统计形式，所以投资方程也是不可识别的。
- 于是，该模型系统不可识别。
- 参数关系体系由3个方程组成，剔除一个矛盾方程，2个方程不能求得4个结构参数的确定值。也证明消费方程与投资方程都是不可识别的。

2. 例题2

$$C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。
- 投资方程仍然是不可识别的，因为第1、第2与第3个方程的线性组合（消去C）构成与它相同的统计形式。
- 于是，该模型系统仍然不可识别。

- 参数关系体系由6个方程组成，剔除2个矛盾方程，由4个方程是不能求得所有5个结构参数的确定估计值。
- 可以得到消费方程参数的确定值，证明消费方程可以识别；因为只能得到它的一组确定值，所以消费方程是恰好识别的方程。
- 投资方程都是不可识别的。
- 注意：与例题1相比，在投资方程中增加了1个变量，消费方程变成可以识别。

3. 例题3

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程仍然是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。
- 投资方程也是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。
- 于是，该模型系统是可以识别的。

- 参数关系体系由9个方程组成，剔除3个矛盾方程，在已知简化式参数估计值时，由6个方程能够求得所有6个结构参数的确定估计值。
- 所以也证明消费方程和投资方程都是可以识别的。
- 而且，只能得到所有6个结构参数的一组确定值，所以消费方程和投资方程都是恰好识别的方程。
- 注意：与例题2相比，在消费方程中增加了1个变量，投资方程变成可以识别。

4. 例题4

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \alpha_3 P_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程和投资方程仍然是可以识别的，因为任何方程的线性组合都不能构成与它们相同的统计形式。
- 于是，该模型系统是可以识别的。

- 参数关系体系由12个方程组成，剔除4个矛盾方程，在已知简化式参数估计值时，由8个方程能够求得所有7个结构参数的确定估计值。
- 所以也证明消费方程和投资方程都是可以识别的。
- 但是，求解结果表明，对于消费方程的参数，只能得到一组确定值，所以消费方程是恰好识别的方程;
- 而对于投资方程的参数，能够得到多组确定值，所以投资方程是过度识别的方程。

- 注意：
- 在求解线性代数方程组时，如果方程数目大于未知数数目，被认为无解；如果方程数目小于未知数数目，被认为有无穷多解。
- 但是在这里，无穷多解意味着没有确定值，所以，如果参数关系体系中有效方程数目小于未知结构参数估计量数目，被认为不可识别。
- 如果参数关系体系中有效方程数目大于未知结构参数估计量数目，那么每次从中选择与未知结构参数估计量数目相等的方程数，可以解得一组结构参数估计值，换一组方程，又可以解得一组结构参数估计值，这样就可以得到多组结构参数估计值，被认为可以识别，但不是恰好识别，而是过度识别。

5. 如何修改模型使不可识别的方程变成可以识别

- 或者在其它方程中增加变量；
- 或者在该不可识别方程中减少变量。
- 必须保持经济意义的合理性。

三、结构式识别条件

1. 结构式识别条件

- 直接从结构模型出发
- 一种规范的判断方法
- 每次用于1个随机方程
- 具体描述为：

联立方程计量经济学模型的结构式

$$\mathbf{BY}+\mathbf{\Gamma X}=\mathbf{N}$$

中的第 i 个方程中包含 g_i 个内生变量（含被解释变量）和 k_i 个先决变量（含常数项），模型系统中内生变量和先决变量的数目仍用 g 和 k 表示，矩阵 $(\mathrm{B}_0\Gamma_0)$ 表示第 i 个方程中未包含的变量（包括内生变量和先决变量）在其它 $g-1$ 个方程中对应系数所组成的矩阵。于是，判断第 i 个结构方程识别状态的结构式条件为：

如果 $R(\mathrm{B}_0\Gamma_0)<g-1$，则第 i 个结构方程不可识别；

如果 $R(\mathrm{B}_0\Gamma_0)=g-1$，则第 i 个结构方程可以识别，并且

如果 $k-k_i=g_i-1$，则第 i 个结构方程恰好识别，

如果 $k-k_i>g_i-1$，则第 i 个结构方程过度识别。

- 一般将该条件的前一部分称为秩条件（Rank Condition），用以判断结构方程是否识别；
- 将后一部分称为阶条件（Order Conditon），用以判断结构方程恰好识别或者过度识别。

2. 例题

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \alpha_3 P_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

$$[\mathrm{B}\Gamma] = \begin{bmatrix} 1 & 0 & -\alpha_1 & -\alpha_0 & 0 & -\alpha_2 & -\alpha_3 \\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0 & 0 \\ -1 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- **判断第1个结构方程的识别状态**

$$\left[\mathrm{B}_0\Gamma_0\right]=\begin{bmatrix}1 & -\beta_2\\ -1 & 0\end{bmatrix} \longrightarrow R(\mathrm{B}_0\Gamma_0)=2=g-1$$

所以，该方程可以识别。

因为

$$k-k_1=1=g_1-1$$

所以，第1个结构方程为恰好识别的结构方程。

- 判断第2个结构方程的识别状态

$$\begin{bmatrix} \mathrm{B}_0 \Gamma_0 \end{bmatrix} = \begin{bmatrix} 1 & -\alpha_2 & -\alpha_3 \\ -1 & 0 & 0 \end{bmatrix} \longrightarrow R(\mathrm{B}_0 \Gamma_0) = 2 = g - 1$$

所以，该方程可以识别。

因为

$$k - k_2 = 2 > g_2 - 1$$

所以，第2个结构方程为过度识别的结构方程。

- 第3个方程是平衡方程，不存在识别问题。
- 综合以上结果，该联立方程模型是可以识别的。
- 与从定义出发识别的结论一致。

四、简化式识别条件

1. 简化式识别条件

- 如果已经知道联立方程模型的简化式模型参数，那么可以通过对简化式模型的研究达到判断结构式模型是否识别的目的。
- 由于需要首先估计简化式模型参数，所以很少实际应用。

对于简化式模型

$$\boldsymbol{Y} = \Pi\boldsymbol{X} + \mathrm{E}$$

简化式识别条件为：

如果 $R(\Pi_2) < g_i - 1$，则第 i 个结构方程不可识别；

如果 $R(\Pi_2) = g_i - 1$，则第 i 个结构方程可以识别，并且

如果 $k - k_i = g_i - 1$，则第 i 个结构方程恰好识别，

如果 $k - k_i > g_i - 1$，则第 i 个结构方程过度识别。

其中 Π_2 是简化式参数矩阵 π 中划去第 i 个结构方程所不包含的内生变量所对应的行和第 i 个结构方程中包含的先决变量所对应的列之后，剩下的参数按原次序组成的矩阵。

2. 例题

- 需要识别的结构式模型

$$\begin{cases} y_{1i} = \alpha_1 + \alpha_2 x_{1i} + \alpha_3 x_{2i} + \mu_{1i} \\ y_{2i} = \beta_1 y_{3i} + \beta_2 x_{3i} + \mu_{2i} \\ y_{3i} = \gamma_1 y_{1i} + \gamma_2 y_{2i} + \gamma_3 x_{3i} + \mu_{3i} \end{cases}$$

- 已知其简化式模型参数矩阵为

$$\Pi = \begin{bmatrix} 4 & -2 & 3 \\ 2 & -1 & 1 \\ 2 & -1 & 0 \end{bmatrix}$$

- 判断第1个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$R(\Pi_2) = 1 = g_1 - 1$$

所以该方程是可以识别的。又因为

$$k - k_1 = 1 = g_1 - 1$$

所以该方程是恰好识别的。

- 判断第2个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix}$$

$$R(\Pi_2) = 1 = g_2 - 1$$

所以该方程是可以识别的。又因为

$$k - k_2 = 2 > g_2 - 1$$

所以该方程是过度识别的。

- 判断第3个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 4 & -2 \\ 2 & -1 \\ 2 & -1 \end{bmatrix} \qquad R(\Pi_2) = 1 < g_3 - 1$$

所以该方程是不可识别的。

- 所以该模型是不可识别的。

- 可以从数学上严格证明，简化式识别条件和结构式识别条件是等价的。

  《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年3月）第104—107页。

- 讨论：阶条件是确定过度识别的充分必要条件吗？（李子奈，《数量经济技术经济研究》，1988年第10期）

五、实际应用中的经验方法

- 当一个联立方程计量经济学模型系统中的方程数目比较多时，无论是从识别的概念出发，还是利用规范的结构式或简化式识别条件，对模型进行识别，困难都是很大的，或者说是不可能的。
- 理论上很严格的方法在实际中往往是无法应用的，在实际中应用的往往是一些经验方法。
- 关于联立方程计量经济学模型的识别问题，实际上不是等到理论模型已经建立了之后再进行识别，而是在建立模型的过程中设法保证模型的可识别性。

- “在建立某个结构方程时，要使该方程包含前面每一个方程中都不包含的至少1个变量（内生或先决变量）；同时使前面每一个方程中都包含至少1个该方程所未包含的变量，并且互不相同。”
- 该原则的前一句话是保证该方程的引入不破坏前面已有方程的可识别性。只要新引入方程包含前面每一个方程中都不包含的至少1个变量，那么它与前面方程的任意线性组合都不能构成与前面方程相同的统计形式，原来可以识别的方程仍然是可以识别的。
- 该原则的后一句话是保证该新引入方程本身是可以识别的。只要前面每个方程都包含至少1个该方程所未包含的变量，并且互不相同。那么所有方程的任意线性组合都不能构成与该方程相同的统计形式。

- 在实际建模时，将每个方程所包含的变量记录在如下表所示的表式中，将是有帮助的。

	变量 1	变量 2	变量 3	变量 4	变量 5	变量 6	…
方程 1	×	×		×			
方程 2		×	×	×	×		
方程 3	×		×	×		×	
方程 4		×	×				×
…							

§6.5-6联立方程模型的单方程估计方法
Single-Equation Estimation Methods

一、狭义的工具变量法（IV）

二、间接最小二乘法(ILS)

三、二阶段最小二乘法(2SLS)

四、三种方法的等价性证明

五、简单宏观经济模型实例演示

六、主分量法的应用

七、其它有限信息估计方法简介

八、k级估计式

- 联立方程计量经济学模型的估计方法分为两大类：单方程估计方法与系统估计方法。
- 所谓单方程估计方法，指每次只估计模型系统中的一个方程，依次逐个估计。
- 所谓系统估计方法，指同时对全部方程进行估计，同时得到所有方程的参数估计量。
- 联立方程模型的单方程估计方法不同于单方程模型的估计方法 。

一、狭义的工具变量法
（IV，Instrumental Variables）

1. 方法思路

- “狭义的工具变量法” 与“广义的工具变量法”
- 解决结构方程中与随机误差项相关的内生解释变量问题。
- 方法原理与单方程模型的IV方法相同。
- 模型系统中提供了可供选择的工具变量，使得IV方法的应用成为可能。

2. 工具变量的选取

- 对于联立方程模型的每一个结构方程，例如第1个方程，可以写成如下形式：

$$Y_1 = \beta_{12}Y_2 + \beta_{13}Y_3 + \cdots + \beta_{1g_1}Y_{g_1} + \gamma_{11}X_1 + \gamma_{12}X_2 + \cdots + \gamma_{1k_1}X_{k_1} + \mathrm{N}_1$$

- 内生解释变量（g_1-1）个，先决解释变量k_1个。
- 如果方程是恰好识别的，有（g_1-1）=（k- k_1）。
- 可以选择（k- k_1）个方程没有包含的先决变量作为（g_1-1）个内生解释变量的工具变量。

3. IV参数估计量

- 方程的矩阵表示为

$$Y_1 = (\boldsymbol{Y}_0, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_0 \\ \Gamma_0 \end{pmatrix} + \mathrm{N}_1$$

- 选择方程中没有包含的先决变量X_0*作为包含的内生解释变量Y_0的工具变量，得到参数估计量为：

$$\boxed{\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{IV} = \left(\begin{pmatrix} \boldsymbol{X}_0^* & \boldsymbol{X}_0 \end{pmatrix}' \begin{pmatrix} \boldsymbol{Y}_0 & \boldsymbol{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \boldsymbol{X}_0^* & \boldsymbol{X}_0 \end{pmatrix}' Y_1}$$

4. 讨论

- 该估计量与OLS估计量的区别是什么？
- 该估计量具有什么统计特性？
- （$k-k_1$）工具变量与（g_1-1）个内生解释变量的对应关系是否影响参数估计结果？为什么？
- IV是否利用了模型系统中方程之间相关性信息？
- 对于过度识别的方程，可否应用IV？为什么？
- 对于过度识别的方程，可否应用GMM？为什么？

二、间接最小二乘法
(ILS, Indirect Least Squares)

1. 方法思路

- 联立方程模型的结构方程中包含有内生解释变量，不能直接采用OLS估计其参数。但是对于简化式方程，可以采用OLS直接估计其参数。
- 间接最小二乘法：先对关于内生解释变量的简化式方程采用OLS估计简化式参数，得到简化式参数估计量，然后通过参数关系体系，计算得到结构式参数的估计量。
- 间接最小二乘法只适用于恰好识别的结构方程的参数估计，因为只有恰好识别的结构方程，才能从参数关系体系中得到唯一一组结构参数的估计量。

2.一般间接最小二乘法的估计过程

$$Y_1 = (\boldsymbol{Y}_0, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_0 \\ \Gamma_0 \end{pmatrix} + \mathrm{N}_1$$

$$\downarrow$$

$$Y_1 - \mathrm{B}_0 \mathbf{Y}_0 - \Gamma_0 \mathbf{X}_0 = \mathrm{N}_1$$

$$\downarrow$$

$$\begin{pmatrix} 1 & -\mathrm{B}_{\mathbf{0}} & -\Gamma_0 \end{pmatrix}\begin{pmatrix} Y_1 \\ \mathbf{Y}_0 \\ \mathbf{X}_0 \end{pmatrix} = \mathrm{N}_1$$

$$\downarrow$$

$$\begin{pmatrix} \mathrm{B}_{00} & \Gamma_{00} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{00} \\ \mathbf{X}_{0} \end{pmatrix} = \mathrm{N}_1$$

$$\mathbf{Y}_{00} = \Pi_{00}\mathbf{X} + \mathrm{E}$$

$$\mathrm{B}_{00}\Pi_{00}\mathbf{X} + \Gamma_{00}\mathbf{X}_0 = 0$$

$$\mathrm{B}_{00}\Pi_{00}\begin{pmatrix} \mathbf{X}_{0} \\ \mathbf{X}_{0}^{*} \end{pmatrix} + \Gamma_{00}\mathbf{X}_0 = 0$$

$$\Pi_{00} = \begin{pmatrix} \Pi_{00}^{1} & \Pi_{00}^{2} \end{pmatrix}$$

$$\begin{cases} \mathrm{B}_{00}\Pi_{00}^{1} = \Gamma_{0} \\ \mathrm{B}_{00}\Pi_{00}^{2} = 0 \end{cases}$$

- 用**OLS**估计简化式模型，得到简化式参数估计量，代入该参数关系体系，先由第**2**组方程计算得到内生解释变量的参数，然后再代入第**1**组方程计算得到先决解释变量的参数。于是得到了结构方程的所有结构参数估计量。

3. 间接最小二乘法也是一种工具变量方法

- **ILS等价于一种工具变量方法：依次选择X作为（Y_0,X_0）的工具变量。**
- 数学证明见《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年3月）第126—128页。
- 估计结果为：

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{ILS} = \left(\mathbf{X}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \mathbf{X}' Y_1$$

三、二阶段最小二乘法
(2SLS, Two Stage Least Squares)

1. 2SLS是应用最多的单方程估计方法

- IV和ILS一般只适用于联立方程模型中恰好识别的结构方程的估计。
- 在实际的联立方程模型中，恰好识别的结构方程很少出现，一般情况下结构方程都是过度识别的。为什么？
- 2SLS是一种既适用于恰好识别的结构方程，又适用于过度识别的结构方程的单方程估计方法。

2. 2SLS的方法步骤

- 第一阶段：对内生解释变量的简化式方程使用OLS。得到：

$$\hat{\boldsymbol{Y}}_0 = \boldsymbol{X}\hat{\Pi}_0 = \boldsymbol{X}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_0)$$

- 用估计量代替结构方程中的内生解释变量，得到新的模型：

$$Y_1 = (\hat{\boldsymbol{Y}}_0, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_0 \\ \Gamma_0 \end{pmatrix} + \mathrm{N}_1$$

- 第二阶段：对该模型应用**OLS**估计，得到的参数估计量即为原结构方程参数的二阶段最小二乘估计量。

$$\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{2SLS} = \left(\begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' \begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' Y_1$$

3. 二阶段最小二乘法也是一种工具变量方法

- 如果用Y_0的估计量作为工具变量，按照工具变量方法的估计过程，应该得到如下的结构参数估计量：

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix} = \left(\begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' \begin{pmatrix} \boldsymbol{Y}_0 & \boldsymbol{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' Y_1$$

- 可以严格证明两组参数估计量是完全等价的，所以可以把2SLS也看成为一种工具变量方法。
- 证明过程见《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年3月）第130—131页。

四、三种方法的等价性证明

1. 三种单方程估计方法得到的参数估计量

$$\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{IV} = \left(\begin{pmatrix} \boldsymbol{X}_0^* & \boldsymbol{X}_0 \end{pmatrix}' \begin{pmatrix} \boldsymbol{Y}_0 & \boldsymbol{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \boldsymbol{X}_0^* & \boldsymbol{X}_0 \end{pmatrix}' Y_1$$

$$\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{ILS} = \left(\mathbf{X}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \mathbf{X}' Y_1$$

$$\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{2SLS} = \left(\begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' \begin{pmatrix} \boldsymbol{Y}_0 & \boldsymbol{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0 \end{pmatrix}' Y_1$$

2. IV与ILS估计量的等价性

- 在恰好识别情况下
- 工具变量集合相同，只是次序不同。
- 次序不同不影响正规方程组的解。

2. 2SLS与ILS估计量的等价性

- 在恰好识别情况下
- ILS的工具变量是全体先决变量。
- 2SLS的每个工具变量都是全体先决变量的线性组合。
- 2SLS的正规方程组相当于ILS的正规方程组经过一系列的初等变换的结果。
- 线性代数方程组经过初等变换不影响方程组的解。

五、简单宏观经济模型实例演示

1. 模型

$$
\begin{cases}
C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t} \\
I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\
Y_t = I_t + C_t + G_t
\end{cases}
$$

- 消费方程是恰好识别的；
- 投资方程是过度识别的；
- 模型是可以识别的。

2. 数据

年份	Y	I	C	G
1978	3606	1378	1759	469
1979	4074	1474	2005	595
1980	4551	1590	2317	644
1981	4901	1581	2604	716
1982	5489	1760	2868	861
1983	6076	2005	3182	889
1984	7164	2469	3675	1020
1985	8792	3386	4589	817
1986	10133	3846	5175	1112
1987	11784	4322	5961	1501
1988	14704	5495	7633	1576
1989	16466	6095	8524	1847
1990	18320	6444	9113	2763
1991	21280	7517	10316	3447
1992	25864	9636	12460	3768
1993	34501	14998	15682	3821
1994	47111	19261	21230	6620
1995	59405	23877	27839	7689
1996	68498	26867	32589	9042

3. 用狭义的工具变量法估计消费方程

用G_t作为Y_t的工具变量

$$\hat{\alpha}_0 = 164.79951$$

$$\hat{\alpha}_1 = 0.3175387$$

$$\hat{\alpha}_2 = 0.3919359$$

- 估计结果显示

Dependent Variable: CC
Method: Two-Stage Least Squares
Date: 04/11/03 Time: 22:06
Sample(adjusted): 1979 1996
Included observations: 18 after adjusting endpoints
Instrument list: C G CC1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	164.8004	95.45182	1.726529	0.1048
Y	0.317539	0.032376	9.807786	0.0000
CC1	0.391935	0.087514	4.478510	0.0004

R-squared	0.999435	Mean dependent var	9875.667
Adjusted R-squared	0.999360	S.D. dependent var	9026.792
S.E. of regression	228.3835	Sum squared resid	782385.2
F-statistic	13200.10	Durbin-Watson stat	2.015655
Prob(F-statistic)	0.000000		

4. 用间接最小二乘法估计消费方程

$$\begin{cases} C_t = \pi_{10} + \pi_{11} C_{t-1} + \pi_{12} G_t + \varepsilon_{1t} \\ Y_t = \pi_{20} + \pi_{21} C_{t-1} + \pi_{22} G_t + \varepsilon_{2t} \end{cases}$$

$$\hat{\pi}_{10} = -63.594002$$
$$\hat{\pi}_{11} = 0.8132890$$
$$\hat{\pi}_{12} = 1.2191863$$

$$\hat{\pi}_{20} = -719.26343$$
$$\hat{\pi}_{21} = 1.3269366$$
$$\hat{\pi}_{22} = 3.8394822$$

$$\hat{\alpha}_1 = \hat{\pi}_{12} / \hat{\pi}_{22} = 0.31753925$$
$$\hat{\alpha}_2 = \hat{\pi}_{11} - \hat{\alpha}_1 \hat{\pi}_{21} = 0.39193422$$
$$\hat{\alpha}_0 = \hat{\pi}_{10} - \hat{\alpha}_1 \hat{\pi}_{20} = 164.800368$$

- **C简化式模型估计结果**

Dependent Variable: CC

Method: Least Squares

Date: 04/11/03 Time: 22:13

Sample(adjusted): 1979 1996

Included observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-63.59400	279.1279	-0.227831	0.8229
CC1	0.813289	0.145306	5.597062	0.0001
G	1.219186	0.402482	3.029167	0.0085

R-squared	0.994079	Mean dependent var	9875.667
Adjusted R-squared	0.993289	S.D. dependent var	9026.792
S.E. of regression	739.4562	Akaike info criterion	16.20072
Sum squared resid	8201931.	Schwarz criterion	16.34911
Log likelihood	-142.8065	F-statistic	1259.163
Durbin-Watson stat	1.542608	Prob(F-statistic)	0.000000

- **Y简化式模型估计结果**

Dependent Variable: Y
Method: Least Squares
Date: 04/11/03 Time: 22:17
Sample(adjusted): 1979 1996
Included observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-719.2634	740.2944	-0.971591	0.3467
CC1	1.326937	0.385377	3.443215	0.0036
G	3.839482	1.067451	3.596869	0.0026

R-squared	0.991131	Mean dependent var	20506.28
Adjusted R-squared	0.989948	S.D. dependent var	19561.13
S.E. of regression	1961.163	Akaike info criterion	18.15147
Sum squared resid	57692390	Schwarz criterion	18.29987
Log likelihood	-160.3633	F-statistic	838.1285
Durbin-Watson stat	1.427616	Prob(F-statistic)	0.000000

5. 用两阶段最小二乘法估计消费方程

$$\hat{Y}_t = -719.26343 + 1.3269366 C_{t-1} + 3.8394822 G_t$$

*代替原消费方程中的**Y**$_t$，应用**OLS**估计*

$$\hat{\alpha}_0 = 164.90009$$
$$\hat{\alpha}_1 = 0.3175580$$
$$\hat{\alpha}_2 = 0.3918794$$

- 比较上述消费方程的3种估计结果，证明这3种方法对于恰好识别的结构方程是等价的。估计量的差别只是很小的计算误差。

- # 第2阶段估计结果

Dependent Variable: CC
Method: Least Squares
Date: 04/11/03 Time: 22:22
Sample(adjusted): 1979 1996
Included observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	164.8004	309.0523	0.533244	0.6017
YF	0.317539	0.104827	3.029167	0.0085
CC1	0.391935	0.283353	1.383203	0.1868

R-squared	0.994079	Mean dependent var	9875.667
Adjusted R-squared	0.993289	S.D. dependent var	9026.792
S.E. of regression	739.4562	Akaike info criterion	16.20072
Sum squared resid	8201931.	Schwarz criterion	16.34911
Log likelihood	-142.8065	F-statistic	1259.163
Durbin-Watson stat	1.542608	Prob(F-statistic)	0.000000

6. 用两阶段最小二乘法估计投资方程

- 投资方程是过度识别的结构方程，只能用**2SLS**估计。估计过程与上述**2SLS**估计消费方程的过程相同。得到投资方程的参数估计量为：

$$\hat{\beta}_0 = -380.11614$$
$$\hat{\beta}_1 = 0.4049326$$

- 至此，完成了该模型系统的估计。

- **2SLS第2阶段估计结果**

Dependent Variable: I

Method: Least Squares

Date: 04/11/03 Time: 22:28

Sample(adjusted): 1979 1996

Included observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-380.2044	427.6175	-0.889123	0.3871
YF	0.404935	0.015324	26.42468	0.0000

R-squared	0.977599	Mean dependent var	7923.500
Adjusted R-squared	0.976199	S.D. dependent var	7975.613
S.E. of regression	1230.436	Akaike info criterion	17.17256
Sum squared resid	24223582	Schwarz criterion	17.27149
Log likelihood	-152.5531	F-statistic	698.2639
Durbin-Watson stat	1.376531	Prob(F-statistic)	0.000000

7. 用GMM估计投资方程

- 投资方程是过度识别的结构方程，也可以用GMM估计。选择的工具变量为c、G、CC1,得到投资方程的参数估计量为：

$$\hat{\beta}_0 = -388.2216$$

$$\hat{\beta}_1 = 0.405241$$

- 与**2SLS**结果比较，结构参数估计量变化不大。残差平方和由**24223582**变为**3832486**，显著减少。为什么？利用了更多的信息。

- **GMM估计结果**

Dependent Variable: I
Method: Generalized Method of Moments
Date: 04/11/03 Time: 22:33
Sample(adjusted): 1979 1996
Included observations: 18 after adjusting endpoints
No prewhitening
Bandwidth: Fixed (2)
Kernel: Bartlett
Convergence achieved after: 2 weight matricies, 3 total coef iterations
Instrument list: C G CC1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-388.2216	82.86703	-4.684874	0.0002
Y	0.405241	0.004748	85.34159	0.0000

R-squared	0.996456	Mean dependent var	7923.500
Adjusted R-squared	0.996234	S.D. dependent var	7975.613
S.E. of regression	489.4184	Sum squared resid	3832486.
Durbin-Watson stat	1.357784	J-statistic	0.002874

六、主分量法的应用

1. 方法的提出

- 主分量方法本身并不是联立方程模型的估计方法，而是配合其它方法，例如**2SLS**使用于模型的估计过程之中。
- 数学上的主分量方法早就成熟，**Kloek**和**Mennes**于**1960**年提出将它用于计量经济学模型的估计。
- **2SLS**是一种普遍适用的联立方程模型的单方程估计方法，但是当它在实际模型估计中被应用时，立刻就会遇到不可逾越的困难。其第一阶段—用**OLS**估计简化式方程，是难以实现的。为什么？

2. 方法的原理

- 所谓主分量方法，就是用较少数目的新变量重新表示原模型中较多数目的先决变量的方法。
- 例如，如果能够找到5个左右的新变量表示宏观经济模型中的30个先决变量，那么只需要15组以上的样本，就可以进行2SLS第一阶段的估计。
- 对充当主分量的变量是有严格要求：一是它必须是先决变量的线性组合，二是它们之间必须是正交的。前一条是保证主分量对先决变量的代表性；后一条是保证主分量之间不出现共线性

3. 主分量的选取

- 用两个主分量表示两个原变量

$$Z_1 = a_{11}X_1 + a_{12}X_2$$

$$Z_2 = a_{21}X_1 + a_{22}X_2$$

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

可以证明，a_1、a_2分别是X'X的2个特征值对应的特征向量。

- 用k个主分量表示k个原变量

$$\mathbf{Z} = \mathbf{X}\mathbf{A}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_k \end{pmatrix}$$

同样可以证明，$\mathbf{a}_1$、$\mathbf{a}_2$、…、$\mathbf{a}_k$分别是X’X的k个特征值对应的特征向量。

- 用f个主分量表示k个原变量

$$\mathbf{Z} = \mathbf{X}\mathbf{A}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_f \end{pmatrix}$$

选择$\mathbf{a}_1$、$\mathbf{a}_2$、…、$\mathbf{a}_f$分别是X'X的f个最大特征值对应的特征向量。

- **在2SLS中主分量的选取**

  对于简化式方程

$$\boldsymbol{Y}_0 = \boldsymbol{X}\Pi_0 + \mathrm{E}_0 = \begin{pmatrix} \boldsymbol{X}_0 & \boldsymbol{X}_0^* \end{pmatrix}\Pi_0 + \mathrm{E}_0$$

一般情况下，结构方程包含的先决解释变量$\mathbf{x}_0$中变量的数目很有限，变量主要集中在结构方程未包含的先决变量$\mathbf{x}_0^*$中。所以只需要选择主分量重新表示$\mathbf{x}_0^*$，就可以有效地减少简化式方程中解释变量的数目，使得在有限样本的支持下模型得到估计。

4. 主分量法在ILS中的应用

- 对于2SLS，直接利用主分量完成第一阶段的估计，得到内生解释变量的估计量。
- 对于ILS，必须求得到简化式参数，进而计算结构式参数。
- 首先估计$Y=Z\Delta+E$，然后将$Z=XA$代入，得到$Y=X\Pi$ 中Π的估计量。

七、其它有限信息估计方法简介
(Limited Information Estimation Methods)

1. 有限信息最大或然法（LIML，**Limited Information Maximum Likelihood**）

- 以最大或然为准则、通过对简化式模型进行最大或然估计，以得到结构方程参数估计量的联立方程模型的单方程估计方法。
- 由**Anderson**和**Rubin**于**1949**年提出，早于两阶段最小二乘法。
- 适用于恰好识别和过度识别结构方程的估计。

- 在该方法中，以下两个概念是重要的：

  一是这里的“有限信息”指的是每次估计只考虑一个结构方程的信息，而没有考虑模型系统中其它结构方程的信息；

  二是这里的“最大或然法”是针对结构方程中包含的内生变量的简化式模型的，即应用最大或然法求得的是简化式参数估计量，而不是结构式参数估计量。

- 具体参见教科书。

2.有限信息最小方差比方法(LVR，**Least Variable Ratio**）

- 估计某一个结构方程参数时，仍然只利用关于该方程的信息，没有利用方程系统的信息，所以是一种有限信息估计方法。
- 参见教科书。

八、k级估计式

1. k级估计式

- 本身不是一种估计方法，而是对上述几种方法得到的估计式的概括。
- 对于联立方程模型中的第1个结构方程：

$$Y_1 = (\boldsymbol{Y}_0, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_{\boldsymbol{0}} \\ \Gamma_{\boldsymbol{0}} \end{pmatrix} + \mathrm{N}_1$$

- **k**级估计式 为：

$$\boxed{\begin{bmatrix} \hat{\mathrm{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix} = ((\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0), \boldsymbol{X}_0)'(\boldsymbol{Y}_0, \boldsymbol{X}_0))^{-1}(\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0), \boldsymbol{X}_0)'Y_1}$$

- 显然，当

  k=0时，即为**OLS**估计式；

  k=1时，即为**2SLS**估计式；

  k等于有限信息估计方法中的时，即为有限信息估计式。

2. k级估计式的性质

- 假设工具变量与随机误差项不相关，即

$$P\lim \tfrac{1}{n}(\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0))\mathrm{N}_1 = 0$$

且先决变量与随机误差项不相关，即

$$P\lim \tfrac{1}{n}(\boldsymbol{X}_0'\mathrm{N}_1) = 0$$

那么，容易证明**k**级估计式是一致性估计式。

- 工具变量与随机误差项不相关，对k是有限制的，必须有（证明见教科书）：

$$P\lim(1-k)=0$$

- 这就是说，只有在**2SLS**或有限信息估计方法中，**k**级估计式是一致性估计式，而在**OLS**方法中，不具有一致性。

§6.7联立方程计量经济学模型的系统估计方法

the Systems Estimation Methods

一、联立方程模型随机误差项方差—协方差矩阵

二、三阶段最小二乘法简介

三、完全信息最大似然法简介

一、联立方程模型随机误差项方差—协方差矩阵

1. 随机误差项的同期相关性

- 随机误差项的相关性不仅存在于每个结构方程不同样本点之间，而且存在于不同结构方程之间。
- 对于不同结构方程的随机误差项之间，不同时期互不相关，只有同期的随机误差项之间才相关，称为具有同期相关性。

2. 具有同期相关性的方差—协方差矩阵

$$\mathbf{Y}\mathrm{B} + \mathbf{X}\Gamma = \mathrm{N} \longrightarrow \mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathrm{N}}$$

$$Y_i = \mathbf{Z}_i \Delta_i + \tilde{\mathrm{N}}_i$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{bmatrix} \qquad Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{bmatrix}$$

