学习文档

李昊[1]　　lihao9153@163.com[2]

吉首大学(湖南)　　My E-mail

[1]感谢所有人。

[2]Thanks for all the people.

目 录

第一章　前言

来自[1]、[2]、[3]、[4]。

1.1　什么是计量经济学

计量经济学定义 弗里希在世界计量经济学会会刊《*Econometric*》的发刊词中写道：

> ”对经济的数量研究有几个方面，其中任何一个方面都不应和计量经济学混为一谈。计量经济学绝不等同于经济统计学，它也不同于我们所说的一般经济理论，计量经济学也不应该视为数学应用于经济学的同义语。经验表明，统计学、经济理论和数学三者对于实际理解现代经济生活的数量关系都是必要的，但其中任何单独一种都是不够的。三者的结合才是有力的工具，正是这三者的统一才构成了计量经济学。”

一般将计量经济学定义为”一门由经济学、统计学和数学结合而成的交叉学科，是一门经济学科”。但由于计算的复杂性，使人们只能建立小规模的简单模型，而随着计算机技术的迅速发展，为计量经济学提供了强有力的工具，使许多复杂的大规模模型得以应用。所以还可以将计量经济学定义为：”计量经济学是以经济理论为指导，以数据事实为依据，以数学为方法、以计算机技术为手段，研究经济关系和经济活动数量规律及其应用，并以建立和应用计量经济学模型为核心的一门经济学学科。”

另外，计量经济学也指以 $R.Fresh$ 等同时代的经济学家基于概率论和数量统计而创建的计量经济方法论。**计量经济学学科的形成** 英文”Econometrics”最早是由 $R.Frish$于 1926年模仿 ”*Biometrics*”（生物计量学）提出的，标志着计量经济学的诞生。但人们一般认为，1930 年 12月～ 29 日世界计量经济学会成立和由她创办的学术刊物《*Econometrica*》于 1933 年正式出版，标志着计量经济学作为一个独立学科正式诞生了。**经济计量学的发展(20 - 40) 年代**

本世纪之前，在错综复杂的经济现象面前，经济工作者主要是使用头脑直接对材料进行归纳、综合和推理。十九世纪欧洲主要国家先后进入资本主义社会。工业化大生产的出现，经济活动规模的不断扩大，需要人们对经济问题做出更精确、深入的分析、解释与判断。这是计量经济学诞生的社会基础。

到本世纪初，数学、统计学理论日趋完善为计量经济学的出现奠定了理论基础。

17 世纪 $Newton-Leibniz$ 提出微积分，19 世纪初（1809年）德国数学家 $Gauss$ 提出最小二乘法，1821 年提出正态分布理论。19 世纪末英国统计学家 $Galton$ 提出”回归”概念。本世纪 20 年代 $FisherR.$（英 $1890-1962$）和 $NeymanJ.D.$（波兰裔美国人）分别提出抽样分布和假设检验理论。至此，数理统计的理论框架基本形成。这时，人们自然想到要用这些知识解释、分析、研究经济问题，从而诞生了计量经济学。

”计量经济学”一词首先由挪威经济学家 $Frisch$仿照生物计量学（*biometrics*）一词于 1926 年提出。1930 年由 $Frisch$，$Tinbergen$ 和 $Fisher$ 等人发起在美国成立了国际计量经济学会。1933年 1 月开始出版”计量经济学”（*Econometrica*）杂志。目前它仍是计量经济学界最权威的杂志。

30 年代计量经济学研究对象主要是个别生产者、消费者、家庭、厂商等。基本上属于微观分析范畴。第二次世界大战后，计算机的发展与应用给计量经济学的研究起了巨大推动作用。从 40 年代起，计量经济学研究从微观向局部地区扩大，以至整个社会的宏观经济体系，处理总体形态的数据，如国民消费、国民收入、投资、失业问题等。但模型基本上属于单一方程形式。

50 - 70 年代 1950 年以 $Koopman$ 发表论文”动态经济模型的统计推断”和 $Koopman-Hood$ 发表论文”线性联立经济关系的估计”为标志，经济计量学理论进入联立方程模型时代。

经济计量学研究经历了从简单到复杂，从单一方程到联立方程的变化过程。进入五十年代人们开始用联立方程模型描述一个国家整体的宏观经济活动。比较著名的是 $Klein$ 的美国经济波动模型（$1921 \sim 1941$，1950 年作）和美国宏观经济模型（$1928 \sim 1950$，1955 年作）后者包括 20 个方程。联立方程模型的应用是经济计量学发展的第二个里程碑。

进入 70 年代西方国家致力于更大规模的宏观模型研究。从着眼于国内发展到着眼于国际的大型经济计量模型。研究国际经济波动的影响，国际经济发展战略可能引起的各种后果，以及制定评价长期的经济政策。70 年代是联立方程模型发展最辉煌的时代。最著名的联立方程模型是"连接计划"（$LinkProject$）。截止 1987 年，已包括 78 个国家 2 万个方程。这一时期最有代表性的学者是 $L.Klein$ 教授。他于 1980 年获诺贝尔经济学奖。

前苏联在本世纪 20 年代也开展过这方面的研究，但到 30 年代就中止了。60 年代中期以来，前苏联及东欧一些国家开始大量编制投入产出模型并取得有益成果。

80 - 90 年代 因为七十年代以前的建模技术都是以"经济时间序列平稳"这一前提设计的，而战后多数国家的宏观经济变量均呈非平稳特征，所以在利用联立方程模型对非平稳经济变量进行预测时常常失败。从 70 年代开始，宏观经济变量的非平稳性问题以及虚假回归问题越来越引起人们的注意。因为这些问题的存在会直接影响经济计量模型参数估计的准确性。$Granger - Newbold$于 1974 年首先提出虚假回归问题，引起了计量经济学界的注意。$Box - Jenkins$ 1967 年出版《时间序列分析，预测与控制》一书。时间序列模型有别于回归模型，是一种全新的建模方法，它是依靠变量本身的外推机制建立模型。由于时间序列模型妥善地解决了变量的非平稳性问题，从而为在经济领域应用时间序列模型奠定了理论基础。人们发现耗费许多财力人力建立的经济计量模型有时竟不如一个简单的时间序列模型预测能力好（$Cooper$ 1972 年专门对两种模型的预测精度作了详细比较）。

此时，计量经济工作者面临三个亟待解决的问题：

(1) 如何检验经济变量的非平稳性

(2) 如何把时间序列模型引入经济计量分析领域

(3) 如何进一步修改传统的经济计量模型

$Dickey - Fuller$ 1979 年首先提出检验时间序列非平稳性（单位根）的 DF检验法，之后又提出 ADF检验法。

$Phillips - Perron$ 1988 年提出 Z 检验法。这是一种非参数检验方法。

$Sargan$ 1964 年提出误差修正模型概念。当初是用于研究商品库存量问题。$Hendry - Anderson$(1977) 和 $Davidson$(1978) 的论文进一步完善了这种模型，并尝试用这种模型解决非平稳变量的建模问题。$Hendry$ 还提出动态回归理论。

1980 年 $Sims$ 提出向量自回归模型（VAR）。这是一种用一组内生变量作动态结构估计的联立模型。这种模型的特点是不以经济理论为基础，然而预测能力很强。以上成果为协整理论的提出奠定基础。

计量经济学发展的第三个里程碑是 1987 年 $Engle - Granger$ 发表论文"协整与误差修正，描述、估计与检验"。该论文正式提出协整概念，从而把计量经济学理论的研究又推向一个新阶段。$Granger$ 定理证明若干个一阶非平稳变量间若存在协整关系，那么这些变量一定存在误差修正模型表达式。反之亦成立。$1988 - 1992$ 年 $Johansen$ （丹麦）连续发表了四篇关于向量自回归模型中检验协整向量，并建立向量误差修正模型（VEC）的文章，进一步丰富了协整理论。协整理论之所以引起经济计量学界的广泛兴趣与极大关注是因为协整理论为当代经济学的发展提供了一种理论结合实际的强有力工具。

经济计量学在我国的发展状况 1960 年中国科学院经济研究所成立了一个经济数学方法研究组。主要搞投入产出、优化研究。

改革开放以后，1979 年三月成立了中国数量经济研究会（1984 年定名为中国数量经济学会，并办有一份杂志，《数量经济技术经济研究》）。

1980 年中国数量经济学会首次举办经济计量学讲习班，邀请 *Klein*等七位美国教授讲课。自此，计量经济学的教学与科研迅速展开，取得许多研究成果。国家信息中心为参加联合国的"连接计划"研制了我国的宏观计量经济模型。吉林大学数量经济研究中心研制了"国家财政模型及经济景气分析系统"。

1998 年教育部规定计量经济学是我国大学经济类专业本科学生的 8 门必修课之一。

思考 1.1 (什么是计量经济学)

CC方法论及其批判 在计量经济学形成的早期，美国的投资家 *A.Cowles* 由于股市的崩溃使他受到损失，从而激起他对计量经济学的兴趣而发起成立以自己名字命名的基金委员会(以下简记为 *CC*)，专门用于资助计量经济学的研究，在 *CC* 的资助下，形成了大量对计量经济学具有奠基意义的成果，构建了计量经济学的概率论框架，因此经典计量经济学，在不严格的意义下，又简称为 *CC* 方法论。

CC 方法论的假设:

(1) 经济行为由联立方程所支配；

(2) 模型的方程对变量和扰动项来说是线性的；

(3) 变量是可观测且观测无误的

(4) 变量的预测值是离散的；

(5) 外生变量和先决变量是设定的；

(6) 联立应变量的系数行列式非零，使得简约形式存在；

(7) 先决变量与扰动项线性独立；

(8) 误差项一般独立同分布；

(9) 通过对结构参数的约束使得结构方程可识别；（可识别的隐含假设）

(10) 方程是动态稳定的。

（假定外生变量在重复抽样中固定，且当样本趋于无限时，外生变量的矩有限以及方程的特征要的绝对值小于 1，即变量为 *I*0）当上述条件满足时，*CC* 方法论是完备而精确的方法论。

以 *CC* 方法论为主要内容的计量经济学常被称为经典计量经济学。针对 *CC* 方法论的批判和取消或弱化 *CC* 方法论的假设而发展起来的计量经济学，即从 1960 后发展起来的，被称为当代或高等计量经济学。

19 世纪 50 年代末期，由于石油危机引发了世界经济的衰退和随之而来的滞胀，以 *CC* 方法论所构建的计量经济模型，几乎均未预测到这次经济的衰退。随后，基于经典计量经济方法论所建立的模型也未能就治理滞胀开出有效的"药方"，由此导致了对 *CC* 方法论的批判，其中 *Lucas*(1976) 批判最具影响。*Lucas* 批判所隐含的意思为: 如果政策反应函数出现变化，这种变化也将改变模型的参数，于是，联立模型的简约形式也将随之而变化，因此，使用联立模型所作出的测预是不可信赖的。*T.Sarget*(1976) 以货币政策为例，重新解析了 *Lucas* 批判。他假定货币政策体制从不变的货币增长政策改变为具有反馈的货币增长政策，因而，联立方程模型具有 2 种简约形式，在常数的货币增长政策之下，简约参数并不随货币政策的改变而改变，而在反馈的货币政策之下，简约参数随货币政策的改变而改变。于是，仅凭简约模型的估计并不能解决哪一种货币政策好这一问题，因而，计量经济学对于评价政策似乎是无能为力的。*Lucas*

是从计量经济学用于政策（而可能导致经济运行的改变）分析而提出的批判，其实质是提出了参数是否随时间而变化的问题，隐含了经典计量经济模型产生不精确预测的重要原因是结构变化问题。

$Sims$(1980)认为：为使结构方程可识别而施加了许多约束，这种约束是不可信的。因此，他建议使用向量自回归（VAR）模型进行预测。随后大量的研究表明，利用 VAR 作预测，其结果优于传统的联立系统所得到的结果。这种学术批判揭示了 CC 方法论直接或间接（隐含）的假设为：①经济行为由联立方程所支配；②模型的方程对变量和扰动来说是线性的；③变量是可观测且观测无误差；④变量的观测值是离散的，即年度（月度）数据等；⑤外生变量和先决变量是设定的；⑥联立应变量的系数行列式非零，使得简约形式存在；⑦先决变量与误差（扰动）项线性独立；⑧误差项是独立同分布的正态变量，其均值为零，方差和协方差为常数，协方差阵非奇异（这一假设实际上在 CC 方法论中已扩展为一般独立同分布扰动），⑨通过对结构参数的约束使得结构方程可识别。⑩方程是动态稳定的。假设⑨⑩是间接的假设，其中⑨是由可识别所隐含的假设，而⑩是最为关键的隐含假设，它通过假定外生变量在重复抽样中固定，且当样本趋于无限时，外生变量的矩有限以及方程的特征根的绝对值小于 1（即变量为 $I(0)$）来实现方程的动态稳定性。我们以后可以看到，正是为取消这一条假设，直接导致了特征根为 1 即变量为单位根过程的研究。

针对 CC 方法论的批判和取消或弱化 CC 方法论的假设而发展起来的计量经济学，即从 19 世纪 60 年代之后发展起来的计量经济学，被称之为当代（高等）计量经济学。

方法论历程 大致内容有：

(1) CC成员 $T.Haavelmo$ 证明，标准的OLS运用于联立模型中的单个方程，将产生有偏估计。

(2) $Haavelmo$，1943，首先提出了间接最小二乘估计。

(3) $TKoopmans$，1949，定义并解决了识别问题。

(4) $Koopmans$ ，$Rubin$，$Leipnik$ 等发展了完全信息的极大似然估计。

(5) $Lucas$，1976，卢卡斯批判：使用计量经济模型预测未来经济政策的变化所产生的效用是不可信的。卢卡期批判是从计量经济学用于政策分析而提出的，其实质是提出了参数是否随时间而变化的问题，隐含了 CLM 产生不精确预测的重要原因是结构变化问题。

(6) $Sims$，1980，认为：为使结构方程可识别而施加了许多约束，这种约束是不可信的。因此，他建议使用向量自回归而回避结构约束问题。从预测结果看，VAR 优于结构联立方程系统，且规模较小。

1.2 计量经济学的研究目的

1．定量描述与分析经济活动，验证经济理论。包括描述宏观、微观经济问题。

2．寻找经济规律、建立经济计量模型，为制定经济政策服务。通过计量模型得到参数（边际系数，弹性系数，技术系数，比率，速率等）的可靠估计值，从而为制定政策，实施宏观调控提供依据

3．做经济预测。这是经济计量学利用模型所要解决的最重要内容，也是最困难的内容。经济计量学的发展史就是谋求对经济变量做出更精确预测的发展史。这要求（1）变量选择要准确，（2）模型形式要合理。

1.3 计量经济模型及其应用

计量经济模型

定义 1.1 (**模型**) 模型，是对现实的描述和模拟。对现实的各种不同的描述和模拟方法，就构成了各种不同的模型，例如，语义模型（也称逻辑模型）、物理模型、几何模型、数学模型和计算机模拟模型等。

(1) 语义模型是用语言来描述现实，例如，对供给不足下的生产活动，我们可以用"产出量是由资本、劳动、技术等投入要素决定的，在一般情况下，随着各种投入要素的增加，产出量也随之增加，但要素的边际产出是递减的"来描述。

(2) 物理模型是用简化了的实物来描述现实，例如，一栋楼房的模型，一架飞机的模型。

(3) 几何模型是用图形来描述现实，例如一个零部件的加工图。

(4) 计算机模拟模型是随着计算机技术而发展起来的一种描述现实的方法，在经济研究中有广泛的应用，例如人工神经元网络技术就是一种计算机模拟技术。

(5) 数学模型是用数学语言描述现实，也是一种重要的模型方法，它能够揭示现实活动中的数量关系。

定义 1.2 (**经济数学模型**) 经济数学模型是用数学方法描述经济活动。根据所采用的数学方法不同、对经济活动揭示的程度不同，构成各类不同的经济数学模型。

定义 1.3 (**数理经济模型**) 数理经济模型揭示经济活动中各个因素之间的理论关系，用确定性的数学方程加以描述。利用数理经济模型，可以分析经济活动中各种因素之间的互相影响，为控制经济活动提供理论指导。但是，数理经济模型并没有揭示因素之间的定量关系。

定义 1.4 (**计量经济模型**) 计量经济模型揭示经济活动中各个因素之间的定量关系，用随机性的数学方程加以描述。

计量经济方法的基本过程

(1) 经济理论

(2) 理论的数学模型

(3) 理论的计量经济学模型

(4) 数据的收集整理

(5) 计量经济模型的参数估计

(6) 假设检验

(7) 预报和预测

(8) 控制或政策制定

计量经济模型的应用

思考 1.2 (计量经济模型有什么用途？)

结构分析 是对经济现象中变量之间相互关系的研究，研究的是当一个变量或几个变量发生变化时会对其它变量以至经济系统产生什么样的影响。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。

弹性，是经济学中一个重要概念，是某一变量的相对变化引起另一变量的相对变化的度量，即是变量的变化率之比。

乘数，也是经济学中一个重要概念，是某一变量的绝对变化引起另一变量的绝对变化的度量，即是变量的变化量之比，也称倍数。

比较静力分析，是比较经济系统的不同平衡位置之间的联系，探索经济系统从一个平衡点到另一个平衡点时变量的变化，研究系统中某个变量或参数的变化对另外变量或参数的影响。

经济预测 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。

政策评价 政策评价是指从许多不同的政策中选择较好的政策予以实行，或者说是研究不同的政策对经济目标所产生的影响的差异。

计量经济学模型用于政策评价，主要有三种方法。一是工具-目标法。给定目标变量的预期值，即我们希望达到的目标，通过求解模型，可以得到政策变量值。二是政策模拟。即将各种不同的政策代入模型，计算各自的目标值，然后比较其优劣，决定政策的取舍。三是最优控制方法。将计量经济学模型与最优化方法结合起来，选择使得目标最优的政策或政策组合。**检验与发展经济理论** 计量经济学模型的两方面功能。一是按照某种经济理论去建立模型，然后用表现已经发生的经济活动的样本数据去拟合，如果拟合很好，则这种经济理论得到了检验。这就是检验理论。二是用表现已经发生的经济活动的样本数据去拟合各种模型，拟合最好的模型所表现出来的数量关系，则是经济活动所遵循的经济规律，即理论。这就是发现和发展理论。

1.4 计量经济学的内容体系

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。

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。**学科角度**

(1) 经典计量经济学，以经济理论为导向，以揭示经济现象中的因果关系为目的，以线性随机方程为理论形式，主要应用回归分析方法估计模型。

(2) 广义计量经济学，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。它既包括几乎与经典的计量经济学同时发展的投入产出分析方法、时间序列分析方法等，也包括近30年来发展的许多新的计量经济学理论方法。

内容角度

(1) 理论计量经济学

(2) 应用计量经济学

模型类型 从模型类型角度，可以将计量经济学模型划分为经典线性模型、非经典线性模型、非线性模型、动态模型和无参数回归模型等。

(1) 经典线性模型是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的线性模型。

(2) 非经典线性模型是经典线性模型在模型结构方面的扩展。例如由经典的常参数线性模型扩展的变参数线性模型；由反映变量之间因果关系的经典线性模型扩展为并不反映因果关系的线性模型，诸如著名的MA、AR、ARMA等时间序列分析模型和线性增长模型；由根据经济理论和经济行为规律设定的经典线性模型扩展为根据对数据的协整分析而设定的误差修正模型；等等。

(3) 非线性模型是一类用非线性方程描述经济变量之间的非线性关系的经济数学模型，包括非线性单方程模型和非线性联立方程模型。非线性模型由于其估计方法的复杂性，构成了高级计量经济学的主要内容。

(4) 这里的动态模型是专指以英国计量经济学家D.F.Hendry为代表的学派所倡导的宏观计量经济模型。Hendry认为，在50至60年代，计量经济学的主导方法论是"结构模型方法"，即以先验给定的经济理论为建立模型的出发点，以模型的参数估计为重心，以参数估计值与其理论预期值相一致为判断标准。这种方法论在70年代后遇到了挑战，所以必须发展新的宏观计量经济模型方法论。在本书中将对它们进行详细的介绍。

(5) 无参数回归模型，顾名思义，这类模型没有明确的函数关系，所以也没有明确的待估参数，只有解释变量和被解释变量以及它们的样本观测值。无参数模型的提出是基于这样的认识：每个参数模型都隐含着一系列的经济学假设，例如C-D生产函数模型的中性技术进步假设、替代弹性不变假设等，而这些假设在实际上是无法满足的，所以参数模型中给定的函数关系实际上是不可靠的。无参数模型利用其适当的估计方法，通过样本观测值，找出被解释变量的变化规律。例如常用的权函数估计，就是通过样本观测值确定权重，将被解释变量的估计描述为被解释变量样本观测值的加权和。由于无参数模型最终也不能给出解释变量和被解释变量之间的结构关系，它在理论计量经济学中的意义大于其实用价值。

估计方法角度

最小二乘法 是经典线性计量经济学模型的最主要的估计方法。例如，在经典线性计量经济学模型满足基本假设时采用的普通最小二乘法，在经典线性计量经济学模型存在异方差性时采用的加权最小二乘法，在经典线性计量经济学模型存在序列相关性时采用的广义最小二乘法，在经典线性计量经济学模型存在随机解释变量时采用的工具变量方法，估计经典线性联立方程计量经济学模型的二阶段最小二乘法、三阶段最小二乘法，等等。

最大似然法 在经典线性计量经济学模型中，存在着一个与最小二乘方法对应的最大似然方法系列，例如与普通最小二乘法对应的最大似然法，与二阶段最小二乘法对应的有限信息最大似然法，与三阶段最小二乘法对应的完全信息最大似然法。

贝叶斯估计方法 主要特点是利用了非样本信息，即前验信息与后验信息。

广义矩方法 广义矩（$GMM, Generalized Method of Moments$）方法是矩方法（$MM, Method of Moments$）的一般化，也是一类依赖样本信息的参数估计方法。一般地，被解释变量的各阶原点矩是待估参数的函数。利用样本数据计算各阶原点矩的估计量，最简单的例如一阶原点矩（即期望）的估计量、二阶原点矩（即方差）的估计量；然后利用该估计量，求解关于待估参数估计量的各阶矩方程，以得到参数估计量。广义矩方法有其广泛的适用性，普通最小二乘法、最大似然法等都可以看成是它的特例。

数据角度

截面数据分析

时序数据分析

平行数据分析 $PanelData$

离散被解释变量数据分析 $ModelwithDiscretedependentVariable$，如 $Probit$ 模型和 $Logit$ 模型

受限被解释变量数据分析 $LimitedDependentVariable$，仅指模型的被解释变量的样本数据是受到某种限制

持续被解释变量数据分析 $DurationModel$，指模型的被解释变量的样本观测值是事件持续的期间长度

经典计量经济学 经典计量经济学（$ClassicalEconometrics$）一般指 20 世纪 70 年代以前发展并广泛应用的计量经济学。由 $R.Frish$ 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：⑴模型类型－随机模型；⑵模型导向－理论导向；⑶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；⑷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；⑸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：⑴应用模型方法论基础－实证分析、经验分析、归纳；⑵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；⑶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 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。

微观计量经济学 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（ $D.McFaddan$）对微观计量经济学作出原创性贡献。

微观计量经济学的主要内容包括：平行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 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。

第二章 估计方法引论

2.1 德尔塔方法

计量经济学家经常希望对模型参数的非线性函数进行推断，因此要求对参数估计非线性函数的标准差进行估计，或者更一般地，对这种函数向量的协方差矩阵进行估计。一种常用方法是以渐近近似为基础的所谓的**德尔塔方法**。

假设已估计出参数 θ，它可能是一个线性回归模型参数。我们感兴趣的是参数 $\gamma = g(\theta)$，$g(\dot{})$ 是连续可导的单调函数。在这种情况下，估计 γ 的方法是利用关系 $\hat{\gamma} = g(\hat{\theta})$。

对于 $g(\theta)$ 是线性函数或仿射函数的情况，我们看如何计算 $Var(\hat{\gamma})$。假定 $\gamma = W^T\theta$，则 $\hat{\gamma} = W^T\hat{\theta}$：

$$\begin{aligned} Var(W^T\hat{\theta}) &= E(W^T(\hat{\theta}-\theta)(\hat{\theta}-\theta)^TW) \\ &= W^TE((\hat{\theta}-\theta)(\hat{\theta}-\theta)^T)W \\ &= \sigma^2W^T(X^TX)^{-1}W \end{aligned}$$