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_0^i & \mathbf{X}_0^i \end{pmatrix}$$

$$\Delta_i = \begin{pmatrix} \mathrm{B}_0^i \\ \Gamma_0^i \end{pmatrix}$$

假设：

- 对于一个结构方程的随机误差项，在不同样本点之间，具有同方差性和序列不相关性。即

$$Cov(\tilde{\mathrm{N}}_i) = \sigma_{ii}^2 \mathbf{I}$$

- 对于不同结构方程的随机误差项之间，具有且仅具有同期相关性。即

$$Cov(\tilde{\mathrm{N}}_i, \tilde{\mathrm{N}}_j) = \sigma_{ij} \mathbf{I}$$

于是，联立方程模型系统随机误差项方差—协方差矩阵为：

$$Cov(\tilde{N}) = \begin{bmatrix} \sigma_{11}^2\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1g}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}^2\mathbf{I} & \cdots & \sigma_{2g}\mathbf{I} \\ \vdots & & & \\ \sigma_{g1}\mathbf{I} & \sigma_{g2}\mathbf{I} & & \sigma_{gg}^2\mathbf{I} \end{bmatrix}$$

$$= \Sigma \otimes \mathbf{I}$$

$$= \Omega$$

二、三阶段最小二乘法简介
(3SLS, Three Stages Least Squares)

1. 概念

- 3SLS是由Zellner和Theil于1962年提出的同时估计联立方程模型全部结构方程的系统估计方法。
- 其基本思路是 3SLS=2SLS+GLS

  即首先用2SLS估计模型系统中每一个结构方程，然后再用GLS估计模型系统。

2. 三阶段最小二乘法的步骤

（1）用2SLS估计结构方程

$$\boxed{Y_i = \mathbf{Z}_i \Delta_i + \tilde{N}_i}$$

得到方程随机误差项的估计值。

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_0^i & \mathbf{X}_0^i \end{pmatrix}$$

$$\boldsymbol{Y}_0^i = \boldsymbol{X}\Pi_0^i + \mathrm{E}_0^i$$

OLS估计

$$\hat{\boldsymbol{Y}}_0^i = \boldsymbol{X}\hat{\Pi}_0^i = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_0^i$$

$$\hat{\boldsymbol{Z}}_i = \begin{pmatrix} \hat{\boldsymbol{Y}}_0^i & \boldsymbol{X}_0^i \end{pmatrix}$$

OLS估计

$$\hat{\Delta}_i = (\hat{\mathbf{Z}}_i'\hat{\mathbf{Z}}_i)^{-1}\hat{\mathbf{Z}}_i'Y_i$$

$$\hat{\hat{Y}}_i = \hat{\boldsymbol{Z}}_i\hat{\Delta}_i$$

$$e_{il} = y_{il} - \hat{\hat{y}}_{il}$$

（2）求随机误差项方差—协方差矩阵的估计量

$$\boldsymbol{e}_i = \begin{pmatrix} e_{i1} & e_{i2} & \cdots & e_{in} \end{pmatrix}'$$

$$\hat{\sigma}_{ij} = \frac{\boldsymbol{e}_i'\boldsymbol{e}_j}{\sqrt{(n-g_i+1-k_i)(n-g_j+1-k_j)}}$$

$$\hat{\Sigma} = (\hat{\sigma}_{ij})$$

$$\boxed{\hat{\Sigma} \otimes \boldsymbol{I} = \hat{\Omega}}$$

(3) 用GLS估计原模型系统

$$\boxed{\mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathrm{N}}}$$

得到结构参数的3SLS估计量为：

$$\boxed{\begin{aligned}\hat{\hat{\Delta}} &= (\hat{Z}'\hat{\Omega}^{-1}\hat{Z})^{-1}\hat{Z}'\hat{\Omega}^{-1}\boldsymbol{Y} \\ &= (\hat{Z}'(\hat{\Sigma}\otimes \boldsymbol{I})^{-1}\hat{Z})^{-1}\hat{Z}'(\hat{\Sigma}\otimes \boldsymbol{I})^{-1}\boldsymbol{Y}\end{aligned}}$$

3. 三阶段最小二乘法估计量的统计性质

(1)如果联立方程模型系统中所有结构方程都是可以识别的，并且非奇异，则3SLS估计量是一致性估计量。

(2) 3SLS估计量比2SLS估计量更有效。为什么？

(3)如果Σ是对角矩阵，即模型系统中不同结构方程的随机误差项之间无相关性，那么可以证明3SLS估计量与2SLS估计量是等价的。

(4)这反过来说明，3SLS方法主要优点是考虑了模型系统中不同结构方程的随机误差项之间的相关性。

三、完全信息最大似然法简介

（FIML, Full Information Maximum Likelihood）

1. 概念

- 另一种已有实际应用的联立方程模型的系统估计方法。
- **Rothenberg和Leenders于1964年提出一个线性化的FIML估计量。**
- **FIML是ML的直接推广，是在已经得到样本观测值的情况下，使整个联立方程模型系统的或然函数达到最大以得到所有结构参数的估计量。**

2.复习：多元线性单方程模型的最大似然估计

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i$$

i=1,2,…,n

$$\boldsymbol{Y} = \boldsymbol{X}\mathrm{B} + \mathrm{N}$$

- Y的随机抽取的**n**组样本观测值的联合概率

$$L(\hat{\mathrm{B}},\sigma_\mu^2)=P(y_1,y_2,\cdots,y_n)$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma_\mu^n}e^{-\frac{1}{2\sigma_\mu^2}\Sigma(y_i-(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\cdots+\hat{\beta}_kx_{ki}))^2}$$

$$=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma_\mu^n}e^{-\frac{1}{2\sigma_\mu^2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\mathrm{B}})'(\boldsymbol{Y}-\boldsymbol{X}\hat{\mathrm{B}})}$$

- 对数或然函数为

$$\boxed{\begin{aligned} L^* &= Ln(L) \\ &= -nLn(\sqrt{2\pi}\sigma_\mu) - \frac{1}{2\sigma_\mu^2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\mathrm{B}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\mathrm{B}}) \end{aligned}}$$

- 参数的最大或然估计

$$\hat{\mathrm{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

3.复习：有限信息最大或然法(LIML，**Limited Information Maximum Likelihood**）

- 以最大或然为准则、通过对简化式模型进行最大或然估计，以得到结构方程参数估计量的联立方程模型的单方程估计方法。
- 由**Anderson**和**Rubin**于**1949**年提出，早于两阶段最小二乘法。
- 适用于恰好识别和过度识别结构方程的估计。

- 在该方法中，以下两个概念是重要的：

  一是这里的“有限信息”指的是每次估计只考虑一个结构方程的信息，而没有考虑模型系统中其它结构方程的信息；

  二是这里的“最大或然法”是针对结构方程中包含的内生变量的简化式模型的，即应用最大或然法求得的是简化式参数估计量，而不是结构式参数估计量。

$$Y_1 = (\boldsymbol{Y}_0, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_{\boldsymbol{0}} \\ \Gamma_{\boldsymbol{0}} \end{pmatrix} + \mathrm{N}_1$$

$$(\boldsymbol{Y}_0^1, \boldsymbol{X}_0)\begin{pmatrix} \mathrm{B}_0^1 \\ \Gamma_{\boldsymbol{0}} \end{pmatrix} + \mathrm{N}_1 = 0$$

$$\boldsymbol{Y}_0^1 = \boldsymbol{X}\Pi_0^1 + \mathrm{E}_0^1$$

$$\boxed{LnL(\boldsymbol{Y}_0^1) = c + \tfrac{n}{2}\, Ln\left|\Omega_0^{-1}\right| - \tfrac{1}{2}\, tr\Omega_0^{-1}(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1)'(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1)}$$

4. 完全信息最大似然函数

$$\mathbf{Y} = \mathbf{Z\Delta} + \tilde{\mathrm{N}}$$

$\tilde{\mathrm{N}}$ ~ 正态（$\boldsymbol{0}$，$\Sigma \otimes \boldsymbol{I}$)

ML的直接推广

$$L(\boldsymbol{Y}) = \frac{1}{(2\pi)^{gn/2}\left|\Sigma \otimes \boldsymbol{I}\right|^{-\frac{1}{2}}}\left|\frac{\partial \tilde{\mathrm{N}}}{\partial \boldsymbol{Y}}\right| e^{-\frac{1}{2}(\boldsymbol{Y}-\mathbf{Z\Delta})'(\Sigma^{-1}\otimes \boldsymbol{I})(\boldsymbol{Y}-\mathbf{Z\Delta})}$$

$$= \frac{1}{(2\pi)^{gn/2}\left|\Sigma \otimes \mathbf{I}\right|^{-\frac{1}{2}}}\left|\mathrm{B}\right|^{n} e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{Z\Delta})'(\Sigma^{-1}\otimes \mathbf{I})(\mathbf{Y}-\mathbf{Z\Delta})}$$

$$L(\boldsymbol{Y})=\frac{1}{(2\pi)^{gn/2}\left|\Sigma\otimes\boldsymbol{I}\right|^{-\frac{1}{2}}}\left|\frac{\partial\tilde{\mathrm{N}}}{\partial\boldsymbol{Y}}\right|e^{-\frac{1}{2}(\boldsymbol{Y}-\mathbf{Z}\Delta)'(\Sigma^{-1}\otimes\boldsymbol{I})(\boldsymbol{Y}-\mathbf{Z}\Delta)}$$

$$-\frac{1}{2}(\mathbf{Y}-\mathbf{Z}\Delta)'(\Sigma^{-1}\otimes\mathbf{I})(\mathbf{Y}-\mathbf{Z}\Delta)$$

- 对数或然函数对于协方差逆矩阵的元素取极大值的一阶条件，得到协方差矩阵的元素的**FIML**估计量；
- 对数或然函数对于待估计参数取极大值的一阶条件，求解该方程系统，即可得到结构参数的**FIML**估计量。
- 研究的重点是如何求解非线性方程系统。

§6.8-9联立方程计量经济学模型的估计方法选择和模型检验

一、模型估计方法的比较

二、为什么普通最小二乘法被普遍采用

三、模型的检验

一、模型估计方法的比较

1. 大样本估计特性的比较

- 在大样本的情况下，各种参数估计方法的统计特性可以从数学上进行严格的证明，因而也可以将各种方法按照各个性质比较优劣。
- 按渐近无偏性比较优劣

除了**OLS**方法外，所有方法的参数估计量都具有大样本下渐近无偏性。因而，除了**OLS**方法最差外，其它方法无法比较优劣。

- 按渐近有效性比较优劣

OLS 非一致性估计，未利用任何单方程外的信息；

IV 利用了模型系统部分先决变量的数据信息；

2SLS、LIML 利用了模型系统全部先决变量的数据信息；

3SLS、FIML 利用了模型系统全部先决变量的数据信息和结构方程相关性信息。

2.小样本估计特性的Monte Carlo试验

- 参数估计量的大样本特性只是理论上的，实际上并没有“大样本”，所以，对小样本估计特性进行比较更有实际意义。
- 而在小样本的情况下，各种参数估计方法的统计特性无法从数学上进行严格的证明，因而提出了一种**Monte Carlo**试验方法。
- **Monte Carlo**试验方法在经济实验中被广泛采用。

- 小样本估计特性的**Monte Carlo**试验过程

第一步：利用随机数发生器产生随机项分布的一组样本；

第二步：代入已经知道结构参数和先决变量观测值的结构模型中；

第三步：计算内生变量的样本观测值；

第四步：选用各种估计方法估计模型的结构参数。

上述步骤反复进行数百次，得到每一种估计方法的参数估计值的序列。

第五步：对每种估计方法的参数估计值序列进行统计分析；

第六步：与真实参数（即试验前已经知道的结构参数）进行比较，以判断各种估计方法的优劣。

- 小样本估计特性实验结果比较

(1)无偏性

OLS 2SLS 3SLS（LIML，FIML）

(2)最小方差性

LIML 2SLS FIML OLS

(3)最小均方差性

OLS LIML 2SLS 3SLS（FIML）

为什么OLS具有最好的最小方差性?

方差的计算公式:

$$V=\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_i-\overline{\hat{\beta}})^2$$

均方差的计算公式:

$$MSE=E(\hat{\beta}-\beta)^2=\frac{1}{n}\sum_{i=1}^{N}(\hat{\beta}_i-\beta)^2$$

前者反映估计量偏离实验均值的程度;后者反映估计量偏离真实值的程度。所以尽管OLS具有最小方差性,但是由于它是有偏的,偏离真实值最为严重,所以它的最小均方差性仍然是最差的。

二、为什么普通最小二乘法被普遍采用

1. 小样本特性

- 从理论上讲，在小样本情况下，各种估计方法的估计量都是有偏的。

2. 充分利用样本数据信息

- 除OLS之外的其它估计方法可以部分地或者全部地利用某个结构方程中未包含的先决变量的数据信息，从而提高参数估计量的统计性质。但是其前提是所有变量具有相同的样本容量。
- 在实际上变量经常不具有相同的样本容量。
- 采用先进估计方法所付出的代价经常是牺牲了该方程所包含的变量的样本数据信息。

3. 确定性误差传递

- 确定性误差：结构方程的关系误差和外生变量的观测误差。
- 采用OLS方法，当估计某一个结构方程时，方程中没有包含的外生变量的观测误差和其它结构方程的关系误差对该方程的估计结果没有影响。
- 如果采用2SLS方法 …
- 如果采用3SLS方法…

4. 样本容量不支持

- 实际的联立方程模型中每个结构方程往往是过度识别的，适宜采用2SLS或3SLS方法，但是在其第一阶段要以所有先决变量作为解释变量，这就需要很大容量的样本。实际上是难以实现的。
- 采用主分量方法等可以克服这个矛盾，但又带来方法的复杂性和新的误差。

5. 实际模型的递推（Recurred）结构

- 应用中的联立方程模型主要是宏观经济计量模型。
- 宏观经济计量模型一般具有递推结构。
- 具有递推结构的模型可以采用OLS。

补充：递推模型（Recursive Model ）

$$\mathrm{B}\boldsymbol{Y}+\Gamma\boldsymbol{X}=\mathrm{N}$$

$$\mathrm{B}=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\beta_{21} & 1 & 0 & \cdots & 0 \\ -\beta_{31} & -\beta_{32} & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ -\beta_{g1} & -\beta_{g2} & -\beta_{g3} & \cdots & 1 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -\gamma_{11} & -\gamma_{12} & \cdots & -\gamma_{1k} \\ -\gamma_{21} & -\gamma_{22} & \cdots & -\gamma_{2k} \\ \vdots & & & \\ -\gamma_{g1} & -\gamma_{g2} & \cdots & -\gamma_{gk} \end{bmatrix}$$

- 可以采用**OLS**依次估计每个结构方程；
- 在估计后面的结构方程时，认为其中的内生解释变量是“先决”的。

三、模型的检验

- 包括单方程检验和方程系统的检验。
- 凡是在单方程模型中必须进行的各项检验，对于联立方程模型中的结构方程，以及应用2SLS或3SLS方法过程中的简化式方程，都是适用的和需要的。
- 模型系统的检验主要包括:

1. 拟合效果检验

- 将样本期的先决变量观测值代入估计后的模型，求解该模型系统，得到内生变量的估计值。将估计值与实际观测值进行比较，据此判断模型系统的拟合效果。
- 模型的求解方法：迭代法。为什么不直接求解?
- 常用的判断模型系统拟合效果的检验统计量是“均方百分比误差”，用**RMS**表示。

$$RMS_i = \sqrt{\sum_{t=1}^{n} e_{it}^2 / n} \qquad e_{it} = (y_{it} - \hat{y}_{it}) / y_{it}$$

- 当**RMS_i=0**，表示第**i**个内生变量估计值与观测值完全拟合。
- 一般地，在g个内生变量中，RMS<5%的变量数目占70%以上，并且每个变量的RMS不大于10%，则认为模型系统总体拟合效果较好。

2. 预测性能检验

- 如果样本期之外的某个时间截面上的内生变量实际观测值已经知道，这就有条件对模型系统进行预测检验。
- 将该时间截面上的先决变量实际观测值代入模型，计算所有内生变量预测值，并计算其相对误差。

$$RE = (y_{i0} - \hat{y}_{i0})/y_{i0}$$

- 一般认为，RE<5%的变量数目占70%以上，并且每个变量的相对误差不大于10%，则认为模型系统总体预测性能较好。

3. 方程间误差传递检验

- 寻找模型中描述主要经济行为主体的经济活动过程的、方程之间存在明显的递推关系的关键路径。
- 在关键路径上进行误差传递分析，可以检验总体模型的模拟优度和预测精度。
- 例如，计算：

$$\left(\sum_{i=2}^{T}(e_i - e_{i-1})^2 \Big/ \sum_{i=1}^{T} e_i^2\right)\frac{T}{T-1}$$

- 称为冯诺曼比，如果误差在方程之间没有传递，该比值为0。

4. 样本点间误差传递检验

- 在联立方程模型系统中，由于经济系统的动态性，决定了有一定数量的滞后内生变量。
- 由于滞后内生变量的存在，使得模型预测误差不仅在方程之间传递，而且在不同的时间截面之间，即样本点之间传递。
- 必须对模型进行滚动预测检验。

- 给定t=1时的所有先决变量的观测值，包括滞后内生变量，求解方程组，得到内生变量Y_1的预测值；
- 对于t=2，只外生给定外生变量的观测值，滞后内生变量则以前一时期的预测值代替，求解方程组，得到内生变量Y_2的预测值；
- 逐年滚动预测，直至得到t=n时的内生变量Y_n的预测值；
- 求出该滚动预测值与实际观测值的相对误差。

- 将**t=n**时的所有先决变量的观测值，包括滞后内生变量的实际观测值，代入模型，求解方程组，得到内生变量Y_n的非滚动预测值；
- 求出该非滚动预测值与实际观测值的相对误差。
- 比较两种结果，二者的差异表明模型预测误差在不同的时间截面之间的传递。

教学基本要求

本章是课程的重点内容之一。通过教学，要求达到：

- 了解（最低要求）：常用的生产函数模型、需求函数模型、消费函数模型的理论模型和估计方法；在中国建立与应用生产函数模型、需求函数模型、消费函数模型过程中实际问题的处理。
- 掌握（较高要求）：常用的生产函数模型、需求函数模型、消费函数模型的理论模型是如何提出与发展的；在实践中自己提出与发展新的模型的方法论基础；其它常用的单方程模型，例如投资函数模型和货币需求函数模型的建模思路。

- 应用（对应用能力的要求）：分别选择一个研究对象，建立中国的实际模型。例如某个行业的生产函数模型、某种商品的需求函数模型、某类消费者的消费函数模型。

§7.1　生产函数模型(Production Function Models，P.F.)

- 几个重要概念
- 以要素之间替代性质的描述为线索的生产函数模型的发展
- 以技术要素的描述为线索的生产函数模型的发展
- 几个重要生产函数模型的参数估计方法
- 生产函数模型在技术进步分析中的应用
- 建立生产函数模型中的数据质量问题

一、几个重要概念

1. 生产函数

(1) 定义

- 描述生产过程中投入的生产要素的某种组合同它可能的最大产出量之间的依存关系的数学表达式。

$$Y = f(A, K, L, \cdots)$$

•投入的生产要素

•最大产出量

(2) 生产函数模型的发展

- 从20年代末，美国数学家Charles Cobb和经济学家Paul Dauglas提出了生产函数这一名词，并用1899-1922年的数据资料，导出了著名的Cobb-Dauglas生产函数。

- 1928年 Cobb, Dauglas C-D生产函数
  1937年 Dauglas,Durand C-D生产函数的改进型
  1957年 Solow C-D生产函数的改进型
  1960年 Solow 含体现型技术进步生产函数
  1967年 Arrow等 两要素CES生产函数
  1967年 Sato 二级CES生产函数

1968年 Sato, Hoffman VES生产函数

1968年 Aigner, Chu 边界生产函数

1971年 Revanker VES生产函数

1973年 Christensen, Jorgenson 超越对数生产函数

1980年 三级CES生产函数

(3) 生产函数是经验的产物

- 生产函数是在西方国家发展起来的，作为西方经济学理论体系的一部分，与特定的生产理论与环境相联系。

- 西方国家发展的生产函数模型可以被我们所应用：

  生产函数反应的是生产中投入要素与产出量之间的技术关系；

  生产函数模型的形式是经验的产物；

  不能照搬。

2. 要素产出弹性（Elasticity of Output）

(1) 要素的产出弹性

- 某投入要素的产出弹性被定义为，当其它投入要素不变时，该要素增加1%所引起的产出量的变化率。

$$E_K = \frac{\Delta Y}{Y} \Big/ \frac{\Delta K}{K} = \frac{\partial f}{\partial K} \Big/ \frac{K}{Y}$$

$$E_L = \frac{\Delta Y}{Y} \Big/ \frac{\Delta L}{L} = \frac{\partial f}{\partial L} \Big/ \frac{L}{Y}$$

- 要素产出弹性的数值区间？为什么？

(2) 规模报酬

- 所有要素的产出弹性之和
- 规模报酬不变
- 规模报酬递增
- 规模报酬递减
- 为什么经常将规模报酬不变作为生产函数必须满足的条件？

3. 要素替代弹性（Elasticity of Substitution）

(1) 要素的边际产量(Marginal Product)

- 其它条件不变时，某一种投入要素增加一个单位时导致的产出量的增加量。用于描述投入要素对产出量的影响程度。

$$MP_K = \partial f / \partial K$$

$$MP_L = \partial f / \partial L$$

$$\vdots$$

- 边际产量不为负。

$$MP_K \geq 0, MP_L \geq 0, \cdots$$

- 边际产量递减。

$$\frac{\partial(MP_K)}{\partial K} = \frac{\partial^2 f}{\partial K^2} \leq 0$$

$$\frac{\partial(MP_L)}{\partial L} = \frac{\partial^2 f}{\partial L^2} \leq 0$$

(2) 要素的边际替代率

(Marginal Rate of Substitution)

- 当两种要素可以互相替代时，就可以采用不同的要素组合生产相同数量的产出量。要素的边际替代率指的是在产量一定的情况下，某一种要素的增加与另一种要素的减少之间的比例。

$$MRS_{K\rightarrow L} = \Delta K / \Delta L$$

- 要素的边际替代率可以表示为要素的边际产量之比。

$$MRS_{K\to L} = MP_L / MP_K$$

$$MRS_{L\to K} = MP_K / MP_L$$

- 从生产函数可以求得要素的边际产量和要素的边际替代率。

(3) 要素替代弹性

- 要素替代弹性定义为两种要素的比例的变化率与边际替代率的变化率之比。

$$\sigma = \frac{d(K/L)}{(K/L)} \Bigg/ \frac{d(MP_L/MP_K)}{(MP_L/MP_K)}$$

- 要素替代弹性是描述生产行为的重要参数，求得要素替代弹性是生产函数的重要应用。
- 要素替代弹性不为负。
- 特殊情况：要素替代弹性为0、要素替代弹性为∞。

4. 技术进步

(1) 广义技术进步与狭义技术进步

- 所谓狭义技术进步，仅指要素质量的提高。
- 狭义的技术进步是体现在要素上的，它可以通过要素的“等价数量”来表示。
- 求得“等价数量”，作为生产函数模型的样本观测值，以这样的方法来引入技术进步因素。
- 所谓广义技术进步，除了要素质量的提高外，还包括管理水平的提高等对产出量具有重要影响的因素，这些因素是独立于要素之外的。
- 在生产函数模型中需要特别处理广义技术进步。

(2) 中性技术进步

- 假设在生产活动中除了技术以外，只有资本与劳动两种要素，定义两要素的产出弹性之比为相对资本密集度，用ω表示。即

$$\omega = E_L / E_K$$

- 如果技术进步使得ω越来越大，即劳动的产出弹性比资本的产出弹性增长得快，则称之为节约劳动型技术进步；如果技术进步使得ω越来越小，即劳动的产出弹性比资本的产出弹性增长得慢，则称之为节约资本型技术进步；如果技术进步前后ω不变，即劳动的产出弹性与资本的产出弹性同步增长，则称之为中性技术进步。
- 在中性技术进步中，如果要素之比不随时间变化，则称为希克斯中性技术进步；如果劳动产出率不随时间变化，则称为索洛中性技术进步；如果资本产出率不随时间变化，则称为哈罗德中性技术进步。

二、以要素之间替代性质的描述为线索的生产函数模型的发展

1. 线性生产函数模型(Linear P.F.)

$$Y = \alpha_0 + \alpha_1 K + \alpha_2 L$$

$$\sigma = \frac{d(K/L)}{(K/L)} \Big/ \frac{d(MP_L / MP_K)}{(MP_L / MP_K)}$$

$$\sigma = \infty$$

- 为什么?
- 如果选择线性生产函数，就意味着承认什么假设?

2. 投入产出生产函数模型(Input-Output P.F.)

$$Y = \min(\frac{K}{a}, \frac{L}{b})$$

$$\sigma = 0$$

- 为什么?
- 如果选择投入产出生产函数，就意味着承认什么假设?

3. C-D生产函数模型

$$Y = AK^{\alpha} L^{\beta}$$

$$E_K = \frac{\partial Y}{\partial K} \bigg/ \frac{K}{Y} = A\alpha K^{\alpha-1} L^{\beta} \frac{Y}{K} = \alpha$$

$$E_L = \frac{\partial Y}{\partial L} \bigg/ \frac{L}{Y} = AK^{\alpha} \beta L^{\beta-1} \frac{Y}{L} = \beta$$

$$\sigma = \frac{d(K/L)}{(K/L)} \Bigg/ \frac{d(MP_L / MP_K)}{(MP_L / MP_K)}$$

$$= d(\ln(\frac{K}{L})) \Bigg/ d(\ln(\frac{MP_L}{MP_K}))$$

$$= d(\ln(\frac{K}{L})) \Bigg/ d(\ln(\frac{\beta K}{\alpha L}))$$

$$= d(\ln(\frac{K}{L})) \Bigg/ d(\ln(\frac{\beta}{\alpha}) + \ln(\frac{K}{L}))$$

$$= 1$$

- 在C-D生产函数中要素的替代弹性是否随研究对象变化？是否合理？为什么？
- 在C-D生产函数中要素的替代弹性是否随样本区间变化？是否合理？为什么？
- 在C-D生产函数中要素的替代弹性是否随样本点变化？是否合理？为什么？
- C-D生产函数中每个参数的数值范围是什么？为什么？

4. CES生产函数模型(Constant Elasticity Of Substitution)

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m}{\rho}}$$

$$\sigma = \frac{d(K/L)}{(K/L)} \Bigg/ \frac{d(MP_L/MP_K)}{(MP_L/MP_K)}$$

$$= d(\ln(\frac{K}{L})) \Bigg/ d(\ln(\frac{MP_L}{MP_K}))$$

$$= \frac{1}{1+\rho}$$

- 替代弹性的推导过程？（独立推导一遍）
- 在CES生产函数中要素的替代弹性是否随研究对象变化？是否合理？为什么？
- 在CES生产函数中要素的替代弹性是否随样本区间变化？是否合理？为什么？
- 在CES生产函数中要素的替代弹性是否随样本点变化？是否合理？为什么？
- CES生产函数中每个参数的数值范围是什么？为什么？

5. VES生产函数模型(Variable Elasticity Of Substitution)

(1) **1968年Sato和Hoffman**

假定

$$\sigma = \sigma(t) = a + b \cdot t$$

得到

$$Y = B(\lambda L^{\frac{\sigma(t)-1}{\sigma(t)}} + (1-\lambda)K^{\frac{\sigma(t)-1}{\sigma(t)}})^{\frac{\sigma(t)}{\sigma(t)-1}}$$

- 与CES有什么联系与区别?

⑵ 1971年 Revankar

假定 $\sigma = a + b \cdot \dfrac{K}{L}$

$$Z = A\exp\int \frac{dk}{k + c\left(\dfrac{k}{a+bk}\right)^{1/a}}$$

其中 $Z = Y/L, k = K/L$

- **当b=0时，**

$$\frac{Y}{L} = A\exp\int\frac{dk}{k + c(\frac{k}{a})^{1/a}} = A\exp(\frac{a}{1-a}\ln\frac{k^{\frac{1-a}{a}}}{1+\frac{c}{a^{1/a}}k^{\frac{1-a}{a}}} + \mu)$$

令 $$\frac{1-a}{a} = \rho, Ae^{\mu} = A'$$

$$\frac{Y}{L} = A'(\frac{a^{1/a} + ck^{\rho}}{a^{1/a}k^{\rho}})^{-\frac{1}{\rho}} = A''(a^{1/a}k^{-\rho} + c)^{-\frac{1}{\rho}}$$

$$Y = A''(a^{1/a}(\frac{K}{L})^{-\rho} + c)^{-\frac{1}{\rho}} \cdot L$$

$$= A''(a^{1/a}K^{-\rho} + cL^{-\rho})^{-\frac{1}{\rho}}$$

- 退化为CES模型。为什么？

- 当b=0，a=1时，

$$\frac{Y}{L} = A\exp\int\frac{dk}{k(1+c)}$$

$$= A'\exp(\frac{\ln k}{1+c}) = A'k^{\frac{1}{1+c}}$$

$$Y = A'K^{\frac{1}{1+c}} \cdot L^{-\frac{1}{1+c}} \cdot L = A'K^{\frac{1}{1+c}} \cdot L^{\frac{c}{1+c}}$$

- 退化为C-D生产函数。为什么？

- **当a=1时，**

$$\sigma = 1 + bk$$

$$Y = AK^{\frac{1}{1+c}}\left(L + \left(\frac{b}{1+c}\right)K\right)^{\frac{c}{1+c}}$$

$$Y = AK^{\left(\frac{1}{1+c}\right)m}\left(L + \left(\frac{b}{1+c}\right)K\right)^{\left(\frac{c}{1+c}\right)m}$$

为实际应用的VES生产函数。

- 为什么是“变替代弹性”？

6. 超越对数生产函数模型（Translog P.F.）

$$\ln Y = \beta_0 + \beta_K \ln K + \beta_L \ln L + \beta_{KK}(\ln K)^2 + \beta_{LL}(\ln L)^2 + \beta_{KL} \ln K \cdot \ln L$$

- 如果 $\beta_{KK} = \beta_{LL} = \beta_{KL} = 0$，表现为何种时常函数?
- 如果 $\beta_{KK} = \beta_{LL} = -\frac{1}{2}\beta_{KL}$，表现为何种时常函数?

7. 多要素生产函数模型

(1) 多要素线性生产函数模型

$$Y = \alpha_0 + \alpha_1 K + \alpha_2 L + \alpha_3 E$$

(2) 多要素投入产出生产函数模型

$$Y = \min(\frac{K}{a}, \frac{L}{b}, \frac{E}{c})$$

(3) 多要素C-D生产函数模型

$$Y = AK^{\alpha} L^{\beta} E^{\gamma}$$

(4) 多要素一级CES生产函数模型

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho} + \delta_3 E^{-\rho})^{-\frac{m1}{\rho}}$$

• 要素之间的替代弹性是否相同？是多大？为什么？

(5) 多要素二级CES生产函数模型

$$Y_{KE} = (a_1 K^{-\rho_1} + a_2 E^{-\rho_1})^{-\frac{1}{\rho_1}}$$

$$Y = A(b_1 Y_{KE}^{-\rho} + b_2 L^{-\rho})^{-\frac{m}{\rho}}$$

• 要素之间的替代弹性是否相同？是多大？为什么？

(6)多要素三级CES生产函数模型

三、以技术进步的描述为线索的生产函数模型的发展

1. 将技术要素作为一个不变参数的生产函数模型

$$Y = AK^{\alpha}L^{\beta}$$

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m}{\rho}}$$

$$Z = A\exp\int \frac{dk}{k + c\left(\frac{k}{a+bk}\right)^{1/a}}$$

2. 改进的C-D生产函数模型

$$Y = A(t)K^{\alpha}L^{\beta}$$

$$Y = A_0(1+\gamma)^t K^{\alpha}L^{\beta}$$

$$Y = A_0 e^{\lambda t} K^{\alpha}L^{\beta}$$

- 参数的经济意义是什么?
- 关于技术进步的假设是什么? 为什么?

3. 改进的CES生产函数模型

$$Y = A_0(1+\gamma)^t(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m1}{\rho}}$$

$$Y = A_0 e^{\lambda t}(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m1}{\rho}}$$

- 关于技术进步的假设是什么？为什么？

4. 含体现型技术进步的生产函数模型

(1) 总量增长方程

$$\frac{\Delta Y}{Y} = \frac{\Delta A}{A} + \alpha \frac{\Delta K}{K} + \beta \frac{\Delta L}{L}$$

(2) 分离资本质量的含体现型技术进步的生产函数模型

$$Y_t = A_t' J_t^{\alpha} L_t^{\beta} \qquad J_t = \sum_{m=0}^{t} K_{mt} (1+\lambda)^m$$

$$\frac{\Delta J}{J} = \frac{\Delta K}{K} + \lambda - \lambda \cdot \Delta \bar{a}$$

$$\frac{\Delta Y}{Y} = \frac{\Delta A'}{A'} + \alpha(\lambda - \lambda \Delta \bar{a} + \frac{\Delta K}{K}) + \beta \frac{\Delta L}{L}$$

(3)分离劳动质量的含体现型技术进步的生产函数模型

$$\frac{\Delta Y}{Y}=\frac{\Delta A''}{A''}+\alpha(\lambda-\lambda\Delta\overline{a}+\frac{\Delta K}{K})$$

$$+\beta(\delta-\delta\Delta\overline{b}+\frac{\Delta L}{L})$$

5. 引入人力资本的生产函数模型

- Lucas（1988）为了解决技术内生问题，提出人力资本的概念，Romer等人（1992）提出包括人力资本的生产函数模型

6. 边界生产函数模型

(1)确定性边界生产函数

$$Y = f(K, L, \cdots)e^{-u} \qquad (u \geq 0)$$

(2)随机边界生产函数

$$Y = f(K, L, \cdots)e^{v-u} = (f(K, L, \cdots)e^{v})e^{-u}$$

四、几个重要生产函数模型的参数估计方法

1. C-D生产函数模型及其改进型的估计

(1)线性估计方法

$$Y = AK^{\alpha}L^{\beta}\mu$$

(2)非线性估计方法

$$Y = AK^{\alpha}L^{\beta} + \mu$$

- 能否线性化，与假设有关。哪个方法更合理?