德尔塔方法是想找出 $g(\theta)$ 的一个线性近似，然后应用上面这个结论。

非线性函数进行线性近似的常用数学工具是**泰勒定理**。它的最简单形式为**中值定理**:

$$f'(x) = \frac{f(b)-f(a)}{b-a}$$

$$f(a+h) = f(a) + hf'(a+\lambda h) \tag{2-1}$$

直到 h 的 p 次方的泰勒定理可表述为：

$$f(a+h) = f(a) + \sum_{i=1}^{p-1}\frac{h^i}{i!}f^{(i)}(a) + \frac{h^p}{p!}f'(a+\lambda h)$$

最常用的是二阶泰勒展开：

$$f(a+h) \cong f(a) + hf'(a) + \frac{h^2}{2}f'(a)$$

两种形式都要求一个正则条件：$f(x)$ 在 $[a, a+h]$ 上的 p 阶导数连续。

假定参数估计 $\hat{\theta}$ 是 $\sqrt{n}$ 一致和渐近正态的，则

$$n^{\frac{1}{2}}(\hat{\theta}-\theta_0) \rightarrow N(0, V(\hat{\theta})) \tag{2-2}$$

对 $g(\hat{\theta})$ 在 θ_0 附近进行泰勒展开：

$$\hat{\gamma} \cong g(\theta_0) + g`(\theta_0)(\hat{\theta}-\theta_0) \tag{2-3}$$

两个确定性量称为**渐近相等**当且仅当 $n \rightarrow \infty$ 时它们具有相同的极限。类似地，两个随机变量称为渐近相等的当且仅当它们以概率趋于相同的极限。通常我们需要因子 n 的适当次幂以使推导能够顺利进行。对上式两边乘以 $n^{1/2}$:

$$n^{1/2}(\hat{\gamma}-\gamma_0)\underline{a}g`(\theta_0)n^{1/2}(\hat{\theta}-\theta_0) \tag{2-4}$$

由上式立即可以导出估计 $\hat{\gamma}$ 标准误差的一个实用方法：

$$S_\gamma \equiv \left|g'(\hat{\theta})\right| S_\theta \tag{2-5}$$

2.2 修正的普通最小二乘估计

修正的普通最小二乘估计（$COLS$）是 $Richmand$ 于 1974 年首先提出的在普通最小二乘估计结果的基础上对常数项进行修正的一种估计方法

用修正的普通最小二乘法估计确定性统计边界生产函数模型，即是首先用最小二乘法估计平均生产函数，然后计算所有样本点的产出量的观测值与平均生产函数估计值之差，取其最大者加到平均生产函数的常数项上，即得到边界生产函数的常数项，进行而得到边界生产函数。

$$
\begin{aligned}
&Y = AK^{\alpha}L^{\beta}e^{-u} \\
&\ln Y = \ln A + \alpha \ln K + \beta \ln L - u \\
&\ln \hat{Y} = (\hat{a} - \hat{u}) + \hat{\alpha} \ln K + \hat{\beta} \ln L \\
&\hat{u} = \max(\ln Y_i - \ln \hat{Y}) \\
&\ln Y* = \hat{u} + \ln \hat{Y}
\end{aligned}
$$

2.3 普通最小二乘估计

残差平方和最小：

$$Q = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2-6}$$

得出参数估计量：

$$
\begin{aligned}
&\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \\
&= \begin{bmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix} \\
&= \frac{1}{T\sum x_t^2 - (\sum x_t)^2} \begin{bmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & T \end{bmatrix} \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix}
\end{aligned} \tag{2-7}
$$

2.4 广义最小二乘估计 Feasible Generalized Least Squares

对于模型

$$\boldsymbol{Y} = \boldsymbol{X}\mathrm{B} + \mathrm{N} \tag{2-8}$$

如果存在序列相关和异方差

$$
\begin{aligned}
&E(\mathrm{N}) = 0 \\
&Cov(\mathrm{NN}') = E(\mathrm{NN}') = \sigma^2\Omega \\
&\Omega = \begin{bmatrix} w_1 & w_{12} & \cdots & w_{1n} \\ w_{21} & w_2 & \cdots & w_{2n} \\ & & \ddots & \\ w_{n1} & w_{n2} & \cdots & w_n \end{bmatrix} = DD'
\end{aligned} \tag{2-9}
$$

用 D^{-1} 左乘方程

$$\boldsymbol{D}^{-1}\boldsymbol{Y} = \boldsymbol{D}^{-1}\boldsymbol{X}\mathrm{B} + \boldsymbol{D}^{-1}\mathrm{N} \tag{2-10}$$

可改写为

$$\boldsymbol{Y}^{*} = \boldsymbol{X}^{*}\mathrm{B} + \mathrm{N}^{*} \tag{2-11}$$

它具有无序列相关及无异方差的特性。如果 Ω 已知，可以用 OLS 法得出参数估计量。如果 Ω 未知，则需要进行估计:

$$\hat{B} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y \tag{2-12}$$

这个估计量称为 $FGLS$ 估计量。

一般地，Ω 可能包括某些参数 θ。如模型存在一阶自相关时，有

$$\Omega = \begin{bmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{bmatrix} \tag{2-13}$$

如果能够得到参数 θ 的一致估计，则 Ω 的 $FGLS$ 估计量是渐近等价的。

对于 Ω 的估计，可以先用 OLS 法估计模型，得出误差项估计量，然后以其作为 Ω 的估计：

$$\hat{\Omega} = \begin{bmatrix} \tilde{e}_1^2 & \tilde{e}_1\tilde{e}_2 & \cdots & \tilde{e}_1\tilde{e}_n \\ \tilde{e}_2\tilde{e}_1 & \tilde{e}_2^2 & \cdots & \tilde{e}_2\tilde{e}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_n\tilde{e}_1 & \tilde{e}_n\tilde{e}_2 & \cdots & \tilde{e}_n^2 \end{bmatrix} \tag{2-14}$$

2.5 分部回归法 Partitioned Regression

将解释变量分成两部分，对应的参数也分成两部分

$$\boldsymbol{Y} = \boldsymbol{X}_1\mathrm{B}_1 + \boldsymbol{X}_2\mathrm{B}_2 + \mathrm{N}$$

在解释变量和误差项不相关的情况下

$$\begin{bmatrix} \boldsymbol{X'}_1\boldsymbol{Y} \\ \boldsymbol{X'}_2\boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x'}_1\boldsymbol{x}_1 & \boldsymbol{x'}_1\boldsymbol{x}_2 \\ \boldsymbol{x'}_2\boldsymbol{x}_1 & \boldsymbol{x'}_2\boldsymbol{x}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathrm{B}}_1 \\ \hat{\mathrm{B}}_2 \end{bmatrix} \tag{2-15}$$

估计参数

$$\begin{aligned} \hat{\mathrm{B}}_1 &= (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{Y} - (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_2\hat{\mathrm{B}}_2 \\ &= (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1(\boldsymbol{Y} - \boldsymbol{X}_2\hat{\mathrm{B}}_2) \end{aligned} \tag{2-16}$$

如果

$$\boldsymbol{X'}_1\boldsymbol{X}_2 = \boldsymbol{0}$$

则

$$\hat{\mathrm{B}}_1 = (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{Y}$$

同理，

$$\hat{\mathrm{B}}_2 = (\boldsymbol{X'}_2\boldsymbol{X}_2)^{-1}\boldsymbol{X'}_2\boldsymbol{Y}$$

结论：如果两组解释变量是正交的，那么相应的参数估计量可以分别在仅以一部分变量为解释变量的情况下加以估计。再进一步，如果模型的所有解释变量都是互相正交的，那么可以将多元线性模型化成若干个一元模型加以分别估计。这就是所谓的"分部回归估计"。当然，上述两组解释变量是完全正交的情况在实际中是很难发现的，或者说是不可能出现的。所以，"分部回归估计"并没有实际意义。在应用主分量法时

在经典计量经济学模型中，读者已经很熟悉如下问题：一是当模型中某些变量被检验为不显著时，就要将这些变量从模型中剔除；二是当发现解释变量出现多重共线性时，一个有效的处理方法是从模型中去掉产生多重共线性的部分变量。这些无疑是可行的。但是，值得注意的是，由于模型原有的变量之间并不是正交的，所以当剔除部分变量后，所保留变量的参数估计值将发生变化，它所反映的不再仅仅是该变量与被解释变量之间的关系。这一点从分部回归中得到了证明。于是，它提醒人们，不能轻易剔除"不显著"的变量，除非它的显著性低到对被解释变量几乎没有影响。更不能轻易采用剔除变量的方法消除多重共线性，应该多采用差分法或主分量法等。

2.6 偏回归估计 Partial Regression

由分部回归内容

$$\boldsymbol{X'}_2\boldsymbol{Y} = \boldsymbol{X'}_2\boldsymbol{X}_1\hat{\mathrm{B}}_1 + \boldsymbol{X'}_2\boldsymbol{X}_2\hat{\mathrm{B}}_2$$

得

$$\begin{aligned}\boldsymbol{X'}_2\boldsymbol{Y} &= \boldsymbol{X'}_2\boldsymbol{X}_1((\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{Y} - (\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_2\hat{\mathrm{B}}_{\boldsymbol{2}}) + \boldsymbol{X'}_2\boldsymbol{X}_2\hat{\mathrm{B}}_2 \\ &= \boldsymbol{X'}_2\boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{Y} - \boldsymbol{X'}_2\boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_2\hat{\mathrm{B}}_{\boldsymbol{2}} + \boldsymbol{X'}_2\boldsymbol{X}_2\hat{\mathrm{B}}_2\end{aligned}$$

推导

$$\boldsymbol{X'}_2(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{Y} = \boldsymbol{X'}_2(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{X}_2\hat{\mathrm{B}}_{\boldsymbol{2}}$$

有

$$(\boldsymbol{X'}_2(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{X}_2)^{-1}\boldsymbol{X'}_2(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{Y} = \hat{\mathrm{B}}_{\boldsymbol{2}}$$

令

$$(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1) = \boldsymbol{M}$$

成立

$$\hat{\mathrm{B}}_{\boldsymbol{2}} = (\boldsymbol{X'}_2\boldsymbol{M}\boldsymbol{X}_2)^{-1}(\boldsymbol{X'}_2\boldsymbol{M}\boldsymbol{Y})$$

因为 $\boldsymbol{M}$ 是等幂矩阵

$$\begin{aligned}\hat{\mathrm{B}}_{\boldsymbol{2}} &= (\boldsymbol{X'}_2\boldsymbol{M'}\boldsymbol{M}\boldsymbol{X}_2)^{-1}(\boldsymbol{X'}_2\boldsymbol{M'}\boldsymbol{M}\boldsymbol{Y}) \\ &= (\boldsymbol{X}_2^*\boldsymbol{X}_2^*)^{-1}(\boldsymbol{X}_2^*\boldsymbol{Y}^*)\end{aligned}$$

其中

$$\begin{aligned}\boldsymbol{Y}^* = \boldsymbol{M}\boldsymbol{Y} &= (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{Y} \\ &= \boldsymbol{Y} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{Y}\end{aligned}$$

是 $\boldsymbol{Y}$ 仅以 $\boldsymbol{X}_1$ 为解释变量建立的模型的残差。

以 $\boldsymbol{X}_2$ 对 $\boldsymbol{X}_1$ 回归：

$$\hat{\boldsymbol{X}}_{\boldsymbol{2}} = \boldsymbol{X}_1((\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_{\boldsymbol{2}})$$

得出残差

$$\begin{aligned}\boldsymbol{X}_{\boldsymbol{2}} - \hat{\boldsymbol{X}}_{\boldsymbol{2}} &= \boldsymbol{X}_{\boldsymbol{2}} - \boldsymbol{X}_1((\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1\boldsymbol{X}_{\boldsymbol{2}}) \\ &= (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X'}_1\boldsymbol{X}_1)^{-1}\boldsymbol{X'}_1)\boldsymbol{X}_{\boldsymbol{2}} \\ &= \boldsymbol{M}\boldsymbol{X}_{\boldsymbol{2}} \\ &= \boldsymbol{X}_{\boldsymbol{2}}^*\end{aligned} \tag{2-17}$$

2.7 两阶段最小二乘

定义 2.1 (排斥性约束) 假定结构模型

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$$

有一个内生解释变量 y_2 和一个外生解释变量 z_1。假定现在我们有两个被排斥在该式之外的外生变量 z_2 和 z_3。它们不出现该式，且与误差项不相关的诸假定称为 $exclusionrestriction$。

外生变量的任何线性组合都是有效的 IV。为寻找最好的 IV，我们选择与 y_2 最高度相关的线性组合。这正是由 y_2 的**诱导型方程**$reducedformequation$ 所给出的：

$$y_2 = \phi_0 + \phi_1 z_1 + \phi_2 z - 2 + \phi_3 z_3 + v_2$$

获得 OLS 估计值 $\hat{y}_2$，用它做为 y_2 的 IV。在复工具条件下，IV 估计量也叫做**两阶段最小二乘估计量**$twostageleastsquaresestimator$。如果 z_2 和 z_3 在上式中不是联合显著的，做 IV 估计是在浪费时间。

2.8 交叉估计方法 Across Regression

交叉估计也是对模型参数进行部分回归的一种估计方法，但是与上述部分回归估计不同的是，它将模型的参数按照其性质分类，然后分别用不同的样本观测值，包括被解释变量的样本观测值，估计各类参数。那么自然地，它只是相对于某类应用模型而言。**问题的提出** 在需求函数模型中，解释变量一般为收入和价格，这两类变量对商品需求量的影响是不同的。按照协整理论，商品需求量和收入为流量指标，一般情况下为一阶单整，它们之间可能存在协整关系，反映了二者之间的长期关系；而价格水平一般是 0 阶单整，它对商品需求量具有短期影响。从直观上也可以看出，收入对商品需求量具有长期影响，价格对商品需求量只具有短期影响。它们的参数分别属于长期弹性和短期弹性，具有不同的性质。而一般说来，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。所以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。即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。

为什么时间序列数据适合于短期弹性的估计，而截面数据适合于长期弹性的估计？结合需求函数模型来看：在截面上，由于价格并不随收入而显著变化，所以对商品需求量起作用的是收入；而且，在同一截面上，不同的消费者的收入差距可能相当大，使得收入的样本观测值数据变化较大。两者综合，说明收入对需求量的影响适宜于用截面数据估计。如果用时间序列数据，由于收入随时间的变化是缓慢的，不同时间的收入的样本观测值数据变化较小，不宜于揭示收入对需求量的长期影响。反过来，价格的时间序列数据适宜于揭示价格对需求量的短期影响。

交叉估计不仅适于需求函数模型的估计，也适用于包含长期影响和短期影响两类解释变量的其它模型的估计。例如居民储蓄方程。居民新增储蓄由收入水平和利率决定，其中收入水平具有长期影响，利率具有短期影响，适合于用交叉估计方法进行分析。再如税收方程，以税基和税率为解释变量，也适合于用交叉估计方法进行估计

估计方法 以对数线性需求函数为例，为了简化，假设解释变量中只包括收入和自价格。对数线性需求函数为：

$$\ln q = \alpha_0 + \alpha_1 \ln I + \alpha_2 \ln p + \mu$$

现有某一年的截面数据，在这个截面上，价格是常数。按收入分组：

$$\ln q_j = a + \alpha_1 \ln I_j + \mu_j \qquad j = 1, 2, \cdots, m$$

采用经典线性单方程模型的估计方法估计得到 $\hat{\alpha}_1$。

当以时间序列数据为样本时，将模型写成：

$$\ln q_t = \alpha_0 + \alpha_1 \ln I_t + \alpha_2 \ln p_t + \mu_t$$

令

$$y_t = \ln q_t - \hat{\alpha}_1 \ln I_t$$

有

$$y_t = \alpha_0 + \alpha_2 \ln p_t + \mu_t$$

采用经典线性单方程模型的估计方法估计得到全部参数。

在截面分析时，样本观测值取自分组人均数据，而在时序分析时，样本观测值一般取总体平均数据，这里存在一致性问题。另外，采用的年份不同，得到的 $\hat{\alpha}_1$ 估计量也不同，这里存在一个任意性问题。所以，交叉估计作为一种实用方法，尚缺少理论计量经济学的支持。

2.9 最大似然估计方法

回归模型的普通的、非线性的、广义最小二乘和工具变量以及 GMM 都可以从矩方法中得出。这里引入另一种基本估计方法:极大似然方法。在回归模型误差项正态分布的假设下，极大似然估计量，简称 ML 估计量，与我们熟悉的各种最小二乘估计量相同。ML 估计量的主要缺点是它比矩方法要求更强的分布假设。

最大似然估计原理 对于最小二乘法，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模型能最好地拟合样本数据。而对于最大或然法，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。

极大似然方法估计的模型必须是**完全设定模型**。

定义 2.2 (**完全设定模型**)　完全设定模型是能够给出明确算法的模型，即只要知道模型参数值就能模拟出因变量值。任何一个完全设定的计量经济模型必须为计算机模拟提供一个明确的算式，如果能够利用模型产生模拟数据，该模型一定是完全设定的。

对于这样的模型，一旦参数值给定，则我们就具有模拟因变量所需要的所有信息。要对因变量进行模拟，必须知道其 PDF，即要知道将每个观测看作随机变量的 PDF，还要知道将所有样本看作随机向量时的联合 PDF。

在很多情况下，样本观测假定为统计独立的。因此，整个样本的联合分布等于各个观测密度的乘积：

$$f(y,\theta)=\prod_{t=1}^{n}f(y_t,\theta) \tag{2-18}$$

习惯上采用对数似然函数。

ML 估计量与 MM 估计量相同的情况十分普遍。但 ML 的一个优势是不需要求期望值。另外，如果一个估计量是 MLE，它将具有很多理想的渐近性质，这些性质使得标准误差的计算和检验统计量的计算更加容易。

我们首先讨论极大似然方法用于古典正态线性模型

$$Y=X\beta+U \qquad U\sim N(0,\sigma^2 I) \quad X\text{外生} \tag{2-19}$$

因此 X 条件下，Y 服从 $N(X\beta,\sigma^2)$。Y_t 的 PDF 为:

$$f_t(y_t,\beta,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}EXP(-\frac{(y_t-X_t\beta)^2}{2\sigma^2} \tag{2-20}$$

对数似然函数等于所有观测贡献的和：

$$\iota(y,\beta,\sigma)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta) \tag{2-21}$$

关于未知参数 β 和 σ 对上式求极大值得出 ML 估计量。第一步是关于参数 σ 对 $\iota(y,\beta,\sigma)$ 求极大值。将对数似然函数对 σ 求导，将导数看作数据和其他参数的函数，从一阶条件中解出 σ 并回代到对数似然函数中。这样得到的似然函数称为**集中似然函数** $concentratedloglikelihoodfunction$。第二步关于 β 对这个函数求极值。从推导中可以看出，极大化集中对数似然函数等价于极小化残差平方和函数（关于 β 的函数）。ML 估计量必定等于 OLS 估计量。β 的 ML 估计量等于 OLS 估计量依赖于误差项的正态分布假设。从不同的分布假设出发，可以得到不同的 ML 估计量。

ML 估计量有两种定义:

第一类 ML 估计量 在集合 Θ (参数 θ 取值的参数空间)上极大化对数似然函数得出的估计量。

第二类 ML 估计量 定义为似然方程的解，似然方程是一阶条件 $g(y,\hat{\theta})=0$。其中，$g(y,\theta)$ 是**梯度向量**，或者称为**得分向量**。其代表元素为

$$g_i(y,\theta) \equiv \frac{\partial\iota(y,\theta)}{\partial\theta_i} = \sum_{t=1}^{n}\frac{\partial\iota(y_t,\theta)}{\partial\theta_i}$$

极大似然估计往往很容易计算。得不到直接表达式时，同其他情形一样，必须采用非线性极大化程序。牛顿法和拟牛顿法经过少许修改就可以用于 ML 估计。牛顿法的基本方程是：

$$\theta_{(j+1)} = \theta_j - H_{(j)}^{-1}g_{(j)} \tag{2-22}$$

拟牛顿法的公式是：

$$\theta_{(j+1)} = \theta_j + \alpha D_{(j)}^{-1}g_{(j)} \tag{2-23}$$

ML 估计量的渐近性质 在相当弱的条件下，ML 估计量是一致的，在稍强一些的假设下是渐近正态的。

一致性的证明：首先证明在参数真值处取值的对数似然函数的期望大于在其他处取值的期望。要证明一致性，还需要有限样本下的可识别条件和渐近可识别条件。

定理 2.1 (*Jensen* **不等式**) 如果 X 是一个实值随机变量，$h(\dot{})$ 是一个凹函数，则 $E(h(X)\leq h(E(X))$。当 $h(\dot{})$ 至少在随机变量 X 的支撑的一部分上是严格凹的，则严格不等式成立。所谓**支撑**是指一个实数集合，X 的密度在这个集合上不为零，支撑包含的点要多于一个。

将这个不等式应用于比值 $L(\theta^*)/L(\theta_0)$，其中 θ_0 是参数真值，θ^* 为模型参数空间中的任意一个向量。因为对数似然函数是非负实数上的严格凹函数，并且似然函数是非负函数，从詹森不等式得出：

$$E_0\log\frac{L(\theta^*)}{L(\theta_0)} < \log E_0\frac{L(\theta^*)}{L(\theta_0)} \tag{2-24}$$

E_0 表示在参数向量 θ_0 刻画的 DGP 下取期望。右边的期望可以用随机向量 y 支撑上的积分表示：

$$E_0\frac{L(\theta^*)}{L(\theta_0)} = \int\frac{L(\theta^*)}{L(\theta_0)}L(\theta_0)dy = 1$$

从而有

$$E_0\log\frac{L(\theta^*)}{L(\theta_0)} = E_0\iota(\theta^*) - E_0\iota(\theta_0) < 0 \tag{2-25}$$

(再对对数似然函数应用大数定律。)它可以得出：

$$p\lim_{n\to\infty}\frac{1}{n}\iota(\theta^*) \leq p\lim_{n\to\infty}\frac{1}{n}\iota(\theta_0) \tag{2-26}$$

对所有的 $\theta^*\neq\theta_0$ 成立，因为极限中的不等式不一定是严格的。因为 MLE 是极大化 $\iota(\theta)$：

$$p\lim_{n\to\infty}\frac{1}{n}\iota(\hat{\theta}) \geq p\lim_{n\to\infty}\frac{1}{n}\iota(\theta_0) \tag{2-27}$$

两个不等式同时成立说明：

$$p\lim_{n\to\infty}\frac{1}{n}\iota(\hat{\theta}) = p\lim_{n\to\infty}\frac{1}{n}\iota(\theta_0)$$

一致性的证明还需要满足**渐近可识别条件**:对所有有 $\theta^*\neq\theta_0$，都有$plimn^{-1}\iota(\theta^*)\neq plimn^{-1}\iota(\theta_0)$。

考虑不独立模型的似然函数和对数似然函数的构建，如当回归函数中包括滞后因变量的情形。对于一个用极大似然方法估计的模型，密度函数依赖于 k 维向量参数 θ:

$$f(y^n,\theta) = \prod_{t=1}^{n}f(y_t|y^{t-1};\theta) \tag{2-28}$$

其中每个观测的边际密度用条件密度代替。对应的对数似然函数具有和式的形式：

$$\iota(y^n,\theta)=\sum_{t=1}^{n}\iota_t(y^t,\theta) \tag{2-29}$$

其中省略了表示整体样本的 y 的上标。

定义表明，梯度向量的每个分量是 n 个贡献之和，当观测不独立仍然正确；但 ι_t 关于 θ_i 的偏导数依赖于 y^t 而不再仅仅是 y_t。将这些偏导数写成一个矩阵会带来方便。定义 $n\times k$ 矩阵 $G(y,\theta)$，其代表性元素为：

$$G_{ti}(y^t,\theta)\equiv\frac{\partial\iota_t(y^t,\theta)}{\partial\theta_i} \tag{2-30}$$

这个矩阵称为**梯度贡献矩阵**。因为:

$$g_i(y,\theta)=\sum_{t=1}^{n}G_{ti}(y^t,\theta) \tag{2-31}$$

因此，梯度向量的每个分量都是矩阵 $G(y,\theta)$ 向量的元素的和。$G(y,\theta)$ 的一个关键性质是，如果 y 是由 θ

有限信息最大似然法 有限信息最大似然法 $(LIML, Limited Information Maximum Likelihood)$ 是一种以最大似然为准则、通过对简化式模型进行最大似然估计，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"最大似然法"是针对结构方程中包含的内生变量的简化式模型的，即应用最大似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。

$$\mathrm{B}\boldsymbol{Y}+\Gamma\boldsymbol{X}=\mathrm{N}$$

每一个方程（以第一个方程为例）可以改写为如下的形式：

$$Y_1=\beta_{12}Y_2+\beta_{13}Y_3+\cdots+\beta_{1g_1}Y_{g_1}+\gamma_{11}X_1+\gamma_{12}X_2+\cdots+\gamma_{1k_1}X_{k_1}+\mathrm{N}_1$$