2. CES生产函数模型及其改进型的估计

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m}{\rho}} \mu$$

$$\ln Y = \ln A - \tfrac{m}{\rho} \ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \varepsilon$$

$$\ln Y = \ln A + \delta_1 m \ln K + \delta_2 m \ln L - \tfrac{1}{2} \rho m \delta_1 \delta_2 (\ln(\frac{K}{L}))^2 + \varepsilon$$

$$Z = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon$$

- 假设?
- 误差?

3. VES生产函数的估计

$$Y = AK^{(\frac{1}{1+c})m}(L+(\frac{b}{1+c})K)^{(\frac{c}{1+c})m} \cdot \mu$$

$$\ln Y = \ln A + \frac{m}{1+c}\ln K + \frac{cm}{1+c}\ln(L+\frac{b}{1+c}K)+\varepsilon$$

$$\ln(L+\frac{b}{1+c}K) = \ln(L+\lambda\cdot K) = Z(\lambda)$$

$$Z(\lambda) = \ln L + \frac{K}{L}\cdot\lambda + o(\lambda)$$

$$\ln Y = \ln A + \frac{m}{1+c}\ln K + \frac{cm}{1+c}\ln L + \frac{cmb}{(1+c)^2}\frac{K}{L}+\varepsilon$$

$$Z = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon$$

4. 二级CES生产函数模型的估计

- 二级CES生产函数为：

$$Y_{KE} = (a_1 K^{-\rho_1} + a_2 E^{-\rho_1})^{-\frac{1}{\rho_1}}$$

$$Y = A(b_1 Y_{KE}^{-\rho} + b_2 L^{-\rho})^{-\frac{m}{\rho}}$$

- 由第**2**级函数展开取近似，得到：

$$\ln Y = \ln A + b_1 m \ln Y_{KE} + b_2 m \ln L - \tfrac{1}{2}\rho m b_1 b_2 (\ln(\frac{Y_{KE}}{L}))^2 + \varepsilon$$

- 由第1级函数展开取近似，得到：

$$\ln Y_{KE} = a_1 \ln K + a_2 \ln E - \tfrac{1}{2}\rho_1 a_1 a_2 (\ln(\frac{K}{E}))^2$$

- 代入前式，得到：

$$\ln Y = \ln A + b_1 m a_1 \ln K + b_1 m a_2 \ln E + b_2 m \ln L$$

$$- \tfrac{1}{2} m b_1 \rho_1 a_1 a_2 (\ln(\frac{K}{E}))^2 - \tfrac{1}{2} \rho m b_1 b_2 (\ln(\frac{K}{L}))^2 + \varepsilon$$

$$Z = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \varepsilon$$

- 代入后的式中有多个二次项，应该选择多少项？为什么？
- 是否造成估计结果的任意性？

5. 含体现型技术进步生产函数模型的估计

- 估计的生产函数为：

$$\frac{\Delta Y}{Y}=\frac{\Delta A'}{A'}+\alpha\left(\lambda-\lambda\Delta\bar{a}+\frac{\Delta K}{K}\right)+\beta\frac{\Delta L}{L}+\varepsilon$$

- 直接作为线性模型估计：

$$Z_t=\alpha_0+\alpha X_{1t}+\beta X_{2t}+\varepsilon_t$$

- 关键是如何得到**X_{1t}**的样本观测值

6. 确定性统计边界生产函数模型的修正的普通最小二乘估计（Corrected OLS,COLS）

- 采用**C-D**生产函数形式：

$$Y = f(K, L, \cdots)e^{-u}$$

$$Y = AK^{\alpha}L^{\beta}e^{-u} \quad (u \geq 0)$$

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L - u$$

- 其中实质上的边界生产函数为：

$$\ln Y' = \ln A + \alpha \ln K + \beta \ln L$$

Y'为理论上的最大产出量。

$$\ln Y = (a-\mu) + \alpha \ln K + \beta \ln L - (u-\mu)$$

$$\ln \hat{Y} = (\widehat{a-\mu}) + \hat{\alpha} \ln K + \hat{\beta} \ln L$$

$$\hat{a} = (\widehat{a-\mu}) + \hat{\mu}$$

将

$$Max(\ln Y_i - \ln \hat{Y}_i) = Max(\ln Y_i - ((\widehat{a-\mu}) + \hat{\alpha} \ln K_i + \hat{\beta} \ln L_i))$$

作为 $\hat{\mu}$ 的值，代入得到。于是所要求的边界生产函数为：

$$\hat{Y}' = e^{\hat{a}} K^{\hat{\alpha}} L^{\hat{\beta}}$$

边界生产函数即是平均生产函数向上平移了$\hat{\mu}$。

五、生产函数模型应用一例：生产函数模型在技术进步分析中的应用

1. 从纵向研究技术进步：测算技术进步速度及其对经济增长的贡献

(1) 技术进步速度的测定

- 从生产函数模型求得要素的产出弹性
- 计算产出和各种要素的平均增长速度
- 利用增长方程计算技术进步速度

(2) 技术进步对增长贡献的测定

(3)实例

2. 从横向研究技术进步：部门之间、企业之间技术进步水平的比较分析

(1) 建立并估计某行业的企业确定性统计边界生产函数模型

(2) 确定技术效率为1的企业

(3) 计算每个企业的技术效率

(4) 实例

六、建立生产函数模型过程中的问题一例：数据质量问题

1. 样本数据的一致性问题

- 一致性问题在生产函数模型中的具体体现
- 为什么建立某个行业的生产函数模型必须采用时间序列数据?
- 为什么建立某个行业的企业生产函数模型必须采用截面数据?
- 为什么建立某个特定企业的生产函数模型必须采用时间序列数据?

2. 样本数据的准确性问题

- 样本数据的准确性的两层含义
- 什么样的要素投入量数据才是“准确”的？
- 用部分的数据代替全体的数据必须满足什么假设？

3. 样本数据的可比性问题

- 可比性的极端重要性
- 如何才能保证产出量数据的可比性?
- 如何才能保证资本投入量数据的可比性?

§7.2需求函数(Demand Function, D.F.)

- 几个重要概念
- 几种重要的单方程需求函数模型及其参数估计
- 线性支出系统需求函数模型及其参数估计
- 几种需求函数模型系统
- 建立与应用需求函数模型中的几个问题

一、几个重要概念

1. 需求函数

(1) 定义

- 需求函数是描述商品的需求量与影响因素，例如收入、价格、其它商品的价格等之间关系的数学表达式。

$$q_i = f(I, p_1, \cdots, p_i, \cdots, p_n)$$

- 特定情况下可以引入其它因素。

- 需求函数与消费函数是两个完全不同的概念。为什么？
- 单方程需求函数模型和需求函数模型系统

  哪类更符合需求行为理论？

(2) 单方程需求函数模型是经验的产物

- 与需求行为理论不符
- 经常引入其它因素
- 参数的经济意义不明确

(3) 需求函数模型系统来源于效用函数

- 由效用函数在效用最大化下导出，符合需求行为理论
- 只包括收入和价格
- 参数有明确的经济意义

2. 从效用函数到需求函数

(1) 从直接效用函数到需求函数

- 直接效用函数为：

$$U = u(q_1, q_2, \cdots, q_n)$$

- 预算约束为：

$$\sum_{i=1}^{n} q_i p_i = I$$

- 在预算约束下使效用最大，即得到需求函数模型。

构造如下的拉格朗日函数：

$$L(q_1,q_2,\cdots,q_n,\lambda)= u(q_1,q_2,\cdots,q_n)$$

$$+\lambda(I-\sum_{i=1}^{n}q_i p_i)$$

极值的一阶条件：

$$\begin{cases} \dfrac{\partial L}{\partial q_i}=\dfrac{\partial u}{\partial q_i}-\lambda p_i=0 \\ \dfrac{\partial L}{\partial \lambda}=I-\sum_{i=1}^{n}q_i p_i=0 \end{cases}$$

求解即得到需求函数模型。

(2) 从间接效用函数到需求函数

- 间接效用函数为：

$$V = v(p_1, p_2, \cdots, p_n, I)$$

- 利用公式

$$q_i = -\frac{\partial V}{\partial p_i} \Big/ \frac{\partial V}{\partial I} \qquad i = 1,2,\cdots,n$$

- 可以得到所求的使效用达到最大的商品需求函数。

3. 需求函数的0阶齐次性

(1) 需求的收入弹性

$$\eta_i = \frac{\Delta q_i}{q_i} \Big/ \frac{\Delta I}{I} \xrightarrow{\Delta \to 0} \frac{\partial q_i}{\partial I} \Big/ \frac{I}{q_i}$$

- 生活必须品的需求收入弹性?
- 高档消费品的需求收入弹性?
- 低质商品的的需求收入弹性?

(2) 需求的自价格弹性

$$\varepsilon_{ii}=\frac{\Delta q_i}{q_i}\Bigg/\frac{\Delta p_i}{p_i}\xrightarrow{\Delta\to 0}\frac{\partial q_i}{\partial p_i}\Bigg/\frac{p_i}{q_i}$$

- 生活必须品的需求自价格弹性?
- 高档消费品的需求自价格弹性?
- “吉芬品” 的的需求收入弹性?

(3) 需求的互价格弹性

$$\varepsilon_{ij} = \frac{\Delta q_i}{q_i} \Bigg/ \frac{\Delta p_j}{p_j} \xrightarrow{\Delta \to 0} \frac{\partial q_i}{\partial p_j} \Bigg/ \frac{p_j}{q_i}$$

- 替代品的需求互价格弹性?
- 互补品的需求互价格弹性?
- 互相独立商品的需求互价格弹性?

(4) 需求函数的0阶齐次性条件

- 当收入、价格、其它商品的价格等都增长倍时，对商品的需求量没有影响。即

$$f(\lambda I, \lambda p_1, \cdots, \lambda p_i, \cdots, \lambda p_n) = \lambda^0 f(I, p_1, \cdots, p_i, \cdots, p_n)$$

- 需求函数模型的重要特征

- 模型的检验

二、几种重要的单方程需求函数模型及其参数估计

1. 线性需求函数模型

$$q_i = \alpha + \sum_{j=1}^{n} \beta_j p_j + \gamma \cdot I + \mu$$

- 经验中存在
- 缺少合理的经济解释
- 不满足0阶齐次性条件
- OLS估计

2. 对数线性需求函数模型

$$\ln q_i = \alpha + \sum_{j=1}^{n} \beta_j \ln p_j + \gamma \ln I + \mu$$

- 经验中比较普遍存在
- 参数有明确的经济意义

  每个参数的经济意义和数值范围？
- 可否用0阶齐次性条件检验？
- OLS估计

3. 耐用品的存量调整模型

- 导出过程

$$S_t^e = \alpha_0 + \alpha_1 p_t + \alpha_2 I_t + \mu_t$$

$$S_t - S_{t-1} = \lambda(S_t^e - S_{t-1})$$

$$S_t = (1-\delta)S_{t-1} + q_t$$

$$q_t = S_t - S_{t-1} + \delta \cdot S_{t-1}$$

$$= \lambda(S_t^e - S_{t-1}) + \delta \cdot S_{t-1}$$

$$= \lambda\alpha_0 + \lambda\alpha_1 p_t + \lambda\alpha_2 I_t + (\delta - \lambda)S_{t-1} + \lambda\mu_t$$

- 常用于估计的模型形式

$$q_t = \beta_0 + \beta_1 p_t + \beta_2 I_t + \beta_3 S_{t-1} + \mu_t$$

- 直接估计。
- 参数估计量的经济意义不明确 。
- 必须反过来求得原模型中的每个参数估计量，才有明确的经济意义。
- 由**4**个参数估计量求原模型的**5**个参数估计量，必须外生给定δ 。

4. 非耐用品的状态调整模型

$$q_t = \beta_0 + \beta_1 p_t + \beta_2 I_t + \beta_3 q_{t-1} + \mu_t$$

- Houthakker和Taylor于1970年建议。
- 反映消费习惯等“心理存量”对需求的影响 。
- 用上一期的实际实现了的需求（即消费）量作为“心理存量”的样本观测值。

三、线性支出系统需求函数模型及其参数估计

(LES，Linear Expenditure System)

1. 线性支出系统需求函数模型

- **Klein、Rubin 1947年 直接效用函数**

$$U = \sum_{i=1}^{n} u_i(q_i) = \sum_{i=1}^{n} b_i \ln(q_i - r_i)$$

该效用函数的含义？

- R. Stone、1954年 在预算约束

$$\sum_{i=1}^{n} q_i p_i = V$$

- 导出需求函数

- 拉格朗日方程

$$L(q_1, q_2, \cdots, q_n, \lambda) = \sum_{i=1}^{n} b_i \ln(q_i - r_i)$$

$$+ \lambda(V - \sum_{i=1}^{n} q_i p_i)$$

- 极值条件

$$\begin{cases} \dfrac{\partial L}{\partial q_i} = \dfrac{b_i}{q_i - r_i} - \lambda \cdot p_i = 0 \\ \dfrac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} q_i p_i - V = 0 \end{cases} \qquad i = 1, 2, \cdots, n$$

- 对于前n个方程，消去λ可得

$$\frac{p_i}{p_j} = \frac{b_i}{b_j} \cdot \frac{q_j - r_j}{q_i - r_i} \qquad i, j = 1,2,\cdots,n$$

$$b_j(p_i q_i - p_i r_i) = b_i(p_j q_j - p_j r_j)$$

$$i = 1,2,\cdots,n \quad i \neq j$$

$$\sum_{i=1}^{n} b_j(p_i q_i - p_i r_i) = \sum_{i=1}^{n} b_i(p_j q_j - p_j r_j)$$

$$b_j \sum_{i=1}^{n} (p_i q_i - p_i r_i) = (p_j q_j - p_j r_j) \sum_{i=1}^{n} b_i$$

$$p_j q_j = p_j r_j + b_j \sum_{i=1}^{n} (p_i q_i - p_i r_i)$$

$$p_j q_j = p_j r_j + b_j (V - \sum_{i=1}^{n} (p_i r_i))$$

$$q_i = r_i + \frac{b_i}{p_i} (V - \sum_j p_j r_j) \quad i = 1,2,\cdots,n$$

- **LES是一个联立方程模型系统**
- 函数的经济意义
- 参数的经济意义
- 模型系统估计的困难是什么？

2. 扩展的线性支出系统需求函数模型

(ELES, Expend Linear Expenditure System)

(1) 模型的扩展

- **1973年 Liuch**

$$q_i = r_i + \frac{b_i}{p_i}(I - \sum_j p_j r_j) \qquad i = 1,2,\cdots,n$$

- 两点扩展
- 扩展后参数的经济意义发生了什么变化?
- 为什么扩展后的模型可以估计?

(2) 扩展的线性支出系统的0阶齐次性证明

$$\eta_i = \frac{\partial q_i}{\partial I}\cdot\frac{I}{q_i} = \frac{b_i I}{p_i q_i}$$

$$\varepsilon_{ii} = \frac{\partial q_i}{\partial p_i}\cdot\frac{p_i}{q_i} = \left(-\frac{b_i I}{p_i^2} + b_i\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{p_j r_j}{p_i^2}\right)\cdot\frac{p_i}{q_i} = \frac{(1-b_i)p_i r_i}{p_i q_i} - 1$$

$$\varepsilon_{ij} = \frac{\partial q_i}{\partial p_j}\cdot\frac{p_j}{q_i} = -\frac{b_i r_j}{p_i}\cdot\frac{p_j}{q_i} = -\frac{b_i p_j r_j}{p_i q_i}$$

$$\eta_i + \varepsilon_{ii} + \sum_{j\neq i}\varepsilon_{ij} = \frac{p_i r_i + b_i\left(I - \sum_{j=1}^{n} p_j r_j\right)}{p_i q_i} - 1 = 0$$

3. 扩展的线性支出系统需求函数模型的估计方法

(1) 迭代法

$$q_i p_i = r_i p_i + b_i(I - \sum_j p_j r_j) + \mu_i$$

$$i = 1,2,\cdots,n$$

$$V_i = r_i p_i + b_i(I - \sum_j p_j r_j) + \mu_i$$

- 首先改写成如下形式：

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{R} + \mathrm{N} \qquad (1)$$

其中

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \qquad \boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad \boldsymbol{R} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

$$Y_i = V_i - b_i I$$

$$X_i = (-b_i p_1, \cdots, -b_i p_{i-1}, (1 - b_i) p_i, -b_i p_{i+1}, \cdots, -b_i p_n)$$

• 再改写成如下形式：

$$\boldsymbol{W} = \dot{\boldsymbol{Z}}\boldsymbol{B} + \mathrm{N} \qquad (2)$$

$$\boldsymbol{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix} \qquad \dot{\mathbf{Z}} = \begin{pmatrix} Z & & & \\ & Z & & \\ & & \ddots & \\ & & & Z \end{pmatrix} \qquad \boldsymbol{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

$$Z = I - \sum_{j=1}^{n} p_j r_j \qquad W_i = V_i - p_i r_i$$

- 迭代过程

给定一组边际消费倾向b的初始值；

计算(1)中X的样本观测值；

采用OLS估计(1)，得到基本需求量r的第一次估计值；

代入(2)中，计算Z和W的样本观测值；

采用OLS估计(2)，得到b的第一次估计值；

重复该过程，直至两次迭代得到的参数估计值满足收敛条件为止。即完成了模型的估计。

- 采用OLS估计(1)时，应该首先将个方程相加，然后对相加得到的方程进行最小二乘估计。为什么？
- 首先给定b的初始值与首先给定r的初始值，不影响估计结果。为什么？

(2) 截面数据作样本时的最小二乘法

$$V_i = r_i p_i - b_i \sum_j p_j r_j + b_i I + \mu_i$$

- 利用截面上价格相同，写成：

$$V_i = a_i + bI_i + \mu_i \qquad i = 1,2,\cdots,n$$

- 对模型采用普通最小二乘法进行估计，得到：

$$\hat{a}_i, \hat{b}_i (i = 1,2,\cdots,n)$$

- 然后利用参数之间的关系计算 $r_i (i = 1,2,\cdots,n)$

四、几种需求函数模型系统

1. Rotterdam模型

- **Theil和Barten于1965、1966年**采用对数线性需求函数的微分形式，描述需求量、收入、价格的相对变化之间的关系。

$$d(\log q_i) = \pi_{i0} d(\log \overline{m}) + \sum_{i=1}^{n} \beta_{ij} d(\log p_i)$$

$$i = 1, 2, \cdots, n$$

- 用ML法估计

2. 超越对数需求函数模型系统(TLS)

- **Christenson 、Jorgenson 和Liu于1975年提出了如下的间接效用函数：**

$$\ln U = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln\left(\frac{p_i}{M}\right) + \sum_{i=1}^{n}\sum_{i=1}^{n} \beta_{ij} \ln\left(\frac{p_i}{M}\right) \ln\left(\frac{p_j}{M}\right)$$

- 得到需求函数模型系统为：

$$\frac{p_i q_i}{M} = \frac{-\alpha_i + \sum_{i=1}^{n} \beta_{ij} \ln\left(\frac{p_i}{M}\right)}{-\sum_{j=1}^{n} \alpha_j + \sum_{k=1}^{n}\sum_{j=1}^{n} \beta_{jk} \ln\left(\frac{p_k}{M}\right)} \qquad i = 1, 2, \cdots, n$$

3. 几乎理想的需求函数模型系统（AIDS，Almost Ideal Demand System ）

- **Deaton和Muellbauer于1980年提出了如下的间接效用函数：**

$$\log M = \alpha_0 + \sum_{k=1}^{n} \alpha_i \log p_k + \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{kj} \log p_k \log p_j + U\beta_0 \prod_{k=1}^{n} p_k^{\beta_k}$$

- 导出需求函数形式为：

$$w_i = \alpha_0 + \sum_{j=1}^{n} \gamma_{ij} \log p_j + \beta_i \log \frac{M}{a} \qquad i = 1,2,\cdots,n$$

$$w_i = \frac{p_i q_i}{M}$$

$$\log a = \alpha_0 + \sum_{k=1}^{n} \alpha_i \log p_k + \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_{kj} \log p_k \log p_j$$

4. Lewbel需求系统(Lewbel Demand System)

- **Lewbel（1989）对AIDS进行了改进，提出了包含AIDS和TLS的Lewbel需求系统**

5. 逆需求函数模型（Inverse Demand System）

- 价格是需求量的函数
- 适用于某些商品
- 根据**Anderson**（**1980**），**Barten**，**Betterdorf**（**1989**），**Holt**（**2002**）等人的研究发现，同常规的需求函数模型系统一样，逆需求函数模型系统也可以通过效用最大化法则推导出来。
- **Anderson**（**1980**），**Huang**（**1988**）和**Eales**（**1994**）等通过应用距离函数推导出了逆需求函数系统。

- 几乎所有需求函数模型系统，都发展了相应的逆需求函数模型系统
- 绝大多数经验研究工作都集中在肉类、鱼类、食品等不易保存的产品市场，这种市场一般带有较浓的买方市场的特征。

五、建立与应用需求函数模型中的几个问题

1. 交叉估计

(1) 问题的提出

- 收入和价格两类变量对商品需求量的影响是不同的。为什么？
- 商品需求量和收入之间存在长期关系；而价格水平一般只对商品需求量具有短期影响。为什么？
- 时间序列数据适合于短期弹性的估计，截面数据适合于长期弹性的估计。
- 用同一组样本数据同时估计需求函数模型的所有参数，在理论上是存在问题的。

- 于是就提出了合并时间序列数据和截面数据的估计方法，即交叉估计方法。
- 用截面数据为样本估计模型中的一部分反映长期影响的参数，然后再用时间序列数据为样本估计模型中的另一部分反映短期影响的参数，分两阶段完成模型的估计。

(2) 估计方法

以对数线性需求函数为例，假设只包括收入和自价格

$$\ln q = \alpha_0 + \alpha_1 \ln I + \alpha_2 \ln p + \mu$$

- 利用第T年的截面数据 $q_j, I_j (j = 1,2,\cdots,m)$
- 在截面上认为价格是常数

$$\ln q_j = a + \alpha_1 \ln I_j + \mu_j \quad j = 1,2,\cdots,m$$

- 估计得到 $\hat{\alpha}_1$

- 当以时间序列数据为样本时，将模型写成：

$$\ln q_t = \alpha_0 + \alpha_1 \ln I_t + \alpha_2 \ln p_t + \mu_t \qquad t = 1,2,\cdots,T$$

- 令 $y_t = \ln q_t - \hat{\alpha}_1 \ln p_t$
- 有 $y_t = \alpha_0 + \alpha_2 \ln p_t + \mu_t \qquad t = 1,2,\cdots,T$
- 估计得到 $\hat{\alpha}_0, \hat{\alpha}_2$

2. 大类商品的数量与价格

(1) 以购买支出额度量数量、以价格指数度量价格

例如：

$$V_i = R_i + b_i(I - \sum_{j=1}^{n} R_j) + \mu_i$$

- 模型是否满足0阶齐次性条件？

(2) 对于具有相同计量单位的类商品的处理

$$\overline{q} = \sum_{i=1}^{l} q_i$$

$$\overline{p} = \sum_{i=1}^{l} p_i q_i \Big/ \sum_{i=1}^{l} q_i$$

(3) 对于具有不同计量单位的类商品的处理

$$\bar{p} = \sum_{i=1}^{l}((p_i q_i)p_i) \Big/ \sum_{i=1}^{l}(p_i q_i)$$

$$\bar{q} = \sum_{i=1}^{l} p_i q_i \Big/ \bar{p}$$

- 一种经验处理方法，缺少理论支持

§7.3 消费函数
（Consumption Function）

- 几个重要的消费函数模型及其参数估计
- 消费函数模型的一般形式
- 中国居民消费行为实证分析

一、几个重要的消费函数模型及其参数估计

1. 绝对收入假设消费函数模型

- 消费是由收入唯一决定的

$$C_t = \alpha + \beta Y_t + \mu_t \qquad t = 1,2,\cdots,T$$

- 参数的经济意义和数值范围？
- 是否反映消费的边际效用递减规律？

- 变参数模型可以较好地反映边际消费倾向递减规律。

$$\beta = \beta_0 + \beta_1 Y_t$$

$$C_t = \alpha + \beta_0 Y_t + \beta_1 Y_t^2 + \mu_t$$

2. 相对收入假设消费函数模型

(1) “示范性”假设消费函数模型

- **Duesenberry**认为，在一个群体收入分布中处于低收入的个体，往往有较高的消费倾向。

$$\frac{C_i}{Y_i} = \alpha_0 + \alpha_1 \frac{\overline{Y}_i}{Y_i}$$

- 消费函数

$$C_i = \alpha_0 Y_i + \alpha_1 \overline{Y}_i + \mu_i \qquad i = 1,2,\cdots,n$$

- 参数的经济意义和数值范围？

(2) “不可逆性”假设消费函数模型

- **Duesenberry**认为当前收入低于曾经达到的最高收入时，往往有较高的消费倾向。

$$\frac{C_t}{Y_t} = \alpha_0 + \alpha_1 \frac{Y_0}{Y_t}$$

- 消费函数

$$C_t = \alpha_0 Y_t + \alpha_1 Y_0 + \mu_t$$

$$t = 1,2,\cdots,T$$

$$C_t = \alpha_0 Y_t + \alpha_1 Y_{t-1} + \mu_t$$

3. 生命周期假设消费函数模型

- **Modigliani，Brumberg和Ando于1954年提出预算约束为**

$$\sum_{t=1}^{T} \frac{C_t}{(1+r)^{t-1}} = \sum_{t=1}^{T} \frac{Y_t}{(1+r)^{t-1}}$$

- 使得效用函数达到最大，消费是各个时期的收入和贴现率的函数 。即

$$C_t = c_t(Y_1, Y_2, \cdots, Y_T, r)$$

- 表示为当前收入和资产存量的函数

$$C_t = \alpha_1 Y_t + \alpha_2 A_t + \mu_t$$

$$t = 1,2,\cdots,T$$

4. 持久收入假设消费函数模型

- **Friedman于1957年提出收入与消费都分为两部分**

$$Y_t = Y_t^p + Y_t^t \qquad C_t = C_t^p + C_t^t$$

- **消费函数**

$$C_t = \alpha_0 + \alpha_1 Y_t^p + \alpha_2 Y_t^t + \mu_t \quad t = 1,2,\cdots,T$$

- **对于时间序列数据，第t时刻的持久收入可表示为**

$$Y_t^p = \lambda Y_t + \lambda(1-\lambda)Y_{t-1} + \lambda(1-\lambda)^2 Y_{t-2} + \cdots$$

- 如何估计？

5. 合理预期的消费函数模型

- 假设第t期的消费是收入预期值的函数，即

$$C_t = \alpha + \beta Y_t^e$$

- 收入预期值是现期实际收入与前一期预期收入的加权和：

$$Y_t^e = (1-\lambda)Y_t + \lambda Y_{t-1}^e$$

$$= (1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots)$$

- 理论假设的合理性?

- 代入得到

$$C_t = \alpha + \beta(1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots)$$

$$C_{t-1} = \alpha + \beta(1-\lambda)(Y_{t-1} + \lambda Y_{t-2} + \lambda^2 Y_{t-3} + \cdots)$$

$$C_t - \lambda C_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)Y_t$$

$$C_t = \alpha(1-\lambda) + \lambda C_{t-1} + \beta(1-\lambda)Y_t + \mu_t$$

$$t = 1,2,\cdots,T$$

6. 适应预期的消费函数模型

$$C_t^e = \alpha + \beta Y$$

$$C_t - C_{t-1} = \lambda(C_t^e - C_{t-1})$$

$$C_t^e = \frac{1}{\lambda} C_t + \frac{\lambda - 1}{\lambda} C_{t-1}$$

$$C_t = \lambda\alpha + (1 - \lambda) C_{t-1} + \lambda\beta Y_t + \mu_t$$

$$t = 1,2,\cdots,T$$

- 理论假设和最终模型与 5. 的异同?