用矩阵形式表达是：

$$Y_1=(\boldsymbol{Y}_0,\boldsymbol{X}_0)\begin{pmatrix}\mathrm{B}_{\mathbf{0}}\\ \Gamma_{\mathbf{0}}\end{pmatrix}+\mathrm{N}_1$$

其中

$$\boldsymbol{Y}_0=\begin{bmatrix}Y_2 & Y_3 & \cdots & Y_{g_1}\end{bmatrix}=\begin{bmatrix}y_{21} & y_{31} & \cdots & y_{g_1 1}\\ y_{22} & y_{32} & & y_{g_1 2}\\ \vdots & \vdots & & \vdots\\ y_{2n} & y_{3n} & & y_{g_1 n}\end{bmatrix}$$

$$\boldsymbol{X}_0=\begin{bmatrix}X_1 & X_2 & \cdots & X_{k_1}\end{bmatrix}=\begin{bmatrix}x_{11} & x_{21} & \cdots & x_{k_1 1}\\ x_{12} & x_{22} & & x_{k_1 2}\\ \vdots & \vdots & & \vdots\\ x_{1n} & x_{2n} & & x_{k_1 n}\end{bmatrix}$$

$$\mathrm{B}_0=\begin{bmatrix}\beta_{12}\\ \beta_{13}\\ \vdots\\ \beta_{1g_1}\end{bmatrix}\quad \Gamma_0=\begin{bmatrix}\gamma_{11}\\ \gamma_{12}\\ \vdots\\ \gamma_{1k_1}\end{bmatrix}$$

$$Y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} \quad \mathrm{N}_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1n} \end{bmatrix}$$

第一个结构方程可以表示为

$$(\boldsymbol{Y}_0^1, \boldsymbol{X}_0) \begin{pmatrix} \mathrm{B}_0^1 \\ \Gamma_{\mathbf{0}} \end{pmatrix} + \mathrm{N}_1 = 0$$

其中

$$\boldsymbol{Y}_0^1 = (Y_1, \boldsymbol{Y}_0) \qquad \mathrm{B}_0^1 = \begin{pmatrix} -1 \\ \mathrm{B}_0 \end{pmatrix} \tag{2-32}$$

该方程包含的内生变量的简化式模型为：

$$\boldsymbol{Y}_0^1 = \boldsymbol{X}\Pi_0^1 + \mathrm{E}_0^1 \tag{2-33}$$

其对数似然函数：

$$Ln\ L(\boldsymbol{Y}_0^1) = c + \frac{n}{2} ln \left|\Omega_0^{-1}\right| - \frac{1}{2} tr(\Omega_0^{-1})(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1)'(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1) \tag{2-34}$$

的最大化等价于广义方差

$$(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1)'(\boldsymbol{Y}_0^1 - \boldsymbol{X}\Pi_0^1) \tag{2-35}$$

的极小化，得到的 $\hat{\Pi}_0^1$ 就是简化式模型的最大似然估计量。

完全信息最大似然估计 完全信息最大似然法($FullInformationMaximumLikelihood, FIML$)是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似然函数达到最大以得到所有结构参数的估计量。

2.10 贝叶斯估计

贝叶斯($Bayes$)统计是由 $T.R.Bayes$ 于 19 世纪创立的数理统计的一个重要分支，20 世纪 50 年代，以 $H.Robbins$ 为代表提出了在计量经济学模型估计中将经验贝叶斯方法与经典方法相结合，引起了广泛的重视，得到了广泛的应用。贝叶斯估计对经典计量经济学模型估计方法的扩展在于，它不仅利用样本信息，同时利用非样本信息。

贝叶斯方法的基本原理 贝叶斯方法是与传统（也称经典的）计量经济学模型的估计方法相对的一种统计学方法。它的基本思路是：认为要估计的模型参数是服从一定分布的随机变量，根据经验给出待估参数的先验分布（也称为主观分布），关于这些先验分布的信息被称为先验信息；然后根据这些先验信息，并与样本信息相结合，应用贝叶斯定理，求出待估参数的后验分布；再应用损失函数，得出后验分布的一些特征值，并把它们作为待估参数的估计量。

贝叶斯方法与经典估计方法的主要不同之处是：

- 关于参数的解释不同:经典估计方法认为待估参数具有确定值，它的估计量才是随机的，如果估计量是无偏的，该估计量的期望等于那个确定的参数；而贝叶斯方法认为待估参数是一个服从某种分布的随机变量。

- 所利用的信息不同:经典方法只利用样本信息；贝叶斯方法要求事先提供一个参数的先验分布，即人们对有关参数的主观认识，被称为先验信息，是非样本信息，在参数估计过程中，这些非样本信息与样本信息一起被利用。

- 对随机误差项的要求不同:经典方法，除了最大或然法，在参数估计过程中并不要求知道随机误差项的具体分布形式，但是在假设检验与区间估计时是需要的；贝叶斯方法需要知道随机误差项的具体分布形式。

- 选择参数估计量的准则不同:经典方法或者以最小二乘，或者以最大或然为准则，求解参数估计量；贝叶斯方法则需要构造一个损失函数，并以损失函数最小化为准则求得参数估计量。

贝叶斯定理

$$g(\theta\,|Y) = \frac{f(Y\,|\theta)g(\theta)}{f(Y)} \tag{2-36}$$

2.11 渐近方法

来自[16]第十四章渐近方法。

渐近分布

定义 2.3 (**渐近分布**) 用 $T_1 < T_2 < \cdots < T_N$ 表示连续递增样本容量，设在每个样本容量 T_i 下重复抽样，则每个 X_{T_i} 都有自己的样本均值和方差利用递增样本可以求得随机样本序列 $X_T = \{X_{T_1}, \cdots, X_{T_N}\}$，其中每个元素是相应样本容量下的一个随机变量。当 $T_N \to \infty$ 时，这些分布收敛于某一分布，则称该分布为渐近分布或极限分布。

相应的，可建立渐近期望和渐近方差概念。

问题 2.1 $\lim E[X_T - E(X_T)]^2 = 0$，即分布退化成一点。

用 T 乘 $Var(X_T)$ 使 $TVar(\hat{X}) \to \sigma^2$。**O、o 记号** 随机变量：$O_p, o_p$；非随机变量: O, o。

设 $\{a_n\}$ 和 $\{b_n\}$ 是两个实数序列：

定义 2.4 $a_n = O(b_n)$，读作 a_n 是大 Ob_n，若比值 $|a_n/b_n|$ 对大的 n 都有界；或，存在一个数 $\boldsymbol{K}$ 和一个整数 $n(\boldsymbol{K})$ 使得当 n 大于 $n(\boldsymbol{K})$ 后总有 $|a_n| > \boldsymbol{K}\,|b_n|$。即，$\{a_n\}$ 和比较的序列 $\{b_n\}$ 有相同的阶。

定义 2.5 $a_n = o(b_n)$，读作 a_n 是小 ob_n，若比值 $|a_n/b_n|$ 收敛于零；或，$\forall \varepsilon > 0$，$\exists n(\varepsilon)$，使得当 $n > n(\varepsilon)$ 时，就有 $|a_n| < \varepsilon\,|b_n|$。即，$\{a_n\}$ 和比较的序列 $\{b_n\}$ 有更小的阶。

想法是比较 $\{b_n\}$ 对 $\{a_n\}$ 近似的阶或量。一些重要的 $\{b_n\}$ 是 $b_n = n^{-1}, b_n = n^{-1/2}, b_n = n, b_n = n\log n$。

阶的比较是涉及序列的"大 N"的性质而与序列的初始值无关。

(1) $\{a_n\}$ 的值是无穷或是对有限个 n 没有定义，它是不受此影响的；

(2) 若 $\|\boldsymbol{a}_n\|$ 表示向量 $\boldsymbol{a}_n$ 的长度，即

$$\|\boldsymbol{a}_n\| = \sqrt{\sum_i a_{ni}^2} \tag{2-37}$$

(2.4)和(2.5)都可用于向量序列 $\{\boldsymbol{a}_n\}$；

(3) 若 c 是非零常数，则 $a_n = O(b_n)$ 与 $a_n = O(cb_n)$ 是等价的；

(4) $a_n = o(1)$ 表示 $a_n \to 0$，$a_n = O(1)$ 表示对某个 $\boldsymbol{K}$，只要 n 充分大就有 $|a_n| \leq \boldsymbol{K}$，即 $|a_n|$ 本质上是界的；

(5) $a_n = O(a_n)$ 总成立；

(6) 乘积规律：

$$\begin{aligned} O(a_n)O(b_n) &= O(a_n b_n) \\ O(a_n)o(b_n) &= o(a_n b_n) \\ o(a_n)o(b_n) &= o(a_n b_n) \end{aligned} \tag{2-38}$$

(7) 求和规律：和数的阶是被加项中最大的的阶。（当被加的项依赖于 n 时就不一定对）

例 2.1　e 的近似

序列 e_n 是：

$$e_n = (1+\frac{1}{n})^n \tag{2-39}$$

可先取对数 $\log(e_n) = n\log(1+n^{-1})$，求取对数后的极限再反求对数得到。令 $f(t) = \log(1+t)$，泰勒一阶展开：

$$\begin{aligned} & f(t) \\ &= f(0) + f'(0)t + o(t) \\ &= t + o(t) \end{aligned} \tag{2-40}$$

所以

$$\begin{aligned} & \log(e_n) \\ &= n\log(1+n^{-1}) \\ &= n(n^{-1} + o(n^{-1}) \\ &= 1 + o(1) \end{aligned} \tag{2-41}$$

因此

$$\log(e_n) \to 1 \tag{2-42}$$

问题 2.2　$\{e_n\}$ 收敛到 e 速度比较慢，为此修正 e_n 并考虑 $f(t)$ 的高阶展开式。

定义序列 x_n 为

$$x_n = \log(1+n^{-1}) \tag{2-43}$$

考虑序列

$$\{(n+c)x_n\} \tag{2-44}$$

其中 c 为待定常数。将序列 x_n 二阶展开：

$$x_n = f(n^{-1}) = n^{-1} - \frac{1}{2}n^{-2} + o(n^{-2}) \tag{2-45}$$

所以

$$\begin{aligned} & (n+c)x_n \\ &= (n+c)(n^{-1} - \frac{1}{2}n^{-2} + o(n^{-2}) \\ &= 1 + \frac{c}{n} - \frac{1}{2}\frac{1}{n} - \frac{1}{2}cn^{-2} + no(n^{-2}) + co(n^{-2}) \\ &= 1 + (c - \frac{1}{2})n^{-1} + o(n^{-1}) \end{aligned} \tag{2-46}$$

选择 $c = \frac{1}{2}$，$(n+c)x_n$ 收敛于 1 的阶从 $o(1)$ 改进到 $o(n^{-1})$。这样定义一个新序列 e_n^*：

$$e_n^* = (1+n^{-1})^{n+\frac{1}{2}} \tag{2-47}$$

它的收敛速度比 e_n 要快。

同理，可按三阶泰勒展开，考虑 $\{(n+c+dn^{-1}\}$ 序列。$f(t)=\log(1+t)$ 的三阶展开是

$$f(t)=t-\frac{1}{2}t^2+\frac{1}{3}t^3+o(t^3) \tag{2-48}$$

则

$$\begin{aligned}&(n+c+dn^{-1})x_n\\&=(n+c+dn^{-1})(n^{-1}-\frac{1}{2}n^{-2}+\frac{2}{3}n^{-3}+o(n^{-3})\end{aligned} \tag{2-49}$$

可以证明 $c=\frac{1}{2}$，$d=-\frac{1}{12}$。

随机序列的收敛性 依分布收敛 渐近方法的众多的重要统计应用之一是计算显著性水平概率的近似值和给出置信区间，依分布收敛是评判这些近似的技术性工具。

若 $\{\boldsymbol{X}_n\}$ 是一元随机变量序列，则它的分布函数是

$$F_n(x)=P\{X_n\le x\} \tag{2-50}$$

若 X_n 是离散型随机变量，则 $F_n(x)$ 是右连续的阶梯函数，只在 X_n 取值的点上有跳跃。

定义 2.6 (**依分布收敛**) 设 X_n 的分布函数是 F_n，X 的分布函数是 F，则当

$$\lim_{n\to\infty}F_n(x)=F(x) \tag{2-51}$$

对 F 的所有连续点 x 都成立，则称 X_n 依分布收敛于 X。

依概率收敛 一个数字的常数 C 总是看成一个退化的随机变量 C，它的分布是全部概率集中于一个点 c 上。它的分布函数是

$$F_c(x)=\begin{cases}0 & x<c\\1 & x\geqslant c\end{cases} \tag{2-52}$$

F_n 是一个右连续函数，只有一个不连续点 $x=c$。

定义 2.7 (**依概率收敛于常数**) 如果对每个 $\varepsilon>0$，有

$$\lim_{n\to\infty}\boldsymbol{P}\{|\boldsymbol{X}_n-c|\le\varepsilon\}=1 \tag{2-53}$$

则称 $\{\boldsymbol{X}_n\}$ 依概率收敛于c，写成 $\boldsymbol{X}_n\overrightarrow{p}c$。

定理 2.2 若 $\boldsymbol{C}$ 是一个退化的随机变量满足 $P\{\boldsymbol{C}=\boldsymbol{c}\}=1$，则 $\mathscr{L}[\boldsymbol{X}_n]\to\mathscr{L}[\boldsymbol{C}]$ 等价于 $\boldsymbol{X}_n\overrightarrow{p}c$。

依分布收敛是依概率收敛于一个常数的推广。

确立依分布收敛的方法 直接证明 $F_n\to F$ 要证明对 F 的所有连续点都有 $F_n(x)\to F(x)$。

例 2.2 长等待时间的分布

设 $\boldsymbol{X}_n$ 有几何分布

$$\boldsymbol{P}\{\boldsymbol{X}_n=k\}=(1-p_n)^{k-1}p_n \tag{2-54}$$

它的分布函数 $G_n(x)$ 是一个阶梯函数，在每一个正整数上有跳跃，在它们之间是常数：

$$G_n(x)=P\{X_n\leqslant x\}=\begin{cases}1-(1-p_n)^{[x]} & x>0\\0 & x\leqslant 0\end{cases} \tag{2-55}$$

等待成功的期望次数是 p^{-1}，对 $p=p_n=\lambda n^{-1},\lambda>0$，当 $n\to\infty$ 时研究几何等待次数的情况。

令 $E(X_n)=\lambda^{-1}n$，

问题 2.3 当 $n\to\infty$ 时，$X_n\to\infty$。

为避免此一情况，考虑规格化变量

$$Y_n = n^{-1} X_n \tag{2-56}$$

Y_n 的分布函数是

$$F_n(x) = P\{Y_n \le x\} = P\{X_n \le nx\} = G_n(nx) \tag{2-57}$$

目的是求 $F_n(x)$ 的极限。对 $x \le 0$，易知 $F_n(x) \to 0$。当 $x > 0$时，

$$F_n(x) = 1 - (1-\lambda n^{-1})^{nx}(1-\lambda n^{-1})^{[nx]-nx} \tag{2-58}$$

易知

$$(1-\lambda n^{-1})^{-1} \ge (1-\lambda n^{-1})^{[nx]-nx} \ge 1 \tag{2-59}$$

因此对任一 $x > 0$：

$$(1-\lambda n^{-1})^{[nx]-nx} \to 1 \tag{2-60}$$

或当 $n \to \infty$ 时，有

$$(1-\lambda n^{-1})^{[nx]-nx} = 1 + o(1) \tag{2-61}$$

又

$$(1-\lambda n^{-1})^{nx} \to e^{-\lambda x} \tag{2-62}$$

所以

$$F_n(x) \to 1 - e^{-\lambda x} \tag{2-63}$$

因此，几何分布的极限分布是指数分布。

用矩母函数间接证明 随机变量或向量的矩母函数 MGF 是离散分布确立依分布收敛最重要的工具。设 $\boldsymbol{X} = (X_1, \cdots, X_T$ 是一个随机向量，则 $\boldsymbol{X}$ 的矩母函数由下式给出：

$$\boldsymbol{X}_X(t_1, \cdots, t_T) = \boldsymbol{E}(e^{t_1 X_1 + \cdots + t_T X_T}) \tag{2-64}$$

定理 2.3 若(2-64)的期望值在 $\mathbf{0} = (0, \cdots, 0)$ 点附近的一个邻域内对 $(t_1, \cdots, t_T)$ 都有限，则

$$E(X_1^{\alpha_1} \cdots X_T^{\alpha_T} = \frac{\partial^{\alpha_1 + \cdots + \alpha_T}}{\partial t_1^{\alpha_1} \cdots \partial t_T^{\alpha_T}} M_X(t_1, \cdots, t_T)|_{t=0} \tag{2-65}$$

定理 2.4 设 $\boldsymbol{X}_n$ 的 MGF 是 $M_n(t)$，$\boldsymbol{X}$ 的 MGF 是 $M(t)$，当 $M_n(t) \to M(t)$ 在 $\mathbf{0} = (0, \cdots, 0)$ 的一个邻域内对所有 t 成立，则 $\mathscr{L}[\boldsymbol{X}_n] \to \mathscr{L}[\boldsymbol{X}]$。

2.12 广义矩方法

GMM 就是极小化

$$q = m(\beta)' W^{-1} m(\beta) \tag{2-66}$$

估计量为

$$\hat{\beta} = \arg\min\,(m(\beta)' W^{-1} m(\beta)) \tag{2-67}$$

权矩阵的选择是矩估计方法的核心问题。

$Hansen's$(1982) 提出最佳的权矩阵为：

若随机误差项存在异方差且不存在自相关，$White$(1980) 提出权矩阵的估计量为：

$$\hat{W} = \frac{1}{n} S_0 \tag{2-68}$$

若随机误差项存在自相关，$Newey$ 和$West$(1987) 提出权矩阵的估计量为：

L 的选取准则为：使得随机误差项滞后大于的序列相关小到可以忽略不计。

GMM 估计在大样本情况下是渐近有效的，在小样本情况下是无效的。所以，只有在大样本情况下，才使用 GMM 方法进行参数估计。

估计步骤 对于模型

$$y_i = h(X_i, \beta) + \varepsilon_i \tag{2-69}$$

(1) 采用 OLS 法估计模型得出 $\hat{\beta}$；

(2) 计算权矩阵的估计量；

(3) 代入(2-67)。

假设检验

2.13 稳健估计 robust estimation method

设第 i 个观察的古典线性回归模型为

$$Y_t = X_t\alpha + \varepsilon_t \qquad t = 1, 2, \cdots, T \tag{2-70}$$

其中随机扰动项 ε_t 为独立同分布，且有分布密度 $f(\cdot)$。**M–估计** 设 $\rho(\cdot)$ 是一个任一给定的上凸函数($convex$)，则称最小值问题：

$$\min_{\alpha} \sum_{i=1}^{T} \rho(Y_t - X_t\alpha) \tag{2-71}$$

的解 $\alpha = \hat{\alpha}$ 为参数 α 的一个 $M-$ 估计。若 $\rho(\cdot)$ 有导函数 $\psi(\cdot)$，则它等价于下面这个方程的解：

$$\sum_{i=1}^{T} X_t \psi(Y_t - X_t\alpha) = 0 \tag{2-72}$$

$$\text{最小绝对偏差估计}\rho(x) = |x|$$

$$\text{最小二乘估计}\rho(x) = \frac{x^2}{2}$$

$$\text{极大似然估计}\rho(x) = -\ln f(x)$$

L–估计 设参数 $0 < \theta < 1$，最小值问题

$$\min \sum_{t:Y_t \geq X_t\alpha}^{T} \theta \left| Y_t - X_t\alpha \right| + \sum_{t:Y_t < X_t\alpha}^{T} (1-\theta) \left| Y_t - X_t\alpha \right| \tag{2-73}$$

若取 $\theta = \frac{1}{2}$，则为最小绝对偏差（$minimumobsolutedeviation$）估计。进一步，若误差项服从一个双边指数分布，即具有分布密度：

$$f(x) = \frac{\lambda}{2} e^{-\lambda |x|}$$

则这个 MAD 估计就是极大似然估计。

2.14 工具变量法

一阶差分或固定效应估计排除了不随时间变化的变量。迄今为止我们已研究出的综列数据法还不能解决与解释变量相关的随同时间变化的遗漏变量的问题。

在存在遗漏变量时，OLS 估计量是有偏误且非一致的。对未能观测到的解释变量给出适宜的代理变量，能消除或减轻遗漏变量偏误。

对于内生性问题的解决，一是工具变量法；一是两阶段最小二乘法。

IV 法可:

- 在存在遗漏变量的情况下，获得一致性估计量
- 解决含误差变量 $errors - in - variable$ 的问题
- 估计联立方程模型

定义 2.8 (**工具变量**)　若简单回归模型：$y = \beta_0 + \beta_1 x + u$ 。假定我们有一个可观测到的变量 z，它满足两个假定：

(1) z 与u 不相关：$Cov(z, u) = 0$

(2) z 与x 相关：$Cov(z, x) \neq 0$

(3) 为用 IV 估计量做统计推断，需要同方差假定$E(u^2|z) = \sigma^2 = Var(u)$

我们称 z 是 x 的 $instrumentalvariable$。

工具变量估计量:

$$\hat{\beta}_1 = \frac{Cov(z, y)}{Cov(z, x)} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})} \tag{2-74}$$

大数定律表明，IV 估计量具有一致性：$plim(\hat{\beta}_1) = \beta_1$。它的渐近方差为：

$$\frac{\sigma^2}{n\sigma_x^2 \rho_{x,z}^2} \tag{2-75}$$

在高斯–马尔科夫假定下，OLS 估计量的方差为 σ^2 / SST_x ，而 IV 估计量类似的计算式是: $\sigma^2 / SST_x R_{x,z}^2$。所以，$IV$ 估计量的渐近方差总是大于 OLS 估计量的渐近方差。

2.15　蒙特卡罗模拟

[style=numbers,caption=SAS:产生AR1 序列]

源代码 2.1

```
OPTIONS MSTORED SASMSTORE=LI;

%MACRO AR1(N=50,A=0.7,B=0.3,SEED=-1,CNAME=AR1SER,DNAME=
    AR1DST);
DATA AR;
    ARRAY OBS OBS1-OBS&N; *** Room is made for observations;

    OBS(1)=RANNOR(&SEED); *** Generated the first observation;

    DO J=2 TO &N;
      OBS(J)=SQRT(&A)*OBS(J-1)+SQRT(&B)*RANNOR(&SEED);
    END;

    OUTPUT;
    KEEP OBS1-OBS&N;
RUN;

PROC TRANSPOSE OUT=&DNAME; DATA &DNAME;
    SET &DNAME;
    RENAME COL1=&CNAME;
RUN;

PROC ARIMA DATA=&DNAME;
    IDENTIFY VAR=&CNAME NLAG=1;
    ESTIMATE P=1;
RUN;
%MEND;

%AR1
```

The multivariate AR model is defined by

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \varepsilon_t$$

What may not be obvious is that a Φ matrix holds not only the AR coefficients for each process,but also cross-lag coefficients existing between the multiple series collected.

Consider the case in which a researcher wants to model three 200-observation series holding cross-lagged relationships.Again,because the first observation in each series equals the first error,we begin by specifying those codes .

Program (2.2) generates three series of 200 observations each with a Multivariate AR1 process with cross-lagged relationships.The cross-lagged relationship between Series A and B is .20;between Series A and C is .10;and between Series B and C is .30.

Program (2.3) accumulated the results of 200 replications.