二、消费函数模型的一般形式

1. 消费函数模型的一般形式

- 形式

$$C_t = f(Y_t, C_{t-1}) + \mu_t$$

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 C_{t-1} + \varepsilon_t$$

- 经济意义解释合理。
- 各种消费函数模型，除绝对收入假设消费函数外，都可以近似表达为这种形式。

- 估计中的问题有哪些？

  共线性问题？

  随机解释变量问题？

2. 各种消费函数模型向一般形式的推导

- “示范性” 相对收入假设消费函数模型已具有相同的统计形式。
- “不可逆性”相对收入假设消费函数模型推导过程中仅忽略收入的两期滞后量的影响。
- 生命周期假设消费函数模型推导过程中仅去掉明显共线性项，引入常数项。
- 持久收入假设消费函数模型推导过程中仅将瞬时消费归入随机项，引入常数项。
- 合理预期假设与适应预期假设消费函数模型已经是相同的统计形式。
- 结论：该一般形式与各种理论假设都相容，具有包容性。

三、中国居民消费行为实证分析

1. 中国的总消费构成

- 总消费=居民消费+政府消费=农业居民消费+非农业居民消费+政府消费
- 总消费构成数据（看统计年鉴）
- 各个消费群体具有不同的消费行为
- 拟按照各自的消费行为建立各自的消费函数模型

2. 农业居民的消费行为分析

- （讨论）
- 关于两种假设的检验：绝对收入假设和生命周期假设。
- 两种假设导致不同的政策选择。

3. 非农业居民的消费行为分析

- （讨论）
- 关于两种假设的检验：绝对收入假设和相对收入假设。
- 两种假设导致不同的政策选择。

4. 政府的消费行为分析

- （讨论）
- 一致认为遵循绝对收入假设

§7.4投资函数
(Investment Function)

- 加速模型
- 利润决定的投资函数模型
- 新古典投资函数模型
- 一个中国的投资函数模型

一、加速模型

1. 常见4类模型形式

- **I、Y、K**分别表示投资、国民收入、固定资产。

$$I_t = f(\Delta Y_t) + \mu_t$$

$$I_t = f(Y_t, K_{t-1}) + \mu_t$$

$$I_t = f(Y_t, Y_{t-1}, I_{t-1}) + \mu_t$$

$$I_t = f(\Delta Y_t, Y_{t-1}, I_{t-1}) + \mu_t$$

- 分别为后面4类加速模型。

2. 原始加速模型(Naïve Accelerator Model)

- **1917年Clark提出**

$$K^e = \alpha Y$$

$$I_t = K_t - K_{t-1} = K_t^e - K_{t-1} = \alpha(Y_t - Y_{t-1})$$

$$I_t = \alpha \Delta Y_t + \mu_t$$

3. 灵活的加速模型（Flexible Accelerator Model）

- **Koyck于1954年**

$$K_t - K_{t-1} = \lambda(K_t^e - K_{t-1})$$

$$K_t = \lambda K_t^e + (1-\lambda)K_{t-1} = \lambda\alpha Y_t + (1-\lambda)K_{t-1}$$

$$K_t = \alpha(\lambda Y_t + \lambda(1-\lambda)Y_{t-1} + \lambda(1-\lambda)^2 Y_{t-2} + \cdots)$$

- 如果考虑到折旧，则有：

$$I_t = K_t - K_{t-1} + \delta K_{t-1} = \alpha\lambda Y_t + (\delta-\lambda)K_{t-1}$$

$$I_t = \alpha\lambda Y_t + (\delta-\lambda)K_{t-1} + \mu_t$$

4. 实用的加速模型

- 利用 $$I_{t-1} = K_{t-1} + (1-\delta)K_{t-2}$$

$$I_t - (1-\delta)I_{t-1} = \alpha\lambda Y_t + (\delta - \lambda)K_{t-1}$$
$$- (1-\delta)\alpha\lambda Y_{t-1} - (1-\delta)(\delta - \lambda)K_{t-2}$$
$$= \alpha\lambda Y_t - (1-\delta)\alpha\lambda Y_{t-1} + (\delta - \lambda)I_{t-1}$$

$$I_t = \alpha\lambda Y_t - (1-\delta)\alpha\lambda Y_{t-1} + (1-\lambda)I_{t-1} + \mu_t$$

5. 利用最新信息的加速模型

- **Hines和Catephores于1970年指出，人们是根据产出水平的最新信息来确定资本存量的期望值，而不是根据尚未可知的实际产出水平。于是有**

$$K_t^e = \alpha Y_{t-n}$$

$$I_t = \alpha\lambda Y_{t-n} - (1-\delta)\alpha\lambda Y_{t-n-1} + (1-\lambda) I_{t-1}$$

$$= \alpha\lambda \Delta Y_{t-n} + \delta\alpha\lambda Y_{t-n-1} + (1-\lambda) I_{t-1}$$

$$I_t = \alpha\lambda\Delta Y_{t-n} + \delta\alpha\lambda Y_{t-n-1} + (1-\lambda) I_{t-1} + \mu_t$$

6. 对加速模型的评价

- 假设

  没有资本闲置

  资本产出比为常数

  不存在自发投资

  采用几何滞后

- 揭示了投资活动的原动力
- 从总体上反映了投资活动中的因果关系
- 具有较大的实际应用价值

二、利润决定的投资函数模型

1、假设

- 加速模型认为投资的原动力是产出的增长。
- 但由于投资活动是一个多周期过程，投资决策必然与资金的回报有关，所以就要考虑市场条件、税率、利率、产品与资本品的价格等因素。
- 所以，资本存量的预期值并不取决于产出水平，而是取决于利润水平。

2、模型

- **Grunfeld于1961年提出了资本存量的预期值与利润水平之间的关系：**

$$K_t^e = \alpha_0 + \alpha_1 V_t$$

- **考虑资本存量的调整过程，投资函数模型为：**

$$I_t = \lambda(K_t^e - K_{t-1}) + \delta K_{t-1}$$

$$= \lambda\alpha_0 + \lambda\alpha_1 V_t + (\delta - \lambda) K_{t-1}$$

- 其计量形态为：

$$I_t = \lambda\alpha_0 + \lambda\alpha_1 V_t + (\delta - \lambda)K_{t-1} + \mu_t$$

- 先验地得到折旧率δ，然后估计模型的其它参数。

三、新古典投资函数模型

1、假设

- 加速模型假设资本产出比为常数，即认为资本与其它要素之间不具有可替代性。
- Jorgenson将新古典生产函数引入投资函数模型，承认在生产函数中要素之间具有可替代性，提出了新古典投资函数模型。

2、模型

- 以利润最大为目标，以新古典生产函数为约束条件，求解如下极值问题：

$$MaxR_t = p_t Y_t - w_t L_t - r_t K_t$$

约束：$Y_t = f(K_t, L_t)$

其中R、p、w、r分别为利润、产品的价格、工资率和资本的租金。

- 求解该极值问题即得到资本的最优存量，以此决定投资。
- 该模型的求解过程利用了边际生产力条件，不适用。

四、一个中国的投资函数模型

1. 模型形式

- 常用的模型形式

$$I_t = \beta_0 K_{t-1} + \beta_1 I_{t-1} + \beta_2 I_{t-2} + \cdots + \beta_l I_{t-l} + \mu_t$$

- 合理的经济解释
- 估计中的问题

2. 推导过程

- 根据经济行为，有

$$I_t = f_1(Y_t)$$

$$K_t = f_2(K_{t-1}, I_{t-1}, I_{t-2}, \cdots, I_{t-l})$$

$$Y_t = f_3(K_t, L_t)$$

- 逐一代入，则得到上面所表示的投资函数模型 。

- 分别采用简单的线性关系表示上述3个函数，有

$$I_t = \alpha_t Y_t$$

$$K_t = u_t K_{t-1} + \lambda_1 I_{t-1} + \lambda_2 I_{t-2} + \cdots + \lambda_l I_{t-l}$$

$$Y_t = e_t K_t$$

第八章 联立方程计量经济学模型的应用——宏观计量经济模型

Macro-Economy Econometrics Model

教学基本要求

本章是课程的非重点内容，可以教学要求选择全部或部分内容。通过教学，使学生达到：

- 了解（最低要求）：计量经济学模型的一个重要研究与应用领域—宏观经济；中国宏观计量经济模型的主要特征、总体结构和主要模块与方程的设计；
- 掌握（较高要求）：宏观计量经济模型的设定理论中的主要要点；不同经济体制、不同发展阶段、不同国民经济核算体系下宏观计量经济模型的异同；
- 应用（对应用能力的要求）：能够看懂已有的宏观计量经济模型。

一、几个概念
二、传统宏观计量经济模型的设定
三、影响宏观计量经济模型设定的几个因素
四、建立宏观计量经济模型的工作程序
五、Hendry学派建模理论简介
六、一个著名的小型宏观计量经济模型
七、中国宏观计量经济模型的主要特征
八、中国宏观计量经济模型中主要模块的设计
九、中国宏观计量经济模型中主要方程的设定

一、几个概念

1. 宏观计量经济学（Macroeconometrics）

- 名称由来已久
- 主要内容和研究方向发生了变化
- 经典的宏观计量经济学—宏观计量经济学模型理论
- 现代宏观计量经济学的主要研究方向——单位根检验、协整理论以及动态计量经济学

2. 宏观经济模型与宏观计量经济模型

- 宏观经济模型是在宏观总量水平上把握和反映经济运动的全面特征，研究宏观经济主要指标间的相互依存关系，描述国民经济和社会再生产过程各环节之间的联系，并可以用以进行宏观经济的结构分析、政策评价、决策研究和发展预测。
- 将应用计量经济学方法建立的宏观经济模型称为宏观计量经济模型，它是宏观经济模型中的一类。

3. 宏观经济计量模型的类型

- 按建模目的分类
- 按建模范围分类
- 按时间长度分类
- 按照经济理论基础分类

4.“方程”与“函数”

- 宏观计量经济模型中描述各种经济行为的结构方程称为“方程”，例如生产方程、需求方程、消费方程等。
- “方程”可以采用“函数”的形式，也可以是其它形式。
- 一般“方程”中包含的变量比“函数”中的变量多，但是关系没有“函数”复杂。

5.“因素”与“变量”

- “因素”通过行为分析确定。
- “变量”用以反映“因素”，具有经济统计学的要求。
- 首先确定“因素”，然后选择“变量”。
- 在大多数情况下，并不完全一致。

二、传统宏观计量经济模型的设定

1. CC方法论

- 1932年成立考利斯委员会（Cowles Commission）
- 1933年正式出版《Econometrica》。
- 1939年更名为考利斯基金委员会（Cowles Foundation Commission）。
- 宏观计量经济模型的基本理论形成于40年代，大部分基础性工作是由CC支持完成的。
- 产生了几位诺贝尔经济学奖得主。

2. 基本理论要点

(1) 依据某种已经存在的经济理论或者已经提出的对经济行为规律的某种解释设定模型的总体结构和个体结构，即模型是建立在已有的经济理论和经济行为规律假设的基础之上的；

(2) 引进概率论思想作为模型研究的方法论基础，选择随机联立线性方程组作为模型的一般形式；

(3) 模型的识别、参数的估计、模型的检验是主要的技术问题；

(4) 以模型对样本数据的拟合优度作为检验模型的主要标准。

3. 模型设定方法

- 从简单到复杂
- 从一般到简单

4. 评价

- 传统宏观计量经济模型的设定理论是在宏观计量经济模型的发展过程中逐渐形成的，反过来又极大地推动了宏观计量经济模型的发展。
- 对于同样的研究对象，不同的研究者只要对理论假设理解不同，仍然可以建立不同的模型。
- 对CC的主要批判源于20世纪70年代初期。

三、影响宏观经济计量模型设定的几个因素

1. 宏观经济环境对模型设定的影响

- 需求不足和供给不足是两类不同的宏观经济环境。
- 在需求不足的环境下，需求成为经济增长的主要制约，刺激需求成为宏观经济政策的主要目标。

讨论：宏观计量经济模型的总体结构特征和个体结构特征。

- 在供给不足的环境下，供给成为经济增长的主要制约，刺激生产成为宏观经济政策的主要目标。

讨论：宏观计量经济模型的总体结构特征和个体结构特征。

2. 宏观经济决策方式对模型设定的影响

- 宏观经济决策方式主要分为以集中决策为主和以分散决策为主两类。
- 讨论：分散决策方式对宏观计量经济模型的总体结构和个体结构都将产生影响。
- 讨论：集中决策方式对宏观计量经济模型的总体结构和个体结构都将产生影响。

3. 经济核算体系对模型设定的影响

- 宏观计量经济模型是在一定的核算体系基础是建立起来的。由指标体系组成的核算体系反映宏观经济的运行过程和状态，是宏观计量经济模型的数据来源，是设定宏观计量经济模型的重要依据。
- 两类核算体系：国民核算体系，简称为**SNA(System of National Accounting)**体系；国民经济平衡表体系，简称为**MPS(Material Product Balance System)**体系。
- 核算体系对宏观计量经济模型的影响在于指标体系以及主要指标的核算方法。

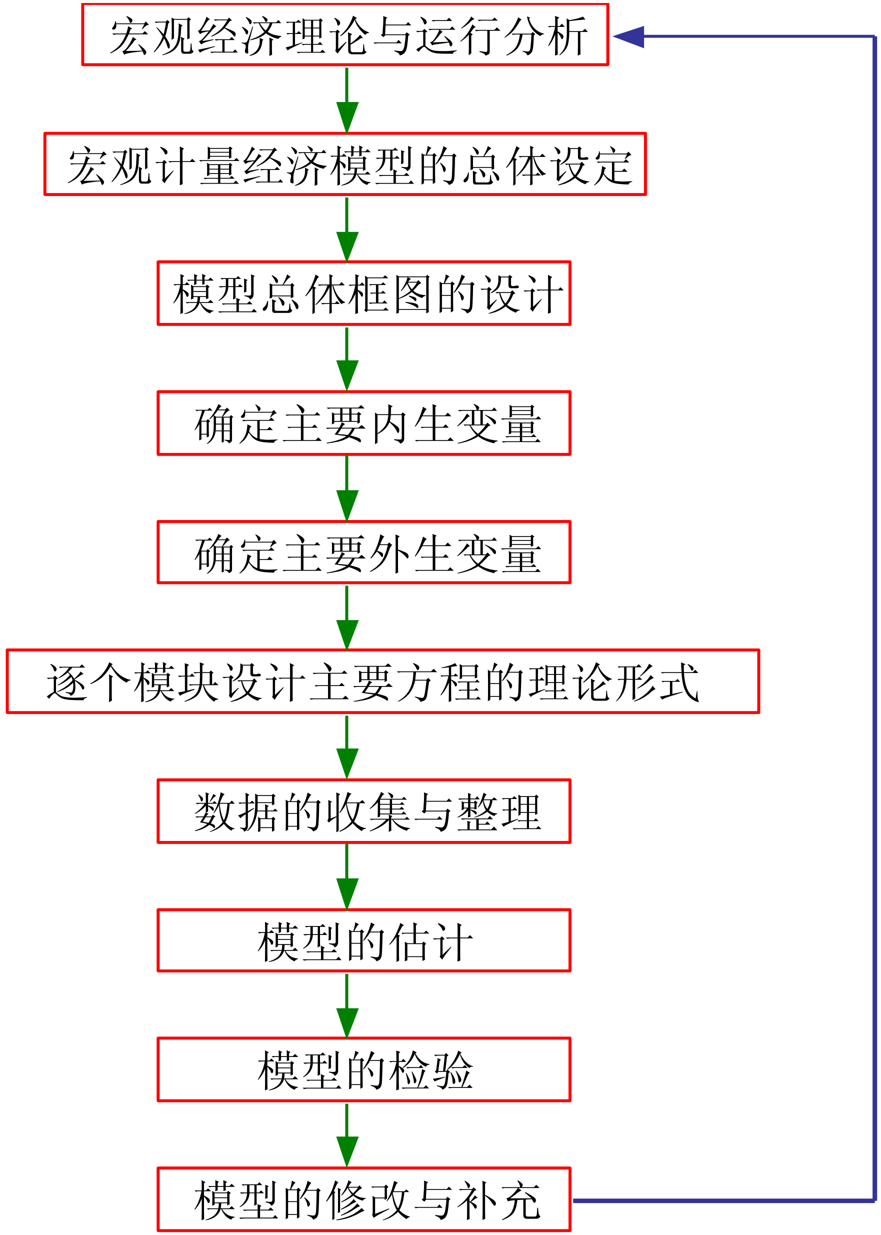
4. 模型外生性程度的决定

- 所谓外生性程度，简单说就是模型中外生变量与内生变量数目之间的比例。
- 影响外生性程度的因素：模型的功能 、决策方式 、可解释性 、样本容量 。
- 较高外生性程度的优点：控制模型规模 、减少方程设定误差 、方便于政策模拟和多方案计算 。
- 较高外生性程度的缺点：需要较大的样本容量 、预测外生变量值的困难 。

5. 模型分解性程度的决定

- 指部门的分解。
- 影响模型分解性程度的因素 ：宏观经济中的结构性变化 、建模目的的影响 、模型规模的限制 。
- 较高分解性程度的优点：模型具有较好的结构功能 、方程能较好地描述经济行为 、模型的样本期模拟精度和样本期外的预测精度都较高、使偏差多样化和分散化。
- 较高分解性程度的缺点：数据收集和调整的工作量和难度增大、模型中包含了更多的方程带来更大的方程设定误差。

四、建立宏观经济计量模型的工作程序


宏观经济理论与运行分析
宏观计量经济模型的总体设定
模型总体框图的设计
确定主要内生变量
确定主要外生变量
逐个模块设计主要方程的理论形式
数据的收集与整理
模型的估计
模型的检验
模型的修改与补充

五、Hendry学派建模理论简介

1. 数据生成过程(DGP, Data Generation Process)

- 建立模型的理论上的起点。
- 数据生成过程(**DGP**)是所有变量联合概率分布的一般表达式。即

$$\prod_{t=1}^{T} D(x_t \mid x_{t-1};\theta)$$

2. 自回归分布滞后模型(ADL, Autoregressive Distributed Lag)

- 建立模型的实际上的起点。
- **ADL**是由**DGP**约化而来的，如果每步约化都有效，即关于关注参数无信息损失，那么它可以近似地代表DGP。

$$y_t = \beta_0 + \sum_{i=0}^{q} \delta_i z_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t$$

3.误差修正模型（ECM, Error Correction Model）

- 关键是对ADL中包含的变量进行单整和协整检验；
- 将约化后的模型写成误差修正模型的形式，即包含变量间长期稳定关系的简单的模型。

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t + \gamma ecm_{t-1} + \varepsilon_t$$

六、一个著名的小型宏观经济计量模型——Klein战争之间模型

1. 变量

Y: 收入

C: 消费

I:净投资

W_P: 私人工资

Π: 利润

K:年末的股本

内生变量

G: 政府非工资开支

W_G: 政府工资

T: 企业税收

t: 时间

外生变量

2. 模型

$$C_t = \alpha_0 + \alpha_1 \prod_t + \alpha_2 \prod_{t-1} + \alpha_3 (W_{Pt} + W_{Gt}) + \varepsilon_{1t}$$

$$I_t = \beta_0 + \beta_1 \prod_t + \beta_2 \prod_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t}$$

$$W_{Pt} = \gamma_0 + \gamma_1 (Y_t + T_t - W_{Gt}) + \gamma_2 (Y_{t-1} + T_{t-1} - W_{Gt-1}) + \gamma_3 t + \varepsilon_{3t}$$

$$Y_t = C_t + I_t + G_t - T_t$$

$$\prod_t = Y_t - W_{Pt} - W_{Gt}$$

$$K_t = I_t + K_{t-1}$$

3. 估计

$$\hat{\alpha}_0 = 16.78 \quad \hat{\alpha}_1 = 0.020 \quad \hat{\alpha}_2 = 0.235 \quad \hat{\alpha}_3 = 0.800$$

$$\hat{\beta}_0 = 17.79 \quad \hat{\beta}_1 = 0.231 \quad \hat{\beta}_2 = 0.546 \quad \hat{\beta}_3 = -0.146$$

$$\hat{\gamma}_0 = 1.60 \quad \hat{\gamma}_1 = 0.420 \quad \hat{\gamma}_2 = 0.164 \quad \hat{\gamma}_3 = 0.135$$

- 以美国两次世界大战之间的**1920~1941**年年度数据为样本, 采用**FIML**估计。

4.应用

- 直接应用结构参数估计结果。

例如分析消费方程, 工资收入是私人工资和政府工资之和, 其消费边际倾向是**0.8**, 即工资增加**1**美元, 消费就增加**0.8**美元; 现期利润的消费边际倾向**0.02**, 而前期利润的边际消费倾向**0.235**。由此可见, 现期工资收入是消费的一个决定性因素。

- 利用简化式参数估计结果。

	Y_{t-1}	Π_{t-1}	K_{t-1}	W_{Gt}	G_t	T_t	t	W_{Gt-1}	T_{t-1}
C_t	0.189	0.743	-0.098	0.666	0.671	-0.188	0.155	-0.189	0.189
I_t	-0.015	0.746	-0.184	-0.052	0.259	-0.296	-0.012	0.015	-0.015
W_{Pt}	0.237	0.626	-0.119	-0.162	0.811	-0.204	0.195	-0.237	0.237
Y_t	0.174	1.489	-0.283	0.614	1.930	-1.484	0.143	-0.174	0.174
Π_t	-0.063	0.363	-0.164	0.224	1.119	-1.281	-0.052	0.063	-0.063
K_t	-0.015	0.746	0.816	-0.052	0.259	-0.296	-0.012	0.015	-0.015

表中政府控制变量列中数据表示这些变量对每一个内生变量的短期影响乘数。

例如, 当税收增加10000美元, 引起消费下降1880美元, 投资减少2960美元;

例如，当政府支出增加10000美元, 引起收入增加19300美元, 如果同时增加税收10000美元, 收入减少14840美元, 二者的平衡预算影响乘数(对收入)为二者之和, 即4460美元。

- 计算并利用长期均衡乘数。

	C	I	W_P	Y	$\prod$	K
W_G	0.536	0	-0.271	0.536	-0.192	-1.024
G	1.323	0	1.358	2.323	0.965	5.123
T	-0.569	0	-0.333	-1.569	-1.237	-6.564

可见, 各政策变量的投资均衡乘数为0, 因为在均衡情况下, 资本存量不变, 没有投资发生; 政府支出对收入的长期影响乘数是2.323, 而短期乘数是1.930, 即80%的影响是当期发生的。

七、中国宏观经济计量模型的主要特征

1. 总体上的供给导向

- 供给导向的模块结构关系；
- 主要方程的供给导向；
- 部分方程的需求导向或者供需双导向。

2. 集中决策与分散决策并存

- 关于投资模块的设计
- 关于价格模块的设计

3.以SNA体系为主，兼顾少数MPS体系指标

- 总体上的SNA体系
- 少数MPS指标的方程
- 数据收集和整理方面的困难

4. 较高的分解性程度

- 经济中结构性问题突出

5. 虚变量的普遍应用

- 政策在经济活动中的作用；
- “奇异点”的较多存在。
- 例如某个模型中的工业生产方程：

$$
\begin{aligned}
\ln(GIS) = & -0.8053 + 0.2335\ln(LSWSI) + 0.336\ln(NKFI_{-1}) \\
& + 0.1996(\ln(NIA_{-1}/PIR_{-1}) + \ln(YW_{-1}/PIV_{-1})) \\
& + 0.4956\ln(ECC + ED - 80DD79 - 40DD80)
\end{aligned}
$$

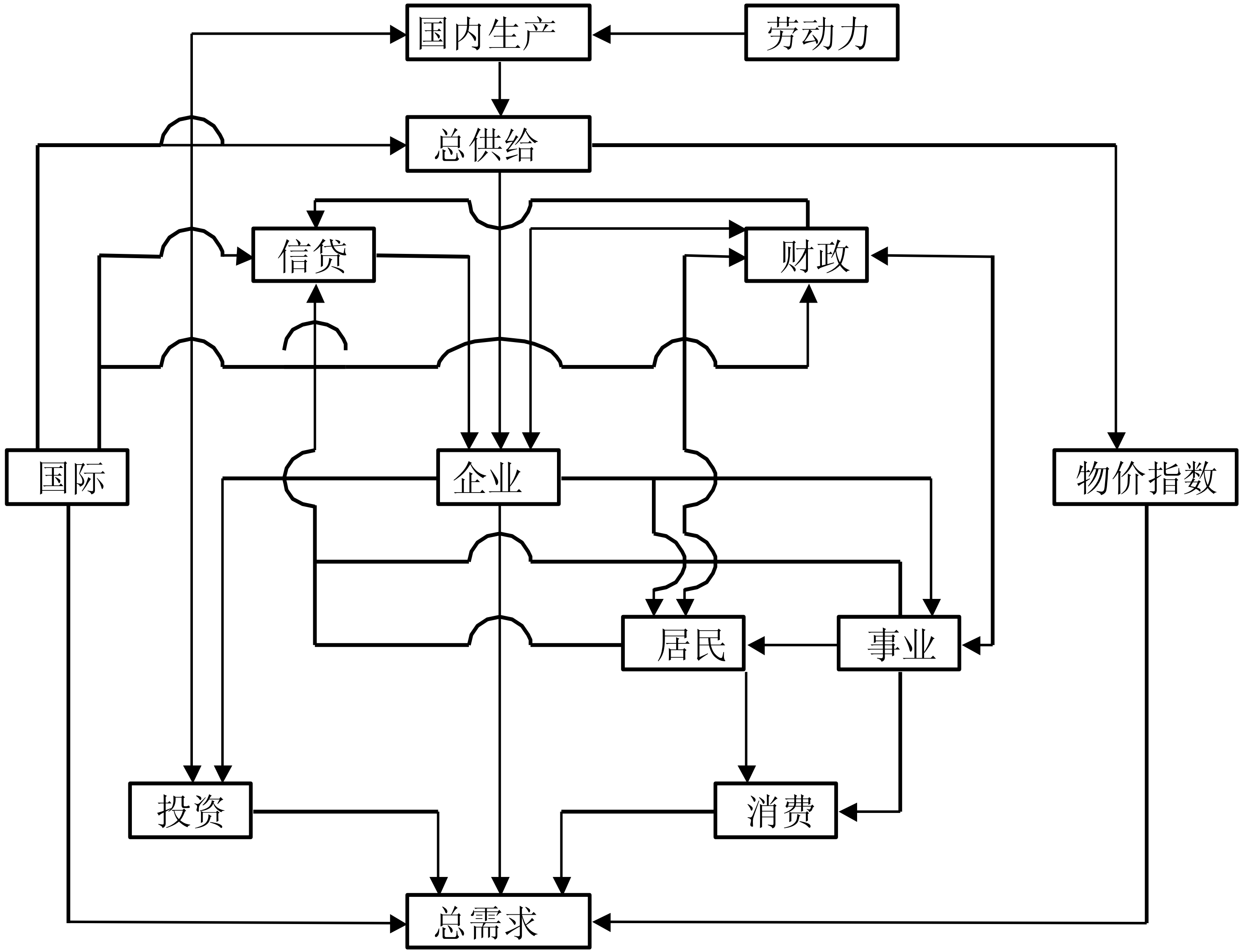
八、中国宏观计量经济模型中主要模块的设计

1. 模块

- 宏观计量经济模型由若干模块组成。
- 每个模块描述某项经济活动或者某个经济主体的经济行为。
- 每个模块由若干结构方程组成。模块中方程的数目取决于分解性程度。
- 模块之间的关系描述了宏观计量经济模型的总体结构。

2. 一个模型示例

- “中国宏观计量经济模型CEMT-1”总体结构框图。
- 共有12个模块：生产、投资、财政、信贷、企业、事业、居民、国际、消费、价格、就业和总供求。
- 模型共包含256个方程，其中随机方程102个，衡等方程154个；模型包括256个内生变量和41个外生变量。


国内生产
劳动力
总供给
信贷
财政
国际
企业
物价指数
居民
事业
投资
消费
总需求

九、中国宏观计量经济模型中主要方程的设定

- 具体略，参见教科书。
- 例如，

某类商品出口额=f（国内供给能力，国际市场需求，价格，汇率，政策，μ）

某类商品进口额=f（国际市场供给，国内需求，外汇支付能力，价格，汇率，政策，μ）

某种商品价格指数=f（相关商品的价格指数，需求因素，供给因素，政策虚变量，μ）

某部门劳动者人数=f（前一年劳动者人数，新增生产规模，μ）

某部门固定资产原值或净值=f（前一年固定资产原值或净值，当年投资额，μ）

城镇居民储蓄存款=f（前一年末城镇居民储蓄存款余额，城镇居民收入，存款利率，μ）

各项财政支出=f（财政支出，该项目前一年实际支出，政策，μ）

第九章

时间序列计量经济学模型的理论与方法

第一节　时间序列的平稳性及其检验

第二节　随机时间序列模型的识别和估计

第三节　协整分析与误差修正模型

§9.1 时间序列的平稳性及其检验

一、问题的引出：非平稳变量与经典回归模型

二、时间序列数据的平稳性

三、平稳性的图示判断

四、平稳性的单位根检验

五、单整、趋势平稳与差分平稳随机过程

一、问题的引出：非平稳变量与经典回归模型

1. 常见的数据类型

到目前为止，经典计量经济模型常用到的数据有：

- 时间序列数据（time-series data)；
- 截面数据(cross-sectional data)
- 平行/面板数据（panel data/time-series cross-section data)

★时间序列数据是最常见，也是最常用到的数据。

2. 经典回归模型与数据的平稳性

- 经典回归分析暗含着一个重要假设：数据是平稳的。
- 数据非平稳，大样本下的统计推断基础——“一致性”要求——被破坏。
- 经典回归分析的假设之一：解释变量X是非随机变量
- 放宽该假设：X是随机变量，则需进一步要求：

(1)X与随机扰动项 μ 不相关:Cov(X,μ)=0

(2) $\sum (X_i - \bar{X})^2 / n$ 依概率收敛：$P\lim_{n\to\infty}(\sum (X_i - \bar{X})^2 / n) = Q$

第（1）条是OLS估计的需要

第（2）条是为了满足统计推断中大样本下的“一致性”特性：$P\lim_{n\to\infty}(\hat{\beta})=\beta$

注意：在双变量模型中：

$$\hat{\beta}=\beta+\frac{\sum x_i u_i}{\sum x_i^2}=\beta+\frac{\sum x_i u_i/n}{\sum x_i^2/n}$$

因此：

$$P\lim_{n\to\infty}\hat{\beta}=\beta+\frac{P\lim\sum x_i u_i/n}{P\lim\sum x_i^2/n}=\beta+\frac{0}{Q}=\beta$$

▲如果X是非平稳数据（如表现出向上的趋势），则（2）不成立，回归估计量不满足“一致性”，基于大样本的统计推断也就遇到麻烦。

3. 数据非平稳，往往导致出现“虚假回归”问题

表现在:两个本来没有任何因果关系的变量，却有很高的相关性（有较高的R^2):

例如：如果有两列时间序列数据表现出一致的变化趋势（非平稳的），即使它们没有任何有意义的关系，但进行回归也可表现出较高的可决系数。

在现实经济生活中：

情况往往是实际的时间序列数据是非平稳的，而且主要的经济变量如消费、收入、价格往往表现为一致的上升或下降。这样，仍然通过经典的因果关系模型进行分析，一般不会得到有意义的结果。

时间序列分析模型方法就是在这样的情况下，以通过揭示时间序列自身的变化规律为主线而发展起来的全新的计量经济学方法论。

时间序列分析已组成现代计量经济学的重要内容，并广泛应用于经济分析与预测当中。

二、时间序列数据的平稳性

时间序列分析中**首先遇到的问题**是关于时间序列数据的**平稳性**问题。

假定某个时间序列是由某一随机过程（stochastic process）生成的，即假定时间序列$\{X_t\}$（t=1, 2, …）的每一个数值都是从一个概率分布中随机得到，如果满足下列条件：

1）均值$E(X_t)=\mu$是与时间t 无关的常数；

2）方差$Var(X_t)=\sigma^2$是与时间t 无关的常数；

3）协方差$Cov(X_t, X_{t+k})=\gamma_k$ 是只与时期间隔k有关，与时间t 无关的常数；

则称该随机时间序列是平稳的（stationary)，而该随机过程是一平稳随机过程（stationary stochastic process）。

例9.1.1. 一个最简单的随机时间序列是一具有零均值同方差的独立分布序列：

$$X_t=\mu_t \quad , \quad \mu_t \sim N(0,\sigma^2)$$

该序列常被称为是一个**白噪声**（**white noise**）。

由于Xt具有相同的均值与方差，且协方差为零, 由定义, **一个白噪声序列是平稳的**。

例9.1.2. 另一个简单的随机时间列序被称为**随机游走**（**random walk**），该序列由如下随机过程生成：

$$X_t=X_{t-1}+\mu_t$$

这里，μ_t是一个白噪声。

容易知道该序列有相同的**均值**：$E(X_t)=E(X_{t-1})$

为了检验该序列是否具有相同的方差，可假设X_t的初值为X_0，则易知

$X_1=X_0+\mu_1$

$X_2=X_1+\mu_2=X_0+\mu_1+\mu_2$

… …

$X_t=X_0+\mu_1+\mu_2+\ldots+\mu_t$

由于X_0为常数，μ_t是一个白噪声，因此$Var(X_t)=t\sigma^2$

即X_t的方差与时间t有关而非常数，它是一非平稳序列。

- 然而，对X取一阶差分（**first difference**）：

$$\Delta X_t = X_t - X_{t-1} = \mu_t$$

由于μ_t是一个白噪声，则序列$\{X_t\}$是平稳的。

后面将会看到:如果一个时间序列是非平稳的，它常常可通过取差分的方法而形成平稳序列。

- **事实上，随机游走过程是下面我们称之为1阶自回归AR(1)过程的特例**

$$X_t = \phi X_{t-1} + \mu_t$$

不难验证:1)|φ|>1时，该随机过程生成的时间序列是发散的，表现为持续上升(φ>1)或持续下降(φ<-1)，因此是非平稳的；

2) $\phi=1$时，是一个随机游走过程，也是非平稳的。

第二节中将证明:只有当$-1<\phi<1$时，该随机过程才是平稳的。

- **1**阶自回归过程**AR(1)**又是如下**k**阶自回归**AR(K)**过程的特例：

$$\mathbf{X_t}= \phi_1 X_{t-1}+\phi_2 X_{t-2}\ldots+\phi_k X_{t-k}$$

该随机过程平稳性条件将在第二节中介绍。

三、平稳性检验的图示判断

- 给出一个随机时间序列，首先可通过该序列的**时间路径图**来粗略地判断它是否是平稳的。
- 一个**平稳的时间序列**在图形上往往表现出一种围绕其均值不断波动的过程；
- 而**非平稳序列**则往往表现出在不同的时间段具有不同的均值（如持续上升或持续下降）。

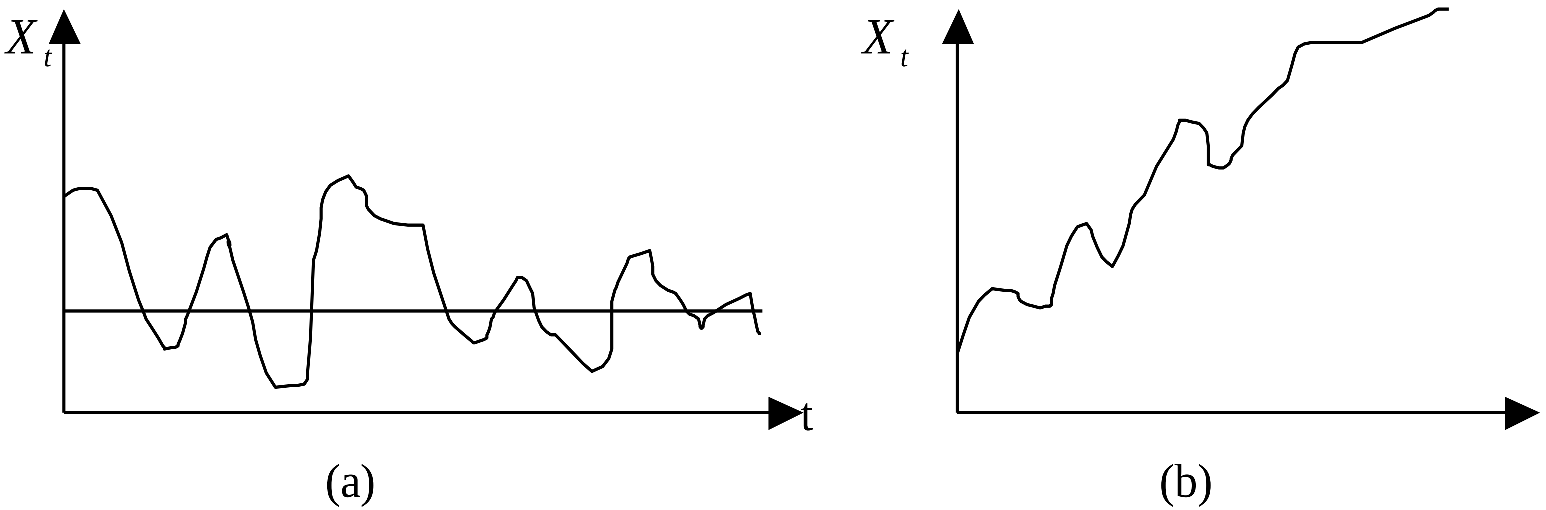


图 9.1　平稳时间序列与非平稳时间序列图

- 进一步的判断:

  检验样本自相关函数及其图形

定义随机时间序列的**自相关函数**（**autocorrelation function, ACF**）如下：

$$\rho_k=\gamma_k/\gamma_0$$

自相关函数是关于滞后期k的递减函数(Why?)。

实际上,对一个随机过程只有一个实现（样本），因此，只能计算样本自相关函数（Sample autocorrelation function）。

一个时间序列的样本自相关函数定义为：

$$r_k = \frac{\sum_{t=1}^{n-k}\left(X_t - \bar{X}\right)\left(X_{t+k} - \bar{X}\right)}{\sum_{t=1}^{n}\left(X_t - \bar{X}\right)^2} \qquad k = 1,2,3,\cdots$$

易知，随着k的增加，样本自相关函数下降且趋于零。但从下降速度来看，平稳序列要比非平稳序列快得多。

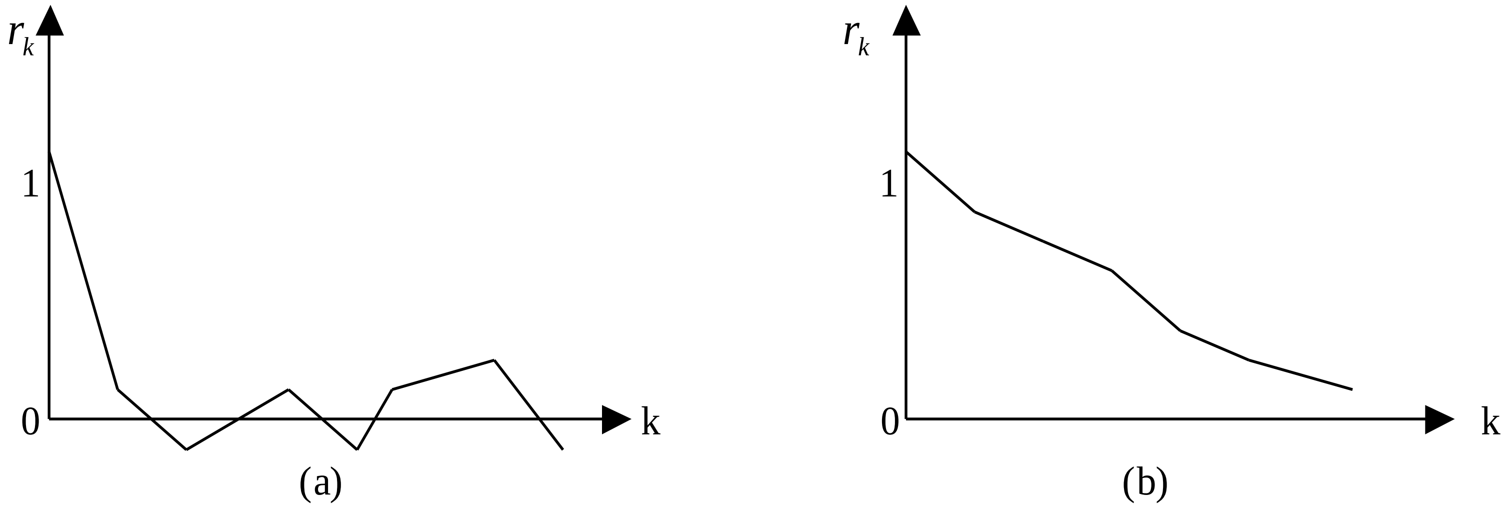


图 9.1.2　平稳时间序列与非平稳时间序列样本相关图

- **注意:**

确定样本自相关函数r_k某一数值是否足够接近于0是非常有用的，因为它可检验对应的自相关函数ρ_k的真值是否为0的假设。

Bartlett曾证明:如果时间序列由白噪声过程生成，则对所有的**k>0**，样本自相关系数近似地服从以**0**为均值，**1/n** 为方差的正态分布，其中**n**为样本数。

也可检验对所有k>0，自相关系数都为0的联合假设，这可通过如下Q_{LB}统计量进行：

$$Q_{LB}=n(n+2)\sum_{k=1}^{m}\left(\frac{r_k^2}{n-k}\right)$$

该统计量近似地服从自由度为m的χ^2分布（m为滞后长度）。

因此:如果计算的Q值大于显著性水平为α的临界值，则有1-α的把握拒绝所有ρ_k(k>0)同时为0的假设。

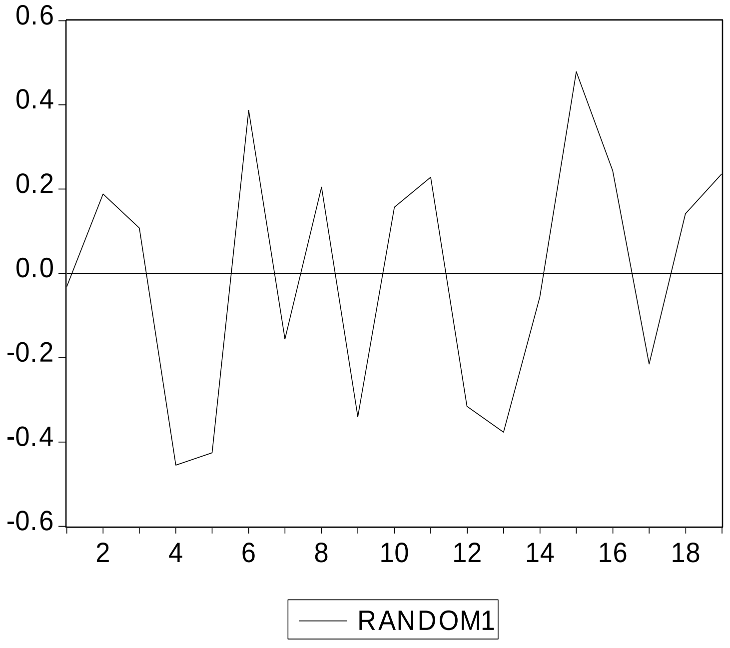
例9.1.3：表9.1.1序列Random1是通过一随机过程（随机函数）生成的有19个样本的随机时间序列。

表 9.1.1　　一个纯随机序列与随机游走序列的检验

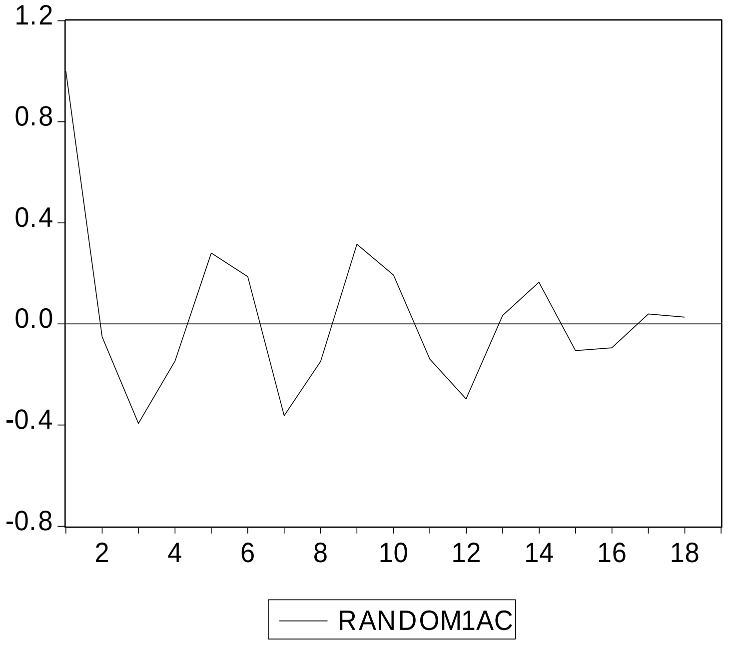
序号	Random1	自相关系数 r_k (k=0, 1, …17)		Q_{LB}	Random2	自相关系数 r_k (k=0, 1, …17)	Q_{LB}
1	-0.031	K=0,	1.000		-0.031	1.000	
2	0.188	K=1,	-0.051	0.059	0.157	0.480	5.116
3	0.108	K=2,	-0.393	3.679	0.264	0.018	5.123
4	-0.455	K=3,	-0.147	4.216	-0.191	-0.069	5.241
5	-0.426	K=4,	0.280	6.300	-0.616	0.028	5.261
6	0.387	K=5,	0.187	7.297	-0.229	-0.016	5.269
7	-0.156	K=6,	-0.363	11.332	-0.385	-0.219	6.745
8	0.204	K=7,	-0.148	12.058	-0.181	-0.063	6.876
9	-0.340	K=8,	0.315	15.646	-0.521	0.126	7.454
10	0.157	K=9,	0.194	17.153	-0.364	0.024	7.477
11	0.228	K=10,	-0.139	18.010	-0.136	-0.249	10.229
12	-0.315	K=11,	-0.297	22.414	-0.451	-0.404	18.389
13	-0.377	K=12,	0.034	22.481	-0.828	-0.284	22.994
14	-0.056	K=13,	0.165	24.288	-0.884	-0.088	23.514
15	0.478	K=14,	-0.105	25.162	-0.406	-0.066	23.866
16	0.244	K=15,	-0.094	26.036	-0.162	0.037	24.004
17	-0.215	K=16,	0.039	26.240	-0.377	0.105	25.483
18	0.141	K=17,	0.027	26.381	-0.236	0.093	27.198
19	0.236				0.000		

- 容易验证：该样本序列的均值为0，方差为0.0789。

从图形看：它在其样本均值0附近上下波动，且样本自相关系数迅速下降到0，随后在0附近波动且逐渐收敛于0。



（a）



（b）

由于该序列由一随机过程生成，可以认为不存在序列相关性，因此该序列为一白噪声。

- 根据Bartlett的理论：$\rho_k \sim N(0, 1/19)$

因此任一r_k(k>0)的95%的置信区间都将是

$$[-Z_{0.025} \bullet \sigma \ , Z_{0.025} \bullet \sigma] = [-1.96 \times \sqrt{1/19}, 1.96 \times \sqrt{1/19}] = [-0.4497, 0.4497]$$

可以看出:k>0时，r_k的值确实落在了该区间内，因此可以接受ρ_k(k>0)为0的假设。

同样地，从Q_{LB}统计量的计算值看，滞后17期的计算值为26.38，未超过5%显著性水平的临界值27.58，因此，可以接受所有的自相关系数ρ_k(k>0)都为0的假设。

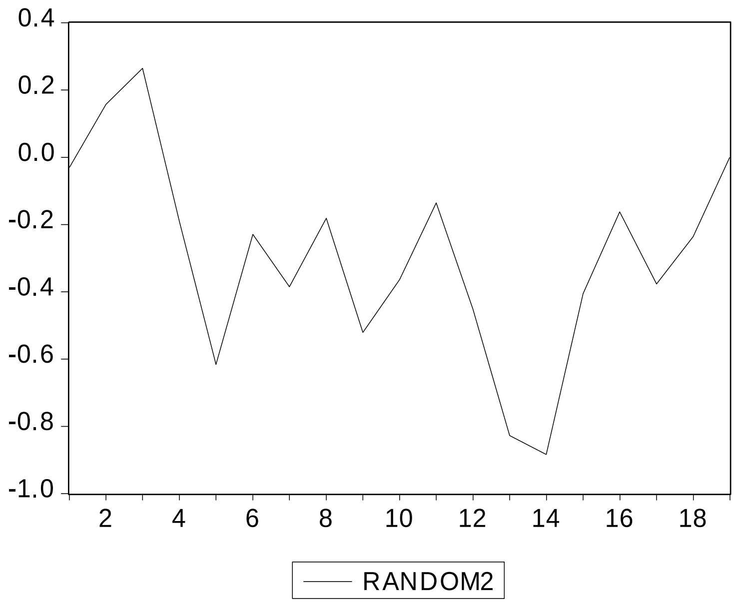
因此，该随机过程是一个平稳过程。

- 序列Random2是由一随机游走过程

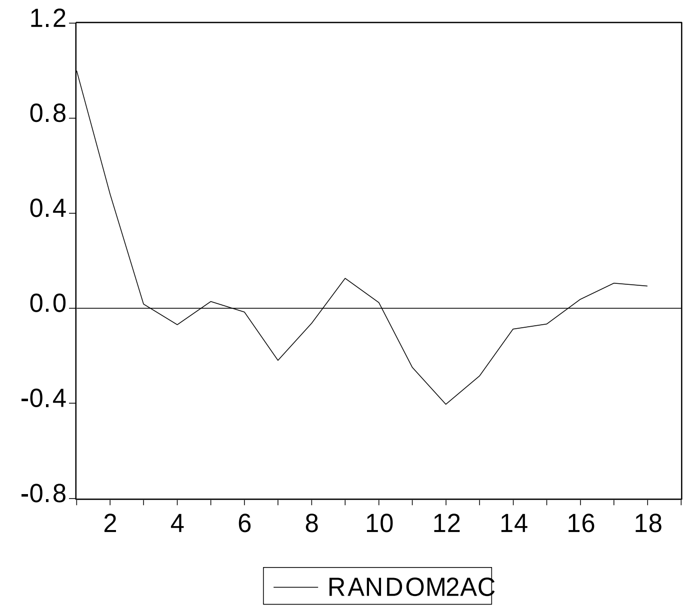
$$X_t=X_{t-1}+\mu_t$$

生成的一随机游走时间序列样本。

其中，第0项取值为0， μ_t是由Random1表示的白噪声。



（a）



（b）

图形表示出：该序列具有相同的均值，但从样本自相关图看，虽然自相关系数迅速下降到0，但随着时间的推移，则在0附近波动且呈发散趋势。

样本自相关系数显示：r_1=0.48，落在了区间[-0.4497，0.4497]之外，因此在5%的显著性水平上拒绝ρ_1的真值为0的假设。

该随机游走序列是非平稳的。

例 9.1.4 检验中国支出法 GDP 时间序列的平稳性。

表 9.1.2　1978~2000 年中国支出法 GDP（单位：亿元）

年份	GDP	年份	GDP	年份	GDP
1978	3605.6	1986	10132.8	1994	46690.7
1979	4073.9	1987	11784	1995	58510.5
1980	4551.3	1988	14704	1996	68330.4
1981	4901.4	1989	16466	1997	74894.2
1982	5489.2	1990	18319.5	1998	79003.3
1983	6076.3	1991	21280.4	1999	82673.1
1984	7164.4	1992	25863.6	2000	89112.5
1985	8792.1	1993	34500.6		

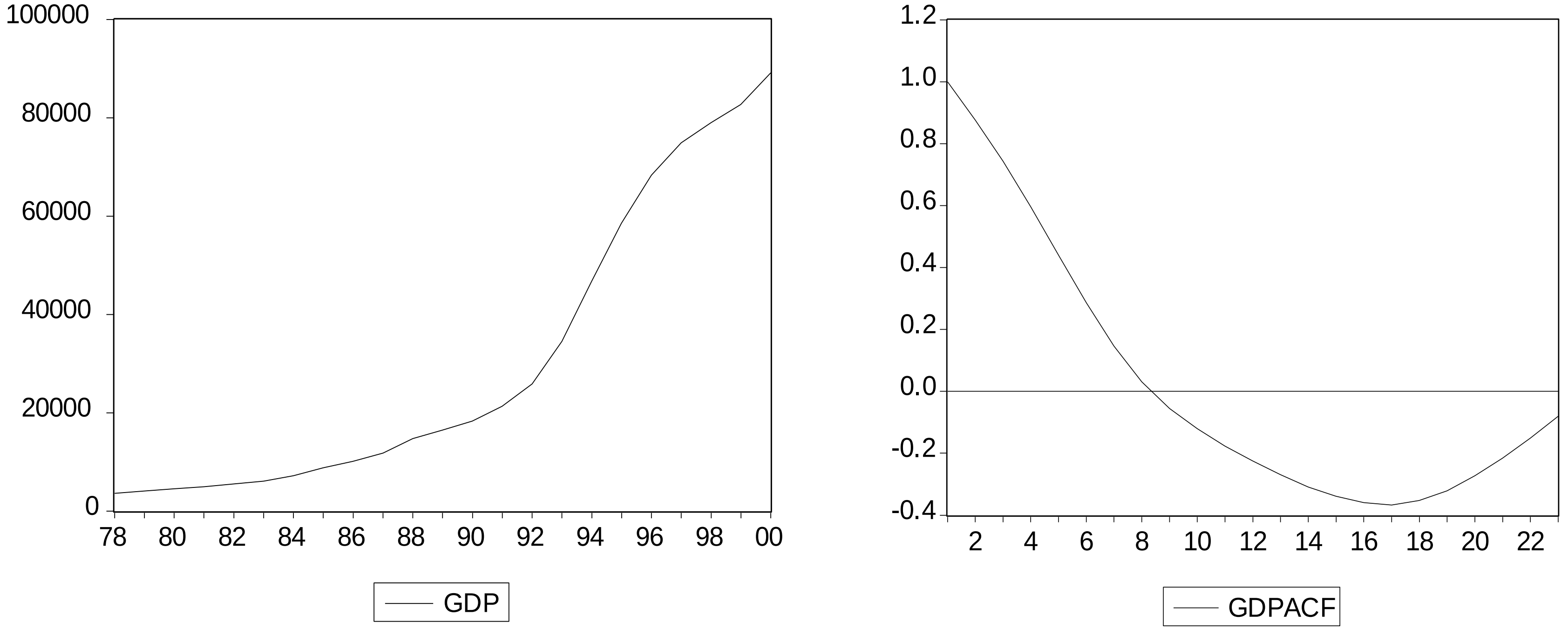


图 9.1.5　1978~2000 年中国 GDP 时间序列及其样本自相关图

- 图形：表现出了一个持续上升的过程，可初步判断是非平稳的。
- 样本自相关系数：缓慢下降，再次表明它的非平稳性。

•**从滞后18期的Q_{LB}统计量看：**

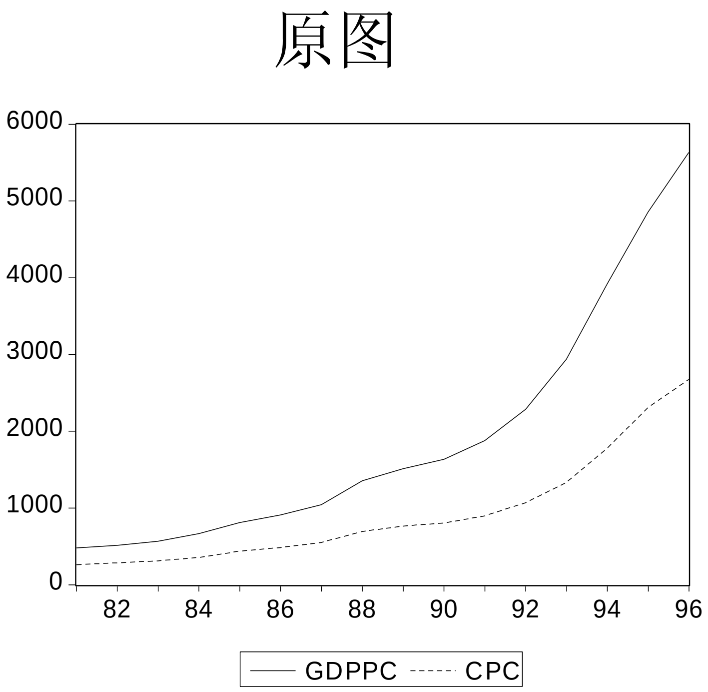
$$Q_{LB}(18)=57.18>28.86=\chi^2_{0.05}$$

拒绝： 该时间序列的自相关系数在滞后1期之后的值全部为0的假设。

结论:

1978~2000年间中国GDP时间序列是非平稳序列。

- 例9.1.5 检验§2.10中关于人均居民消费与人均国内生产总值这两时间序列的平稳性。



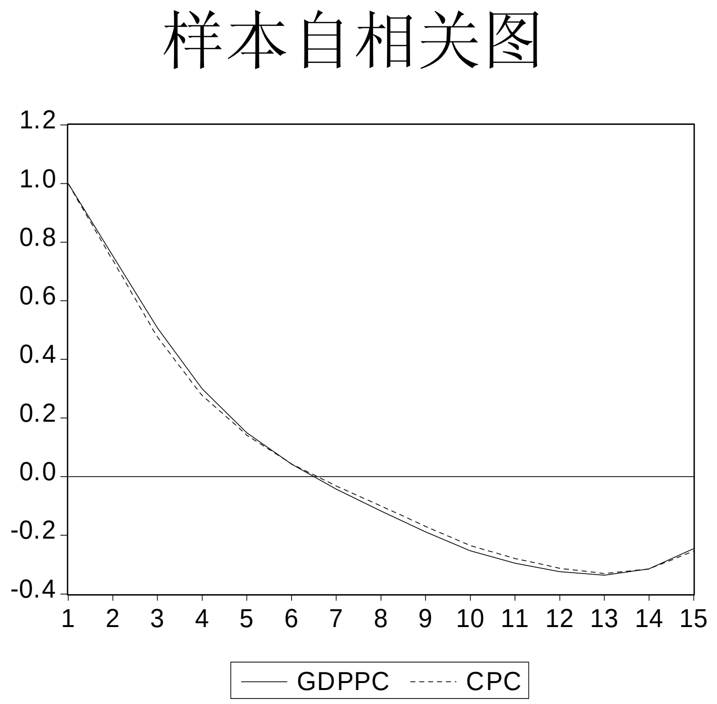


图 9.1.6　1981~1996 中国居民人均消费与人均 GDP 时间序列及其样本自相关图

- **从图形上看：**人均居民消费（CPC）与人均国内生产总值（GDPPC）**是非平稳的**。

- **从滞后14期的Q_{LB}统计量看：**

　　CPC与GDPPC序列的统计量计算值均为57.18，超过了显著性水平为5%时的临界值23.68。再次**表明它们的非平稳性。**

　　就此来说，运用传统的回归方法建立它们的回归方程是无实际意义的。

　　不过，第三节中将看到，如果两个非平稳时间序列是协整的，则传统的回归结果却是有意义的，而这两时间序列恰是协整的。

四、平稳性的单位根检验

对时间序列的平稳性除了通过图形直观判断外，运用统计量进行统计检验则是更为准确与重要的。

单位根检验（unit root test）是统计检验中普遍应用的一种检验方法。

1、DF检验

我们已知道，随机游走序列

$$X_t=X_{t-1}+\mu_t$$

是非平稳的，其中μ_t是白噪声。

而该序列可看成是随机模型

$$X_t=\rho X_{t-1}+\mu_t$$

中参数ρ=1时的情形。

也就是说，我们对式

$$X_t=\rho X_{t-1}+\mu_t \qquad （*）$$

做回归，如果确实发现ρ=1，就说随机变量Xt有一个**单位根**。

- （*）式可变形式成差分形式：

$$\Delta X_t=(1-\rho)X_{t-1}+\mu_t$$

$$=\delta X_{t-1}+\mu_t \qquad (**)$$

检验（*）式是否存在单位根ρ=1，也可通过（**）式判断是否有δ =0。

一般地：

- 检验一个时间序列Xt的平稳性，可通过检验带有截距项的一阶自回归模型

$$X_t=\alpha+\rho X_{t-1}+\mu_t \qquad (*)$$

中的参数ρ是否小于1。

或者：检验其等价变形式

$$\Delta X_t=\alpha+\delta X_{t-1}+\mu_t \qquad (**)$$

中的参数δ是否小于0 。

在第二节中将证明，（*）式中的参数ρ>1或ρ=1时，时间序列是非平稳的；

对应于（**）式，则是δ>0或δ =0。

- 因此，针对式 $\Delta X_t=\alpha+\delta X_{t-1}+\mu_t$
  我们关心的检验为：**零假设 H0：$\delta=0$**。
  备择假设 H1：$\delta<0$

上述检验可通过OLS法下的t检验完成。

然而，在零假设（序列非平稳）下，即使在大样本下t统计量也是有偏误的（向下偏倚），通常的t检验无法使用。

Dicky和Fuller于1976年提出了这一情形下t统计量服从的分布（这时的t统计量称为**τ统计量**），即**DF分布**（见表9.1.3）。

由于t统计量的向下偏倚性，它呈现围绕小于零值的偏态分布。

表 9.1.3　DF 分布临界值表

显著性水平	样本容量 25	50	100	500	∝	t分布临界值（n=∝）
0.01	-3.75	-3.58	-3.51	-3.44	-3.43	-2.33
0.05	-3.00	-2.93	-2.89	-2.87	-2.86	-1.65
0.10	-2.63	-2.60	-2.58	-2.57	-2.57	-1.28

- 因此，可通过OLS法估计

$$\Delta X_t=\alpha+\delta X_{t-1}+\mu_t$$

并计算t统计量的值，与DF分布表中给定显著性水平下的临界值比较：

如果：**t<临界值，则拒绝零假设H_0：δ =0，**

认为时间序列不存在单位根，是平稳的。

- 注意：在不同的教科书上有不同的描述，但是结果是相同的。

例如："如果计算得到的t统计量的绝对值大于临界值的绝对值，则拒绝$\rho=0$"的假设，原序列不存在单位根，为平稳序列。

2、ADF检验

进一步的问题：在上述使用

$$\Delta X_t=\alpha+\delta X_{t-1}+\mu_t$$

对时间序列进行平稳性检验中，**实际上假定了时间序列是由具有白噪声随机误差项的一阶自回归过程AR(1)生成的**。

但在实际检验中，时间序列可能由更高阶的自回归过程生成的，或者随机误差项并非是白噪声，这样用**OLS法进行估计均会表现出随机误差项出现自相关**（autocorrelation），导致DF检验无效。

另外，如果时间序列包含有明显的随时间变化的某种趋势（如上升或下降），则也容易导致上述检验中的**自相关随机误差项问题**。

为了保证DF检验中随机误差项的白噪声特性，Dicky和Fuller对DF检验进行了扩充，形成了**ADF（Augment Dickey-Fuller ）检验**。

ADF检验是通过下面三个模型完成的：

模型 1： $$\Delta X_t = \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \qquad (*)$$

模型 2： $$\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \qquad (**)$$

模型 3： $$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \qquad (***)$$

- **模型3 中的t是时间变量**，代表了时间序列随时间变化的某种趋势（如果有的话）。
- **检验的假设都是：针对H1：δ<0,检验 H0：δ=0，即存在一单位根**。模型1与另两模型的差别在于是否包含有常数项和趋势项。

- 实际检验时从模型3开始，然后模型2、模型1。

何时检验拒绝零假设，即原序列不存在单位根，为平稳序列，何时检验停止。否则，就要继续检验，直到检验完模型1为止。

检验原理与DF检验相同，只是对模型1、2、3进行检验时，有各自相应的临界值。

表9.1.4给出了三个模型所使用的ADF分布临界值表。

表 9.1.4　不同模型使用的 ADF 分布临界值表

模型	统计量	样本容量	0.01	0.025	0.05	0.10
1	τ_δ	25	-2.66	-2.26	-1.95	-1.60
		50	-2.62	-2.25	-1.95	-1.61
		100	-2.60	-2.24	-1.95	-1.61
		250	-2.58	-2.23	-1.95	-1.61
		500	-2.58	-2.23	-1.95	-1.61
		>500	-2.58	-2.23	-1.95	-1.61
2	τ_δ	25	-3.75	-3.33	-3.00	-2.62
		50	-3.58	-3.22	-2.93	-2.60
		100	-3.51	-3.17	-2.89	-2.58
		250	-3.46	-3.14	-2.88	-2.57
		500	-3.44	-3.13	-2.87	-2.57
		>500	-3.43	-3.12	-2.86	-2.57
	τ_α	25	3.41	2.97	2.61	2.20
		50	3.28	2.89	2.56	2.18
		100	3.22	2.86	2.54	2.17
		250	3.19	2.84	2.53	2.16
		500	3.18	2.83	2.52	2.16
		>500	3.18	2.83	2.52	2.16
3	τ_δ	25	-4.38	-3.95	-3.60	-3.24
		50	-4.15	-3.80	-3.50	-3.18
		100	-4.04	-3.73	-3.45	-3.15
		250	-3.99	-3.69	-3.43	-3.13
		500	-3.98	-3.68	-3.42	-3.13
		>500	-3.96	-3.66	-3.41	-3.12
	τ_α	25	4.05	3.59	3.20	2.77
		50	3.87	3.47	3.14	2.75
		100	3.78	3.42	3.11	2.73
		250	3.74	3.39	3.09	2.73
		500	3.72	3.38	3.08	2.72
		>500	3.71	3.38	3.08	2.72
	τ_β	25	3.74	3.25	2.85	2.39
		50	3.60	3.18	2.81	2.38
		100	3.53	3.14	2.79	2.38
		250	3.49	3.12	2.79	2.38
		500	3.48	3.11	2.78	2.38
		>500	3.46	3.11	2.78	2.38

一个简单的检验过程：

同时估计出上述三个模型的适当形式，然后通过ADF临界值表检验零假设H0：$\delta=0$。

1）只要其中有一个模型的检验结果拒绝了零假设，就可以认为时间序列是平稳的；

2）当三个模型的检验结果都不能拒绝零假设时，则认为时间序列是非平稳的。

这里所谓**模型适当的形式**就是在每个模型中选取适当的滞后差分项，以使模型的残差项是一个白噪声（主要保证不存在自相关）。

例9.1.6 检验1978~2000年间中国支出法GDP时间序列的平稳性。

1）经过偿试，模型3取了2阶滞后：

$$\Delta GDP_t = -1011.33 + 229.27T + 0.0093GDP_{t-1} + 1.50\Delta GDP_{t-1} - 1.01\Delta GDP_{t-2}$$

（-1.26） (1.91) (0.31) (8.94) (-4.95)

通过**拉格朗日乘数检验**（Lagrange multiplier test）对随机误差项的自相关性进行检验：

LM（1）=0.92， LM（2）=4.16，

小于5%显著性水平下自由度分别为1与2的χ^2分布的临界值，可见不存在自相关性，因此该模型的设定是正确的。

从δ的系数看，t>临界值，不能拒绝存在单位根的零假设。

时间T的t统计量小于ADF分布表中的临界值，因此不能拒绝不存在趋势项的零假设。**需进一步检验模型2**。

2）经试验，模型2中滞后项取2阶：

$$\Delta GDP_t = 357.45 + 0.057 GDP_{t-1} + 1.65 \Delta GDP_{t-1} - 1.15 \Delta GDP_{t-2}$$

（-0.90）　(3.38)　(10.40)　(-5.63)

LM（1）=0.57　LM（2）=2.85

LM检验表明模型残差不存在自相关性，因此该模型的设定是正确的。

从GDP_{t-1}的参数值看，其t统计量为正值，大于临界值，**不能拒绝存在单位根的零假设**。

常数项的t统计量小于AFD分布表中的临界值，**不能拒绝不存常数项的零假设。**需进一步检验模型1。

3) 经试验，模型1中滞后项取2阶：

$$\Delta GDP_t = 0.063\,GDP_{t-1} + 1.701\,\Delta GDP_{t-1} - 1.194\,\Delta GDP_{t-2}$$

（4.15）　　(11.46)　　(-6.05)

LM（1）=0.17　　LM（2）=2.67

LM检验表明模型残差项不存在自相关性，因此模型的设定是正确的。

从GDP_{t-1}的参数值看，其t统计量为正值，大于临界值，不能拒绝存在单位根的零假设。

- **可断定中国支出法GDP时间序列是非平稳的。**

- **例9.1.7** 检验§2. 10中关于人均居民消费与人均国内生产总值这两时间序列的平稳性。

1)对**中国人均国内生产总值GDPPC**来说，经过偿试，三个模型的适当形式分别为

模型 3:

$$\Delta GDPPC_t = -75.08 + 45.36t - 0.15GDPPC_{t-1} + 1.03\Delta GDPPC_{t-1}$$

（-0.75） (1.93) (-1.04) (2.31)

LM（1）=2.88 LM（2）=1.86

模型 2:

$$\Delta GDPPC_t = -192.02 + 0.652\,GDPPC_{t-1} + 0.040\,\Delta GDPPC_{t-1} - 1.425\,\Delta GDPPC_{t-2}$$

（-1.78） (3.26) (0.08) (-2.96)

$$-0.412\,\Delta GDPPC_{t-3} - 1.403\,\Delta GDPPC_{t-4}$$

(-0.67) (-2.20)

LM（1）=1.67 LM（2）=1.71 LM(3)=6.28 LM（4）=10.92

模型 1：

$$\Delta GDPPC_t = 0.196\,GDPPC_{t-1} + 0.875\,\Delta GDPPC_{t-1} - 0.975\,\Delta GDPPC_{t-2}$$

（2.63）　　(2.61)　　(-2.72)

LM（1）=0.20　　LM（2）=3.53

- 三个模型中参数的估计值的t统计量均大于各自的临界值，因此**不能拒绝存在单位根的零假设**。
- 结论：**人均国内生产总值（GDPPC）是非平稳的。**

2）对于人均居民消费CPC时间序列来说，三个模型的适当形式为

模型 3：

$$\Delta CPC_t = -26.23 + 34.98t - 0.3646\, CPC_{t-1} + 1.4627\, \Delta CPC_{t-1}$$

(-0.477)　(2.175)　(-1.478)　(2.318)

LM(1)=1.577　LM(2)=1.834

模型 2：

$$\Delta CPC_t = -79.88 + 0.545\, CPC_{t-1} + 0.508\, \Delta CPC_{t-1} - 1.655\, \Delta CPC_{t-2} - 0.027\, \Delta CPC_{t-3}$$

(-1.37)　(3.37)　(1.16)　(-3.44)　(-0.05)

$$-1.824\, \Delta CPC_{t-4}$$

(-3.03)