源代码 2.2 Generating Multivariate AR1

```
DATA GENERATE;
        ***each creates room for 200 observations;
        ARRAY SERIEA SERIEA1-SERIEA200;
        ARRAY SERIEB SERIEB1-SERIEB200;
        ARRAY SERIEC SERIEC1-SERIEC200;
        ARRAY VAR_A VAR_A1-VAR_A200;
        ARRAY VAR_B VAR_B1-VAR_B200;
        ARRAY VAR_C VAR_C1-VAR_C200;

        ***generates the  first  occasion;
         SerieA(1)=rannor(-1);
         SerieB(1)=rannor(-11);
         SerieC(1)=rannor(-111);

         VAR_A(1)=0;
         VAR_B(1)=0;
         VAR_C(1)=0;

        DO J=2 TO 200;
           *** random normal deviates are generated;
           VAR_A(J) = rannor(-1111);
           VAR_B(J) = rannor(-11111);
           VAR_C(J) = rannor(-111111);
           *** the equations which generate the process  follow;
           SerieA(J)=VAR_A(J)*SQRT(.20) + SerieA(J-1)*SQRT(.80);
           SerieB(J)=VAR_B(J)*SQRT(.20) + SerieB(J-1)*SQRT(.60) + SerieA(J-1)*
                (.20);
           SerieC(J)=VAR_C(J)*SQRT(.50) + SerieC(J-1)*SQRT(.45) + SerieA(J-1)*
                (.10) + SerieB(J-1)*(.30);
      END;

      KEEP SERIEA1-SERIEA200
              SERIEB1-SERIEB200
              SERIEC1-SERIEC200;
      OUTPUT;
RUN;
```

源代码 2.3 Macro for Multivariate AR1

```
OPTIONS LINESIZE=100 NOSOURCE NOSOURCE2 NONOTES;
lIBNAME MULTITS 'D:\SAS\DST\MYLIB';

%MACRO MULTITS (REPS,N,VARA,AR_AA,VARB,AR_BB,AR_BA,VARC,AR_
    CC,AR_CA,AR_CB,SCENARI);

%DO J=1 %TO &REPS; *** 200 replications are undertaken;

DATA GENERATES; *** creates space for the observations;
      ARRAY SERIEA SERIEA1-SERIEA&N;
      ARRAY SERIEB SERIEB1-SERIEB&N;
      ARRAY SERIEC SERIEC1-SERIEC&N;
      ARRAY VAR_A VAR_A1-VAR_A&N;
      ARRAY VAR_B VAR_B1-VAR_B&N;
      ARRAY VAR_C VAR_C1-VAR_C&N;

      SERIEA(1)=RANNOR(-1);
      SERIEB(1)=RANNOR(-11);
      SERIEC(1)=RANNOR(-111);
      VAR_A(1)=0;
      VAR_B(1)=0;
      VAR_C(1)=0;

      DO J=2 TO &N;

      VAR_A(J)=RANNOR(-1111);
      VAR_B(J)=RANN0R(-11111);
      VAR_C(J)=RANNOR(-111111);

      *** generate the mulitvariate process with cross-lagged relationships;
      SERIEA(J)=VAR_A(J)*SQRT(&VARA) + SERIEA(J-1)*SQRT(&AR_AA);
      SERIEB(J)=VAR_B(J)*SQRT(&VARB) + SERIEB(J-1)*SQRT(&AR_BB)
               + SERIEA(J-1)*(&AR_BA);
      SERIEC(J)=VAR_C(J)*SQRT(&VARC) + SERIEC(J-1)*SQRT(&AR_CC)
               + SERIEA(J-1)*(&AR_CA) + SERIEB(J-1)*(&AR_CB);

      END;
      KEEP SERIEA1-SERIEA&N
           SERIEB1-SERIEB&N
           SERIEC1-SERIEC&N;
      OUTPUT;

RUN;
*** transpose series a to prepare it for analysis;
DATA GENERAT1;
      SET GENERATE&J;
```

```
        KEEP SERIEA1-SERIEA&N;
RUN;
PROC TRANSPOSE OUT=GENERAT1;
RUN;
DATA GENERAT1;
        SET GENERAT1;
        ID=_N_;
        SERIESA=COL1;
        OUTPUT;
        DROP COL1;
RUN;

*** transpose series b to prepare it for analysis;
DATA GENERAT2;
        SET GENERATE&J;
        KEEP SERIEB1=SERIEB&N;
PROC TRANSPOSE OUT=GENERAT2;
DATA GENERAT2;
        SET GENERAT2;
        ID=_N_;
        SERIESB=COL1;
        OUTPUT;
        DROP COL1;

*** transpose series c to prepare it for analysis;
DATA GENERAT3;
        SET GENERATE&J;
        KEEP SERIEC1-SERIEC&N;
PROC TRANSPOSE OUT=GENERAT3;
DATA GENERAT3;
        SET GENERAT3;
        ID=_N_;
        SERIESC=COL1;
        OUTPUT;
        DROP COL1;
*** merge the three series by observation;
DATA GENERATE&J;
        MERGE GENERAT1 GENERAT2 GENERAT3;
        BY ID;
        KEEP ID SERIESA SERISEB SERIESC;
*** fits the multivariate model to the series data;
PROC STATESPACE INTERVAL=DAY MAXIT=200 ARMAX=1 LAGMAX=1 NOPRINT
                    OUTMODEL=ARCOEFFS;
    *** outputs the collected statistics var seriesa seriesb seriesc;
    TITLE 'MULTIVARIATE ARIMA (1,0,0) CROSSLAGGED RESULTS';

```

```
    *** plucks parameter estimates for series a out of statespace output;
DATA GENERAT1;
     SET ARCOEFFS;
     IF _N_=1;
     KEEP F_1 F_2 F_3 SIG_1 SIG_2 SIG_3;
     RESCOVAA=SIG_1; RESCOVAB=SIG_2; RESCOVAC=SIG_3;
     AR_AA=F_1**2; AR_AB=F_2**2; AR_AC=F_3;
     KEEP AR_AA AR_AB AR_AC RESCOVAA RESCOVAB RESCOVAC;
     OUTPUT;

   *** plucks parameter estimates for series B out of statespace output;
DATA GENERAT2;
     SET ARCOEFFS;
     IF _N_=3;
     KEEP F_1 F_2 F_3 SIG_1 SIG_2 SIG_3;
     RESCOVBA=SIG_1; RESCOVBB=SIG_2; RESCOVBC=SIG_3;
     AR_BA=F_1**2; AR_BB=F_2**2; AR_BC=F_3;
     KEEP AR_BA AR_BB AR_BC RESCOVBA RESCOVBB RESCOVBC;
     OUTPUT;

   *** plucks parameter estimates for series C out of statespace output;
DATA GENERAT3;
     SET ARCOEFFS;
     IF _N_=1;
     KEEP F_1 F_2 F_3 SIG_1 SIG_2 SIG_3;
     RESCOVCA=SIG_1; RESCOVCB=SIG_2; RESCOVCC=SIG_3;
     AR_CA=F_1**2; AR_CB=F_2**2; AR_CC=F_3;
     KEEP AR_CA AR_CB AR_CC RESCOVCA RESCOVCB RESCOVCC;
     OUTPUT;

    *** merges the parameter estimate for each of the three series ;
DATA ARCOEFFS;
     MERGE GENERAT1 GENERAT2 GENERAT3 GENERATE&J;
     SCENARIO=&SCENARI;
     OUTPUT;
     KEEP AR_AA AR_AB AR_AC RESCOVAA RESCOVAB RESCOVAC
          AR_BA AR_BB AR_BC RESCOVBA RESCOVBB RESCOVBC
          AR_CA AR_CB AR_CC RESCOVCA RESCOVCB RESCOVCC
          SCENARIO M_A M_B M_C VAR_A VAR_B VAR_C;

     *** frees some of the memory space for SAS;
     PROC DELETE DATA=GENERATE&J;
     PROC DELETE DATA=GENERAT1;
     PROC DELETE DATA=GENERAT2;
     PROC DELETE DATA=GENERAT3;

     *** adds the parameter information to an output file for one of 200
```

```
                  iterations ;
PROC APPEND BASE=MULTITS.RESULT (CNTLLEV=MEMBER);

%END;
%MEND COMPUTE;

*** the following  macro statements feed the macro above the parameters for each
        condition under study;
%MULTITS(200,100,.20,.80,
                    .20,.60,.20,
                    .15,.45,.10,.30,1)  ;

%MULTITS(200,500,.20,.80,
                    .20,.60,.20,
                    .15,.45,10,.30,2)  ;
%MULTITS(200,1000,.20,.80,
                     .20,.60,.20,
                     .15,.45,.10,.30,3)  ;

PROC SORT;
     BY SCENARIO;

*** summmary summarizes the results for each condition;
PROC SUMMARY PRINT VARDEF=N MAXDEC=2 FW=8;
     CLASS SCENARIO;
     VAR AR_AA AR_AB AR_AC
         AR_BA AR_BB AR_BC
         AR_CA AR_CB AR_CC
         RESCOVAA RESCOVAB RESCOVAC
         RESCOVBA RESCOVBB RESCOVBC
         RESCOVCA RESCOVCB RESCOVCC;

         TITLE1 '****␣SCENARIO␣1␣(N=100)␣****';
         TITLE2 '****␣SCENARIO␣2␣(N=500)␣****';
         TITLE3 '****␣SCENARIO␣3␣(N=1000)␣****';
         TITLE5 'AR_AA=.80␣AR_BB=.60␣AR_CC=.45';
         TITLE6 'AR_BA=.20␣AR_CA=.10␣AR_CB=.30';
         TITLE7 '␣';
RUN;
```

源代码 2.4 PROC ARIMA Macro Example for a Monte Carlo Study

```
LIBNAME AUTOREG 'D:\MYDOCUMENT\DATASET\SAS';

OPTIONS LINESIZE=100 NOSOURCE NOSOURCE2 NONOTES; *** Log file will
report errors only;

/*****************************************************************************************/
/*This program was written to consider only six different research
 situations.                    */ /*Parametes:
*/ /* N        The number of observations in the series.
*/ /* ARLAG The squared value of the lag relationship desired.
*/ /* VARIANCE The squared value of the error variance desired.
*/ /* SCENARI The designated number id of a given condition.
*/ /* REP      The replications of programme.
*/
/******************************************************************************************/
*** Lables six conditions;
 PROC FORMAT;
        VALUE SCENE
        1 = 'ARLAG**2 = .40, N = 100'
        2 = 'ARLAG**2 = .60, N = 100'
        3 = 'ARLAG**2 = .80, N = 100'
        4 = 'ARLAG**2 = .40, N = 500'
        5 = 'ARLAG**2 = .60, N = 500'
        6 = 'ARLAG**2 = .80, N = 500'
        ;
*** Macro begins;
 %MACRO AR1(N,ARLAG,VARIANCE,SCENARI,REP);
      %DO J=1 %TO &REP; ***  replications ;

      DATA SEM&J;
           ARRAY SERIE SERIE1-SERIE&N;***Room is made for observations;

           SERIE(1)=RANNOR(-1); *** generate AR1 data;
           DO J=2 TO &N;
               SERIE(J) = SQRT(&ARLAG)*SERIE(J-1) + SQRT(&VARIANCE)*RANNOR(-1);
           END;
           OUTPUT;
           KEEP SERIE1-SERIE&N;
        RUN;

        PROC TRANSPOSE OUT=D2;***Transposes series to a column;
        RUN;

```

```
        DATA D2;
            SET D2;
        RUN;

        PROC ARIMA DATA=D2;
            IDENTIFY VAR=COL1 NALG=1 NOPRINT;
            ESTIMATE P=1 OUTSTAT=ARIMA1 OUTMODEL=ARIMA2
                NOPRINT MAXIT=5000;
        RUN;

        DATA ARIMA1;
            SET ARIMA1;
            KEEP _VALUE_ _STAT_;
            IF 1<=_N_<=2;
        PROC TRANSPOSE OUT=ARIMA1;

        DATA ARIMA1;
            SET ARIMA1;
            AIC=COL1;
            SBC=COL2;
            OUTPUT;
            KEEP AIC SBC;

        DATA ARIMA2;
            SET ARIMA2;
            IF _N_=6;
            ARLAG=_VALUE_;
            ARLAGSTD=_STD_;
            TTEST=ARLAG/ARLAGSTD;
            OUTPUT;
            KEEP ARLAG TTEST;

         DATA COMBINE;
             MERGE ARIMA1 ARIMA2;
             SCENARIO=&SCENARI;
             OUTPUT;

          PROC APPEND BASE=AUTOREG.RESULT1 (CNTLLEV=MEMBER);

    %END;
  %MEND ;***End of Macro;

  %AR1(100,.40,.60,1,10);
  %AR1(100,.60,.40,2,10);
  %AR1(100,.80,.20,3,10);
  %AR1(100,.40,.60,4,10);
  %AR1(100,.60,.40,5,10);
```

```
%AR1(100,.80,.20,6,10);

FORMAT SCENARIO SCENE;

PROC SORT;
     BY SCENARIO;

PROC SUMMARY PRINT VARDEF=N MAXDEC=2 FW=8;
     CLASS SCENARIO;
     VAR AIC SBC ARLAG TTEST;
RUN;
```

2.16 例子

需求与GDP

例 2.3 以我国国内生产总值(GDP)为被解释变量，居民消费总额(COM)、政府消费额(GOV)、投资总额(INV)为解释变量建立线性模型，用以分析各项最终需求与用生产法计算的国内生产总值之间的统计关系。

源代码 2.5 SAS:PartialRegression

```
proc reg data=li.partial;
      OLS: model gdp=com gov inv;
proc reg data=li.partial;
      E:   model gdp=inv;
      output out=a p=ahat;
data a(keep=e);
      set a;
      e=gdp-ahat;
proc reg data=li.partial;
      E1:  model com=inv;
      output out=b p=bhat;
data b(keep=e1);
      set b;
      e1=com-bhat;
run; proc reg data=li.partial;
      E2:  model gov=inv;
      output out=c p=chat;
run; data c(keep=e2);
      set c;
      e2=gov-chat;
run; data temp;
      merge a b c;
run; proc reg data=temp;
      model e=e1 e2;
run;
```

第三章　经济计量检验

$G.B.Wetherill$(1986) 指出，古典线性回归模型的实际应用所引起的问题主要有二类：一是由模型设定及随机扰动项的问题所导致，如非线性、异方差性、序列相关性、模型设定误差与非正态误差；另一是关于回归变量及其数据质量的问题，如随机回归变量、小样本、多重共线性、异常值。

3.1　异方差问题

异方差的表现与来源 经济时间序列中的异方差常为递增型异方差。金融时间序列中的异方差常表现为自回归条件异方差。递增型异方差的来源主要是因为随着解释变量值的增大，被解释变量取值的差异性增大。

(1) 因变量的测量误差；

(2) 省略某些变量；

(3) 模型数学形式设定错误；

(4) 随机系数模型；

(5) 异常值；

(6) 经验教训的吸取及资料收集技术的改进。

异方差的形式 第一种是分组型异方差（$groupwiseheteroscedasticity$）。假定 T 个观察值可以分成 G 组，每一组观察值都是同方差。对 G 个模型进行合并，则形成异方差。

第二种是函数形式异方差（$functionalheteroscedasticity$）。又称 $Breusch--Pagan--Godfrey(1978,1979)$ 假定，是对扰动项方差所做的函数形式假定：

$$\sigma_t^2 \equiv Var(\varepsilon_t) = h(Z_t'\beta) \tag{3-1}$$

其中 $h(\cdot)$ 是一元的二次可微函数， Z_t 是由一组非随机外生变量的第 t 观测所作成的 $r+1$ 维向量，其第一分量取 1。β 是 $r+1$ 维未知参数向量。常见形式有：平方项，一次项，指数项。**异方差的后果** 普通最小二乘法参数估计量不再有效；显著性检验失效；预测精度下降。**异方差的检验 图解法**

(1) 对数据进行 OLS 回归得出残差平方；

(2) 作出残差平方和关于解释变量或拟合值的散点图；

(3) 进行观测，看残差是否有规律性。

源代码 3.1 SAS：残差散点图

```
PROC REG DATA=LI.MACROECO NOPRINT;
      MODEL Y=I G C;
      OUTPUT OUT=ABC RESIDUAL=R PREDICTED=P;
RUN;

DATA TEMP;
      MERGE LI.MACROECO ABC;
      KEEP I G C R P;
RUN; DATA TEMP;
      SET TEMP;
      R2=R*R;
RUN; PROC GPLOT DATA=TEMP;
      PLOT R*P='.';
      PLOT R2*P='+';
RUN;
```

斯皮尔曼等级相关检验 异方差的实质是随机扰动项与解释变量相关，这种相关是不能用皮尔逊线性相关度量的，因为此时它恒为零。

$$\rho = 1 - 6\frac{\sum_{t=1}^{T} d_t^2}{T(T^2-1)} \tag{3-2}$$

选用统计量

统计量 3.1 (**斯皮尔曼**)

$$t^* = \sqrt{T-2}\frac{\rho}{\sqrt{1-\rho^2}}$$

它服从自由度为 $T-2$ 的 t 分布。

原假设是不存在斯皮尔曼相关。

源代码 3.2 SAS：斯皮尔曼等级相关检验

```
PROC REG DATA=LI.MACROECO NOPRINT;
      MODEL Y=I G C;
      OUTPUT OUT=ABC RESIDUAL=R PREDICTED=P;
RUN;

DATA TEMP;
      MERGE LI.MACROECO ABC;
      ABSR=ABS(R);
      KEEP ABSR P I G C;
RUN; PROC CORR DATA=TEMP SPEARMAN NOSIMPLE;
      VAR P I G C;
      WITH ABSR;
RUN;
```

Bartlett 检验 用于探测($detection$)分组型异方差。

$$\begin{aligned} &H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_G^2 \\ &H_1 : \exists i, j \in \{1, 2, \cdots, G\}, \sigma_i^2 \neq \sigma_j^2 \end{aligned} \tag{3-3}$$

统计量 3.2 $M = \frac{(T-G)\ln\hat{\sigma^2} - \sum_{g=1}^{G}(T_g - p)\ln\hat{\sigma_g^2}}{1+\frac{1}{3}(G-1)(\sum_{g=1}^{G}(T_g-p)^{-1} - (T-G)^{-1}}$

渐近服从自由度为 $G-1$ 的卡方分布。其中 p 为解释变量个数，$\hat{\sigma^2} = \frac{1}{T-G}\sum_{g=1}^{T}(T_g - p)\hat{\sigma_g^2}$

源代码 3.3 SAS：三公司间的异方差分析

```
%MACRO BART(DA=);
PROC REG DATA=LI.&DA ;
      &DA: MODEL Y=X1 X2;
      OUTPUT OUT=&DA PREDICTED=Y_&DA.HAT RESIDUAL=R_&DA;
RUN;

PROC UNIVARIATE DATA=&DA NORMAL;
      VAR R_&DA;
RUN; DATA &DA;
      SET &DA;
      AR_&DA=ABS(R_&DA);
      OUTPUT;
RUN; PROC CORR DATA=&DA NOSIMPLE SPEARMAN;
      VAR Y_&DA.HAT X1 X2;
      WITH AR_&DA;
RUN;
%MEND;

%BART(DA=GM)
%BART(DA=GE)
%BART(DA=WH)

DATA _NULL_;
      P=1-PROBCHI(34.94,2);
      PUT P=;
RUN;
```

将相应的数据代入 *Bartlett* 检验公式。

本质上这个检验在一个在假定随机扰动项服从正态分布，各组样本中的扰动项相互独立成立的前提下的似然比(LR)检验。在实际应用中，存在许多问题：应如何分组及有可能不能将样本分组。

$Goldfeld - Quandt$ **检验** 它是一个将样本分成三组的 *Bartlett* 检验，但又属于一种函数型异方差检验法。该方法是基于异方差 σ_4^2 与某一解释变量成正相关的假定：

$$\sigma_t^2 = \sigma^2 X_{it}^2$$

该方法有三个前提条件：$T > 2p$；正态误差；误差不相关。检验步骤是：第一，将样本按某个指定解释变量的升序排序，然后去掉中间 c 个样本值（一般取为 $T/4$）。将剩余样本分成两个子样本。第二步是对两个子样本分别进行回归，得出残差平方和。第三步构造 F 统计量：

$$F = \frac{ESS_1/([\frac{T-c}{2} - (p+1)])}{ESS_2/([\frac{T-c}{2} - (p+1)])}$$

当摸型含有多个解释变量时，应以每一个解释变量为基准检验异方差。此法只适用于递增型异方差。对于截面样本，计算 F 统计量之前，必须先把数据按解释变量的值从小到大排序。

已经证明，此检验的势取决于 c 的取值。$Goldfeld\&Quandt$(19651972) 建议若 T 为 30 左右，c 取 4；若 T 为 60 左右，c 取 10。**怀特检验** 由 $H.White$ 于 1980 年提出，属于 LM 检验。它不需要对观测值进行排序，也不依赖于随机误差项服从正态分布。首先用普通最小二乘法估

计模型，得出残差项；用残差对解释变量，解释变量平方项、解释变量交叉项进行回归，注意要包括常数项；构造 LM 统计量：

$$TR^2 \sim \chi^2(k) \tag{3-4}$$

Glejser检验 看 $|\hat{u}|$ 是否与解释变量有关系。

自回归条件异方差（$ARCH$）检验

表 3-1 函数型异方差的探测方法

检验法	σ_t^2、$auxilliaryregression$、$statistic$
Park	$\sigma_t^2 = \sigma^2 X_t^{\beta}$ $\ln \hat{\varepsilon}_t^2 = \beta_0 + \beta_1 \ln X_t + v_t$ Ft
$Glejs$	$\sigma_t^2 = (\beta_0 + \beta_1 X_{jt}^h)^2$ $\lvert\hat{\varepsilon_t}\rvert = \beta_0 + \beta_1 X_{jt}^h + v_t$ $h = 1, -1, 1/2, -1/2$ Ft
BPC	$\sigma_t^2 = f(\beta_0 + \beta_1 Z_{1t} + \cdots + \beta_m Z_{mt})$ $\hat{\varepsilon_t^2} - \hat{\sigma^2} = f(\cdot) + v_t$ $TR^2 \sim \chi^2(m)$
$White$	$\sigma_t^2 = \beta_0 - \sum_{i=1}^{p} \beta_i X_{it} + \sum_{i,j=1}^{p} \beta_{ij} X_{it} X_{jt}$ $\hat{\varepsilon}_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i X_{it} + \sum_{i,j=1}^{p} \beta_{ij} X_{it} X_{jt} + v_t$ $TR^2 \sim \chi^2(\frac{(p+1)(p+2)}{2} - 1)$

例子

源代码 3.4 SAS:异方差检验

```
data cpss;
    set li.hetska;
run; proc reg data=cpss;
    model lnwage=grade potexp exp2 union;
    output out=c01 p=predict r=residual;
run; proc print data=c01; run; data tt;
    set c01;
    rsq=residual*residual;
run; proc plot data=tt;
    plot rsq*predict="*";
run;
```

源代码 3.5 SAS:white test

```
data wt;
    set tt;
    grade2=grade*grade;
    exp4=exp2*exp2;
    exp3=potexp*exp2;
    gx2=grade*exp2;
    gu=grade*union;
    pu=potexp*union;
    eu=exp2*union;
    gp=grade*potexp;
run; proc reg data=wt;
    model rsq=grade potexp exp2 union grade2 exp4 exp3 gx2 gp gu pu eu;
run; data t1;
    chis1=cinv(0.95,12);
run; proc print data=t1; run; proc reg data=cpss;
    model lnwage=grade potexp exp2 union/spec;
run;
```

源代码 3.6 SAS:BG test

```
proc reg data=wt;
        model rsq=grade potexp union;
run; data wt1;
        set wt;
        rsqadjust=rsq/0.2099;
run; proc reg data=wt1;
        model rsqadjust=grade potexp union;
run; data t2;
        chis2=cinv(0.95,3);
run; proc print data=t2; run;
```

源代码 3.7 SAS:GQ test

```
proc sort data=wt;
        by potexp;
run; data w001;
        set wt;
        if 36=<_n_<=65 then delete;
run; proc print data=w001; run; data w01;
        set w001;
        if 36=<_n_<=70 then delete;
run; data w02;
        set w001;
        if 1=<_n_<=35 then delete;
run; proc reg data=w01;
        model lnwage=grade potexp exp2 union;
        output out=c011 p=predict1 r=residual1;
run; proc reg data=w02;
        model lnwage=grade potexp exp2 union;
        output out=c012 p=predict2 r=residual2;
run; data t3;
        fvalue=finv(0.95,30,30);
run; proc print data=t3; run;
```

3.2 序列相关性 Serial Correlation

什么是序列相关

$$E(u_i, u_j) \neq 0 \tag{3-5}$$

即随机扰动项的取值与其前后期的值相关联。特别在，当 $|t-s|=1$ 时，称为一创序列相关或自相关($auto-correlation$)。一般经验告诉我们，对于采用时间序列数据作样本的计量经济学问题，由于在不同样本点上解释变量以外的其他因素在时间上的连续性，带来它们对被解释变量的影响的连续性，所以往往存在序列相关性。随机误差项之间的相关性主要表现为一阶序列相关。但是，连续的一阶序列相关实际上构成了多阶序列相关。

如考虑误差的自回归模型：

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \qquad |\rho| < 1 \quad v_t\text{白噪声} \tag{3-6}$$

则有

$$
\begin{aligned}
Var(\varepsilon_t) \equiv \sigma_\varepsilon^2 = \frac{\sigma_v^2}{1-\rho^2} \\
Var(\varepsilon_t, \varepsilon_{t-s}) = \rho^s \sigma_\varepsilon^2
\end{aligned}
\tag{3-7}
$$

序列相关的原因

(1) 经济变量的惯性($Inertia$)，即本期值与前期值有关，如企业固定资产的形成，不仅与本期投资相关，而且与前几年的投资相关。

(2) 模型设定偏倚。由于建模时对次要解释变量归于误差项，从而由次要解释变量产生的惯性将表现在随机扰动项中；

(3) 错误的函数形式的影响。当进行线性近似时，余项被包含在随机扰动项。

(4) 蛛网现象$cobwebphenomenon$。

(5) 当采用内插或修匀等方法处理删失值时易产生变量相关。

序列相关的后果

(1) 参数估计量非有效：当计量经济学模型出现序列相关性，其普通最小二乘法参数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$E(NN‘) = \sigma^2 I$

(2) 变量的显著性检验失去意义: t分布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。如果出现了序列相关性，t 检验就失去意义。采用其他检验也是如此。

(3) 模型的预测失效:由于上述后果，使得模型不具有良好的统计性质。所以，当模型出现序列相关性时，它的预测功能失效。

序列相关性的检验 思路

(1) 采用普通最小二乘法估计模型，求出残差估计值$\hat{e}_i = y_i - \hat{y}_{ols}$

(2) 分析残差估计值之间的相关性

回归检验法 以 $\hat{e}_i$ 为被解释变量，以各种可能的相关量，如 $\hat{e}_{i-1}$ 等为解释变量建立方程。如果存在某种函数形式，使得议程显著成立，则认为存在序列相关性。具体应用时需要反复试算。回归检验法的优点是一旦确定了模型存在序列相关性，也就同时知道了相关的形式，而且它适用于任何类型的序列相关性问题的检验。

冯诺曼比法

$$
\frac{\sum\limits_{i=2}^{n} (\hat{e}_i - \hat{e}_{i-1})^2 \Big/ (n-1)}{\sum\limits_{i=1}^{n} (\hat{e}_i - \bar{\hat{e}}_{i-1})^2 \Big/ n}
\tag{3-8}
$$

当样本容量足够大时（大于 30），该统计量近似服从正态分布。计算该统计量的值，将它与具有正态分布的理论分布值进行比较，如果大于临界值，表示不存在序列相关，如果小于临界值，表示存在序列相关。

DW **法**

$$
\frac{\sum\limits_{i=2}^{n} (\hat{e}_i - \hat{e}_{i-1})^2}{\sum\limits_{i=1}^{n} \hat{e}_i^2}
\tag{3-9}
$$

经验表明，如果不存在一阶自相关，一般也不存在高阶序列相关。所以在实际应用中，对于序列相关问题一般只进行 $D.W.$检验。

BG **检验** DW 统计量只适用于一阶自相关检验，而对于高阶自相关检验并不适用。利用 BG 统计量可建立一个适用性更强的自相关检验方法，既可检验一阶自相关，也可检验高阶自相关。BG 检验由 $Breusch-Godfrey$ 于 1978 年提出。BG 检验是通过一个辅助回归式完成的，既可检验一阶自相关，也可检验高阶自相关。具体步骤如下。对于多元回归模型

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_{2t} x_{2t} + \cdots + \beta_k x_{kt} + u_t \tag{3-10}$$

考虑误差项为 p 阶自回归形式：

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + v_t \tag{3-11}$$

其中 v_t 为随机扰动项，符合各种假定。

零假设为：

$$H_0: \quad \rho_1 = \rho_2 = \cdots = \rho_p = 0 \tag{3-12}$$

表示不存在序列相关。

用估计(3-10)得到的残差建立辅助回归式：

$$\hat{u}_t = \hat{\rho}_1 u_{t-1} + \hat{\rho}_2 u_{t-2} + \cdots + \hat{\rho}_p u_{t-p} + \beta_0 + \beta_1 x_{1t} + \beta_{2t} + \cdots + \beta_k x_{kt} + v_t \tag{3-13}$$

估计上式并计算 R^2。构造 LM 统计量：

$$LM = TR^2 \tag{3-14}$$

在零假设成立的情况下，LM 统计量近似服从 $\chi^2(p)$ 分布。如果零假设成立，LM 统计量的值将很小。

有滞后因变量时的检验：$Durbinh$ **统计量** 在有一个或多个滞后内生变量时，即使是误差项确实存在序列相关，DW 的值也会接近于 2。

$Durbinh$ 统计量对于大样本严格有效，也能够用于小样本。假设我们用普通最小二乘法对下面这个公式进行估计：

$$Y_t = \alpha + \beta Y_{t-1} + \gamma X_t + \epsilon_t \tag{3-15}$$

检验用统计量定义为：

$$h = \hat{\rho}\sqrt{\frac{T}{1 - TVar(\hat{\beta}}} \tag{3-16}$$

$\hat{\rho}$ 可以直接从 DW 统计量的估计得到，因为

$$DW \approx 2(1 - \hat{\rho}) \tag{3-17}$$

$Durbin$ 证明了 h 统计量近似地服从方差为 1 的正态分布。

源代码 3.8 SAS:序列自相关检验

```
data ice;
      set  li.autcor;
run; proc reg data=ice;
      model cons=price income temp/dw;
      output out=ice1 p=consp r=resid;
run; symbol1 i=none v=dot c=blue h=.5; proc gplot data=ice1;
      plot  resid*time=1/vref=0;
run;
```

$rh0 = 0.40063$，$R - square = 0.1541$，$t = 1$，$R - square = 0.1792$；

源代码 3.10 SAS:yule-walker estimates

```
proc autoreg data=ice;
      model cons=price income temp/nlag=1 method=yw ;
run;
```

源代码 3.9 SAS:BG test

```
data tt1;
        set ice1;
        resid1=lag(resid);
run; proc reg data=tt1;
        model resid=resid1 /noint;
run;

data bgt;
        bg=29*0.1541;
        chisq=cinv(0.95,1);
        if bg>chisq then t=1;
                          else t=0;
        put t=;
run;

data tt2;
        set ice1;
        resid1=lag(resid);
        resid2=lag(resid1);
        resid3=lag(resid2);
run; proc reg data=tt2;
        model resid=resid1 resid2 resid3/noint;
run;

data bgt2;
        bg=(29-3)*0.1792;
        chisq=cinv(0.95,3);
        if bg>chisq then t=1;else t=0;
        put t= chisq= bg=;
run;
```

源代码 3.11 SAS:COCHRANE-ORCUTT estimates

```
proc reg data=ice;
      model cons=price income temp/dw;
      output out=tt p=chat r=res;
run; proc print data=tt; run; data tt;
      set tt;
      relag=Lag(res);
run; proc print data=tt; run; proc reg data=tt outest=b1;
      model res=relag/noint;
run;

data pp;
     set tt;
     c1=lag(cons);
     t1=lag(temp);
     i1=lag(income);
     p1=lag(price);
run; proc print data=pp; run; data pp1;
     set pp;
     if _n_=1 then delete;
     c2=cons-0.40063*c1;
     t2=temp-0.40063*t1;
     i2=income-0.40063*i1;
     p2=price-0.40063*p1;
run; proc print data=pp1; run; proc reg data=pp1;
      MODEL c2=t2 i2 p2/dw;
run;
```

源代码 3.12 Box-Pierce test

```
data tt;
      set ice1;
      r1=lag(resid);
      r2=lag(r1);
      r3=lag(r2);
run; proc reg data=tt;
      model resid=r1/noint;
      model resid=r2/noint;
      model resid=r3/noint;
run; data test;
      q1=29*(0.44196*0.44196+0.05569*0.05569+0.01345*0.01345);
      q2=29*31*(0.44196*0.44196+0.05569*0.05569+0.01345*0.01345)/26;
      chis=cinv(0.95,3);
      if q1>chis then d1=1;
                  else d1=0;
      if q2>chis then d2=2;
                  else d2=0;
```

```
        put d1= d2=;
run;
```

序列相关性下的估计方法 如果模型的误差项存在自相关，首先应分析产生自相关的原因。如果自相关是由于错误地设定模型的数学形式所致，那么就应当修改模型的数学形式。怎样查明自相关是由于模型数学形式不妥造成的？一种方法是用残差 e_t 对解释变量的较高次幂进行回归，然后对新的残差作 DW 检验，如果此时自相关消失，则说明模型的数学形式不妥。

如果自相关是由于模型中省略了重要解释变量造成的，那么解决办法就是找出略去的解释变量，把它做为重要解释变量列入模型。怎样查明自相关是由于略去重要解释变量引起的？一种方法是用残差 e_t 对那些可能影响因变量但又未列入模型的解释变量回归，并作显著性检验，从而确定该解释变量的重要性。如果是重要解释变量，应该列入模型。

只有当以上两种引起自相关的原因都消除后，才能认为误差项 u_t "真正"存在自相关。在这种情况下，解决办法是变换原回归模型，使变换后的随机误差项消除自相关，进而利用普通最小二乘法估计回归参数。这种变换方法称作广义最小二乘法。

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和差分法。**广义最小二乘法** 广义最小二乘法，顾名思义，是最具有普遍意义的最小二乘法，普通最小二乘法和加权最小二乘法是它的特例。

$$\hat{B} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \tag{3-18}$$

若 $\Omega = \sigma^2 I$ 则是普通最小二乘法；若 Ω除主对角线外元素全为零，则为加权最小二乘法。

经验方法，即并不对原模型进行异方差性检验和序列相关性检验，而是直接选择广义最小二乘法。如果确实存在异方差性和序列相关性，则被有效地消除了；如果不存在，则广义最小二乘法等价于普通最小二乘法。**广义差分法** 广义差分法可以克服所有类型的序列相关带来的问题，一阶差分法是它的一个特例。如果原模型存在

$$\mu_i = \rho_1\mu_{i-1} + \rho_2\mu_{i-2} + \cdots + \rho_i\mu_{i-l} + \epsilon_i \tag{3-19}$$

模型可变换为：

$$\begin{aligned}
&y_i - \rho_1 y_{i-1} - \cdots - \rho_l y_{i-l} \\
&= \beta_0(1-\rho_1-\cdots-\rho_l) + \beta_1(x_{1i} - \rho_1 x_{1i-1} - \cdots - \rho_l x_{1i-l} \\
&+\cdots \\
&+\beta_k(x_{ki} - \rho_1 x_{ki-1} - \cdots - \rho_l x_{ki-l} + \epsilon_i \\
&i = 1+l, 2+l, \cdots, n
\end{aligned} \tag{3-20}$$

为广义差分模型，该模型不存在序列相关问题。采用普通最小二乘法估计该模型得到的参数估计量，即为原模型参数的无偏的、有效的估计量。

为避免差分变换时的信息损失，$K.R.Kadiyala$ 建议在应用差分变换

$$Y_t^* = \psi(L)Y_t \quad X_{it}^* = \psi(L)X_{it} \quad i = 1, 2, \cdots, k \tag{3-21}$$

的同时，对第一个观察进行 $Prais - Winstern$ **变换**：

$$Y_1^* = Y_1\sqrt{1-\rho^2} \quad X_{j1}^* = X_{j1}\sqrt{1-\rho^2} \quad j = 1, 2, \cdots, k \tag{3-22}$$

这样的做的目的是为了使其对应的扰动项能同方差：

$$Var(\varepsilon_1^*) = Var(\sqrt{1-\rho^2}\varepsilon_1) = (1-\rho^2)Var(\varepsilon_1) = \sigma^2 \tag{3-23}$$

广义差分法

$$\begin{aligned}
Y^* &= DY \\
X^* &= DX \\
\varepsilon^* &= D\varepsilon
\end{aligned} \tag{3-24}$$

其中变换矩阵：

$$D=\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & & \\ -\rho & 1 & 0 & & \\ 0 & -\rho & & & \\ & & & 1 & 0 \\ 0 & 0 & & -\rho & 1 \end{pmatrix}$$

变换后的新模型满足 OLS 法古典假定。 **随机误差项自相关系数 ρ 的估计** 应用广义差分法，必须已知不同样本点之间随机误差项的相关系数 $\rho_1,\cdots,\rho_l$。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。于是发展了许多估计方法，诸如迭代法、杜宾两步法等。其基本思路是采用普通最小二乘法估计原模型，得到随机误差项的“近似估计值”，然后利用该“近似估计值”求得随机误差项相关系数的估计量。不同的方法旨在力图使得这些估计量更加逼近实际。**利用 DW 统计量**

$$\hat{\rho}\approx 1-\frac{1}{2}DW \tag{3-25}$$

在小样本情况下，该法是不准确的。可考虑**Theil-Nagar 估计**:

$$\hat{\rho}=\frac{T^2(1-DW/2)+(k+1)^2}{T^2-(k+1)^2} \tag{3-26}$$

Cochrane-Oreutt 迭代法 iterative procedure

(1) 据原始数据，用 OLS 法对原模型（线性回归模型）进行参数估计，得拟合值和残差；

(2) 以残差根据 $e_t=\rho e_{t-1}+v_t$ 用 OLS 法进行参数估计得$\hat{\rho}=\sum\limits_{t=2}^{T}e_te_{t-1}/\sum\limits_{t=2}^{T}e_{t-1}^2$。若 $\hat{\rho}$ 满足所需要的精度，则停止；否则，转入下一步；

(3) 进行广义差分变换：$Y_t^*=Y_t-\hat{\rho}Y_{t-1}$,$X_{it}^*=X_{it}-\hat{\rho}X_{i,t-1}$,$\varepsilon_t^*=\varepsilon_t-\hat{\varepsilon}_{t-1}$,$i=1,2,\cdots,k$,$t=2,3,\cdots,T$。求得变换后的模型的 OLS 残差，重复第二步；

杜宾两步法

(1) 对原始模型进行广义差分变换应用普通最小二乘法求得系数估计；

(2) 用系数估计值代入变换后的模型进行普通最小二乘估计。

虚假序列相关问题 上述检验和克服序列相关问题的方法，是针对真实序列相关问题发展起来的。如果模型的序列相关性是一种虚假序列相关，那么首先必须排除虚假性。所谓虚假序列相关问题，是指模型的序列相关性是由于省略了显著的解释变量而引致的。例如，在生产函数模型中，如果省略了资本这个重要的解释变量，资本对产出量的影响被归入随机误差项。由于资本在时间序列上的连续性，以及对产出量影响的连续性，必然导致随机误差项的序列相关。在这种情况下，首先要找出引起虚假序列相关性的原因，把显著的变量引入到解释变量之中。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。

3.3 多重共线性

"多重共线性"一词由 $R.Frisch$1934 年提出，它原指模型的解释变量间存在线性关系。现在所指的多重共线性，除了完全相关外，特指一种随机的线性相关，是样本的质量。古典线性模型要求回归变量之间无多重共线性，这等价于要求矩阵 X 列满秩。在这个假定下，设计矩阵 $X'X$ 可逆，可唯一求出 OLS 估计值 $\hat{\alpha}=(X'X)^{-1}(X'Y)$。此时我们称回归系数 α 是可识别的。**多重共线性的原因**

(1) 经济变量间的内在关系。由于经济变量之间的内在相关性而导致，在这个意义上，它是一个不可避免的问题；

(2) 经济变量在时间上有同方向变动的趋势，受同一决定因素的影响。如经济增长时期，收入、消费、投资、价格、就业都上升。

(3) 分布滞后模型的广泛运用。一个经济变量前后期之值可能高度线性相关。

(4) 所使用的数据收集及计算方法。

(5) 模型设定偏误。如自变量的多项式项易导致这一问题。

(6) 过分确定模型。如解释变量个数多于观测数。

后果 多重共线性尽管不会改变普通最小二乘估计的无偏性，但仍将对模型的估计、检验与预测带来不良后果：

(1) 具有较大的方差与协方差，难以得到精确的估计；

(2) 参数估计不稳健，对异常值敏感；

(3) 参数估计值标准误增大，从而使 t 检验得出误导性结果；

(4) 产生有偏的预测置信区间。

多重共线性的检验：相关系数和可决系数法 样本决定系数检验法 如果拟合优度很高，F 值显著，但 t 值不显著，且参数估计量方差较大，可能存在多重共线性。其严重程度可应用 $Theil$ 所提出的多重共线性效应系数来度量：

$$R^2 - \sum_{h=1}^{p}(R^2 - R_h^2) \tag{3-27}$$

其中 R_h^2 是指因变量在去掉 X_h 后对其他 $p-1$ 个解释变量进行回归的 R^2 值。若该系数接近于零，则可认为多重共线性不存在。若接近于 1 时，意味着存在多重共线性。**样本相关系数检验法** 为综合地探测所有解释变量间的多重共线性，可以考虑相关矩阵 R 行列式的假设：

$$H_0 : det(R) = 1 \leftrightarrow H_1 : det(R) \neq 1 \tag{3-28}$$

的显著性检验。其检验统计量及其分布为

$$FG = -(T - 1 - \frac{1}{6}(2p+5))\log(det(R)) \sim \chi^2_{p(p-1)/2} \tag{3-29}$$

若拒绝原假设，则认为存在多重共线性。**辅助回归检验法**

$$F_i = \frac{R_i^2/(p-1)}{(1-R_i^2)/(T-p)} \tag{3-30}$$

R_i^2 是第 i 个解释变量 X_i 关于其余解释变量的回归的决定系数，即复相关系数。$Klein$ **经验准则**(1962)：如果某解释变量间的偏相关系数高于所有解释变量间的复相关系数，即可决系数，则此两个变量间的多重共线性是有害的。**偏相关系数检验法**

$$\begin{aligned} H_0 : \rho_{ij} = 0 &\leftrightarrow H_1 : \rho_{ij} \neq 0 \\ t_i = \sqrt{T-p-1}&\frac{\gamma_{ij}}{\sqrt{1-\gamma_{ij}^2}} \end{aligned} \tag{3-31}$$

综合样本相关系数法、辅助回归法和偏相关系数检验法的就是**Farrar–Glauber 检验**。

源代码 3.13 Farrar-Glauber 检验

```
TITLE 'Calculate␣Correlation␣Matrix␣R'; PROC CORR DATA=LI.PARTIAL
OUTP=COR NOSIMPLE;
      VAR COM GOV INV;
RUN; PROC PRINT; RUN; TITLE 'Calculate␣the␣Determinant␣of
Correlation␣Matrix␣R␣Showed␣on␣LOG'; PROC IML;
      R={1 .99 1,.99 1 1,1  1  1};
      D=DET(R);
      RUN;
      PRINT R D;
      QUIT;

TITLE 'Calculate␣the␣FG␣Statistic␣and␣its␣P␣value'; DATA TEMP;
      INPUT N K D;
      FG=-(N-1-1/6*(2*K+5))*LOG(D);
      DF=K*(K-1)/2;
      P=1-PROBCHI(FG,DF);
      DATALINES;
      20 3  -0.0001
      ;
PROC PRINT DATA=TEMP; RUN;

TITLE 'Calculate␣Multiple␣Correlation␣Coefficients␣of␣Independents
Variable'; PROC REG DATA=LI.PARTIAL;
      C:MODEL COM=GOV INV;
      G:MODEL GOV=COM INV;
      I:MODEL INV=COM GOV;
RUN;

TITLE 'Partial␣Correlation␣Test'; PROC CORR DATA=LI.PARTIAL
NOSIMPLE;
      VAR INV COM;
      PARTIAL GOV;
RUN;
```

多重共线性检验：特征值法 这类方法所依据的是矩阵行列式与矩阵特征值的关系：

$$det(X'X) = \left|X'X\right| = \lambda_1\lambda_2\cdots\lambda_p \tag{3-32}$$

较小的行列式意味着存在一个或几个较小的特征值，而较小的特征值表明矩阵 X 各列变量之间可能存在某种很近似的线性相关关系。**条件数检验法** 造成多重共线性的原因，主要在于模型的信息矩阵 $X'X$ 的可逆性很差。这种可逆性的程度可由与该矩阵的特征值 λ_i 有关的条件指数刻画。条件指数 $conditionindexCI_i$:

$$CI_i = \sqrt{\frac{max\lambda_j}{\lambda_i}} \tag{3-33}$$

条件数 $conditionnumber$

$$k = maxCI_i \tag{3-34}$$

它刻画了特征值散布程度，较大的条件数意味着设计矩阵有较强的共线性。经验法则：若条件数大于 30 则多重共线性问题严重；若小于 10 则轻微。

在对全部 p 个解释变量进行主成分分析时，每个标准化后的解释变量相应估计值的方差（数值为1）被分解到 p 个主成分变量上去，每个主成分变量分得的方差称为该解释变量的**方差分量**（variace proportion）:对大的条件数，即 k 值，同时有 2 个以上变量的方差分量超过 50%，则意味着这些变量间有一定程度的相关。

源代码 3.14 SAS:条件数检验

```
PROC REG DATA=LI.PARTIAL;
      MODEL GDP=INV COM GOV /COLLIN COLLINOINT;
RUN;
```

方差膨胀因子检验法 基于复相关系数 R_i^2 可以定义各个解释变量 X_i 的容忍度($tolerance$):

$$\text{容忍度}TOL_i = 1 - R_i^2 \tag{3-35}$$

该值越小，意味着变量 X_i 不由其余解释变量说明的部分相对越小，即 X_i 与其余解释变量关系越密切。

方差膨胀因子($VIFvarianceinflationfactor$)：

$$VIF = \frac{1}{1 - R_i^2} \tag{3-36}$$

若最大方差膨胀因子大于等于 10 则认为多重共线问题严重；若小于 5，则认为轻微。

源代码 3.15 SAS:方差膨胀因子

```
PROC REG DATA=LI.PARTIAL;
      MODEL GDP=INV COM GOV /TOL VIF;
RUN;
```

源代码 3.16 SAS:多重共线性检验

```
data ex01;
      set li.mullinear;
run; proc corr data=ex01;
      var x1-x3;
run;
```

源代码 3.17 SAS:theil test

```
proc reg data=ex01;
      equation: model y=x1 x2 x3;
      equation3:model y=x1 x2;
      equation2:model y=x1 x3;
      equation1:model y=x2 x3;
run; data theil;
      rsq=0.9919;r1s=0.9913;r2s=0.9473;r3s=0.9828;
      theil=rsq-(3*rsq-(r1s+r2s+r3s));
      put theil=;
run;
```

源代码 3.18 SAS:辅助回归检验

```
proc reg data=ex01;
      equation3:model x3=x1 x2;
      equation2:model x2=x1 x3;
      equation1:model x1=x2 x3;
run;
```

源代码 3.19 SAS:FG test

```
proc corr data=ex01 outp=corr nosimple;
      var x1-x3;
run; proc print data=corr; run;

proc iml;
      R={1.000 0.026 0.997,0.026 1 0.036,0.9152 0.6306 1};
      d=det(R);
      print d;
run; data fg;
      n=11;p=3;d=0.081371;
      fg=-(n-1-1/6*(2*p+5))*log(d);
      df=p(p-1)/2;
      p=1-probchi(fg,df);
      put fg= p=;
run;

proc reg data=ex01 outest=ex012 graphics outvif;
      model y=x1-x3/ridge=0.0 to 0.1 by 0.01;
      plot/ridgeplot;
run; proc print data=ex012; run;

proc reg data=ex01 outest=ex103;
      model y=x1-x3/pcomit=1,2 outvif;
run; proc print data=ex103; run;
```

Frisch 综合分析法或逐步分析估计法

(1) 取 Y 关于每个解释变量的简单回归并从中挑选一个最优回归；

(2) 以这个最优回归为基础，逐一引入其他解释变量，重新再作回归；

(3) 若回归结果能提高拟合优度，且每个参数显著，符号正确，则该变量予以保留；反之，则舍弃。

(4) 若提高了拟合优度，但其它参数的符号和数值有明显变化，则认为存在多重共线性。

主成分法 构造某些人为的正交变量，如确定解释变量的正交多项式，或来自某些解释变量的线性组合，从而把多重共线性的 X 变换为正交变数。**在模型中引入附加方程** 研究者可依据专业知识发现那些形成多重共线性的变量间具有经济意义的关系，并明确列出这些关系式。这样与原方程一起便形成一个联立方程模型。若这一模型可识别，则可对联立方程进行估计。进一步，只要模型是恰当识别，则用简约式就可以回避原方程中的多重共线性问题。若为过度识别，则可以用某些参数的额外信息，从简约式参数中求得保留参数的唯一估计值。