LM(1)=3.57　LM(2)= 4.10　LM(3)=4.89　LM(4)=10.99

模型 1：

$$\Delta CPC_t = 0.37\,CPC_{t-1} + 0.88\,\Delta CPC_{t-1} - 1.48\,\Delta CPC_{t-2} + 0.08\,\Delta CPC_{t-3} - 1.71\,\Delta CPC_{t-4}$$

(3.60)　　(2.37)　　(-2.97)　　(0.12)　　(-2.68)

LM(1)=1.83　　LM(2)= 1.84　　LM(3)=2.00　　LM(4)=2.33

- 三个模型中参数CPC_{t-1}的t统计量的值均比ADF临界值表中各自的临界值大，**不能拒绝该时间序列存在单位根的假设**，
- 因此,**可判断人均居民消费序列CPC是非平稳的。**

五、单整、趋势平稳与差分平稳随机过程

1. 单整

随机游走序列

$$X_t = X_{t-1} + \mu_t$$

经差分后等价地变形为

$$\Delta X_t = \mu_t$$

由于μ_t是一个白噪声，因此**差分后的序列$\{\Delta X_t\}$是平稳的。**

如果一个时间序列经过一次差分变成平稳的，就称原序列是一阶单整（**integrated of 1**）序列，记为**I(1)**。

一般地，如果一个时间序列经过**d**次差分后变成平稳序列，则称原序列是**d** 阶单整（**integrated of d**）序列，记为**I(d)**。

显然，**I(0)**代表一平稳时间序列。

现实经济生活中**:**

1)只有少数经济指标的时间序列表现为平稳的，如利率等**;**

2)大多数指标的时间序列是非平稳的，如一些价格指数常常是**2**阶单整的，以不变价格表示的消费额、收入等常表现为**1**阶单整。

大多数非平稳的时间序列一般可通过一次或多次差分的形式变为平稳的。

但也有一些时间序列，无论经过多少次差分，都不能变为平稳的。这种序列被称为非单整的（**non-integrated**）。

例9.1.8 中国支出法GDP的单整性。

经过试算，发现**中国支出法GDP是1阶单整的**，适当的检验模型为

$$\Delta^2 GDP_t = 1174.08 + 261.25t - 0.495\Delta GDP_{t-1} + 0.966\Delta^2 GDP_{t-1}$$

(-1.99)　(4.23)　（-5.18）　(6.42)

R^2 =0.7501　LM(1)=0.40　LM(2)=1.29

例9.1.9 中国人均居民消费与人均国内生产总值的单整性。

经过试算，发现**中国人均国内生产总值GDPPC是2阶单整的**，适当的检验模型为

$$\Delta^3 GDPPC_t = -0.60\Delta^2 GDPPC_{t-1}$$

（-2.17）

R^2 =0.2778，　　LM(1)=0.31　　LM(2)= 0.54

同样地，**CPC也是2阶单整的**，适当的检验模型为

$$\Delta^3 CPC_t = -0.67\Delta^2 CPC_{t-1}$$

（-2.08）

R^2 =0.2515　　LM(1)=1.99　　LM(2)= 2.36

2. 趋势平稳与差分平稳随机过程

前文已指出，一些非平稳的经济时间序列往往表现出共同的变化趋势，而这些序列间本身不一定有直接的关联关系，这时对这些数据进行回归，尽管有较高的R^2，但其结果是没有任何实际意义的。这种现象我们称之为**虚假回归**或**伪回归**（**spurious regression**）。

如：用中国的劳动力时间序列数据与美国GDP时间序列作回归，会得到较高的R^2，但不能认为两者有直接的关联关系，而只不过它们有共同的趋势罢了，这种回归结果我们认为是虚假的。

为了避免这种虚假回归的产生，通常的做法是引入作为趋势变量的时间，这样包含有时间趋势变量的回归，可以消除这种趋势性的影响。

然而这种做法，只有当趋势性变量是**确定性的（deterministic）**而非**随机性的（stochastic）**，才会是有效的。

换言之，**如果一个包含有某种确定性趋势的非平稳时间序列，可以通过引入表示这一确定性趋势的趋势变量，而将确定性趋势分离出来。**

考虑如下的含有一阶自回归的随机过程：

$$X_t=\alpha+\beta t+\rho X_{t-1}+\mu_t \qquad (*)$$

其中:μ_t是一白噪声，t为一时间趋势。

1)如果$\rho=1$，$\beta=0$，则（*）式成为**一带位移的随机游走过程**：

$$X_t=\alpha+X_{t-1}+\mu_t \qquad (**)$$

根据α的正负，X_t表现出明显的上升或下降趋势。这种趋势称为**随机性趋势（stochastic trend）**。

2)如果$\rho=0$，$\beta\neq 0$，则（*）式成为一带时间趋势的随机变化过程：

$$X_t=\alpha+\beta t+\mu_t \qquad (***)$$

根据β的正负，X_t表现出明显的上升或下降趋势。这种趋势称为**确定性趋势（deterministic trend）**。

3) 如果$\rho=1$，$\beta\neq0$，则Xt包含有**确定性与随机性两种趋势。**

判断一个非平稳的时间序列，它的趋势是随机性的还是确定性的，可通过ADF检验中所用的第3个模型进行。

该模型中已引入了表示确定性趋势的时间变量t，即分离出了确定性趋势的影响。

因此，**(1)如果检验结果表明所给时间序列有单位根，且时间变量前的参数显著为零，则该序列显示出随机性趋势;**

(2)如果没有单位根，且时间变量前的参数显著地异于零，则该序列显示出确定性趋势。

随机性趋势可通过差分的方法消除

如：对式

$$X_t=\alpha+X_{t-1}+\mu_t$$

可通过差分变换为

$$\Delta X_t= \alpha+\mu_t$$

该时间序列称为**差分平稳过程（difference stationary process）**；

确定性趋势无法通过差分的方法消除，而只能通过除去趋势项消除，

如：对式

$$X_t=\alpha+\beta t+\mu_t$$

可通过除去βt变换为

$$X_t - \beta t =\alpha+\mu_t$$

该时间序列是平稳的，因此称为趋势平稳过程（**trend stationary process**）。

最后需要说明的是，趋势平稳过程代表了一个时间序列长期稳定的变化过程，因而用于进行长期预测则是更为可靠的。

§9.2 随机时间序列分析模型

一、时间序列模型的基本概念及其适用性

二、随机时间序列模型的平稳性条件

三、随机时间序列模型的识别

四、随机时间序列模型的估计

五、随机时间序列模型的检验

- 经典计量经济学模型与时间序列模型
- 确定性时间序列模型与随机性时间序列模型

一、时间序列模型的基本概念及其适用性

1、时间序列模型的基本概念

随机时间序列模型（time series modeling）是指仅用它的过去值及随机扰动项所建立起来的模型，其一般形式为

$$X_t=F(X_{t-1}, X_{t-2}, \ldots, \mu_t)$$

建立具体的时间序列模型，需解决如下三个问题：

(1)模型的具体形式

(2)时序变量的滞后期

(3)随机扰动项的结构

例如，取线性方程、一期滞后以及白噪声随机扰动项（ $\mu_t=\varepsilon_t$），模型将是一个**1阶自回归过程AR(1):**

$$X_t=\varphi X_{t-1}+\varepsilon_t$$

这里， ε_t特指一白噪声。

一般的**p阶自回归过程AR(p)**是

$$X_t=\varphi_1X_{t-1}+\varphi_2X_{t-2}+\ldots+\varphi_pX_{t-p}+\mu_t \qquad (*)$$

(1)如果随机扰动项是一个白噪声($\mu_t=\varepsilon_t$)，则称(*)式为一**纯AR(p)过程（pure AR(p) process）**，记为

$$X_t=\varphi_1X_{t-1}+\varphi_2X_{t-2}+\ldots+\varphi_pX_{t-p}+\varepsilon_t$$

(2)如果μ_t不是一个白噪声，通常认为它是一个q阶的**移动平均（moving average）过程MA(q)**:

$$\mu_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q}$$

该式给出了一个**纯MA(q)过程（pure MA(p) process）**。

将纯AR(p)与纯MA(q)结合，得到一个一般的**自回归移动平均（autoregressive moving average）过程ARMA（p,q）**：

$$X_t=\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\ldots+\varphi_p X_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q}$$

该式表明：

（1）一个随机时间序列可以通过一个自回归移动平均过程生成，即该序列可以由其自身的过去或滞后值以及随机扰动项来解释。

（2）如果该序列是平稳的，即它的行为并不会随着时间的推移而变化，**那么我们就可以通过该序列过去的行为来预测未来。**

这也正是随机时间序列分析模型的优势所在。

2、时间序列分析模型的适用性

- **经典回归模型的问题：**
- **迄今为止，**对一个时间序列Xt的变动进行解释或预测，是通过某个单方程回归模型或联立方程回归模型进行的，由于它们以因果关系为基础，且具有一定的模型结构，因此也常称为**结构式模型（structural model）**。
- **然而，**如果Xt波动的主要原因可能是我们无法解释的因素，如气候、消费者偏好的变化等，则利用结构式模型来解释Xt的变动就比较困难或不可能，因为要取得相应的量化数据，并建立令人满意的回归模型是很困难的。
- **有时，**即使能估计出一个较为满意的因果关系回归方程，但由于对某些解释变量未来值的预测本身就非常困难，甚至比预测被解释变量的未来值更困难，这时因果关系的回归模型及其预测技术就不适用了。

在这些情况下，我们采用另一条预测途径：通过时间序列的历史数据，得出关于其过去行为的有关结论，进而对时间序列未来行为进行推断。

例如，时间序列过去是否有明显的增长趋势，如果增长趋势在过去的行为中占主导地位，能否认为它也会在未来的行为里占主导地位呢？

或者**时间序列显示出循环周期性行为，**我们能否利用过去的这种行为来外推它的未来走向？

●**随机时间序列分析模型，就是要通过序列过去的变化特征来预测未来的变化趋势。**

使用时间序列分析模型的另一个原因在于：

如果经济理论正确地阐释了现实经济结构，则这一结构可以写成类似于ARMA(p,q)式的时间序列分析模型的形式。

例如，对于如下最简单的宏观经济模型：

$$C_t = \alpha_0 + \alpha_1 Y_1 + \alpha_2 C_{t-1} + \mu_t$$

$$Y_t = C_t + I_t$$

这里，Ct、It、Yt分别表示消费、投资与国民收入。

Ct与Yt作为内生变量，它们的运动是由作为外生变量的投资It的运动及随机扰动项μt的变化决定的。

上述模型可作变形如下：

$$C_t=\frac{\alpha_2}{1-\alpha_1}C_{t-1}+\frac{\alpha_0}{1-\alpha_1}+\frac{\alpha_1}{1-\alpha_1}I_t+\frac{1}{1-\alpha_1}\mu_t$$

$$Y_t=\frac{\alpha_2}{1-\alpha_1}Y_{t-1}+\frac{\alpha_0}{1-\alpha_1}+\frac{1}{1-\alpha_1}I_t-\frac{\alpha_2}{1-\alpha_1}I_{t-1}+\frac{1}{1-\alpha_1}\mu_t$$

- 两个方程等式右边除去第一项外的剩余部分可看成一个综合性的随机扰动项，其特征依赖于投资项I_t的行为。
- **如果I_t是一个白噪声**，则消费序列C_t就成为一个**1阶自回归过程AR(1)**，而收入序列Y_t就成为一个**(1,1)阶的自回归移动平均过程ARMA(1,1)**。

二、随机时间序列模型的平稳性条件

1、AR(p)模型的平稳性条件

自回归移动平均模型（ARMA）是随机时间序列分析模型的普遍形式，自回归模型（AR）和移动平均模型（MA）是它的特殊情况。

关于这几类模型的研究，是**时间序列分析的重点内容**：**主要包括模型的平稳性分析**、**模型的识别和模型的估计**。

随机时间序列模型的平稳性，可通过它所生成的随机时间序列的平稳性来判断。

如果一个p阶自回归模型AR(p)生成的时间序列是平稳的，就说该AR(p)模型是平稳的，

否则，就说该AR(p)模型是非平稳的。

考虑p阶自回归模型AR(p)

$$X_t=\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\ldots+\varphi_p X_{t-p}+\varepsilon_t \qquad (*)$$

- 引入**滞后算子（lag operator ）L：**

$$LX_t=X_{t-1},\ L^2X_t=X_{t-2},\ \ldots,\ L^pX_t=X_{t-p}$$

(*)式变换为

$$(1-\varphi_1 L-\varphi_2 L^2-\ldots-\varphi_p L^p)X_t=\varepsilon_t$$

记$\Phi(L)=(1-\varphi_1 L-\varphi_2 L^2-\ldots-\varphi_p L^p)$，则称多项式方程

$$\Phi(z)=(1-\varphi_1 z-\varphi_2 z^2-\ldots-\varphi_p z^p)=0$$

为AR(p)的特征方程(characteristic equation)。

可以证明，如果该特征方程的所有根在单位圆外（根的模大于1），则AR(p)模型是平稳的。

例9.2.1 AR(1)模型的平稳性条件。

对1阶自回归模型AR(1)

$$X_t = \varphi X_{t-1} + \varepsilon_t$$

方程两边平方再求数学期望，得到Xt的方差

$$E(X_t^2) = \varphi^2 E(X_{t-1}^2) + E(\varepsilon_t^2) + 2E(X_{t-1}\varepsilon_t)$$

由于X_t仅与ε_t相关，因此，$E(X_{t-1}\varepsilon_t)=0$。如果该模型稳定，则有$E(X_t^2)=E(X_{t-1}^2)$，从而上式可变换为：

$$\gamma_0 = \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$$

在稳定条件下，该方差是一非负的常数，从而有 $|\varphi|<1$。

而AR(1)的特征方程

$$\Phi(z) = 1 - \varphi z = 0$$

的根为 $z=1/\varphi$

AR(1)稳定，即 $|\varphi| < 1$，意味着特征根大于1。

例9.2.2 AR(2)模型的平稳性。

对AR(2)模型

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$$

方程两边同乘以Xt，再取期望得：

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + E(X_t \varepsilon_t)$$

又由于

$$E(X_t\varepsilon_t)=\varphi_1 E(X_{t-1}\varepsilon_t)+\varphi_2 E(X_{t-2}\varepsilon_t)+E(\varepsilon_t^2)=\sigma_\varepsilon^2$$

于是

$$\gamma_0=\varphi_1\gamma_1+\varphi_2\gamma_2+\sigma_\varepsilon^2$$

同样地，由原式还可得到

$$\gamma_1=\varphi_1\gamma_0+\varphi_2\gamma_1$$

$$\gamma_2=\varphi_1\gamma_1+\varphi_2\gamma_0$$

于是方差为

$$\gamma_0=\frac{(1-\varphi_2)\sigma_\varepsilon^2}{(1+\varphi_2)(1-\varphi_1-\varphi_2)(1+\varphi_1-\varphi_2)}$$

由平稳性的定义，该方差必须是一不变的正数，于是有

$$\varphi_1+\varphi_2<1, \quad \varphi_2-\varphi_1<1, \quad |\varphi_2|<1$$

这就是**AR(2)的平稳性条件**，或称为**平稳域**。它是一顶点分别为（-2,-1），（2,-1），（0,1）的三角形。

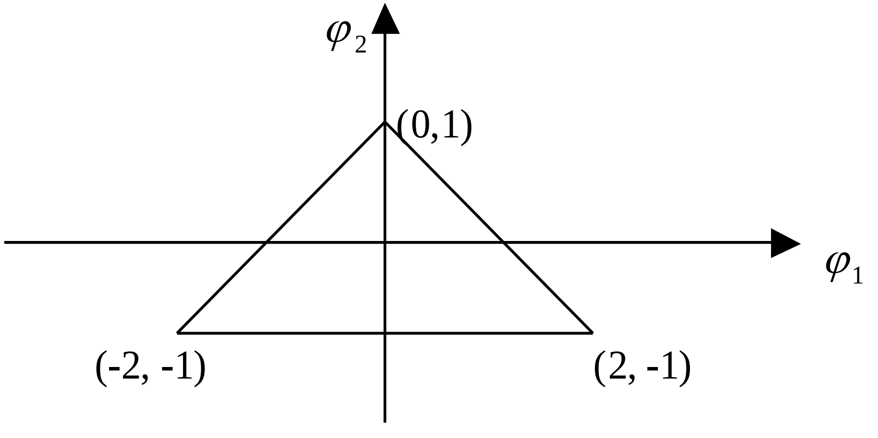


图 9.2.1　　AR(2)模型的平稳域

AR(2)模型

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$$

对应的特征方程$1-\varphi_1 z-\varphi_2 z^2=0$ 的两个根z_1、z_2满足：

$$z_1 z_2 = -1/\varphi_2 \quad , \quad z_1+z_2 = -\varphi_1/\varphi_2$$

解出φ_1，φ_2

$$\varphi_2 = -\frac{1}{z_1 z_2} \qquad \varphi_1 = \frac{z_1 + z_2}{z_1 z_2}$$

由AR(2)的平稳性，$|\varphi_2|=1/|z_1||z_2|<1$，则至少有一个根的模大于1，不妨设$|z_1|>1$，有

$$\varphi_1 + \varphi_2 = \frac{z_1 + z_2}{z_1 z_2} - \frac{1}{z_1 z_2} = 1 - (1-\frac{1}{z_1})(1-\frac{1}{z_2}) < 1$$

$$(1-\frac{1}{z_1})(1-\frac{1}{z_2}) > 0$$

于是$|z_2|>1$。由 $\varphi_2 - \varphi_1 < 1$可推出同样的结果。

对高阶自回模型AR(p)来说，多数情况下没有必要直接计算其特征方程的特征根，但有**一些有用的规则可用来检验高阶自回归模型的稳定性**：

(1)AR(p)模型稳定的必要条件是：

$$\varphi_1+\varphi_2+\ldots+\varphi_p<1$$

(2) 由于φ_i (i=1, 2, …p) 可正可负，**AR(p)模型稳定的充分条件是**：

$$|\varphi_1|+|\varphi_2|+\ldots+|\varphi_p|<1$$

2、MA(q)模型的平稳性

对于移动平均模型MR(q)：

$X_t=\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}$

其中ε_t是一个白噪声，于是

$$E(X_t) = E(\varepsilon_t) - \theta_1 E(\varepsilon_{t-1}) - \cdots - \theta_q E(\varepsilon_q) = 0$$

$$\gamma_0 = \operatorname{var}\left(X_t\right) = (1+\theta_1^2+\cdots+\theta_q^2)\sigma_\varepsilon^2$$

$$\gamma_1 = \operatorname{cov}(X_t, X_{t-1}) = (-\theta_1+\theta_1\theta_2+\theta_2\theta_3+\cdots+\theta_{q-1}\theta_q)\sigma_\varepsilon^2$$

……

$$\gamma_{q-1} = \operatorname{cov}(X_t, X_{t-q+1}) = (-\theta_{q-1}+\theta_1\theta_q)\sigma_\varepsilon^2$$

$$\gamma_q = \operatorname{cov}(X_t, X_{t-q}) = -\theta_q\sigma_\varepsilon^2$$

当滞后期大于q时，Xt的自协方差系数为0。

因此:**有限阶移动平均模型总是平稳的**。

3、ARMA(p,q)模型的平稳性

由于ARMA (p,q)模型是AR(p)模型与MA(q)模型的组合：

$$X_t=\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\ldots+\varphi_p X_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q}$$

而MA(q)模型总是平稳的，因此ARMA (p,q)模型的平稳性取决于AR(p)部分的平稳性。

当AR(p)部分平稳时，则该ARMA(p,q)模型是平稳的，否则，不是平稳的。

最后

（1）一个平稳的时间序列总可以找到生成它的平稳的随机过程或模型；

（2）一个非平稳的随机时间序列通常可以通过差分的方法将它变换为平稳的，对差分后平稳的时间序列也可找出对应的平稳随机过程或模型。

因此，如果我们将一个非平稳时间序列通过d次差分，将它变为平稳的，然后用一个平稳的ARMA(p,q)模型作为它的生成模型，则我们就说该原始时间序列是一个自回归单整移动平均（autoregressive integrated moving average）时间序列，记为ARIMA(p,d,q)。

例如，一个ARIMA(2,1,2)时间序列在它成为平稳序列之前先得差分一次，然后用一个ARMA(2,2)模型作为它的生成模型的。

当然，一个ARIMA(p,0,0)过程表示了一个纯AR(p)平稳过程；一个ARIMA(0,0,q)表示一个纯MA(q)平稳过程。

三、随机时间序列模型的识别

所谓随机时间序列模型的识别，就是对于一个平稳的随机时间序列，找出生成它的合适的随机过程或模型，即判断该时间序列是遵循一纯AR过程、还是遵循一纯MA过程或ARMA过程。

所使用的工具主要是时间序列的**自相关函数**（autocorrelation function，**ACF**）及**偏自相关函数**（partial autocorrelation function，**PACF**）。

1、AR(p)过程

(1) 自相关函数ACF

1阶自回归模型AR(1)

$$X_t=\varphi X_{t-1}+\varepsilon_t$$

的k阶滞后自协方差为：

$$\gamma_k = E(X_{t-k}(\varphi X_{t-1}+\varepsilon_t)) = \varphi\gamma_{k-1} = \varphi^k\gamma_0 \qquad \kappa=1,2,\ldots$$

因此，AR(1)模型的自相关函数为

$$\rho_k = \gamma_k/\gamma_0 = \varphi^k \qquad \kappa=1,2,\ldots$$

由AR(1)的稳定性知$|\varphi|<1$，因此，$k\to\infty$时，呈指数形衰减，直到零。这种现象称为拖尾或称AR(1)有无穷记忆（infinite memory）。

注意，$\varphi<0$时，呈振荡衰减状。

2 阶自回归模型AR(2)

$$X_t=\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\varepsilon_t$$

该模型的方差γ_0以及滞后1期与2期的自协方差γ_1，γ_2分别为

$$\gamma_0=\varphi_1\gamma_1+\varphi_2\gamma_2+\sigma_\varepsilon^2 \qquad \begin{aligned}\gamma_1&=\varphi_1\gamma_0+\varphi_2\gamma_1\\ \gamma_2&=\varphi_1\gamma_1+\varphi_2\gamma_0\end{aligned}$$

类似地, 可写出**一般的k期滞后自协方差**:

$$\gamma_k=E(X_{t-k}(\varphi_1X_{t-1}+\varphi_2X_{t-2}+\varepsilon_t))=\varphi_1\gamma_{k-1}+\varphi_2 r_{k-2} \qquad (K=2,3,\ldots)$$

于是, AR(2)的**k 阶自相关函数**为:

$$\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2} \qquad (K=2,3,\ldots)$$

其中 :$\rho_1=\varphi_1/(1-\varphi_2)$, $\rho_0=1$

如果AR(2)稳定，则由$\varphi_1+\varphi_2<1$知$|\rho_k|$衰减趋于零，呈拖尾状。

至于衰减的形式，要看AR(2)特征根的实虚性，若为实根，则呈单调或振荡型衰减，若为虚根，则呈正弦波型衰减。

一般地，**p阶自回归模型AR(p)**

$X_t=\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\ldots\ \varphi_p X_{t-p}+\varepsilon_t$

k期滞后协方差为:

$$\gamma_k=E(X_{t-K}(\varphi_1 X_{t-1}+\varphi_2 X_{t-2}+\cdots+\varphi_p X_{t-p}+\varepsilon_t))$$
$$=\varphi_1\gamma_{k-1}+\varphi_2\gamma_{k-2}+\cdots+\varphi_p\gamma_{k-p}$$

从而有自相关函数：

$$\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}+\cdots+\varphi_p\rho_{k-p}$$

可见，**无论k有多大，ρ_k的计算均与其1到p阶滞后的自相关函数有关**，因此**呈拖尾状**。

如果AR(p)是稳定的，则$|\rho_k|$递减且趋于零。

事实上，自相关函数

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}$$

是一p阶差分方程，其通解为 $\rho_k = \sum_{i=1}^{p} C_i z_i^k$

其中：$1/z_i$是AR(p)特征方程$\Phi(z)=0$的特征根，由AR(p)平稳的条件知，$|z_i|<1$;

因此，当$1/z_i$均为实数根时，ρ_k呈几何型衰减（单调或振荡）；

当存在虚数根时，则一对共扼复根构成通解中的一个阻尼正弦波项，ρ_k呈正弦波衰减。

（2）偏自相关函数

自相关函数ACF(k)给出了X_t与X_{t-1}的总体相关性，但总体相关性可能掩盖了变量间完全不同的隐含关系。

例如，在AR(1)随机过程中，X_t与X_{t-2}间有相关性可能主要是由于它们各自与X_{t-1}间的相关性带来的：

$$\rho_2 = \varphi^2 = \rho_1^2 = E(X_t X_{t-1})E(X_{t-1}X_{t-2})$$

即自相关函数中包含了这种所有的“间接”相关。

与之相反，X_t与X_{t-k}间的偏自相关函数(partial autocorrelation，简记为PACF)则是消除了中间变量X_{t-1}，…，X_{t-k+1} 带来的间接相关后的直接相关性，它是在已知序列值X_{t-1}，…，X_{t-k+1}的条件下，X_t与X_{t-k}间关系的度量。

在AR(1)中，

从X_t中去掉X_{t-1}的影响，则只剩下随机扰动项ε_t，显然它与X_{t-2}无关，因此我们说X_t与X_{t-2}的**偏自相关系数**为零，记为

$$\rho_2^* = Corr(\varepsilon_t, X_{t-2}) = 0$$

同样地，**在AR(p)过程中**，对所有的k>p，X_t与X_{t-k}间的**偏自相关系数**为零。

AR(p)的一个主要特征是:k>p时，$\rho_k^*=Corr(X_t,X_{t-k})=0$

即ρ_k^*在p以后是截尾的。

一随机时间序列的识别原则:

若Xt的偏自相关函数在p以后截尾，即k>p时，$\rho_k^*=0$，而它的自相关函数ρ_k是拖尾的，则此序列是自回归AR(p)序列。

需指出的是，

在实际识别时，由于样本偏自相关函数r_k^*是总体偏自相关函数ρ_k^*的一个估计，由于样本的随机性，当**k>p**时，r_k^*不会全为0，而是在0的上下波动。但可以证明，当**k>p**时，r_k^*服从如下渐近正态分布：

$$r_k^* \sim N(0,1/n)$$

式中n表示样本容量。

因此，如果计算的r_k^*满足

$$|r_k^*| \prec \frac{2}{\sqrt{n}}$$

我们就有95.5%的把握判断原时间序列在p之后截尾。

2、MA(q)过程

对MA(1)过程

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}$$

可容易地写出它的**自协方差系数**：

$$\gamma_0 = (1+\theta^2)\sigma_\varepsilon^2$$

$$\gamma_1 = -\theta\sigma_\varepsilon^2$$

$$\gamma_2 = \gamma_3 = \cdots = 0$$

于是，MA(1)过程的**自相关函数**为：

$$\rho_1 = \frac{-\theta}{(1+\theta^2)}$$

$$\rho_2 = \rho_3 = \cdots = 0$$

可见，当**k>1时**，**ρ_k>0**，即**X_t与X_{t-k}不相关**，**MA(1)自相关函数是截尾的。**

MA(1)过程可以等价地写成ε_t关于无穷序列X_t，X_{t-1}，… 的线性组合的形式：

$$\varepsilon_t = X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \cdots$$

或

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \cdots + \varepsilon_t \qquad (*)$$

(*)是一个AR(∞)过程，它的偏自相关函数非截尾但却趋于零，因此**MA(1)的偏自相关函数是非截尾但却趋于零的。**

注意：

(*)式只有当$|\theta|<1$时才有意义，否则意味着距Xt越远的X值，对Xt的影响越大，显然不符合常理。

因此，我们**把$|\theta|<1$称为MA(1)的可逆性条件(invertibility condition)或可逆域。**

一般地，q阶移动平均过程MA(q)

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}$$

其自协方差系数为

$$r_k = E(X_t X_{t-k}) = \begin{cases} \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2) & \text{当} k=0 \\ \sigma_\varepsilon^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q) & \text{当} 1 \le k \le q \\ 0 & \text{当} k > q \end{cases}$$

相应的自相关函数为

$$\rho_k = \frac{r_k}{r_0} = \begin{cases} 1 & \text{当} k=0 \\ (-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)/(1+\theta_1^2+\cdots+\theta_q^2) & \text{当} 1 \le k \le q \\ 0 & \text{当} k > q \end{cases}$$

可见，当k>q时， X_t与X_{t-k}不相关，即存在截尾现象，因此，当k>q时， ρ_k=0是MA(q)的一个特征。

于是：可以根据自相关系数是否从某一点开始一直为0来判断MA(q)模型的阶。

与MA(1)相仿，可以验证MA(q)过程的偏自相关函数是非截尾但趋于零的。

MA(q)模型的识别规则：若随机序列的自相关函数截尾，即自q以后，$\rho_k=0$（ k>q）；而它的偏自相关函数是拖尾的，则此序列是滑动平均MA(q)序列。

同样需要注意的是：在实际识别时，由于样本自相关函数r_k是总体自相关函数ρ_k的一个估计，由于样本的随机性，当**k>q**时，r_k不会全为0，而是在0的上下波动。但可以证明，当**k>q**时，r_k服从如下渐近正态分布：

$$r_k \sim N(0,1/n)$$

式中n表示样本容量。

因此，**如果计算的r_k满足：** $|r_k| < \dfrac{2}{\sqrt{n}}$

我们**就有95.5%的把握判断原时间序列在q之后截尾**。

3、ARMA(p，q)过程

ARMA(p,q)的自相关函数，可以看作MA(q)的自相关函数和AR(p)的自相关函数的混合物。

当p=0时，它具有截尾性质；

当q=0时，它具有拖尾性质；

当p、q都不为0时，它具有拖尾性质

从识别上看，通常：

ARMA(p，q)过程的偏自相关函数（PACF）可能在p阶滞后前有几项明显的尖柱（spikes），但从p阶滞后项开始逐渐趋向于零；

而它的自相关函数（ACF）则是在q阶滞后前有几项明显的尖柱，从q阶滞后项开始逐渐趋向于零。

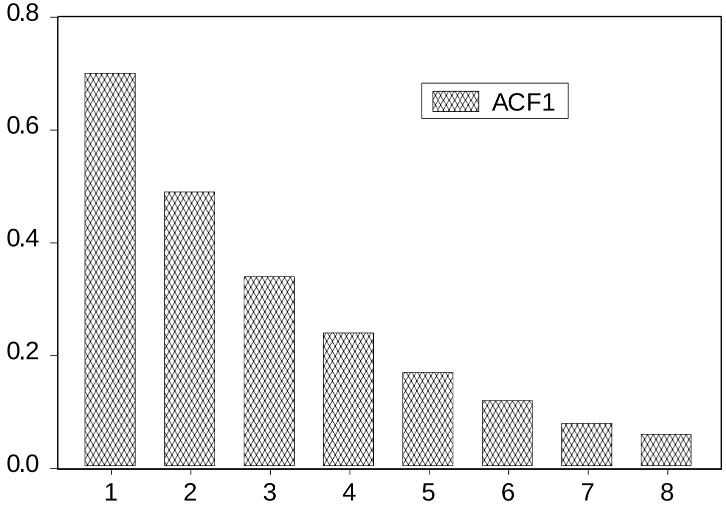
表 9.2.1　　　ARMA(p,q)模型的 ACF 与 PACF 理论模式

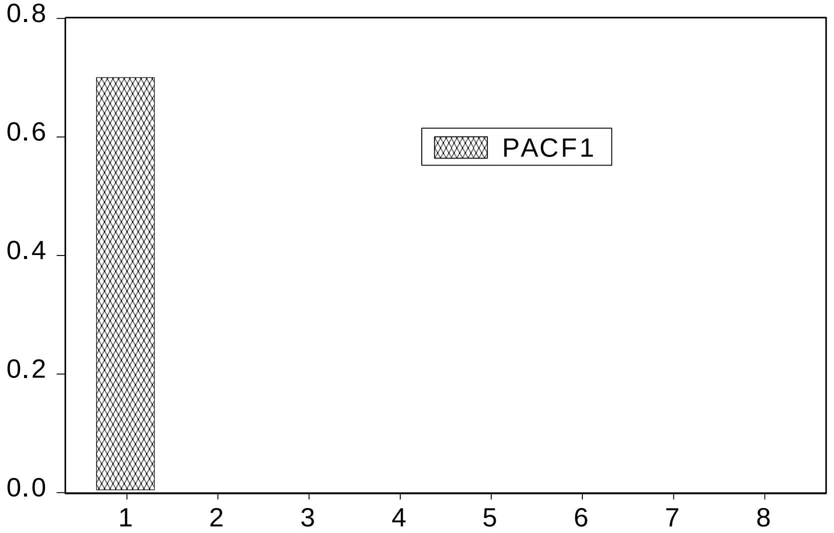
模型	ACF	PACF
白噪声	$\rho_k = 0$	$\rho_k^* = 0$
AR(p)	衰减趋于零（几何型或振荡型）	P 阶后截尾：$\rho_k^* = 0$ ，k>p
MA(q)	q 阶后截尾：，$\rho_k = 0$ ，k>q	衰减趋于零（几何型或振荡型）
ARMA(p,q)	q 阶后衰减趋于零（几何型或振荡型）	p 阶后衰减趋于零（几何型或振荡型）