3.4 非正态误差问题

来源

(1) 经济变量的本质属性使观察数据不服从正态分布，如大多数生产函数有非正态误差。

(2) 模型设定不正确导致设定模型的误差反映在非正态列差中；

(3) 观察样本容量不足或异常值使观察数据不能反映模型的正态误差；

(4) 由于进行数据变换，使原来的正态误差变为非正态误差。

这导致二乘法估计量不有效。**探测** 函数型异方差的控测方法都需要考虑残差平方的辅助回归 $auxiliaryregression$，故又称辅助回归检验法。它可分为辅助回归系数显著性检验，如 $Park, Glejser, BPG$ 检验：零假设是辅助回归系数全为零，备择假设是至少有一回归系数非零。另一是探测模型设定误差的 $White$ 检验（$White, 1980$），零假设是同方差且模型设定正确，备择假设是异方差且模型设定错误。

怀特检验是一种似然比检验，不需要对观测值排序也不依赖于正态性假定，但它要求大样本。$Q-Q$ **图** 一种检验样本观察值 $x_1, x_2, \cdots, x_n$ 是否来自正态总体的图示法，又称直线检验法。其做法是：将样本重排得 $x_{(1)} \leq x_{(2)}, \cdots, x_{(n)}$，若样本来自正态总体 $N(\mu, \sigma^2)$，则总体分布函数

$$F(x) = \Phi(\frac{x-\mu}{\sigma}) \approx F_n(x)$$

$F_n(x)$ 是经验分布函数。因此

$$\frac{x_{(i)}-\mu}{\sigma} = \Phi^{-1}(F_n(x_{(i)})$$

以”连续性修正“ $(i-0.5)/n$ 代替经验分布函数 $F_n(x_{(i)} = i/n$，记 $q_{(i)} = \Phi^{-1}((i-0.5)/n)$，$q_{(i)}$ 为标准正态分布的 $(i-0.5)/n$ 分位数，则应有 $x_{(i)} \approx \sigma q_{(i)} + \mu$。若样本来自正态总体，则在 Oqx 平面上，点 $(q_{(i)}, x_{(i)})$ 应近似在一条直线上。

检验方法 3.1 (W **正态性**) $Shapiro, S.S. \& Wilk M.B.(1965)$ 提出。**解析法**

统计量 3.3 (W) $W = \frac{(\sum_{i=1}^{[T/2]} \alpha_{T-i+1}(\hat{e}_{T-i+1} - \hat{e}_i))^2}{\sum_{i=1}^{T} (\hat{e}_i - \bar{e})^2}$

检验方法 3.2 (JB **正态性**) JB 是拉氏乘子检验。$Jarque \& Bera(1987)$

统计量 3.4 (JB) $JB = T(\frac{skew^2}{6} + \frac{kurt^2}{24})$ 它具有渐近卡方分布，自由度为 2。

注：NORMAL 选项输出正态性检验的统计量及 P 值，当样本容量小于 2000 时，输出 W 统计量；若大于时，输出 $Kolmogorov$ 统计量。

源代码 3.20 SAS：正态性检验

```
PROC REG DATA=LI.MACROECO;
     MODEL Y=I G C;
     OUTPUT OUT=ABC RESIDUAL=R;
RUN;

PROC UNIVARIATE DATA=ABC NORMAL PLOT;
     VAR R;
     OUTPUT OUT=TEMP SKEWNESS=S N=N KURTOSIS=K;
RUN;

DATA TEMP;
     SET TEMP;
     JB=N*(S*S/6+K*K/24);
     P=1-PROBCHI(JB,2);
PROC PRINT DATA=TEMP; RUN;
```

补救 remedial measures 一是进行模型变换。

定义 3.1 (Box-Cox **变换**) 线性回归模型需要满足四个基本假定：线性性($E(y)$ 是 X_i 的线性函数)；误差项同方差；误差项无序列相关；误差项服从正态分布。如果数据不满足这几个条件，则可以进行 $Box-Cox$ 变换以使数据满足这四个条件：

$$y^* = \frac{y^\lambda - 1}{\lambda}$$

或

$$y^* = \ln y$$

例如，设随机扰动项的方差未知，但其形式已知，为 $var(\varepsilon_t) = \sigma_t^2 = k^2 f(x_t)$，其中 k 为已知常数，此时，可考虑模型变换

$$\frac{Y_t}{k\sqrt{f(x_t)}} = \frac{\alpha}{k\sqrt{f(x_t)}} + \beta k\sqrt{f(x_t)} + \frac{\varepsilon_t}{k\sqrt{f(x_t)}} \tag{3-37}$$

对此，可应用 OLS，即可消除扰动项中的异方差。二是重新设定模型。

三是采用稳健估计方法。

四是采用加权最小二乘法。当随机扰动项的方差已知时，以其倒数为权重对残差平方加。

例 3.1 (**异方差探测**) 美国 1988 年$R\&D$ 支出，给出 18 个行业的研究（research）与开发（*development*）支出及有关的销售、利润资料，要求进行线性回归并考虑异方差检验。

源代码 3.21 SAS:异方差探测

```
PROC REG DATA=LI.RD88;
      MODEL RD=S /SSE;
      OUTPUT OUT=ABC R=R;
RUN; \* ESS=7612927*\ DATA TEMP;
      MERGE LI.RD88 ABC;
      LN_S=LOG(S);
      LN_R2=LOG(R*R);
      ABSR=ABS(R);
      SQ_S=SQRT(S);
      INV_S=1/S;
      INV_SQS=1/SQ_S;
      R2=R*R;
      RR2=R2-7612927;
      OUTPUT;
RUN; PROC REG DATA=TEMP;
      Park: MODEL LN_R2=LN_S;
      Glejeser1:MODEL ABSR=S;
      Glejeser2:MODEL ABSR=SQ_S;
      Glejeser3:MODEL ABSR=INV_S;
      Glejeser4:MODEL ABSR=INV_SQS;
      BPG:MODEL RR2=S;
RUN; \*White test*\ PROC REG DATA=LI.RD88;
      OLS2:MODEL RD=S P;
      OUTPUT OUT=ABC2 R=R;
RUN; DATA TEMP1;
      MERGE LI.RD88 ABC2;
      RR=R*R;
      OUTPUT;
RUN; PROC GLM DATA=TEMP1;
      White:MODEL RR=S P S*S S*P P*P;
RUN;
```

例 3.2 根据上例的结果，分别用模型变换法、WLS 法、GLS 法对存在异方差的模型进行修正。

模型变换

$$\frac{RD_i}{\sqrt{S_i}} = \alpha_0 \frac{1}{\sqrt{S_i}} + \alpha_1 \sqrt{S_i} + \frac{\varepsilon_i}{\sqrt{S_i}}$$

源代码 3.22 SAS：异方差修正

```
PROC REG DATA=LI.RD88;
      OLS:MODEL RD=S;
      OUTPUT OUT=ABC R=R;
RUN; DATA TEMP;
      MERGE LI.RD88 ABC;

      /*model transform*/
      SQ_S=SQRT(S);
      INV_SQS=1/SQ_S;
      RD_SQS=RD/SQ_S;

      /*weight  coefficient */
      R2=R*R;
      W=1/R2;

      /*general method*/
      RD_INVR=RD/ABS(R);
      INV_R=1/ABS(R);
      S_INVR=S/ABS(R);
      OUTPUT;
RUN; PROC REG DATA=TEMP;
      T_RD:MODEL RD_SQS=INV_SQS SQ_S/NOITNT;***NOTE;
      GLS:MODEL RD_INVR=INV_R S_INVR/NOINT;
RUN; PROC REG DATA=TEMP;
      WLS:MODEL RD=S;
          WEIGHT W;
RUN;
```

3.5 随机解释变量问题

单方程线性计量经济学模型假设解释变量是确定性变量，并且与随机误差项不相关。违背这一基本假设的问题被称为随机解释变量问题。

3.6 模型设定误差及模型确认检验

设定误差的形式 设定误差($specificationerorrs$)是指因错误设定模型所导致的误差。它主要有如下几种形式：遗漏重要解释变量（过低拟合模型$underfittingamodel$）回归模型中缺少某个重要的变量能使估计量成为有偏和非一致的；过份拟合模型（$overfittingamodel$）不相关变量的存在会导致有效性的降低；采用错误的函数形式；测量误差。

采用错误的函数形式多来自于不加选择地进行了对数变换，或应进行变换而没有变换。或由于将非线性模型近似地线性化导致。**后果** 遗漏重要解释变量则：

(1) 参数估计量有偏且非一致；

(2) 扰动项方差不能被正确估计；

(3) 置信区间及假设检验对原参数估计的显著性可能给出误导性结论。

而增加多余变量却只会增加回归系数估计的方差，降低统计推断的精确度。**变量是否应当从线性回归模型中删除 F 检验** 若干系数的联合 F 检验来确认这些系数是否显著地为零，从而得出判断。由有限制条件模型的残差平方和是否显著地大于无限制条件模型的残差平方和确定某几个变量是否应当保留。**T 检验** 检验某个系数是否为零，从而确定某个变量是否应当保留在模型中。**似然比检验法** 它既不用最小二乘估计也不依赖于误差项正态分布假设。似然比检验可以用于更一般的情形。无论对模型的参数有什么限制条件，下面的结论都是成立的：

$$-2[L(\text{有限制条件}) - L(\text{无限制条件})] \sim \chi_r^2$$

是否存在测量误差 对于一元线性回归模型

$$y_i = \beta x_i + \varepsilon_i \tag{3-38}$$

如果 X 有测量误差:$x_i = x_i^* - v_i$，则实际的最小二乘回归应当是

$$y_i = \beta x_i^* + \varepsilon_i^* \tag{3-39}$$

其中 $\varepsilon_i^* = \varepsilon_i - \beta v_i$。

对此问题，可利用与 x^* 相关但与 ε 和 v 不相关的工具变量 z，可以得到 β 的一致估计量。假设 z 和 x^* 之间的关系是

$$x_i^* = \gamma z_i + w_i \tag{3-40}$$

采用最小二乘估计，这个关系是

$$\hat{x}_i^* = \hat{\gamma} z_i \tag{3-41}$$

或

$$x_i^* = \hat{x}_i^* + \hat{w}_i \tag{3-42}$$

将(3-42)代入(3-42)得到

$$y_i = \beta \hat{x}_i^* + \beta \hat{w}_i + \varepsilon_i^* \tag{3-43}$$

以 δ 代表上式中变量 $\hat{w}_i$ 的系数，将 $\hat{x}_i^* = x_i^* - \hat{w}_i$ 代入，得

$$y_i = \beta x_i^* + (\delta - \beta)\hat{w}_i + \varepsilon_i^* \tag{3-44}$$

若没有测量误差，则 $\delta = \beta$；反之，这个系数将不为零。

我们可以通过一个简单的两步法对测量误差进行检验。首先，将 x^* 对 z 回归得到残差 $\hat{w}$。然后将 y 对 x^* 和 $\hat{w}$ 进行回归，对 $\hat{w}$ 变量的系数进行 t 检验。如果在多元回归模型中，我们怀疑多个变量存在测量误差，可以采用等价的 F 检验。**豪斯曼检验 Hausman** 它是豪斯曼确认检验的一个特例。

检验方法 3.3 (**豪斯曼确认检验**) $Hausman"sspecificationtest$，$orm-statistic$，它用于检验估计量的偏差或非一致性。对于两个估计量 β_0 和 β_1，零假设是：两个估计量都一致但只有 $\hat{\beta}_0$ 渐近有效；备择假设是：$\hat{\beta}_1$ 是一致的。m 统计量是：

$$m = \hat{q}'(\hat{V}_1 - \hat{V}_0)^{-}\hat{q} \tag{3-45}$$

$\hat{V}_1$ 和 $\hat{V}_0$ 代表 $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 渐近协方差矩阵的一致估计，另，

$$q = \hat{\beta}_1 - \hat{\beta}_0$$

m 统计量服从卡方分布，自由度为 $(\hat{V}_1 - \hat{V}_0)$ 的秩 k。

豪斯曼检验可用于确认是否有必要使用工具变量法而不是最小二乘法；用于比较 $2SLS$、$3SLS$和 SUR。

例 3.3 [检验公共开支模型中的测量误差] 美国各洲和地方政府开支 EXP 随洲和地方的不同差异很大。

用普通最小二乘法回归。但变量 AID，联邦政府的拨款量，可能存在测量误差，用豪斯曼检验来检验测量误差是否存在。用小学和中学在校学生人数（PS）作为工具变量。检验分两步：第一步，将 AID 对 PS 回归，计算残差变量 $\hat{w}_i$；第二步，将残差变量加入原回归模型，从而修正测量误差。看残差变量系数的 t 统计值的显著性来判断是否存在测量误差。

源代码 3.23 SAS：检验测量误差

```
PROC IMPORT OUT=LI.HAUSEMAN
            DATAFILE="D:\MYDOCUMENT\DATASET\PINDYCK\EX73.XLS"
                 REPLACE;
            GETNAMES YES;
RUN; PROC REG DATA=LI.HAUSEMAN ;
     MODEL AID=PS ;
     OUTPUT OUT=W R=WHAT;
RUN; DATA W;
     SET W;
     KEEP WHAT;
RUN; DATA TEMP;
     MERGE LI.HAUSEMAN W;
RUN; PROC REG DATA=TEMP;
     MODEL EXP=AID INC POP WHAT;
RUN;
```

例子 *Davidson&Mackinnon*(1993)：对于线性回归模型，不管它的误差是不是服从正态分布，都不需要过问 LM, LR, W。因为我们不能从这些统计量中得到任何不能为 F 统计量所包含的信息。

美国鸡肉需求分析

例 3.4　美国鸡肉的需求（来源：*Gujarati, BasicEconometrics*）。考察了美国 1960 到 1982 年的鸡肉需求量。数据中 y 是人均鸡肉消费量（磅），$x1$ 是人均可支配收入（美元），$x2, x3, x4$ 分别是鸡肉，猪肉，牛排的价格（美元/磅）。为了研究鸡肉的需求量，考察如下需求函数：$lnY(t) = b0 + b1lnX1(t) + b2lnX2(t) + b3lnX3(t) + b4lnX4(t) + u(t)$ 根据经济学理论，系数分别为需求弹性。并且 $b1 > 0, b2 < 0$。还有人认为鸡肉的需求和牛排，猪肉不相关。验证这个命题。

源代码 3.24　SAS:美国鸡肉的需求分析

```
data ex1;
     set li.snlrgn;
run; /*unrestricted regression*/; proc reg data=ex1;
         model ly=lx1 lx2 lx3 lx4 /cli dw vif;
         test lx3,lx4;
         var year;
         output out=ex2 r=resid p=pred l95=l95 u95=u95;
          title 'reg_ly_vs_lx';
run; proc print data=ex2;
      title 'print_ex2';
run;

proc gplot data=ex2;
     plot l95*year u95*year ly*year pred*year /overlay legend;

     symbol1 v=dot cv=black i=join ci=black ;
     symbol2 v=circle cv=red i=join ci=red;
     symbol3 v=triangle cv=green i=join ci=green;
     symbol4 v=hash cv=blue i=join ci=blue;
     run;

     plot resid*year;
     run;
```

英国失业率分析

例 3.5 给出了英国 1958 年 4 季度到 1972 年 2 季度的失业率 UN 与工作空缺率 V。从散点图发现，从 1966 年 4 季度开始，UN, V 的关系似乎向上平移了。这意味着同样的工作空缺率下的失业人数更多了。研究发现这种上移的原因是，1966 年 10 月，英国通过维护短期失业人员的法律，短期失业人员可以得到的社会保障提高了。这样，失业人员就不会太急于找到工作。失业率就增加。下面用数据验证这一说法并进行结构检验。

源代码 3.25 SAS:英国失业率分析

```
data unemploy;
        d=0;
        set  li .unemploy;
        if  year>1966.3 then d=1;
        dv=d*v;
run; proc print  data=unemploy; run; proc reg data=unemploy;
        model un=d v dv;
        test  d,dv;
        title  'demo2_dummy_variable';
run;
 /*structure  test*/;
data new1 new2;
        set  unemploy;
        if  year<1966.4 then output new1;
                            else  output new2;
run; proc reg  data=new1;
        model un=v;
        title '_new1';
run; proc reg  data=new2;
        model un=v;
        title  'new2';
run; proc reg  data=unemploy;
        model un=v;
        title  ' total ';
run; data tt;
        es1=0.23040;es2=0.61852;es=4.06233;
        f=47/2*(es−es1−es2)/(es1+es2);
        fcritical =finv (0.95,2,47) ;
        if  f> fcritical  then d=1;else d=0;
        put f=  fcritical = d=;
run;
```

3.7 格兰杰因果性检验

这个概念首先由 *Granger*1969 提出。*Sims*1972 也提出因果性定义。这两个定义是一致的：

如果由 y_t 和 x_t 滞后值所决定的 y_t 的条件分布与仅由 y_t 滞后值所决定的条件分布相同，即

$$\{y_t|y_{t-1},\cdots,x_{t-1},\cdots\}=\{y_t|y_{t-1},\cdots\} \tag{3-46}$$

则称 x_{t-1} 与 y_t 存在格兰杰非因果性。它的另一种表述是，若其它条件不变，加上 x_t 的滞后值后对 y_t 的预测精度不存在显著性变化，则称 x_{t-1} 对 y_t 存在格兰杰非因果性关系。

格兰杰非因果性检验是 VAR 模型的一个副产品，x_t 对 y_t 是否存在因果关系的检验可通过检验 VAR 模型以 y_t 为被解释变量的方程中是否可以把 x_t 的全部滞后变量剔除而完成。

$$y_t = \sum_{i=1}^{k} \alpha_i y_{t-k} + \sum_{i=1}^{k} \beta_i x_{t-i} + u_i \tag{3-47}$$

3.8 内生性 $Hausman$ 检验

$Hausman(1978)$ 首先提出关于变量内生性的检验用统计量。$Davison\&MacKinnon(1989, 1993)$ 又提出一种借助辅助回归进行 $Hausman$ 检验的方法。假定需要作如下回归：

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 z_t + u_t \tag{3-48}$$

其中 z_t 变量有可能是由 y_t 变量决定的内生变量。那么对上式的 OLS 估计量一定是有偏的和不一致的。

内生性 $Hausman$ 检验通过如下两次回归完成。第一个回归式是用待检验变量 z_t 对(3-48)中的全部外生变量和选定的工具变量（假定是） x_{t3}, x_{t4} 进行回归：

$$z_t = \alpha_0 + \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3 x_{t3} + \alpha_4 x_{t4} + v_t \tag{3-49}$$

用 OLS 法估计并提取残差序列 $\hat{v}_t$。第二步是把 $\hat{v}_t$ 作为附加变量加入到(3-48):

$$y_t = \gamma_0 + \gamma_1 x_{t1} + \gamma_2 x_{t2} + \gamma_3 z_t + \gamma_4 \hat{v}_t + u_t \tag{3-50}$$

用 OLS 法估计。如果 z_t 具有外生性，或 $\beta_0, \beta_1, beta_2, \beta_3$ 具有一致性，那么回归系数 γ_4 应该没有显著性。

例如，影响当期名义货币供应量的主要因素有：基础货币、物价水平、产出水平和上期名义货币供应量。按照货币主义观点，产出水平可能是受名义货币供应量决定的内生变量，因此需要进行内生性检验。可以考虑以失业率作为工具变量，该变量与产出水平高度相关，而与当期名义货币供应量不直接相关。

3.9 检验白噪声的 Q 统计量

序列 y_t 的估计的自相关函数（相关图）是：

$$r_k = \frac{\frac{1}{T-k} \sum\limits_{t=k+1}^{T} (y_t - \bar{y}_t)(y_{t-k} - \bar{y}_t)}{\frac{1}{T} \sum\limits_{t=1}^{T} (y_t - \bar{y}_t)^2}$$

其中 r_k 表示 y_t 与 y_{t-k} 估计的自相关函数，是对自相关系数 ρ_k 的估计。

模型残差序列是否为白噪声的检验是用 $Box-Pierce1970$ 提出的 Q 统计量完成的。Q 检验的零假设是：

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$$

Q 统计量定义为：

统计量 3.5 (Q 统计量)

$$Q = T \sum_{k=1}^{K} r_k^2$$

随着 $T \to \infty$，Q 渐近服从 $\chi^2(k-p-q)$ 分布。$Ljung\&Box$ 认为由上面定义的统计量相应值偏小，提出修正的 Q 统计量：

$$Q = T(T+2)\sum_{k=1}^{K}\frac{r_k^2}{T-k}$$

3.10 回归系数的稳定性检验

在样本 T 基础上求出回归模型系数的估计值后，再增加 n 个观测值从而考查原参数估计值是否稳定时，可采用 $Chow$ 检验法。

首先对同一形式模型分别用样本 T 和 $T+n$ 进行回归：

原假设与备择假设：

$$H_0: \beta_j = \alpha_j$$
$$H_1: \beta_j\text{与}\alpha_j\text{不全对应相等}$$

所用的统计量定义为：

$$F = \frac{(SSE_2 - SSE_1)/n)}{SSE_1/(T-k)} \tag{3-51}$$

3.11 方差不等结构变化的检验

运用邹检验时的一个重要假设是，两个或所有回归中的方差是相同的。若约束模型的方差是异方差的，则古典结果不再适用。若样本适当大，我们有一个不论干扰项方差是否相同都有效的一个检验。假定 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是基于独立样本的一个参数的两个正态分布估计量，且它们的方差矩阵为 V_1 和 V_2。则 $\hat{\theta}_1 - \hat{\theta}_2$ 的均值为零，方差为 $V_1 + V_2$。这样，沃尔德统计量：

$$W = (\hat{\theta}_1 - \hat{\theta}_2)`(V_1 + V_2)^{-1}(\hat{\theta}_1 - \hat{\theta}_2)$$

服从自由度为 K 的卡方分布。

3.12 模型稳定性检验

结构变化检验基本上基于模型对用于估计它的观测值之外的数据的预测能力。

布朗、德宾和伊万斯(1975)基于递归残差的模型稳定性检验可处理结构变化点未知的问题，原假设是系数向量各项相等，备择假设是它（或扰动方差）并不相等。它与邹检验相比，势相当有限。这一技术适用于时间序列数据。

第 t 个递归残差是：

$$e_t = y_t - X`_t b_{t-1}$$

这个残差的预测方差是：

$$Var(e_t) = \sigma^2[1 + x`_t(X`_{t-1}X)^{-1}x`_t$$

第 r 个标准化残差是：

$$w_r = \frac{e_r}{\sqrt{1 + x`_r(X`_{r-1}X_{r-1})^{-1}x_r}}$$

假设整个样本期内系数保持不变，则

$$w_r \sim N[0, \sigma^2]$$

模型	样本容量	残差平方和	相应自由度	回归系数
1	T	SSE_1	$T-k$	β_j
2	$T+n$	SSE_2	$T+n-k$	α_j

且对所有 $r \neq s$，w_r 和 w_s 独立。w_r 的分布随同时间而变对模型的假设不利。

基于 w_r 有两种检验，$CUSUM$ 检验基于残差累积和:

$$W_t = \sum_{i=K+1}^{t} \frac{w_r}{\hat{\sigma}}$$

其中

$$\hat{\sigma}^2 = \frac{1}{T-K-1} \sum_{i=K+1}^{T} (w_r - \bar{w})^2$$

和

$$\bar{w} = \frac{1}{T-K} \sum_{r=K=1}^{T} w_r$$

在原假设下，W_t 的均值为零，方差近似为累积残差的数目（每一项的方差为 1）。

另一个可选检验基于递归残并的平方，$CUSUMSQ$ 检验：

$$S_t = \frac{\sum_{r=K+1}^{r=t} w_r^2}{\sum_{r=K+1}^{T} w_r^2}$$

两项中的每一个都是近似自由度为 1 的卡方变量的和。所以

$$E[S_t] = \frac{t-K}{T-K}$$

由哈维和科利尔(1977)的提出的检验直接基于均值，在模型稳定性假设下，$\bar{w}$ 服从均值为零方差为 $\sigma^2/(T-K)$ 的正态分布。从而可以构造一个 t 检验:

$$t[n-k-1] = \frac{(\sqrt{T-K})\bar{w}}{s}$$

其中

$$s^2 = \frac{1}{T-K-1} \sum_{r=K+1}^{r=T} (w_r - \bar{w})^2$$

3.13 非嵌套假设检验

我们将注意力放在两个竞争模型中：

$$H_0 : Y = X\beta + \varepsilon_0$$

和

$$H_1 : Y = Z\gamma + \varepsilon_1$$

由于 H_1 不能写作 H_0 的参数约束形式，所以系数约束假设检验不适用。一个方便方法是考虑人工嵌套模型。令 $\bar{X}$ 代表 X 中不属于 Z 的向量，$\bar{Z}$ 亦如此。W 代表在 X 和 Z 都存在的向量。则人工模型是：

$$y = \bar{X}\bar{\beta} + \bar{Z}\bar{\gamma} + W\delta + \varepsilon$$

在理论上，若用一个传统的 F 检验发现 $\bar{\gamma} = 0$，则 H_1 被拒绝；若发现 $\bar{\beta} = 0$，则 H_0 被拒绝。这种方法有两个问题：