图 9.2.2　ARMA(p,q)模型的 ACF 与 PACF 理论模式

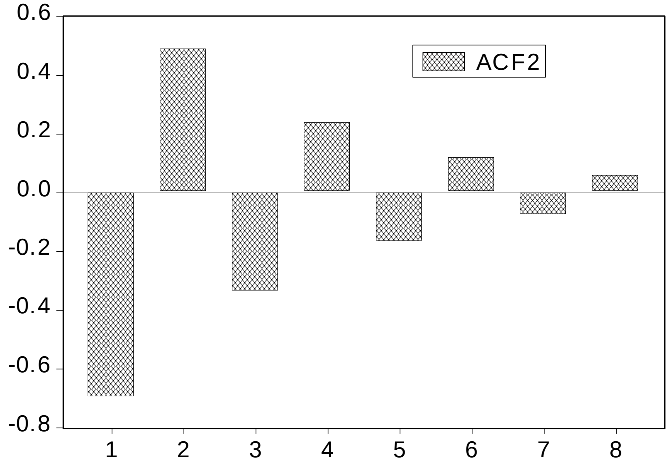
ACF　　　　　　　　PACF

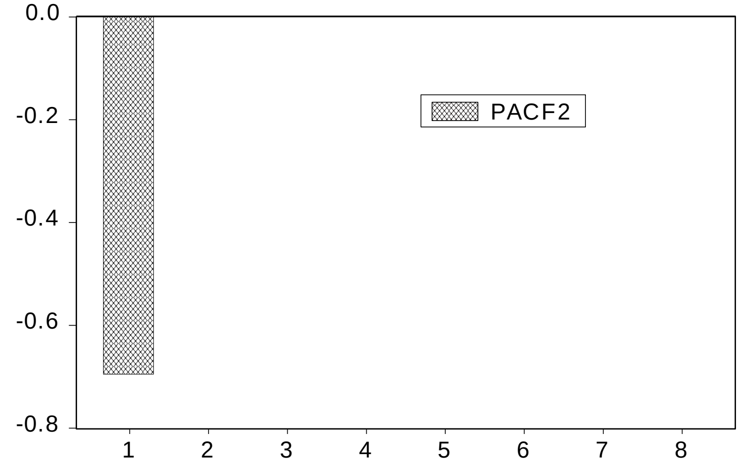
模型 1： $X_t = 0.7X_{t-1} + \varepsilon_t$



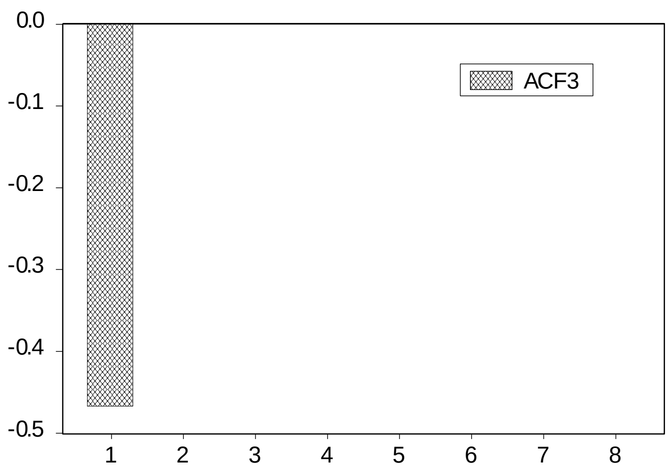


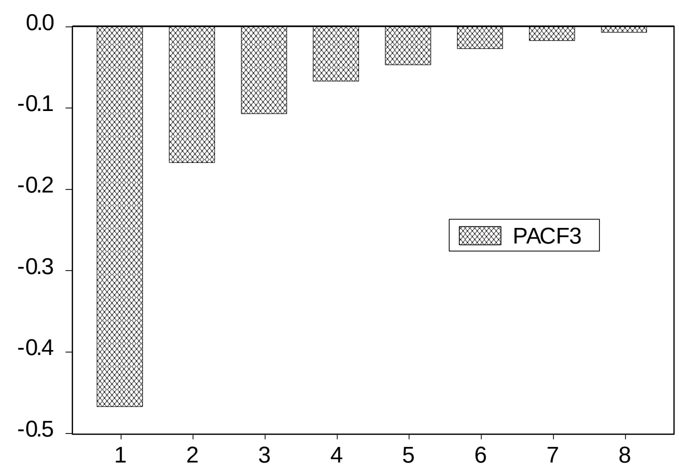
模型 2:　$X_t = -0.7X_{t-1} + \varepsilon_t$



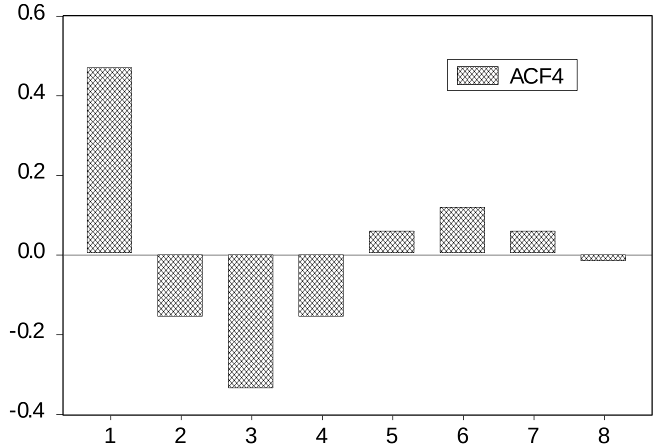


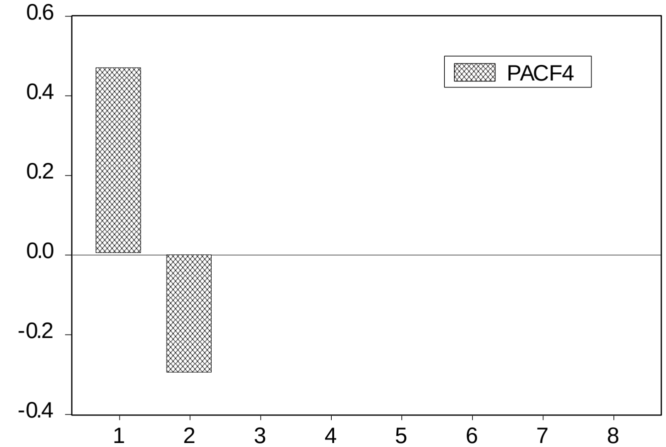
模型 3:　$X_t = \varepsilon_t - 0.7\varepsilon_{t-1}$



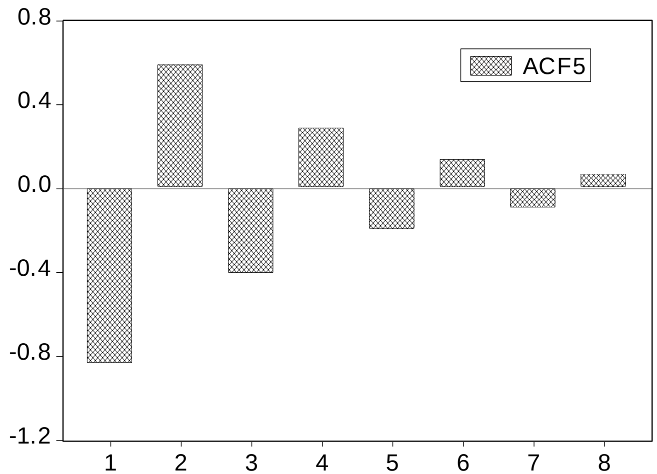


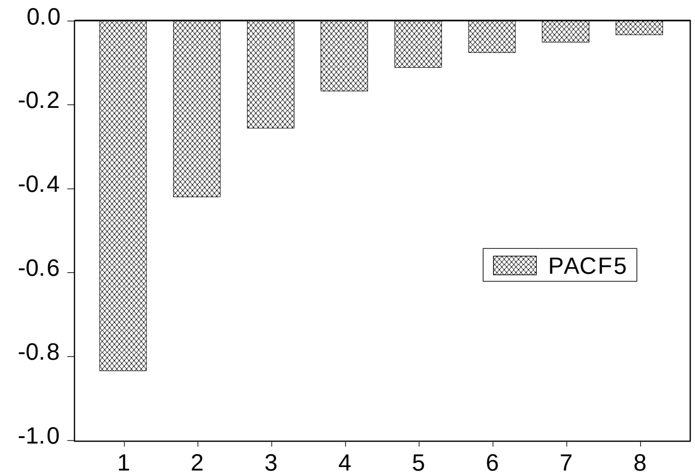
模型 4：$X_t = 0.7X_{t-1} - 0.49X_{t-2} + \varepsilon_t$



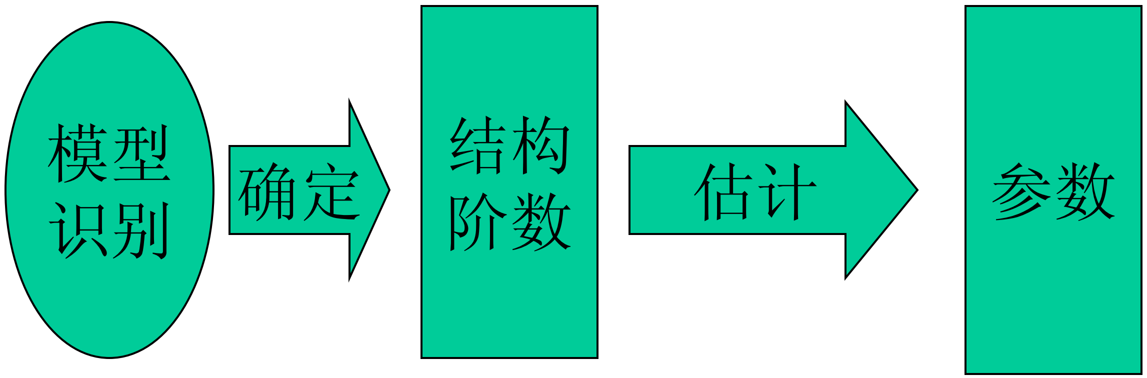


模型 5：$X_t = -0.7X_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-1}$





四、随机时间序列模型的估计



AR(p)、MA(q)、ARMA(p,q)模型的估计方法较多，**大体上分为3类：**

（1）最小二乘估计；

（2）矩估计；

（3）利用自相关函数的直接估计。

下面有选择地加以介绍。

1. AR(p)模型的Yule Walker方程估计

在AR(p)模型的识别中，曾得到

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \cdots + \varphi_p\rho_{k-p}$$

利用$\rho_k=\rho_{-k}$，得到如下方程组：

$$\rho_1 = \varphi_1 + \varphi_2\rho_1 + \cdots + \varphi_p\rho_{p-'1}$$

$$\rho_2 = \varphi_1\rho_1 + \varphi_2 + \cdots + \varphi_p\rho_{p-2}$$

……

$$\rho_p = \varphi_1\rho_{p-1} + \varphi_2\rho_{p-1} + \cdots\varphi_p\rho_{p-k}$$

此方程组被称为Yule Walker方程组。该方程组建立了AR(p)模型的模型参数$\varphi_1, \varphi_2, \ldots, \varphi_p$与自相关函数$\rho_1, \rho_2, \ldots, \rho_p$的关系，

利用实际时间序列提供的信息，首先求得自相关函数的估计值 $\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_p$

然后利用Yule Walker方程组，求解模型参数的估计值 $\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{p-2} \\ \vdots & & & \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \hat{\rho}_0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}$$

由于 $\varepsilon_t = X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}$ 于是

$$\sigma_\varepsilon^2 = E\varepsilon_t^2 = \cdots = \gamma_0 - \sum_{i,j=1}^{p} \varphi_i \varphi_j \gamma_{j-i}$$

从而可得**σ_ε^2的估计值** $\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0 - \sum_{i,j=1}^{p} \hat{\varphi}_i \hat{\varphi}_j \hat{\gamma}_{j-i}$

在具体计算时，$\hat{\rho}_k$ 可用样本自相关函数r_k替代。

2. MA(q)模型的矩估计

将MA(q)模型的自协方差函数中的各个量用估计量代替，得到：

$$\hat{\gamma}_k = \begin{cases} \hat{\sigma}_\varepsilon^2(1+\hat{\theta}_1^2+\hat{\theta}_2^2+\cdots+\hat{\theta}_q^2) & \text{当} k=0 \\ \hat{\sigma}_\varepsilon^2(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q) & \text{当} 1\le k\le q \\ 0 & \text{当} k>q \end{cases} \qquad (*)$$

首先求得自协方差函数的估计值，(*)是一个包含(q+1)个待估参数

$$\hat{\theta}_1,\hat{\theta}_2\cdots\hat{\theta}_q,\hat{\sigma}_\varepsilon^2$$

的非线性方程组，可以用直接法或迭代法求解。

常用的迭代方法有线性迭代法和Newton-Raphsan迭代法。

（1）MA(1)模型的直接算法

对于MA(1)模型，（*）式相应地写成

$$\begin{cases}\hat{\gamma}_0 = \hat{\sigma}_\varepsilon^2(1+\hat{\theta}_1^2) \\ \hat{\gamma}_1 = -\hat{\sigma}_\varepsilon^2\hat{\theta}_1\end{cases}$$

于是

$$\hat{\theta}_1 = -\hat{\gamma}_1 / \hat{\sigma}_\varepsilon^2$$

有 $\hat{\sigma}_\varepsilon^4 - \hat{\gamma}_0\hat{\sigma}_\varepsilon^2 + \hat{\gamma}_1^2 = 0$ 或 $\hat{\gamma}_0^{-1}\hat{\sigma}_\varepsilon^4 - \hat{\sigma}_\varepsilon^2 + \hat{\rho}_1^2 = 0$

于是有解

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\gamma}_0}{2}(1 \pm \sqrt{1-4\hat{\rho}_1^2})$$

$$\hat{\theta}_1 = -\hat{\gamma}_1 / \hat{\sigma}_\varepsilon^2 = -2\hat{\rho}_1 / (1 \pm \sqrt{1-4\hat{\rho}_1^2})$$

由于参数估计有两组解，可根据可逆性条件$|\theta_1|<1$来判断选取一组。

（2）MA(q)模型的迭代算法

对于q>1的MA(q)模型，一般用迭代算法估计参数：

由（*）式得

$$\begin{cases} \hat{\sigma}_\varepsilon^2 = \dfrac{\hat{\gamma}_0}{1+\hat{\theta}_1^2+\cdots+\hat{\theta}_q^2} \\ \hat{\theta}_k = -\left(\dfrac{\hat{\gamma}_k}{\hat{\sigma}_\varepsilon^2} - \hat{\theta}_1\hat{\theta}_{k+1} - \hat{\theta}_2\hat{\theta}_{k+2} - \cdots - \hat{\theta}_{q-k}\hat{\theta}_q\right) \end{cases} \qquad (**)$$

第一步，给出 $\hat{\sigma}_\varepsilon^2, \hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ 的一组初值，比如

$$\hat{\sigma}_\varepsilon^2(0) = \hat{\gamma}_0 \qquad \hat{\theta}_1(0) = \hat{\theta}_2(0) = \cdots \hat{\theta}_k(0) = 0$$

代入（**）式，计算出第一次迭代值

$$\hat{\sigma}_\varepsilon^2(1) = \hat{\gamma}_0 \qquad \hat{\theta}_k(1) = -\hat{\gamma}_k / \hat{\gamma}_0$$

第二步，将第一次迭代值代入（**）式，计算出第二次迭代值

$$\hat{\sigma}_{\varepsilon}^{2}(2) = \hat{\gamma}_0 /(1 + \hat{\theta}_1^2(1) + \cdots + \hat{\theta}_q^2(1))$$

$$\hat{\theta}_k(2) = -(\hat{\gamma}_k \big/ \hat{\gamma}_0 - \hat{\theta}_1(1)\hat{\theta}_{k+1}(1) - \cdots - \hat{\theta}_{q-k}(1)\hat{\theta}_q(1))$$

按此反复迭代下去，直到第m步的迭代值与第m-1步的迭代值相差不大时（满足一定的精度），便停止迭代，并用第m步的迭代结果作为（**）的近似解。

3. ARMA(p,q)模型的矩估计

在ARMA(p,q)中共有(p+q+1)个待估参数$\varphi_1,\varphi_2,\ldots,\varphi_p$与$\theta_1,\theta_2,\ldots,\theta_q$以及$\sigma_\varepsilon^2$，其估计量计算步骤及公式如下：

第一步，估计$\varphi_1,\varphi_2,\ldots,\varphi_p$

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p+1} \\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p} \\ \vdots & & & \\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \cdots & \hat{\rho}_q \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_{q+1} \\ \hat{\rho}_{q+2} \\ \vdots \\ \hat{\rho}_{q+p} \end{bmatrix}$$

$\hat{\rho}_k$ 是总体自相关函数的估计值，可用样本自相关函数r_k代替。

第二步，改写模型，求$\theta_1,\theta_2,\ldots,\theta_q$以及$\sigma_\varepsilon^2$的估计值

将模型

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

改写为：

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} - \cdots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \qquad (*)$$

令 $\tilde{X}_t = X_t - \hat{\varphi}_1 X_{t-1} - \hat{\varphi}_2 X_{t-2} - \cdots - \hat{\varphi}_p X_{t-p}$

于是(*)可以写成：

$$\tilde{X}_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

构成一个MA模型。按照估计MA模型参数的方法，可以得到$\theta_1,\theta_2,\ldots,\theta_q$以及$\sigma_\varepsilon^2$的估计值。

4. AR(p)的最小二乘估计

假设模型AR(p)的参数估计值已经得到，即有

$$X_t = \hat{\varphi}_1 X_{t-1} + \hat{\varphi}_2 X_{t-2} + \cdots + \hat{\varphi}_p X_{t-p} + \hat{\varepsilon}_t$$

残差的平方和为：

$$S(\hat{\varphi}) = \sum_{t=p+1}^{n} \hat{\varepsilon}_t^2 = \sum_{t=p+1}^{n} (X_t - \hat{\varphi}_1 X_{t-1} - \hat{\varphi}_2 X_{t-2} - \cdots - \hat{\varphi}_p X_{t-p})^2 \qquad (*)$$

根据最小二乘原理，所要求的参数估计值是下列方程组的解：

$$\frac{\partial S}{\partial \hat{\varphi}_j} = 0$$

即

$$\sum_{t=p+1}^{n} (X_t - \hat{\varphi}_1 X_{t-1} - \hat{\varphi}_2 X_{t-2} - \cdots - \hat{\varphi}_p X_{t-p}) X_{t-j} = 0 \qquad j=1,2,\ldots,p \qquad (**)$$

解该方程组，就可得到待估参数的估计值。

为了与AR(p)模型的Yule Walker方程估计进行比较，将(**)改写成：

$$\frac{\hat{\varphi}_1}{n}\sum_{t=p+1}^{n}X_{t-1}X_{t-j}+\frac{\hat{\varphi}_2}{n}\sum_{t=p+1}^{n}X_{t-2}X_{t-j}+\cdots+\frac{\hat{\varphi}_p}{n}\sum_{t=p+1}^{n}X_{t-p}X_{t-j}=\frac{1}{n}\sum_{t=p+1}^{n}X_tX_{t-j}$$

j=1,2,…,p

由自协方差函数的定义，并用自协方差函数的估计值

$$\hat{\gamma}_k=\frac{1}{n}\sum_{t=p+1}^{n-k}X_{t+k}X_t$$

代入，上式表示的方程组即为：

$$\hat{\varphi}_1\hat{\gamma}_{j-1}+\hat{\varphi}_2\hat{\gamma}_{j-2}+\cdots+\hat{\varphi}_p\hat{\gamma}_{j-p}=\hat{\gamma}_j \qquad \text{j=1,2,…,p}$$

或

$$\hat{\varphi}_1 r_{j-1}+\hat{\varphi}_2 r_{j-2}+\cdots+\hat{\varphi}_p r_{j-p}=r_j \qquad \text{j=1,2,…,p}$$

解该方程组，得到：

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & \cdots & r_{p-2} \\ \vdots & & & \\ r_{p-1} & r_{p-2} & \cdots & r_0 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}$$

即为参数的最小二乘估计。

Yule Walker方程组的解

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{p-2} \\ \vdots & & & \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \hat{\rho}_0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}$$

比较发现，当n足够大时，二者是相似的。 σ_ε^2的估计值为：

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p} \sum_{t=p+1}^{n} \varepsilon_t^2 = \frac{S}{n-p}$$

需要说明的是，**在上述模型的平稳性、识别与估计的讨论中，ARMA(p,q)模型中均未包含常数项。**

如果包含常数项，该常数项并不影响模型的原有性质，因为通过适当的变形，可将包含常数项的模型转换为不含常数项的模型。

下面以一般的ARMA(p,q)模型为例说明。

对含有常数项的模型

$$X_t = \alpha + \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

方程两边同减$\alpha/(1-\varphi_1-\ldots-\varphi_p)$，则可得到

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

其中 $x_i = X_i - \alpha/(1-\varphi_1 - \cdots - \varphi_p) \qquad i = t, t-1, \cdots, t-p$

五、模型的检验

1、残差项的白噪声检验

由于ARMA(p,q)模型的识别与估计是在假设随机扰动项是一白噪声的基础上进行的，因此，**如果估计的模型确认正确的话，残差应代表一白噪声序列**。

如果通过所估计的模型计算的样本残差不代表一白噪声，则说明模型的识别与估计有误，需重新识别与估计。

在实际检验时，主要检验残差序列是否存在自相关。

可用QLB的统计量进行χ2检验：在给定显著性水平下，可计算不同滞后期的**QLB**值，通过与**χ2**分布表中的相应临界值比较，来检验是否拒绝残差序列为白噪声的假设。

若大于相应临界值，则应拒绝所估计的模型，需重新识别与估计。

2、AIC与SBC模型选择标准

另外一个遇到的问题是，在实际识别ARMA(p,q)模型时，需多次反复偿试，有可能存在不止一组（p,q）值都能通过识别检验。

显然，**增加p与q的阶数，可增加拟合优度，但却同时降低了自由度。**

因此，**对可能的适当的模型，存在着模型的“简洁性”与模型的拟合优度的权衡选择问题。**

常用的模型选择的判别标准有：赤池信息法（Akaike information criterion，简记为**AIC**）与施瓦兹贝叶斯法（Schwartz Bayesian criterion，简记为**SBC**）：

$$AIC = T\ln(RSS) + 2n$$

$$SBC = T\ln(RSS) + n\ln(T)$$

其中，n为待估参数个数（p+q+可能存在的常数项），T为可使用的观测值，RSS为残差平方和（Residual sum of squares）。

在选择可能的模型时，**AIC与SBC越小越好**

显然，如果添加的滞后项没有解释能力，则对RSS值的减小没有多大帮助，却增加待估参数的个数，因此使得AIC或SBC的值增加。

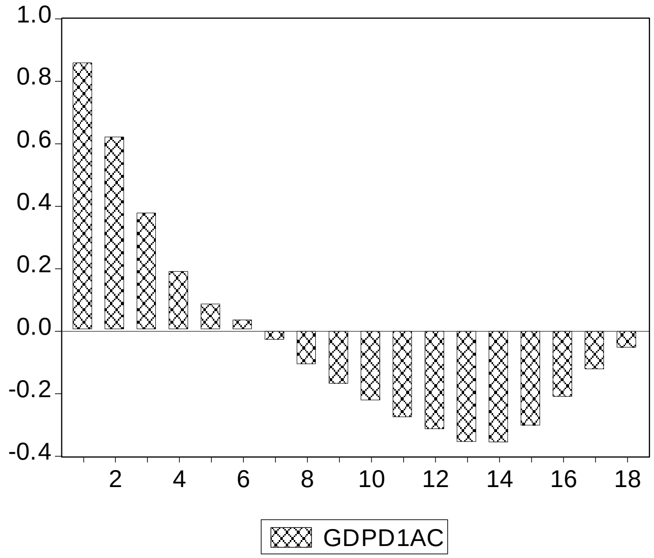
需注意的是：在不同模型间进行比较时，必须选取相同的时间段。

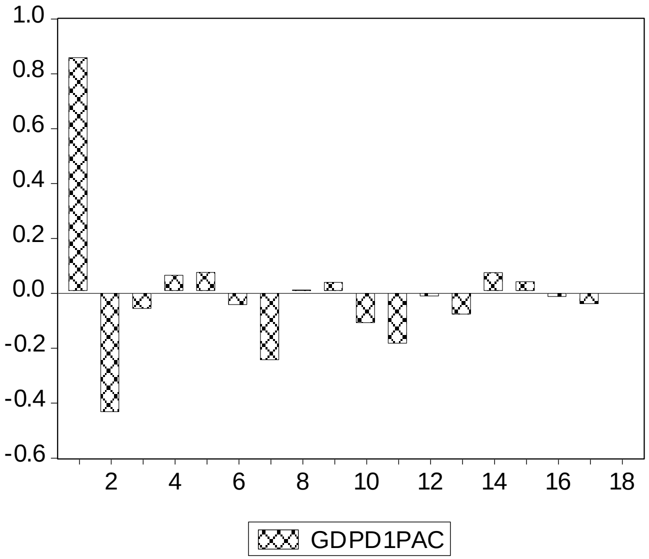
例9.2.3 中国支出法GDP的ARMA(p,q)模型估计。

由第一节知：中国支出法GDP是非平稳的，但它的一阶差分是平稳的，即支出法GDP是I(1)时间序列。

可以对经过一阶差分后的GDP建立适当的ARMA(p,q)模型。

记GDP经一阶差分后的新序列为GDPD1，该新序列的样本自相关函数图与偏自相关函数图如下：





图形：样本自相关函数图形呈正弦线型衰减波，而偏自相关函数图形则在滞后两期后迅速趋于0。因此**可初步判断该序列满足2阶自回归过程AR(2)。**

自相关函数与**偏自相关函数**的**函数值：**

相关函数具有明显的拖尾性；

偏自相关函数值在k>2以后，$|r_k^*| < 2/\sqrt{22} \approx 0.426$

可认为：偏自相关函数是截尾的。再次验证了一阶差分后的GDP满足AR(2)随机过程。

表 9.2.2　中国 GDP 一阶差分序列的样本自相关函数与偏自相关函数

k	r_k	r_k^*	k	r_k	r_k^*	k	r_k	r_k^*
1	0.859	0.859	7	-0.034	-0.252	13	-0.361	-0.086
2	0.622	-0.441	8	-0.112	0.012	14	-0.363	0.076
3	0.378	-0.065	9	-0.175	0.04	15	-0.308	0.043
4	0.191	0.066	10	-0.228	-0.117	16	-0.216	-0.022
5	0.087	0.077	11	-0.282	-0.192	17	-0.128	-0.048
6	0.036	-0.051	12	-0.32	-0.02	18	-0.059	-0.002

设序列GDPD1的模型形式为

$$GDPD1_t = \varphi_1 GDPD1_{t-1} + \varphi_2 GDPD1_{t-2} + \varepsilon_t$$

有如下Yule Walker 方程：

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0.859 \\ 0.859 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.859 \\ 0.622 \end{pmatrix}$$

解为： $\hat{\varphi}_1 = 1.239, \hat{\varphi}_2 = -0.442$

用OLS法回归的结果为：

$$GDPD1_t = 1.593 GDPD1_{t-1} - 0.653 GDPD1_{t-2} + \varepsilon_t$$

（7.91) (-3.60)

r^2=0.8469 R^2=0.8385 DW=1.15

有时，在用回归法时，也可加入常数项。

本例中加入常数项的回归为：

$$GDPD1_t = 909.59 + 1.495 GDPD1_{t-1} - 0.678 GDPD1_{t-2} + \varepsilon_t$$

（1.99） （7.74） （-3.58）

r^2 =0.8758 R^2 =0.8612 DW.=1.22

- **模型检验**

下表列出三模型的残差项的自相关系数及Q_{LB}检验值。

模型1与模型3的残差项接近于一白噪声，但模型2存在4阶滞后相关问题，Q统计量的检验也得出模型2拒绝所有自相关系数为零的假设。因此：

模型1与3可作为描述中国支出法GDP一阶差分序列的随机生成过程。

表 9.2.3　模型残差项的自相关系数及 Q 检验值

	模型1		模型2		模型3	
K	Resid-ACF	Q	Resid-ACF	Q	Resid-ACF	Q
1	0.382	3.3846	0.258	1.5377	0.257	1.5263
2	0.014	3.3893	-0.139	2.0077	-0.040	1.5646
3	-0.132	3.8427	-0.246	3.5677	-0.059	1.6554
4	-0.341	7.0391	-0.529	11.267	-0.328	4.6210
5	-0.170	7.8910	-0.300	13.908	-0.151	5.2864
6	0.253	9.9097	0.271	16.207	0.345	9.0331
7	0.144	10.613	0.158	17.051	0.155	9.8458
8	0.057	10.730	0.116	17.541	0.076	10.059
9	-0.019	10.745	0.097	17.914	0.011	10.064
10	-0.146	11.685	-0.036	17.969	-0.123	10.728
11	-0.233	14.329	-0.136	18.878	-0.230	13.319
12	-0.049	14.461	0.064	19.104	-0.012	13.328

- **用建立的AR(2)模型对中国支出法GDP进行外推预测。**

模型1可作如下展开：

$$GDP_t - GDP_{t-1} = \varphi_1(GDP_{t-1} - GDP_{t-2}) + \varphi_2(GDP_{t-2} - GDP_{t-3})$$

$$GDP_t = (1+\varphi_1)GDP_{t-1} + (\varphi_2 - \varphi_1)GDP_{t-2} - \varphi_2 GDP_{t-3}$$

于是，当已知t-1、t-2、t-3期的GDP时，就可对第t期的GDP作出外推预测。

模型3的预测式与此相类似，只不过多出一项常数项。

对2001年中国支出法GDP的预测结果（亿元）

	预测值	实际值	误差
模型1	95469	95933	-0.48%
模型3	97160	95933	1.28%

例9.2.4 中国人均居民消费的ARMA(p, q)模型

由于中国人均居民消费（CPC）与人均国内生产总值（GDPPC）这两时间序列是非平稳的，因此不宜直接建立它们的因果关系回归方程。

但它们都是I(2)时间序列，因此可以建立它们的ARIMA(p, d, q)模型。

下面只建立中国人均居民消费（CPC）的随机时间序列模型。

中国人均居民消费（CPC）经过二次差分后的新序列记为CPCD2，其自相关函数、偏自相关函数及Q统计量的值列于下表：

表 9.2.4　CPCD2 序列的自相关函数、偏自相关函数与 Q 统计量值

k	ACF	PACF	Q	k	ACF	PACF	Q
1	0.125	0.125	0.269	7	0.196	0.014	6.286
2	-0.294	-0.314	1.882	8	-0.218	-0.335	8.067
3	-0.034	0.060	1.906	9	-0.010	0.024	8.072
4	-0.213	-0.350	2.919	10	0.102	-0.147	8.650
5	-0.258	-0.193	4.576	11	-0.071	0.001	9.025
6	0.131	0.017	5.057	12	0.006	-0.119	9.029

在5%的显著性水平下，通过Q统计量容易验证该序列本身就接近于一白噪声，因此可**考虑采用零阶MA(0)模型**：

$$CPCD2_t = \varepsilon_t$$

由于k=2时，$|r_2| = |-0.29| > 1/\sqrt{14}$

因此，也可考虑采用下面的MA模型：

$$CPCD2_t = \varepsilon_t - \theta_2 \varepsilon_{t-2}$$

当然，还可观察到自相关函数在滞后4、5、8时有大于0.2的函数值，因此，可考虑在模型中增加MA(4)、MA(5)、MA(8)。不同模型的回归结果列于表9.2.5。

可以看出:在纯MA模型中，模型4具有较好的性质，但由于MA(5)的t检验偏小，因此可选取模型3。

表 9.2.5　中国居民人均消费水平的 ARMA 模型

模型	a	MA(2)	MA(4)	MA(5)	MA(8)	AR(1)	R2	SSR	AIC
1	24.57						0	93137.4	8.94
2	32.4	-0.89					0.42	53699.9	8.54
	(3.62)	(-7.43)							
3	14.07	-0.72	-1.71				0.7	28128.8	8.03
	(8.75)	(-3.07)	(-5.08)						
4	11.73	-1.09	-1.99	-1.3			0.82	17480.8	7.7
	(17.81)	(-3.38)	(-4.61)	(-1.58)					
5	11.79	-1.07	-1.91	-1.25	-0.34		0.81	17402.7	7.84
	(14.93)	(-3.10)	(-2.56)	(-1.42)	(-0.15)				
6	14.95	-0.66	-1.27		-1.99		0.75	22924.2	7.97
	(5.16)	(-2.14)	(-1.77)		(-1.29)				
7	214.25	-2.53	-2.45		-6.52	1.39	0.99	8943.7	7.06
	(63.83)	(-2.25)	(-2.53)		(-2.23)	(98.26)			

最后，给出通过模型3的外推预测。

模型3的展开式为：

$$\Delta^2 CPC_t = \Delta CPC_t - \Delta CPC_{t-1} = (CPC_t - CPC_{t-1}) - (CPC_{t-1} - CPC_{t-2})$$
$$= CPC_t - 2CPC_{t-1} + CPC_{t-2} = 14.07 + \varepsilon_t - 0.72\,\varepsilon_{t-2} - 1.71\varepsilon_{t-4}$$

即 $CPC_t = 2CPC_{t-1} - CPC_{t-2} + 14.07 + \varepsilon_t - 0.72\varepsilon_{t-2} - 1.71\varepsilon_{t-4}$

由于ε_t表示预测期的随机扰动项，它未知，可假设为0，于是t期的预测式为：

$$CPC_t = 2CPC_{t-1} - CPC_{t-2} + 14.07 - 0.72\hat{\varepsilon}_{t-2} - 1.71\hat{\varepsilon}_{t-4}$$

$\hat{\varepsilon}_{t-2}$ $\hat{\varepsilon}_{t-4}$ 为模型3中滞后2期与滞后4期的相应残差项的估计值。

表9.2.6列出了采用模型3对中国居民人均居民消费水平的2期外推预测。

为了对照，表中也同时列出了采用§2.10的模型的预测结果。

表 9.2.6　　中国居民人均消费水平 2 期外推预测比较（单位：元）

	实际值	ARMA模型		因果关系模型	
		预测值	相对误差（%）	预测值	相对误差（%）
1997	2834	3048	7.6	2822	-0.4
1998	2972	3407	14.6	2977	0.2

§9.3 协整与误差修正模型

一、长期均衡关系与协整

二、协整检验

三、误差修正模型

一、长期均衡关系与协整

0、问题的提出

- **经典回归模型（classical regression model）**是建立在稳定数据变量基础上的，对于非稳定变量，不能使用经典回归模型，否则会出现**虚假回归**等诸多问题。
- 由于许多经济变量是非稳定的，这就给经典的回归分析方法带来了很大限制。
- 但是，如果变量之间有着长期的稳定关系，**即它们之间是协整的（cointegration)，则**是可以使用经典回归模型方法建立回归模型的。
- 例如，**中国居民人均消费水平与人均GDP变量的例子中：**

因果关系回归模型要比ARMA模型有更好的预测功能，

其原因在于，从经济理论上说，人均GDP决定着居民人均消费水平，而且它们之间有着长期的稳定关系，即它们之间是协整的（cointegration）。

1、长期均衡

经济理论指出，某些经济变量间确实存在着长期均衡关系，这种均衡关系意味着经济系统不存在破坏均衡的内在机制，如果变量在某时期受到干扰后偏离其长期均衡点，则均衡机制将会在下一期进行调整以使其重新回到均衡状态。

假设X与Y间的长期“均衡关系”由式描述

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$

式中：μt是随机扰动项。

该均衡关系意味着：给定X的一个值，Y相应的均衡值也随之确定为 $\alpha_0+\alpha_1 X$。

在t-1期末，存在下述三种情形之一：

（1）Y等于它的均衡值：$Y_{t-1}= \alpha_0+\alpha_1 X_t$；

（2）Y小于它的均衡值：$Y_{t-1}< \alpha_0+\alpha_1 X_t$；

（3）Y大于它的均衡值：$Y_{t-1}> \alpha_0+\alpha_1 X_t$；

在时期t，假设X有一个变化量ΔX_t，如果变量X与Y在时期t与t-1末期仍满足它们间的长期均衡关系，则Y的相应变化量由式给出：

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t$$

式中，$v_t=\mu_t-\mu_{t-1}$。

实际情况往往并非如此

如果t-1期末，发生了上述第二种情况，即Y的值小于其均衡值，则Y的变化往往会比第一种情形下Y的变化ΔY_t大一些;

反之，如果Y的值大于其均衡值，则Y的变化往往会小于第一种情形下的ΔY_t。

可见，如果$Y_t=\alpha_0+\alpha_1 X_t+\mu_t$正确地提示了X与Y间的长期稳定的“均衡关系”，则意味着Y对其均衡点的偏离从本质上说是“临时性”的。

因此，**一个重要的假设就是：随机扰动项μ_t必须是平稳序列。**

显然，如果μ_t有随机性趋势（上升或下降），则会导致Y对其均衡点的任何偏离都会被长期累积下来而不能被消除。

式$Y_t=\alpha_0+\alpha_1X_t+\mu_t$中的随机扰动项也被称为**非均衡误差（disequilibrium error）**，它是变量X与Y的一个线性组合：

$$\mu_t = Y_t - \alpha_0 - \alpha_1 X_t \qquad (*)$$

因此，如果$Y_t=\alpha_0+\alpha_1X_t+\mu_t$式所示的X与Y间的长期均衡关系正确的话，（*）式表述的非均衡误差应是一平稳时间序列，并且具有零期望值，即是具有0均值的I(0)序列。

从这里已看到，非稳定的时间序列，它们的线性组合也可能成为平稳的。

例如：假设$Y_t=\alpha_0+\alpha_1X_t+\mu_t$式中的X与Y是I(1)序列，如果该式所表述的它们间的长期均衡关系成立的话，则意味着由非均衡误差（*）式给出的线性组合是I(0)序列。这时我们**称变量X与Y是协整的（cointegrated）**。

2. 协整

如果序列$\{X_{1t}, X_{2t}, \ldots, X_{kt}\}$都是d阶单整，存在向量$\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \ldots, \alpha_k)$，使得

$$\mathbf{Z}_t = \boldsymbol{\alpha}\mathbf{X}^T \sim I(d-b)$$

其中，$b>0$，$\mathbf{X}=(X_{1t}, X_{2t}, \ldots, X_{kt})^T$，则认为序列$\{X_{1t}, X_{2t}, \ldots, X_{kt}\}$是(d,b)阶协整，记为$\mathbf{X}_t \sim \mathbf{CI(d,b)}$，$\boldsymbol{\alpha}$为**协整向量（cointegrated vector）**。

在中国居民人均消费与人均GDP的例中，该两序列都是2阶单整序列，而且可以证明它们有一个线性组合构成的新序列为0阶单整序列，于是认为该两序列是(2,2)阶协整。

由此可见:**如果两个变量都是单整变量，只有当它们的单整阶数相同时，才可能协整；如果它们的单整阶数不相同，就不可能协整。**

三个以上的变量，如果具有不同的单整阶数，有可能经过线性组合构成低阶单整变量。

例如，如果存在：

$$W_t \sim I(1), V_t \sim I(2), U_t \sim I(2)$$

并且

$$P_t = aV_t + bU_t \sim I(1)$$

$$Q_t = cW_t + eP_t \sim I(0)$$

那么认为：

$$V_t, U_t \sim CI(2,1)$$

$$W_t, P_t \sim CI(1,1)$$

从协整的定义可以看出：

（d,d）阶协整是一类非常重要的协整关系，它的经济意义在于：两个变量，虽然它们具有各自的长期波动规律，但是如果它们是（d, d）阶协整的，则它们之间存在着一个长期稳定的比例关系。

例如：前面提到的中国CPC和GDPPC，它们各自都是2阶单整，并且将会看到，它们是(2, 2)阶协整，说明它们之间存在着一个长期稳定的比例关系，从计量经济学模型的意义上讲，建立如下居民人均消费函数模型

$$CPC_t = \alpha_0 + \alpha_1 GDPPC_t + \mu_t$$

变量选择是合理的，随机误差项一定是“白噪声”（即均值为0，方差不变的稳定随机序列），模型参数有合理的经济解释。

这也解释了尽管这两时间序列是非稳定的，但却可以用经典的回归分析方法建立回归模型的原因。

- 从这里，我们已经初步认识到：检验变量之间的协整关系，在建立计量经济学模型中是非常重要的。

　　而且，从变量之间是否具有协整关系出发选择模型的变量，其数据基础是牢固的，其统计性质是优良的。

二、协整检验

1、两变量的Engle-Granger检验

为了检验两变量Y_t,X_t是否为协整，Engle和Granger于1987年提出两步检验法，也称为EG检验。

第一步，用OLS方法估计方程 $\mathbf{Y_t}=\alpha_0+\alpha_1 X_t+\mu_t$

并计算非均衡误差，得到：

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t$$

$$\hat{e}_t = Y_t - \hat{Y}_t$$

称为**协整回归(cointegrating)**或**静态回归(static regression)**。

第二步，检验$\hat{e}_t$的单整性。如果$\hat{e}_t$为稳定序列，则认为变量Y_t, X_t为(1,1)阶协整；如果$\hat{e}_t$为1阶单整，则认为变量Y_t, X_t为(2,1)阶协整；…。

$\hat{e}_t$ **的单整性的检验方法仍然是DF检验或者ADF检验。**

由于协整回归中已含有截距项，则检验模型中无需再用截距项。如使用模型1

$$\Delta e_t = \delta\, e_{t-1} + \sum_{i=1}^{p} \theta_i \Delta e_{t-i} + \varepsilon_t$$

进行检验时，**拒绝零假设H_0：δ=0**，意味着误差项e_t是平稳序列，从而**说明X与Y间是协整的**。

需要注意是，这里的DF或ADF检验是针对协整回归计算出的误差项 $\hat{e}_t$ 而非真正的非均衡误差μ_t进行的。

而OLS法采用了残差最小平方和原理，因此估计量δ是向下偏倚的，这样将导致拒绝零假设的机会比实际情形大。

于是对e_t平稳性检验的DF与ADF临界值应该比正常的DF与ADF临界值还要小。

- MacKinnon(1991)通过模拟试验给出了协整检验的临界值，表9.3.1是双变量情形下不同样本容量的临界值。

表 9.3.1　　双变量协整 ADF 检验临界值

样本容量	显著性水平		
	0.01	0.05	0.10
25	−4.37	−3.59	−3.22
50	−4.12	−3.46	−3.13
100	−4.01	−3.39	−3.09
∞	−3.90	−3.33	−3.05

- **例9.3.1 检验中国居民人均消费水平CPC与人均国内生产总值GDPPC的协整关系。**

在前文已知CPC与GDPPC都是I(2)序列，而§2.10中已给出了它们的回归式

$$CPC_t = 49.764106 + 0.45831GDPPC_t \qquad R^2=0.9981$$

通过对该式计算的残差序列作ADF检验，得适当检验模型

$$\Delta\hat{e}_t = -1.55\hat{e}_{t-1} + 1.49\Delta\hat{e}_{t-1} + 2.27\Delta\hat{e}_{t-3}$$

（-4.47）　(3.93)　(3.05)

LM(1)=0.00　LM(2)=0.00

t=-4.47<-3.75=$ADF_{0.05}$，拒绝存在单位根的假设，残差项是稳定的，因此**中国居民人均消费水平与人均GDP是(2,2)阶协整的，说明了该两变量间存在长期稳定的“均衡”关系。**

2、多变量协整关系的检验—扩展的E-G检验

多变量协整关系的检验要比双变量复杂一些，主要在于协整变量间可能存在多种稳定的线性组合。

假设有4个I(1)变量Z、X、Y、W，它们有如下的长期均衡关系：

$$Z_t = \alpha_0 + \alpha_1 W_t + \alpha_2 X_t + \alpha_3 Y_t + \mu_t \qquad (*)$$

其中，非均衡误差项μ_t应是I(0)序列：

$$\mu_t = Z_t - \alpha_0 - \alpha_1 W_t - \alpha_2 X_t - \alpha_3 Y_t \qquad (**)$$

- 然而，如果Z与W，X与Y间分别存在长期均衡关系：

$$Z_t = \beta_0 + \beta_1 W_t + v_{1t}$$
$$X_t = \gamma_0 + \gamma_1 Y_t + v_{2t}$$

则非均衡误差项v_{1t}、v_{2t}一定是稳定序列I(0)。于是它们的任意线性组合也是稳定的。例如

$$v_t = v_{1t} + v_{2t} = Z_t - \beta_0 - \gamma_0 - \beta_1 W_t + X_t - \gamma_1 Y_t \qquad (***)$$

一定是I(0)序列。

由于v_t象（**）式中的μ_t一样，也是Z、X、Y、W四个变量的线性组合，由此（***）式也成为该四变量的另一稳定线性组合。

（1，$-\alpha_0, -\alpha_1, -\alpha_2, -\alpha_3$）是对应于（**）式的协整向量，（$1, -\beta_0-\gamma_0, -\beta_1, 1, -\gamma_1$）是对应于（***）式的协整向量。

检验程序：

对于多变量的协整检验过程，基本与双变量情形相同，即需检验变量是否具有同阶单整性，以及是否存在稳定的线性组合。

在检验是否存在稳定的线性组合时，需通过设置一个变量为被解释变量，其他变量为解释变量，进行OLS估计并检验残差序列是否平稳。

如果不平稳，则需更换被解释变量，进行同样的OLS估计及相应的残差项检验。

当所有的变量都被作为被解释变量检验之后，仍不能得到平稳的残差项序列，则认为这些变量间不存在（d,d）阶协整。

同样地，检验残差项是否平稳的DF与ADF检验临界值要比通常的DF与ADF检验临界值小，而且该临界值还受到所检验的变量个数的影响。

表9.3.2给出了MacKinnon(1991)通过模拟试验得到的不同变量协整检验的临界值。

表 9.3.2　多变量协整检验 ADF 临界值

样本容量	变量数=3 显著性水平			变量数=4 显著性水平			变量数=6 显著性水平		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
25	-4.92	-4.1	-3.71	-5.43	-4.56	-4.15	-6.36	-5.41	-4.96
50	-4.59	-3.92	-3.58	-5.02	-4.32	-3.98	-5.78	-5.05	-4.69
100	-4.44	-3.83	-3.51	-4.83	-4.21	-3.89	-5.51	-4.88	-4.56
∞	-4.30	-3.74	-3.45	-4.65	-4.1	-3.81	-5.24	-4.7	-4.42

2、多变量协整关系的检验—JJ检验

- **Johansen于1988年，以及与Juselius于1990年提出了一种用极大或然法进行检验的方法，通常称为JJ检验。**
- **《高等计量经济学》（清华大学出版社，2000年9月）P279-282.**
- **E-views中有JJ检验的功能。**

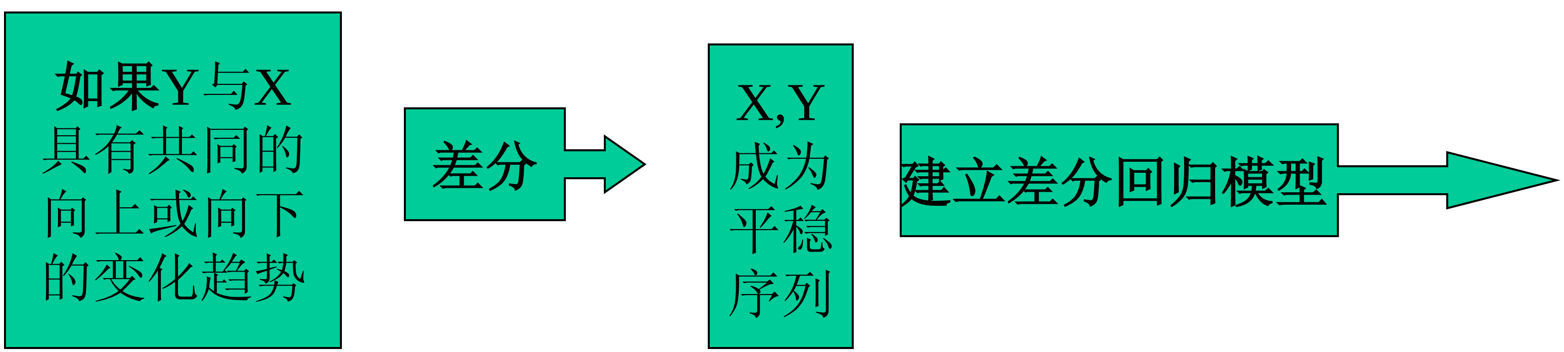
三、误差修正模型

1、误差修正模型

前文已经提到，**对于非稳定时间序列，可通过差分的方法将其化为稳定序列，然后才可建立经典的回归分析模型。**

如：建立人均消费水平（Y）与人均可支配收入（X）之间的回归模型：

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t$$



$$\Delta Y_t = \alpha_1 \Delta X_t + v_t$$

式中，$v_t = \mu_t - \mu_{t-1}$

然而，**这种做法会引起两个问题**：

(1)如果X与Y间存在着长期稳定的均衡关系

$$Y_t=\alpha_0+\alpha_1 X_t+\mu_t$$

且误差项μ_t不存在序列相关，则**差分式**

$$\Delta Y_t=\alpha_1\Delta X_t+\nu_t$$

中的ν_t是一个**一阶移动平均时间序列**，因而**是序列相关的**；

(2)如果采用差分形式进行估计，则关于变量水平值的重要信息将被忽略，**这时模型只表达了X与Y间的短期关系，而没有揭示它们间的长期关系**。

因为，从长期均衡的观点看，Y在第t期的变化不仅取决于X本身的变化，还取决于X与Y在t-1期末的状态，尤其是X与Y在t-1期的不平衡程度。

另外，使用差分变量也往往会得出不能令人满意回归方程。

例如，使用$\Delta Y_t=\alpha_1\Delta X_t+v_t$回归时，很少出现截距项显著为零的情况，即我们常常会得到如下形式的方程：

$$\Delta Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 \Delta X_t + v_t \qquad \hat{\alpha}_0 \neq 0 \qquad (*)$$

在X保持不变时，如果模型存在静态均衡（static equilibrium），Y也会保持它的长期均衡值不变。

但如果使用（*）式，即使X保持不变，Y也会处于长期上升或下降的过程中(Why?)，这意味着**X与Y间不存在静态均衡**。

这与大多数具有静态均衡的经济理论假说不相符。

可见，简单差分不一定能解决非平稳时间序列所遇到的全部问题，因此，误差修正模型便应运而生。

误差修正模型（Error Correction Model，简记为**ECM）是一种具有特定形式的计量经济学模型**，它的主要形式是由Davidson、 Hendry、Srba和Yeo于1978年提出的，**称为DHSY模型**。

为了便于理解，我们通过一个具体的模型来介绍它的结构。

假设两变量X与Y的长期均衡关系为:

$$\mathbf{Y_t}=\alpha_0+\alpha_1 X_t+\mu_t$$

由于现实经济中X与Y很少处在均衡点上，因此实际观测到的只是X与Y间的短期的或非均衡的关系，假设具有如下(1, 1)阶分布滞后形式

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

该模型显示出第t期的Y值，不仅与X的变化有关，而且与t-1期X与Y的状态值有关。

由于变量可能是非平稳的，因此不能直接运用OLS法。对上述分布滞后模型适当变形得

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} - (1-\mu) Y_{t-1} + \varepsilon_t$$

$$= \beta_1 \Delta X_t - (1-\mu)\left(Y_{t-1} - \frac{\beta_0}{1-\mu} - \frac{\beta_1 + \beta_2}{1-\mu} X_{t-1} \right) + \varepsilon_t$$

或

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t \qquad (**)$$

式中， $\lambda = 1-\mu$　　$\alpha_0 = \beta_0/(1-\mu)$　　$\alpha_1 = (\beta_1 + \beta_2)/(1-\mu)$

如果将（**）中的参数，与 $Y_t=\alpha_0+\alpha_1 X_t+\mu_t$ 中的相应参数视为相等，则（**）式中括号内的项就是t-1期的非均衡误差项。

（**）式表明：Y的变化决定于X的变化以及前一时期的非均衡程度。同时，（**）式也弥补了简单差分模型 $\Delta Y_t=\alpha_1 \Delta X_t+v_t$ 的不足，因为该式含有用X、Y水平值表示的前期非均衡程度。因此，Y的值已对前期的非均衡程度作出了修正。

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda(Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t \qquad (**)$$

称为**一阶误差修正模型** (**first-order error correction model**)。

（**）式可以写成：

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda ecm + \varepsilon_t \qquad (***)$$

其中：ecm表示**误差修正项**。由**分布滞后模型**

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \mu Y_{t-1} + \varepsilon_t$$

知，一般情况下$|\mu|<1$，由关系式$\lambda=1-\mu$得$0<\lambda<1$。**可以据此分析ecm的修正作用：**

(1) 若(t-1)时刻Y大于其长期均衡解$\alpha_0+\alpha_1X$，ecm为正，则(-λecm)为负，使得ΔY_t减少；

(2) 若(t-1)时刻Y小于其长期均衡解$\alpha_0+\alpha_1X$，ecm为负，则(-λecm)为正，使得ΔY_t增大。

（***）体现了长期非均衡误差对的控制。

需要注意的是：在实际分析中，变量常以对数的形式出现。

其主要原因在于变量对数的差分近似地等于该变量的变化率，而经济变量的变化率常常是稳定序列，因此适合于包含在经典回归方程中。

于是:(1)**长期均衡模型**

$$Y_t=\alpha_0+\alpha_1 X_t+\mu_t$$

中的α_1可视为Y关于X的**长期弹性（long-run elasticity）**

(2)**短期非均衡模型**

$$Y_t=\beta_0+\beta_1 X_t+\beta_2 X_{t-1}+\mu Y_{t-1}+\varepsilon_t$$

中的β_1可视为Y关于X的**短期弹性（short-run elasticity）**。

更复杂的误差修正模型可依照一阶误差修正模型类似地建立。

如具有季度数据的变量，可在短期非均衡模型

$$Y_t=\beta_0+\beta_1X_t+\beta_2X_{t-1}+\mu Y_{t-1}+\varepsilon_t$$

中引入更多的滞后项。

引入二阶滞后的模型为

$$Y_t=\beta_0+\beta_1X_t+\beta_2X_{t-1}+\beta_3X_{t-2}+\mu_1Y_{t-1}+\mu_2Y_{t-2}+\varepsilon_t$$

经过适当的衡等变形，可得如下二阶误差修正模型

$$\Delta Y_t=-\mu_2\Delta Y_{t-1}+\beta_1\Delta X_t-\beta_3\Delta X_{t-1}-\lambda(Y_{t-1}-\alpha_0-\alpha_1X_{t-1})+\varepsilon_t \qquad (*)$$

式中，$\lambda=1-\mu_1-\mu_2$，$\alpha_0=\beta_0/\lambda$，$\alpha_1=(\beta_1+\beta_2+\beta_3)/\lambda$

引入三阶滞后项的误差修正模型与（*）式相仿，只不过模型中多出差分滞后项ΔY_{t-2}，ΔX_{t-2}，。

多变量的误差修正模型也可类似地建立。

如三个变量如果存在如下长期均衡关系

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_t$$

则其一阶非均衡关系可写成

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \gamma_1 Z_t + \gamma_2 Z_{t-2} + \mu Y_{t-1} + \varepsilon_t$$

于是它的一个误差修正模型为

$$\Delta Y_t = \beta_1 \Delta X_t + \gamma_1 \Delta Z_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1} - \alpha_2 Z_{t-1}) + \varepsilon_t$$

式中，$\lambda = 1 - \mu$，$\alpha_0 = \beta_0 / \lambda$，$\alpha_1 = (\beta_1 + \beta_2)/\lambda$，$\alpha_2 = (\gamma_1 + \gamma_2)/\lambda$

2、误差修正模型的建立

（1）Granger 表述定理

误差修正模型有许多明显的**优点**：如

a）一阶差分项的使用消除了变量可能存在的趋势因素，从而避免了虚假回归问题；

b）一阶差分项的使用也消除模型可能存在的多重共线性问题；

c）误差修正项的引入保证了变量水平值的信息没有被忽视；

d）由于误差修正项本身的平稳性，使得该模型可以用经典的回归方法进行估计，尤其是模型中差分项可以使用通常的t检验与F检验来进行选取；等等。

因此，**一个重要的问题就是**：**是否变量间的关系都可以通过误差修正模型来表述**？

就此问题，Engle 与 Granger 1987年提出了著名的 Grange表述定理（Granger representaion theorem）：

如果变量X与Y是协整的，则它们间的短期非均衡关系总能由一个误差修正模型表述:

$$\Delta Y_t = lagged(\Delta Y, \Delta X) - \lambda \mu_{t-1} + \varepsilon_t \qquad 0<\lambda<1 \qquad (*)$$

式中，μ_{t-1}**是非均衡误差项**或者说成是**长期均衡偏差项**，λ是**短期调整参数**。

对于(1,1)阶自回归分布滞后模型

$$Y_t=\beta_0+\beta_1 X_t+\beta_2 X_{t-1}+\mu Y_{t-1}+\varepsilon_t$$

如果 Yt~I(1), Xt~I(1) ； 那么

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda(Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t$$

的左边ΔY_t~I(0)，右边的ΔX_t~I(0) ，因此，只有Y与X协整，才能保证右边也是I(0)。

因此，**建立误差修正模型**，需要

首先对变量进行协整分析，以发现变量之间的协整关系，即长期均衡关系，并以这种关系构成误差修正项。

然后建立短期模型，将误差修正项看作一个解释变量，连同其它反映短期波动的解释变量一起，建立短期模型，即误差修正模型。

注意，由于

$$\Delta Y = \text{lagged}(\Delta Y, \Delta X) + \lambda\mu_{t-1} + \varepsilon_t \qquad 0<\lambda<1$$

中没有明确指出Y与X的滞后项数，因此，可以是多个；

同时，由于一阶差分项是I(0)变量，因此模型中也允许使用X的非滞后差分项ΔX_t 。

Granger表述定理可类似地推广到多个变量的情形中去。

（2）Engle-Granger两步法

由协整与误差修正模型的的关系，**可以得到误差修正模型建立的E-G两步法：**

第一步，进行协整回归（OLS法），检验变量间的协整关系，估计协整向量（长期均衡关系参数）；

第二步，若协整性存在，则以第一步求到的残差作为非均衡误差项加入到误差修正模型中，并用OLS法估计相应参数。

需要注意的是：在进行变量间的协整检验时，如有必要可在协整回归式中加入趋势项，这时，对残差项的稳定性检验就无须再设趋势项。

另外，第二步中变量差分滞后项的多少，可以残差项序列是否存在自相关性来判断，如果存在自相关，则应加入变量差分的滞后项。

（3）直接估计法

也可以**采用打开误差修整模型中非均衡误差项括号的方法直接用OLS法估计模型**。

但仍需事先对变量间的协整关系进行检验。

如对双变量误差修正模型

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \varepsilon_t$$

可打开非均衡误差项的括号直接估计下式：

$$\Delta Y_t = \lambda \alpha_0 + \beta_1 \Delta X_t - \lambda Y_{t-1} + \lambda \alpha_1 X_{t-1} + \varepsilon_t$$

这时短期弹性与长期弹性可一并获得。

需注意的是，用不同方法建立的误差修正模型结果也往往不一样。

例9.3.2 中国居民消费的误差修正模型

经济理论指出，居民消费支出是其实际收入的函数。

以中国国民核算中的**居民消费支出经过居民消费价格指数缩减得到中国居民实际消费支出时间序列（C）；**

以支出法**GDP对居民消费价格指数缩减近似地代表国民收入时间序列(GDP)**

时间段为1978~2000（表9.3.3）

表 9.3.3 1978~1998 年间中国实际居民消费与实际 GDP 数据（单位：亿元，1990 年价）

年份	C	GDP	年份	C	GDP	年份	C	GDP
1978	3810	7809	1985	7579	14521	1992	11325	23509
1979	4262	8658	1986	8025	15714	1993	12428	27340
1980	4581	8998	1987	8616	17031	1994	13288	29815
1981	5023	9454	1988	9286	17889	1995	14693	31907
1982	5423	10380	1989	8788	16976	1996	16189	34406
1983	5900	11265	1990	9113	18320	1997	17072	36684
1984	6633	12933	1991	9977	20581	1998	18230	39008

（1）对数据lnC与lnGDP进行单整检验

容易验证lnC与lnGDP是一阶单整的，它们适合的检验模型如下：

$$\Delta^2 \ln C_t = 0.056 - 0.744\,\Delta \ln C_{t-1}$$

(2.76)　　(-3.23)

LM(1)=0.929　　　LM(2)=1.121

$$\Delta^2 \ln GDP_t = 0.13 - 1.54\,\Delta \ln GDP_{t-1} + 0.81\Delta^2 \ln GDP_{t-1} + 0.59\Delta^2 \ln GDP_{t-2} + 0.58\Delta^2 \ln GDP_{t-3}$$

(3.81)（-4.01）　　　（2.66）　　　（2.26）　　　（2.54）

LM(1)=0.38　　LM(2)=0.67　　LM(3)=2.34　　LM(4)=2.46

（2）检验lnC与lnGDP的协整性，并建立长期均衡关系

首先，建立lnC与lnGDP的回归模型

$$\ln C_t = 0.047 + 0.923 \ln GDP_t$$

（0.30） (57.48)

R^2=0.994　　　DW=0.744

发现有残关项有较强的一阶自相关性。考虑加入适当的滞后项，得lnC与lnGDP的分布滞后模型

$$\ln C_t = 0.152 + 0.698 \ln GDP_t + 0.622 \ln C_{t-1} - 0.361 \ln GDP_{t-1} \quad (*)$$

(1.63) (6.62）　（4.92）　（-2.17）

R^2=0.994　DW=1.92　LM(1)=0.00　LM(2)=2.31

自相关性消除，因此可初步认为是lnC与lnGDP的长期稳定关系。

残差项的稳定性检验：

$$\Delta\hat{e}_t = -0.9975\hat{e}_{t-1}$$

（-4.32）

$R^2=0.994$　DW=2.01　LM(1)=0.04　LM(2)=1.34

$t=-4.32<-3.64=ADF_{0.05}$

说明lnC与lnGDP是（1，1）阶协整的，（*）式即为它们长期稳定的均衡关系：

$$\ln C_t = 0.152 + 0.698\ln GDP_t + 0.622\ln C_{t-1} - 0.361\ln GDP_{t-1} \qquad (*)$$

（3）建立误差修正模型

- 以稳定的时间序列 $\hat{e}_t$ 做为误差修正项，可建立如下

误差修正模型:

$$\Delta \ln C_t = 0.686\Delta \ln GDP_t + 0.784\Delta \ln C_{t-1} - 0.484\Delta \ln GDP_{t-1} - 1.163\hat{e}_{t-1} \qquad (**)$$

（6.96）　　(2.96)　　(-1.91)　　(-3.15)

R^2=0.994　　DW=2.06　LM(1)=0.70　LM(2)=2.04

由(*)式 $\ln C_t = 0.152 + 0.698\ln GDP_t + 0.622\ln C_{t-1} - 0.361\ln GDP_{t-1}$

可得lnC关于lnGDP的长期弹性：

(0.698-0.361)/(1-0.622)=0.892；

由（**）式可得lnC关于lnGDP的短期弹性：0.686

下面用**打开误差修正项括号的方法直接估计误差修正模型**，适当估计式为：

$$\Delta \ln C_t = 0.153 + 0.698\Delta \ln GDP_t - 0.378 \ln C_{t-1} + 0.337 \ln GDP_{t-1}$$

（1.63）(6.62）　　(-2.99)　　(2.88)

R^2=0.791　=0.0064　DW=1.93　LM(2)=2.31　LM(3)=2.78

写成误差修正模型的形式如下

$$\Delta \ln C_t = 0.698\Delta \ln GDP_t - 0.378(\ln C_{t-1} - 0.405 - 0.892 \ln GDP_{t-1}) \quad (***)$$

由（***）式知，lnC关于lnGDP的短期弹性为0.698，长期弹性为0.892。

可见两种方法的结果非常接近。

（4）预测

由(*)式

$$\ln C_t = 0.152 + 0.698\ln GDP_t + 0.622\ln C_{t-1} - 0.361\ln GDP_{t-1}$$

给出1998年关于长期均衡点的偏差：

$$\hat{e}_{98} = \ln(18230) - 0.152 - 0.698\ln(39008) - 0.662\ln(17072)$$

$$+0.361\ln(36684) = 0.0125$$

由（**）式

$$\Delta\ln C_t = 0.686\Delta\ln GDP_t + 0.784\Delta\ln C_{t-1} - 0.484\Delta\ln GDP_{t-1} - 1.163\hat{e}_{t-1}$$

预测1999年的短期波动：

$$\Delta\ln C_{99} = 0.686(\ln(41400) - \ln(39008)) + 0.784(\ln(18230) - \ln(17072))$$

$$-0.484(\ln(39008) - \ln(36684)) - 1.163 \times 0.0125 = 0.048$$

于是

$$\ln C_{99} = 0.048 + \ln C_{98} = 0.048 + \ln(18230) = 9.859$$

$$C_{99} = e^{9.859} = 19125$$

按照（*** ）式

$$\Delta \ln C_t = 0.698 \Delta \ln GDP_t - 0.378 (\ln C_{t-1} - 0.405 - 0.892 \ln GDP_{t-1})$$

预测的结果为:

$\Delta \ln C_{99}$=0.698（ln(41400)-ln(39008))-0.378(ln(18230)-0.405

-0.892ln(39008))=0.051

于是

$$\ln C_{99} = 0.051 + \ln C_{98} = 0.051 + \ln(18230) = 9.861$$

$$C_{99} = e^{9.861} = 19176$$

以当年价计的1999年实际居民消费支出为39334亿元，用居民消费价格指数（1990=100）紧缩后约为19697亿元，因此：**两个预测结果的相对误差分别为2.9%与2.6%。**