(1) δ 是 γ 和 β 的混合体，其中之一部分是为零并未被 F 检验证实。因此这个检验并未真正区分 H_0 和 H_1，它只是区分了 H_1 和混合模型。

(2) 复合模型存在很多的回归量，在一个时间序列背景下，多重共线性的问题会比较突出

戴维森和麦金农(1981)提出的J **检验**本质上类似于嵌套策略，但克服了其问题。备择模型是：

$$Y=(1-\alpha)X\beta+\alpha Z\gamma+\varepsilon$$

对 $\alpha=0$ 的接受就是拒绝了 H_1。α 在这个模型中不能被估计。他们的方法是，首先对 H_1 用 OLS 估计；接下来用 Y 对 X 的 $Z\hat{\gamma}$ 进行 OLS 估计。对 H_1 的一个渐近有效的检验就是检验 $\alpha=0$ 的 t 检验 $t=\hat{\alpha}/se(\hat{\alpha})$。问题是，这种检验在检验 H_0 对 H_1 及其相反时，有四种可能性。这是有限样本的问题。

一组基于最大似然比的相关检验由考克斯(1961,1962)导出。检验 X 是正确的回归量而 Z 不是的考克斯统计量是：

$$c_{12}=\frac{n}{2}\ln\left[\frac{s_z^2}{s_x^2+(1/n)\,b'X'M_zbX}\right]=\frac{n}{2}\ln\frac{s_z^2}{s_{zx}^2}$$

其中

$$M_z=I-Z\left(Z'Z\right)^{-1}Z'\ b=\left(X'X\right)^{-1}X'Y$$

检验统计量是：

$$q=\frac{c_{12}}{\sqrt{Est.Var(c_{12})}}=\frac{c_{12}}{\sqrt{\left(s_x^2/s_{zx}^4\right)b'X'M_xM_zM_xbX}}$$
$$s_{zx}^2=s_x^2+(1/n)\,b'X'M_zbX$$

$Harvey$(1990) 将检验非嵌套假设的方法概括为二类:一是基于一些拟合准则在两个或两个以上的竞争模型中选择一个模型的判别法($discriminationapproach$);二是利用其他模型的信息来确认模型的辨认法($discerningapproach$)。

判别法如数值越大越好的 R^2，数值越小越好的 AIC,SBC。辨认法如非嵌套的 F 检验与 $Davison--Mackinnon$ 的 J 检验。

对于非嵌套模型：

$$\text{模型}C\qquad Y_i=\alpha_0+X_{1i}\alpha_1+u_i$$
$$\text{模型}D:\qquad Y_i=\beta_0+Z_{1i}\beta_1+v_i$$

进行 $Davidson--MacKinnon$ J 检验的步骤如下：

(1) 对模型 D，进行拟合得拟合值 $\hat{Y}_i^D$ 并考虑嵌套模型 C 的模型：

$$Y_i=\alpha_0+X_{1i}\alpha_1+\alpha_2\hat{Y}_i^D+u_i$$

(2) 对模型 C 实行同样操作：

$$Y_i=\beta_0+X_{1i}\beta_1+\beta_2\hat{Y}_i^C+v_i$$

(3) 分别对以上两个模型进行显著性检验，根据系数 α_2,β_2 的显著性来选择模型。

例 3.6　美国 1970 − −1991 年期间各年按 1987 的不变价格计算的人均个人消费支出 $PPCE$ 及人均个人可支配收入 $PDPI$ 的统计资料如数据集。考虑二个竞争模型：

$$\text{模型}A\qquad PPCE_t=\alpha_0+\alpha_1PDPI_t+\alpha_2PDPI_{t-1}+u_i$$
$$\text{模型}B:\qquad PPCE_t=\beta_0+\beta_1PDPI_t+\beta_2PPCE_{t-1}+v_i$$

这里，模型 A 是一个分布滞后模型，B 是一个自回归模型。应用 J 检验对模型作出选择。若模型都被拒绝，则检查模型中的多重共线性并做适当修正。

运用普通最小二乘法进行拟合，再估计分别嵌套模型 AB 的模型 c,D。

源代码 3.26

```
DATA EX;
    INPUT PPCE PDPI @@;
    DATALINES;
8842 9875 9022 10111 9425 10414 9752 11013 9602 10832 9711 10906
10121 11192 10425 11406 10744 11851 10876 12039 10746 12005 10770
12156 10782 12146 11179 12349 11617 13029 12015 13258 12336 13552
12568 13545 12903 13890 13029 14005 13044 14068 12824 13886 ; DATA
EX;
    RETAIN YEAR 1969;
    SET EX;
    YEAR+1;
    PDPI1=LAG(PDPI);
    PPCE1=LAG(PPCE);
    D_PDPI=DIF(PDPI);
RUN; TITLE 'OLS of Selected Models:A and B'; PROC REG DATA=EX;
    A:MODEL PPCE=PDPI PDPI1 /DW VIF;
      OUTPUT OUT=A P=AHAT;
    B:MODEL PPCE=PDPI PPCE1 /DW VIF;
      OUTPUT OUT=B P=BHAT;
RUN; DATA TEMP;
    MERGE A B;
TITLE 'J Test for Model A & B'; PROC REG DATA=TEMP;
    C:MODEL PPCE=PDPI PDPI1 BHAT /DW VIF;
    D:MODEL PPCE=PDPI PPCE1 AHAT /DW VIF;
RUN;
```

为修正模型中的多重共线性，在模型中引入差分项 $\Delta PDPI_t$。

3.14 Leamer 的EBA 分析法

先把解释变量分为二类，一是自由的（关键的）变量，二是可疑的（次要的）变量；然后基于关键变量的回归，逐步引入可疑变量（考虑各种可能的组合），取回归。这样，由关键变量系数估计值的变动范围便得到关键变量的极值界限($extremebound$)或全距。如这一值相当狭窄，则认为所得数据产生了相当稳健($sturdy$)的关于关键变量的信息；否则，关于关键变量系数的估计是脆弱（$fragile$）的。

3.15 异常值 outlier

设

$$\varepsilon_t \triangleq Y_t - \hat{Y}_t \qquad t = 1, 2, \cdots, T \tag{3-52}$$

$$SE \triangleq \left(\frac{1}{T-(p+1)} \sum_{i=1}^{T} e_i^2\right)^{1/2} \tag{3-53}$$

分别为第 t 个观察值 (X_t, Y_t) 的残差与模型标准差。一般地，若

$$\left|\frac{e_t}{SE}\right| > 2or3 \tag{3-54}$$

则称对应的观察值为异常值。

异常值的存在，将破坏模型中误差项为零均值的假定，引起多重共线性等。出现的原因主要有两个：一是经济过程中外来突发性因素；一是误差分布的非正态性。对异常值的补救措施一般是引入虚拟变量，有多少个异常值，就引进多少个虚拟变量。

3.16 β_j 的联合置信区间

全部 β_j 的联合置信区间接受:

$$F = \frac{1}{k}(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})/s^2 \sim F_{\alpha(k,T-k)} \tag{3-55}$$

$$(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) \leq s^2 k F_{\alpha(k,T-k)} \tag{3-56}$$

它是一个 k 维椭球。

3.17 预测的评价指标

- 预测误差，是对单点预测误差大小的测量

$$e_t = \hat{y}_t - y_t$$

- 相对误差 $PE(PercentageError)$，是单点预测相对误差大小的测量

$$PE = \frac{\hat{y}_t - y_t}{y_t}$$

- 误差均方根 $RMSERRORRootMeanSquaredError$ 是通过若不个预测值对预测效果进行综合评价

$$RMSError = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}$$

- 绝对误差平均 $MAE(MeanAbsoluteError$

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|\hat{y}_t - y_t|$$

- 相对误差绝对值平均 $MAPE$
- $Theil$ 系数：

$$\frac{\sqrt{SSE}}{\sqrt{Y_t} + \sqrt{\hat{Y}_t}}$$

3.18 协方差分析检验

在 SAS 中可用 $ANONVA$ 和 GLM 过程完成。

定义 3.2 (**方差分析**) 方差又称均方，它由离均差平方和除以自由度而得。方差分析时将离均差平方和分即总变异分为几个组成部分，其自由度也相应分解，故方差分析又称变异数分析，其目的是推断两组或多组资料的总体均数是否相同，检验两个或多个样本均数的差异是否有统计学上的显著性意义。

应用条件：

- 可加性；
- 可比性；
- 正态性；转换方法有平方根法、对数法、倒数变换、平方根反正弦变换。
- 方差齐性；可用 *Barttlet* 法进行组间方差齐性检验。

3.19 模型筛选准则

来自[23] 拟合优度

$$R^2 = 1 - \frac{\hat{\varepsilon}'\varepsilon}{(y-\bar{y})'(y-\bar{y})}$$

某些情况下拟合优度可能为负值：

- 回归方程没有包括常数项
- 系数有约束回归
- 估计方法是二阶段最小二乘法或 *ARCH*

为解决增加回归变量提高拟合优度问题，用调整的拟合优度：

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

假设检验对检验非嵌套模型并不特别适合。参数过多的不利之处在于：随着参数估计导致误差的增大，预测误差的方差也增大。即，小模型的样本区间外预测效果会优于大模型的。

有限预测误差准则 FPE 要求最小化 1 步均方预测误差 $MSPE$。以 $AR(p)$ 为例

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t$$

$MSPE$ 为

$$(1 + \frac{p}{T})\sigma^2$$

我们不知道真实的方差 σ^2，可用其无偏估计 $SSR/(T-p)$ 替代：

$$FPE = [1 + \frac{p}{T}][\frac{SSR}{T-p}]$$

选择 p 使 FPE 最小，采用取对数方法，并用 p/T 近似表示 $ln(1+p/T)$：

$$T\ln(SSR) + 2p$$

$$AIC = T\ln(SSR) + 2p$$
$$SBC = T\ln(SSR) + p\cdot(\ln T)$$

- SBC 与 AIC 相比，会选择一个更加简练的模型
- SBC 有更优的大样本特性，且为渐近一致
- AIC 倾向于选择参数更多的模型，小样下情况下，AIC 优于 SBC

3.20 一般的线性假设检验:沃尔德检验、似然比检验与拉格朗日检验

一般地，对某一特定系数的显著性检验就是检验回归模型的一组系数是否满足某一线性约束。极大似然估计给出了假设检验的三种不同方法：似然比检验、沃尔德检验和拉格朗日检验，统称**三种古典检验**。三种方法是渐近等价的，即当样本容量趋于无限大时，三种不同检验统计量趋于同一个随机变量（在原假设下，且是对接近原假设的 DGP）。如果等式约束个数为 r，这个随机变量服从 $\chi^2(r)$ 分布。**沃尔德检验** $Wald$ 检验只需要检验无约束模型估计，检验统计量是向量 $r(\hat{\theta}$ 和其协方差矩阵估计的逆形成的二次型。

采用德尔塔方法得出：

$$Var(r(\hat{\theta})) \overset{a}{=} R(\theta_0)Var(\hat{\theta})R^T(\theta_0) \tag{3-57}$$

$R(\theta)$ 代表元素为 $\partial r_i(\theta)/\partial\theta_j$

对系数参数 q 个线性约束的零假设为：

$$H_0 : R\beta = r \tag{3-58}$$

由于 $\hat{\beta}$ 服从正态分布，所以：

$$R\hat{\beta} \sim N[R\beta, \sigma^2 R(X'X)^{-1}R'] \tag{3-59}$$

因而，在零假设下，统计量

$$\frac{1}{\sigma^2}(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r) \sim \chi^2(q) \tag{3-60}$$

因此，检验的备择假设是 $H_1 : R\beta \neq r$，需要使用如下的统计量：

$$\xi_w = \frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)}{q\hat{s}^2} \sim F(q, n-k) \tag{3-61}$$

注意，若 $q = 1, r = 0$，R 为第 j 个元素为 1 其余元素为零的 $1 \times k$ 向量，则上式简化为

$$t = \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2(X'X)^{-1}_{jj}}}{\sqrt{\hat{s}^2/\sigma^2}} \tag{3-62}$$

因为，

$$F(1, n-k) = t^2(n-k) \tag{3-63}$$

注意：

$$(y - X\beta)'(y - X\beta) = \hat{u}'\hat{u} + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \tag{3-64}$$

用 $\hat{\beta}_c$ 代替 β 代入上式，则(3-61)的分子等于有约束回归和无约束回归 $Y = X\beta + U$ 的残差平方和之差，(3-61)可写为

$$\frac{\hat{u}_c'\hat{u}_c - \hat{u}'\hat{u}}{q\hat{s}^2} \sim F(q, n-k) \tag{3-65}$$

一般地，检验统计量(3-61)称为沃尔德统计量($WaldStatistic$)。

有很多证据表明，$Wald$ 统计量的有限样本性质和检验的效果往往不如人意，一个原因是它对于约束条件的再表达（即通过恒等变形将约束条件表示为其他形式）不具有不变性。约束条件的某种表达形式能够得出很好的 $Wald$ 检验，而另外一些表达形式却会导致严重的过度拒绝或拒绝不足（很少发生）。

似然比检验 还可以根据似然原则推导出零假设和备择假设的检验统计量。

令 $L(y;\hat{\theta}_c)$ 和 $L(y;\hat{\theta})$ 分别为在零假设和备择假设下的对数似然函数，其中

$$\theta' = (\beta', \sigma^2)$$

是 $k+1$ 个参数向量。在一般条件下，当 $n \to \infty$ 时，似然比统计量在零假设下服从自由度为 q 的极限

$$\chi^2$$

分布。

似然比检验统计量等于似然函数的无约束极大值和约束极大值之差两倍：

$$LR = -2(L(y, \hat{\theta}_c) - L(y, \hat{\theta}_u)) \tag{3-66}$$

对于具有正态误差项的回归模型，极大化对数似然函数等价于极小化残差平方和函数 SSR:

$$l(y, \hat{\beta}, \hat{\sigma}) = -\frac{n}{2}(1 + \log 2\pi - \log n) - \frac{n}{2} \log SSR(\hat{\beta}) \tag{3-67}$$

$$\begin{aligned}
LR \quad &= 2(l(\hat{\theta}_c) - l(\tilde{\theta}_u)) \\
&= 2(\tfrac{n}{2} \log SSR(\tilde{\beta}_u) - \tfrac{n}{2} \log SSR(\hat{\beta}_c)) \\
&= n \log \left[\tfrac{SSR(\tilde{\beta}_u)}{SSR(\hat{\beta}_c)}\right] \\
&= n \log \left(1 + \tfrac{SSR(\tilde{\beta}_u) - SSR(\hat{\beta}_c)}{SSR(\hat{\beta}_c)}\right) \\
&= n \log \left(1 + \tfrac{r}{n-k} F\right) \\
&\approx rF
\end{aligned}$$

不管原假设是否为真，LR 统计量都是 F 统计量的确定性严格递增函数。对于古典正态线性模型，如果不采用自助法，用 LR 检验替代 F 检验没有太大意义，因为 F 检验在有限样本下是精确检验，而 LR 不是。

拉格朗日检验 拉氏乘子检验是以约束极大值问题的拉氏乘子向量为基础的。但在实际中，LM 检验很少以这种方式计算，相反，通常是在无约束对数似然函数的梯度向量或者得分向量的基础上进行的。LM 检验通常通过人工回归进行计算:

$$LM = g^T(\tilde{\theta}) \tilde{I}^{-1} g(\tilde{\theta}) \tag{3-68}$$

约束条件 $r(\theta) = 0$ 下对数似然函数 $l(\theta)$ 极大化的一个方法是采用拉氏函数：

$$l(\theta) - r^T(\theta)\lambda$$

极值问题解的一阶条件是 $k+r$ 个方程：

$$\begin{gathered}
g(\tilde{\theta}) - R^T(\tilde{\theta})\tilde{\lambda} = 0 \\
r(\tilde{\theta}) = 0
\end{gathered}$$

可以将 LM 统计量用拉氏乘子 λ 表达:

$$LM = \tilde{\lambda}^T \tilde{R} \tilde{I}^{-1} \tilde{R}^T \tilde{\lambda} \tag{3-69}$$

在实际中，常用其他估计量取代 $\tilde{I}$，如负的海赛矩阵或者 OPG 估计量。如果采用 OPG 估计量，统计量变为：

$$\tilde{g}^T (\tilde{G}^T \tilde{G})^{-1} \tilde{g} \tag{3-70}$$

它可以采用人工回归的方法进行计算，对应的回归称为 $OPG8$，其一般形式为：

$$\tau = G(\theta)c + \text{残差} \tag{3-71}$$

在似然比检验中，对有约束的对数似然函数，其拉格朗日条件为：

$$H = L(y; \theta) - \lambda' h(\theta) \tag{3-72}$$

拉格朗日乘数度量每个约束的影子价格，如果影子价格高，则拒绝约束与数据不一致的假定。这就是 $AitchesonSilvey1958$ 提出的拉格朗日乘数检验。

因此，建立在拉格朗日乘子基础上的检验与拉奥(Rao)的分数检验($Rao'sscoretest, 1948$)等价，其定义如下：

$$-\frac{\partial L(y;\theta)}{\partial\theta'}\bigg|_{\hat{\theta}_c}[E\frac{\partial^2 L(y;\theta)}{\partial\theta\partial\theta'}]^{-1}\frac{\partial L(y;\theta)}{\partial\theta}\bigg|_{\hat{\theta}_c} \tag{3-73}$$

在零假设下它服从自由度为 q 的极限卡方分布。

Wald 、LR 和LM 的关系 就线性回归模型而言，线性约束的零假设意味着：

$$\begin{cases} \frac{\partial H}{\partial\beta}=\frac{\partial L(y;\theta)}{\partial\beta}-R'\lambda \\ \frac{\partial H}{\partial\sigma^2}=\frac{\partial L(y;\theta)}{\partial\sigma^2}=-\frac{n}{2\hat{\sigma}_c^2}+\frac{1}{2\hat{\sigma}_c^4}(y-X\hat{\beta}_c)'(y-X\hat{\beta}_c) \\ \frac{\partial H}{\partial\lambda}=R\hat{\beta}_c-r \end{cases}$$

因此，在零假设下，σ^2 的极大似然估计量等于：

$$\hat{\sigma}_c^2=\frac{1}{n}\hat{u}_c'\hat{u}_c$$

在备择假设下，等于

$$\hat{\sigma}_c^2=\frac{1}{n}\hat{u}\hat{u}$$

。

将上两式代入各自似然函数，指数项等于 $exp(-n/2)$。因此，(**??**)的似然比统计量变为：

$$\begin{aligned}\xi_{LR}&=n\log\frac{\hat{u}_c'\hat{u}_c}{\hat{u}'\hat{u}}\\&=n\log(\frac{\hat{u}_c'\hat{u}_c-\hat{u}'\hat{u}}{\hat{u}'\hat{u}}+\mathbf{1})\\&=n\log(1+\frac{q}{n-k}\xi_W)\end{aligned}$$

在零假设下，

$$\frac{\partial H}{\partial\beta}=\frac{1}{\hat{\sigma}_c^2}X'(y-X\hat{\beta}_c)-R'\hat{\lambda}=0$$

，对其左乘 $R(X'X)^{-1}$，并对 λ 求解得

$$\begin{aligned}\hat{\lambda}&=\frac{1}{\hat{\sigma}_c^2}[R(X'X)R']^{-1}R(X'X)^{-1}\frac{\partial L(y;\theta)}{\partial\beta_c}\\&=\frac{1}{\hat{\sigma}_c^2}[R(X'X)R']^{-1}R(\hat{\beta}-\hat{\beta}_c)\end{aligned}$$

将约束条件代入：

$$\hat{\lambda}=\frac{1}{\hat{\sigma}_c^2}[R(X'X)R']^{-1}(R\hat{\beta}-r)$$

如果约束是正确的，则当 $n\to\infty$ 时，σ_c^2 依概率收敛于 σ^2，$\hat{\lambda}$ 有极限正态分布：

$$\hat{\lambda}\sim\mathbf{N[0,[\sigma^2}R(X'X)^{-1}R']^{-1}]$$

因此，将 σ_c^2 代入 σ^2 得到的二次形式就是拉格朗日乘数检验：

$$\xi_{LM}=\frac{1}{\hat{\sigma}_c^2}(R\hat{\beta}-r)'[R(X'X)R']^{-1}(R\hat{\beta}-r)$$

近似服从自由度为 q 的 χ^2 分布，可写为：

$$\xi_{LM}=(\frac{nq}{n-k}\xi_W)(1+\frac{q}{n-k}\xi_W)^{-1}$$

从本质上说，LM 始于零假设并探求备择假设，而Wald 始于备择假设，而考虑零假设。Engle 1984 讨论得出，极大似然方法则在同一基础上比较两个假设。对有限样本，有：

$$\xi_{LM} \leqslant \xi_{LR} \leqslant \xi_{W}$$

。

3.21 计量经济模型设定检验

对一个错误设定的模型进行估计得出的结果是有偏的和不一致的。我们讨论过的许多检验都可用于设定检验：

- 对变量遗失的检验和系数相等的 F 检验
- 对非线性回归模型和 IV 回归进行的类似检验
- 异方差性检验
- 序列相关检验
- 公因子约束检验
- DHW 检验
- 过度识别检验
- 对极大似然方法估计的模型三种古典检验

3.22 多元回归模型的假设检验

我们检验一个假定的方法将是构造一个含有假定作为对参数约束的统计模型，当一个理论含有某些对模型的可检验约束时，称为**具有可检验含义**$testableimplications$。

例如，考虑一个投资模型：

$$\textbf{投资} = \beta_0 + \beta_1\text{利率} + \beta_2\text{通货膨胀率} + \varepsilon$$

这表明投资者对名义利率和通货膨胀率是敏感的。另一个备择理论说“投资者关心的是实际利率”，则备择模型是：

$$\textbf{投资} = \beta_0 + \beta_1\text{利率- 通货膨胀率} + \beta_2\text{通货膨胀率} + \varepsilon$$

虽然这个模型体现了理论，但它不具备可检验意义。然而，考虑更强的假设“投资者只关心实际利率”：

$$\textbf{投资} = \beta_0 + \beta_1\text{利率- 通货膨胀率} + \varepsilon$$

现在受约束了：

$$\beta_1 + \beta_2 = 0$$

假设检验可从两个角度出发，一是计算了一组参数估计后，我们可以问，参数估计是否合理地接近约束；或，参数估计对约束的偏离是来自抽样误差还是来自系统误差。另一是强加上约束，根据定义知最小二乘估计是“最小平方”的，所以这个强加会导致拟合的损失。我们可以确定这个损失是否仅仅来自抽样误差，或它是否大到让我们怀疑约束的可信性。

一种方法用 F 统计量：

$$F[J, N-K] = \frac{(R\hat{\beta} - q)'[Rs^2(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)}{J}$$

假定我们在回归中明确加入假设的约束，约束最小二乘回归估计量可作为下面的解而得到：

$$\min_{\tilde{\beta},\lambda} = \left(y - X\tilde{\beta}\right)' \left(y - X\tilde{\beta}\right)$$
$$s.t. \qquad R\tilde{\beta} = q$$

这个问题的拉格朗日解是

$$S = \left(y - X\tilde{\beta}\right)' \left(y - X\tilde{\beta}\right) + 2\lambda' \left(R\tilde{\beta} - q\right)$$

$$\frac{\partial S}{\partial \tilde{\beta}} = -2X'(y - X\tilde{\beta}) + 2R'\lambda = 0$$
$$\frac{\partial S}{\partial \lambda} = 2\left(R\tilde{\beta} - q\right) = 0$$

展开可以得到分块矩阵方程：

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} X'y \\ q \end{bmatrix}$$

或

$$Wd^* = V$$

约束最小二乘估计量是：

$$d^* = W^{-1}V$$

若叉积矩阵非奇异，可得到一显示解：

$$\tilde{\beta} = \hat{\beta} - (X'X)^{-1}R'[(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)$$
$$\lambda = [R(X'X)^{-1}R']^{-1}(R\hat{\beta} - q)$$

$\tilde{\beta}$ 的协方差矩阵简单就是 σ^2 乘以 W 矩阵的左上角。在 $X`X$ 非奇异情况下可得到一个显示解：

$$Var(b^*) = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

方程右边第二项是一正定矩阵，对这一方差的减少的解释是约束是包含了有价值的信息。

我们知道，约束最小二乘解对方程的拟合将劣于无约束解的拟合：

$$\begin{aligned} e_c &= Y - X\beta_c \\ &= Y - X\beta_u - X(\beta_c - \beta_u) \\ &= e_u - X(\beta_c - \beta_u) \end{aligned}$$

约束离差平方和是

$$e_c'e_c = e_u'e_u + (\beta_c - \beta_u)'X'X(\beta_c - \beta_u)$$

我们可以将一个约束检验基于回归拟合的损失：

$$e_c'e_c - e_u'e_u = (R\beta - q)'[R(X'X)^{-1}R']^{-1}(R\beta - q)$$

这是 F 统计量的分子形式，补上分母：

$$F[J, n-k] = \frac{(e_c'e_c - e_u'e_u)/J}{e_u'e_u/(n-k)} = \frac{(R_u^2 - R_c^2)/J}{(1 - R_u^2)/(n-k)}$$

第四章　线性回归模型

4.1　古典线性模型

y 对 x 的**总体线性回归模型**:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

构成古典线性模型的五个假定是:

(1) 函数形式

(2) 扰动项零均值

(3) 扰动项同方差

(4) 扰动项无自相关

(5) 解释变量和扰动项独立

对干扰项应用中心极限定理，可得出一个额外假定：

$$\varepsilon_i \sim N[0, \sigma^2]$$

干扰项正态性假定对于确切地表达估计的特性和假设检验是有益的，但古典结论只需要假定1-5就可以了。

在回归的环境中，线性指的参数进入模型的方式，而不一定是变量 x 和 y 之间的关系。

一个常用的函数形式是对数线性模型：

$$y = Ax^{\beta}$$

取对数得：

$$lny = a + \beta lnx$$

这称作**不变弹性形式**。对数线性模型常用来估计需求函数和生产函数。

对任何大于 1 的 n，必须有：

$$\frac{1}{n} S_{XX} = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

为一个有限正数，它被称作**识别条件**。若该条件不被满足，则所有数据落入一条垂直于 X 轴的直线上。它没有违背古典假定，但它不可估计。这来自数据集的缺陷。

若 x 是随机变量，则假定一成为对 x 和 y 之间联合分布的陈述。

每一个干扰项都是一个连续随机变量的实现值。

非自相关假定并不意味着 y_i 和 y_j 不相关，它是观测值与其期望值的离差不相关。

a 和 β 是 y_i 概率分布中的未知参数，我们希望利用样本数据估计出来，这是一个统计推断问题。

最小二乘：

- 作为代数问题的数据回归拟合：由残差平方和最小的一阶条件解出参数，用似合优度进行衡量
- 作为统计推断的回归拟合，可以计算出参数的均值和方差。参数是因变量的函数。
- 马尔可夫定理

原始的拟合标准–残差平方和，也提供了回归直线对数据的拟合度量。但对 y 的值用相应的比例因子相乘后，残差平方和可任意缩放。既然回归的拟合值基于 x，可以考虑用 x 的变差作为 y 的变差的一个不错的预测极值。

标准差 $standarddeviation$

标准误 $standarderror$

方差 $variance$

残差 $residual$

离差 $deviation\ y_i - \bar{y}$

变差 $variation\ \sum_i (y_i - \bar{y})^2$

在不同情况下，回归直线的拟合的相对质量很难比较。如一个线性还是对数线性模型对数据拟合得更好。因为两个模型的变差是不同的。R^2 是对 x 和 y 间线性关系的度量是值得关注的。R^2 解释为变差的可解释部分依赖于用最小二乘法来计算拟合。写出下式是永远正确的：

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + \varepsilon_i$$

但是，在计算两边平方和时，仅当用最小二乘法计算拟合值且包含常数项时叉积项才消失。

最小二乘估计是一个代数问题。还有许多估计方法，如最大值与最小值连线估计斜率等。它们可以在参数估计的统计推断基础上进行比较，如无偏性，有效性和一致性。但这些比较依赖于 DGP 的分布情况。

斜率系数的最小二乘估计量是

$$\hat{\beta} = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})\, y_i}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} = \sum_{i=1}^{n} \omega_i y_i$$

它是因变量的线性函数，以自变量函数为权数。因而是一个线性估计量。它的均值是 β，方差是 σ^2/S_{xx}，是无偏估计量。利用马尔科夫定理，从而它是最佳线性无偏估计量。加入正态分布假定，则可以得到：

$$\begin{bmatrix} \hat{\alpha}_{OLS} \\ \hat{\beta}_{OLS} \end{bmatrix} \sim N\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \sigma^2 \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{S_{XX}} & -\frac{\bar{x}}{S_{XX}} \\ -\frac{\bar{x}}{S_{XX}} & \frac{1}{S_{XX}} \end{pmatrix}\right]$$

通常的参数估计过程包括置信区间和构造和参数的假设检验。这需要参数的真正样本方差的估计，即估计未知的 σ^2。σ^2 是ε^2 的期望值，而 e_i 是 ε_i 的一个估计。写出

$$e_i = y_i - a - bx_i$$

变形可得到：

$$e_i = \varepsilon_i - \bar{\varepsilon}_i - \bar{\varepsilon} - (x_i - \bar{x})(b - \beta)$$

进一步有

$$e_i = \varepsilon_i - \bar{\varepsilon}_i - \bar{\varepsilon} - (x_i - \bar{x}) \sum c_j \varepsilon_j$$

我们对某一个别干扰项的估计受两种因素的扭曲，所有干扰项的样本平均和 β 并非是完美估计这一事实所造成的影响。(所有干扰项是独立的，则 $E\varepsilon_j\varepsilon_i = 0$。)对上式两边平方再期望值，然后再对下标求和可得：

$$E[\sum e_i^2] = (n-2)\sigma^2$$

这表明 σ^2 的一个无偏估计量是

$$s^2 = \frac{\sum e_i^2}{n-2}(1œ2gC\backslash\tilde{}) \tag{4-1}$$

这样可以得到 β 的一个抽样方差的估计：

$$Est.Var[\beta] = \frac{s^2}{S_{XX}}$$

这个数的平方根是 β 的估计的标准误差，这通常被简单的称为$standarderror$。

β 的置信区间将由此得出：

$$Prob\left(\hat{\beta} - t_{\alpha/2}s_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{\alpha/2}s_{\hat{\beta}}\right) = 1-\alpha$$

我们也可以构造干扰项方差 σ^2 的 95% 的置信区间

$$\frac{(n-2)\,s^2}{\chi^2_{0.975}} \leqslant \sigma^2 \leqslant \frac{(n-2)\,s^2}{\chi^2_{0.025}}$$

另一个相关过程是检验参数是否为一给定值。常见的检验方式是检验参数 β 是否明显的异于零。若

$$\left|\frac{\hat{\beta}}{s_{\hat{\beta}}}\right| > t_{\frac{\alpha}{2}}$$

则假设被拒绝，系数为“统计显著的”。

一类不同的假设检验基于回归的拟合。由于 $\hat{\beta}$ 的估计基于最小化离差平方和，又 $R^2 = 1-\frac{ESS}{TSS}$，所以 $\hat{\beta}$ 选取意味着最小化离差平方和，等同于最大化线性拟合优度 R^2。若 $\hat{\beta}=0$，则 $R^2=0$。那么 $\hat{\beta}=0$ 基于 t 比率可以构造一个是否导致拟合显著损失的一个检验。考虑 t 比率的平方：

$$t^2 = \frac{\hat{\beta}^2}{\hat{\sigma}^2} = \frac{\hat{\beta}^2 S_{XX}}{s^2} = \frac{(S_{XY}/S_{XX})^2\,S_{XX}}{S_{YY}(1-R^2)/\,(n-2)} = F(1,n-2)$$

它提供了又一种检验系数是否为零的检验方法。

我们可以从两种角度来检验回归系数是否为零的假设。一种是 t 检验，它基于参数最小二乘估计和假设值之差，其原理是若假设正确，则在抽样偏差范围内最小二乘估计将满足这一假设。第二种是 F 检验，其原理是若假设正确，它将不会降低回归直线拟合程度。

当我们进行预测时，必须接受两类误差。一类来自参数估计的抽样误差；一类来自干扰项的随机扰动

$$\hat{y} = \hat{\alpha} + \hat{\beta}$$

预测误差是

$$\hat{e} = y - \hat{y}$$

它可以重写为

$$\hat{e} = (\alpha-\hat{\alpha}) + (\beta-\hat{\beta})x + \varepsilon$$

预测误差均值为零，从而在这个意义上最小二乘预测是无偏的。方差是：

$$\begin{aligned}
Var(e) &= Var(\hat{\alpha}) + x^2Var(\hat{\beta}) + Var(\varepsilon) + 2xVar(\hat{\alpha},\hat{\beta}) \\
&= \sigma^2\left((\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}) + \frac{x^2}{S_{XX}} + 1 + 2x(-\frac{\bar{x}}{S_{XX}})\right) \\
&= \sigma^2\left(1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{XX}}\right)
\end{aligned}$$

通过样本估计 σ^2 可以为 y 构造一个置信区间。

若一个多元回归中的变量是无关，则多元回归的斜率和在各个简单回归中的斜率相同。对数据矩阵 X 的任一列，有 $x`_k e = 0$。若该列全为 1，则最小二乘残差和为零。

$$Y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

利用分块矩阵可得：

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y$$
$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 Y$$

所以 β_2 来自一个回归的系数集合。这个回归的被解释变量是 Y 单独对 X_1 回归的残差 M_1Y，解释变量是 X_2 的每一列分别对 X_1 回归的得的残差 M_1X_2。这个过程通常被称作排除或筛除 X_1 的影响，部分地由于这个原因，多元回归的系数称作**偏回归系数**。运用相同原理，可以计算**偏相关系数**。它有一个简便公式：

$$r_{yk} = \frac{t_k^2}{t_k^2 + \text{自由度}}$$

任何 z 对 i 的回归系数是 $(i`i)^{-1}i`z = \bar{z}$，残差是 $z - \bar{z}$。利用**重期望律**

$$E\hat{\beta} = E_X\left[E\left(\hat{\beta}|X\right)\right] = E_X\left(\beta + \left(X'X\right)^{-1}X'\varepsilon\right) = \beta$$

从而OLS 的无偏性对关于 X 的矩阵是稳健的。对方差进行分解，

$$Var[\hat{\beta}] = E_X[Var(\hat{\beta}|X)] + Var_X[E(\hat{\beta}|X)] = \sigma^2 E[(X`X)^{-1}]$$

对于多元回归，可知：

$$\hat{\beta} \sim N[\beta, \sigma^2(X`X)^{-1}]$$

最小二乘残差是

$$e = M(X\beta + \varepsilon) = M\varepsilon$$

σ^2 的估计量将基于残差平方和：

$$e`e = (M\varepsilon)`M\varepsilon$$

这个二次型的期望值是

$$E(e`e) = E(\varepsilon`M\varepsilon)$$

其迹是：

$$E(tr(\varepsilon`M\varepsilon)) = E(tr(M)\varepsilon`\varepsilon) = \sigma^2 tr(M) = (n-K)\sigma^2$$

σ^2 的一个无偏估计量是

$$s^2 = \frac{e`e}{n-K}$$

可以导出

$$\frac{(n-K)s^2}{\sigma^2} = \frac{e'e}{\sigma^2} = \left(\frac{\varepsilon}{\sigma}\right)' M \left(\frac{\varepsilon}{\sigma}\right)$$

是一个标准正态向量 ε/σ 的幂等二次型。所以它服从

$$\text{秩}\{M\} = \text{迹}\{M\} = n - K$$

的卡方分布。

$$t_k = \frac{\left(\hat{\beta}_k - \beta_k\right)\Big/\sqrt{\sigma^2 S^{kk}}}{\{[(n-k)\,s^2/\sigma^2]\Big/(n-k)\}^{\frac{1}{2}}}$$
$$= \frac{\hat{\beta}_k - \beta_k}{\sqrt{s^2 S^{kk}}}$$

假定我们的假设是：

$$H_0 : R\beta = q$$

样本估计是：

$$R\hat{\beta}=\hat{q}$$

若 $\hat{q}$ 显著地异于 q，则样本数据与假设不一致。一个自然的检验方法是：

$$t=\frac{\hat{q}-q}{Var(\hat{q})}$$

这需要估计 $\hat{q}$ 的方差。$\hat{q}$ 是 $\hat{\beta}$ 的线性函数：

$$Est.Var(\hat{q})=R`Var(\hat{q})R=R`[s^2(X`X)^{-1}]R$$

给定最小二乘估计量，我们的兴趣集中在差异向量

$$d=R\hat{\beta}-q$$

我们可以将对原假设的检验基于沃尔德准则：

$$W=\chi^2(J)=d`[Var(d)]^{-1}d$$

d 越大，则最小二乘满足约束的可能性越小。

$$Var(d)=\sigma^2R`(X`X)^{-1}R$$

用 s^2 代替 σ^2：

$$F=\frac{\left(R\hat{\beta}-q\right)'\left[\sigma^2R(X'X)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)\Big/J}{[(n-k)\,s^2/\sigma^2]/(n-k)}$$

与 R^2 有关的两个问题是：

- 一是增加变量可以提高拟合优度。
- 二是若回归模型没有常数项，结果不可预测。因为 $M^0e=e$ 的结果依赖于有一列数据全为 1。它意味着增加常数项可以提高似合优度。

检验方程整体显著性的 F 检验是：

$$F(k-1,n-k)=\frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

在 X 为随机回归量的情况下，获得 $\hat{\beta}$ 的统计特性的一个方便方法是，首先求得对 X 的条件期望结果，这等同于非随机回归量的情况，然后，通过“平均”（即积分）条件分布得到无条件分布结果。此论点的关键是，如果我们对任意 X 都可以得到条件无偏性，我们就可以对 X 进行平均求得一个无条件结果：

$$E[\hat{\beta}|X]=\beta+(X`X)^{-1}X1E[\varepsilon|X]=\beta$$

在推导回归分析中许多检验统计量起中心作用的一个结论是：

命题 4.1 若 ε 服从正态分布，最小二乘系数估计量 $\hat{\beta}$ 统计独立残差向量 e 及包括 s^2 在内的任何函数。

$$\frac{\hat{\beta}-\beta}{\sigma}=\frac{(X'X)^{-1}X'\varepsilon}{\sigma}=\left(X'X\right)^{-1}X'(\frac{\varepsilon}{\sigma})$$

它独立于一个秩等于迹的幂等二次型（它服从卡方分布）：

$$\frac{(n-k)\,s^2}{\sigma^2}=\frac{e'e}{\sigma^2}=\left(\frac{\varepsilon}{\sigma}\right)'M\left(\frac{\varepsilon}{\sigma}\right)$$

所以

$$t_K=\frac{\left(\hat{\beta}-\beta\right)\Big/\left(\sqrt{\sigma^2S^{kk}}\right)}{\left\{\left((n-k)s^2/\sigma^2\right)\Big/(n-k)\right\}^{\frac{1}{2}}}=\frac{\hat{\beta}_K-\beta_K}{\sqrt{s^2S^{KK}}}$$

服从自由度为 $n-k$ 的 t 分布。

命题 4.2　t 的边际分布仍是 t_{n-k}，不论 X 的分布是什么，也不论 X 是随机是还是非随机的。

命题 4.3　若干扰项是正态分布的，我们可以在我们的过程不加变化地进行检验和构造参数置信区间，而不去考虑回归量是随机的，非随机的，还是它们的混合。

命题 4.4　仅在一个观测值上取值为 1 的虚拟变量具有从最小二乘计算中删除该观测值的作用。

基于预测残差的两个度量是：

- **均方根误差** $=$ **RMSE** $= \sqrt{\frac{1}{n_0} sum_i (y_i - \hat{y}_i)^2}$
- **平均绝对误差** $=$ **MAE** $= \frac{1}{n_0} \sum_i |y_i - \hat{y}_i|$

这两个度量都存在标度问题，不存在标度问题的是**泰尔的** U **统计量**:

$$U = \frac{\sqrt{\frac{1}{n_0}\left(y_i - \hat{y}_i\right)^2}}{\sqrt{\frac{1}{n_0} y_i^2}}$$

另一个可选度量有以 y 的变化来度量：

$$U = \frac{\sqrt{\frac{1}{n_0}\left(\Delta y_i - \Delta\hat{y}_i\right)^2}}{\sqrt{\frac{1}{n_0} \Delta y_i^2}}$$

其中

$$\Delta y_i = y_i - y_{i-1} \ \Delta y_i = \frac{y_i - y_{i-1}}{y_{i-1}}$$
$$\Delta \hat{y}_i = \hat{y}_i - y_{i-1} \ \Delta \hat{y}_i = \frac{\hat{y}_i - y_{i-1}}{y_{i-1}}$$

这些度量具有跟踪数据中转折点的能力。

4.2　一元线性回归模型

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{4-2}$$

它可以分为两部分：回归函数部分 $E(y_t) = \beta_0 + \beta_1 x_t$ 和随机部分 u_t。

问题 4.1　经济研究中有一个重要假定：控制其它因素不变。如研究消费与收入关系时假设价格等其他影响因素保持不变。而在实际经济活动中，价格等其他因素是变动的，构成了扰动因素。

思考 4.1　为什么要引入随机误差项？

线性回归模型的十一个假定 对一般的线性回归模型

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 \\ &\quad + \cdots + \alpha_p X_p + \epsilon \end{aligned} \tag{4-3}$$

对其进行 T 次观测，得

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} \\ &\quad + \cdots + \alpha_p X_{pt} + \epsilon_t \end{aligned} \tag{4-4}$$

其中 $t = 1, 2, \cdots, \boldsymbol{T}$。对应的矩阵形式为

$$\boldsymbol{Y} = \boldsymbol{X}\alpha + \epsilon \tag{4-5}$$

为了得到较好的统计性质，如**BLUE**，需要对**OLS** 法作如下简单假定（引自$D.N.Gujarati 1995 pp.313 - 314$）：

假定1 回归模型关于参数是线性的；

假定2 在重复抽样中，回归变量 $\boldsymbol{X}_i\ i=1,2,\cdots,p$）的取值是非随机的；

假定3 零均值

假定4 同方差

假定5 无序列相关

假定6 若在回归变量中有随机的 $\boldsymbol{X}_i$，则随机扰动项与这个 $\boldsymbol{X}_i$ 独立或至少是不相关的；

假定7 观测次数**T**必须大于回归变量的个数，即 $\mathbf{T}>p$;

假定8 回归变量的取值必须有足够的变异；

假定9 回归模型已被正确地设定；

假定10 无多重共线性

假定11 随机扰动项 ϵ_t 服从正态分布。

二级检验 这些假定是为了数学处理方便而从大量经济实践中抽象出来的理想状况，而实际研究的具体经济现象却不一定能够满足。因此有必要对这 11 个假定进行再讨论，讨论所研究的经济现象在多大程度上满足这 11 个假定，这就是经济计量检验，又称为二级检验。

$G.B.Wetherill$1986 所指出，古典线性回归模型的实际应用所引起的问题主要有二类：

(1) 一是由假定 $1\sim5$，9与11 的违反所引起的关于模型设定及随机扰动项的问题，如非线性、异方差性、序列相关性、模型设定误差及随机扰动项的问题，如非线性、异方差性、序列相关性、模型设定误差与非正态误差等；

(2) 另一是由假定 $6\sim8$ 与10 的违反的所引起的关于回归变量及其数据质量的问题，如随机回归变量、小样本、多重共线性、异常值等。

随机干扰项的意义 随机扰动项是从模型中省略下来的而又集体地影响着 Y 的全部变量的替代物。显然的问题是：为什么不把这些变量明显地引进到模型中来，而以随即扰动项来替代？理由是多方面的：（1）理论的含糊性：理论不能完全说明影响因变量的所有影响因素。//（2）数据的欠缺：无法获得有关数据。//（3）核心变量与周边变量：希望能找到与有较大影响的核心变量的关系。//（4）内在随机性：因变量具有内在的随机性。//（5）替代变量：用来代替不可观测变量的替代变量选择，造成一定误差。//（6）省略原则：研究中尽可能使回归式简单。//（7）错误的函数形式：回归式的的选择是主观的。

最小二乘估计量的性质

希望当样本容量变得非常大时，$\hat{\beta}$ 不等于 β 的概率变得非常小：如果 $\hat{\beta}$ 依概率收敛于 β，则为一致估计。**高斯–马尔可夫定理 最小二乘回归直线的性质**

(1) 残差和等于零

(2) 直线过 $(\bar{x},\bar{y}$;

(3) 拟合值平均值等于观测值平均值 $\bar{\hat{y}}=\bar{y}$;

(4) 残差与解释变量不相关 $Cov(\hat{u},x)=0$;

(5) 残差与被解释变量拟合值不相关 $Cov(\hat{u},\hat{y})=0$。

参数的估计与分布

$$y_t \sim N(\beta_0 + \beta_1 x_t, \sigma^2) \tag{4-6}$$

$$Var(\hat{\beta}_1) = \frac{1}{\sum (x_t - \bar{x})^2} \hat{\sigma}^2 \tag{4-7}$$

误差均方：

$$\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{T-k} = \frac{\hat{e}'\hat{e}}{T-k} \tag{4-8}$$

$$Var(\hat{\beta}_0) = \frac{\sum x_t^2}{T\sum (x_t - \bar{x})^2} \hat{\sigma}^2 \tag{4-9}$$

参数的显著性检验与置信区间 点预测与区间预测

$$Var(y_F) = (1 + \frac{1}{T} + \frac{(x_F - \bar{x})^2}{\sum (x_t - \bar{x})^2}) \hat{\sigma}^2 \tag{4-10}$$

$$Var(E(y_F)) = (\frac{1}{T} + \frac{(x_F - \bar{x})^2}{\sum (x_t - \bar{x})^2}) \hat{\sigma}^2 \tag{4-11}$$

例子

源代码 4.1 SAS:一元线性回归

```
proc import out=tem
        datafile =”D:\SAS\dst\Mylib\exceldata\singlereg.xls” replace;
        getnames=no;
run; data tem;
      set tem;
      rename F1=X F2=Y;
run; proc reg data=tem corr outest=out1;
      model Y=X /dw r;
      output out=out2 p=yhat r=error;
run; title ’Scatter’; proc gplot data=tem;
      plot y*x;
run; title ’Estimation’; proc print data=out1; run; title ’Fitted
Value␣and␣Residual’; proc print data=out2(keep=yhat error); run;
data temp;
      merge out2 tem;
run; title ’True␣Value␣and␣Fitted␣Value’; proc print
data=temp(keep=yhat y); run;
```

4.3 虚拟变量 dummy variable

定义

定义 4.1 (**虚拟变量**) 虚拟变量是一种离散结构，是用来描述所研究的变量的发展或变异而建立的一种特殊变量。在经济计量学中，所考虑的变量除了可直接度量的数量变量，如收入、产出，还有不能直接度量的质量变量，如性别、政府经济政策的改变。根据质量变量的属性类型，而令其取值为”0”（比较类型、否定类型）或”1”（基础类型、肯定类型），称为虚拟变量。在经济计量模型，它即可作解释变量，又可作被解释变量。

定义 4.2 (**虚拟变量模型**) 含虚拟变量的经济计量模型。

定义 4.3 (**方差分析模型**) $analysisofvariancemodel$:如果解释变量全部都是虚拟变量，则线性回归模型称为方差分析模型。

定义 4.4 (**协方差分析模型**) $analysisofcovariancemodel$:如果解释变量中既有数量变量，又有虚拟变量，则线性回归模型称为协方差分析模型。

引入虚拟变量的作用

- 分离异常因素的影响。如观测我国社会总产值的时间趋势，必须考虑三年自然灾害。
- 检验不同属性类型对因变量的影响。如工资模型中文化程度对工资增长率的影响。
- 提高模型精度。引入虚拟变量后，相当于把不同属性类型的样本合并，即相当于扩大样本容量，从而提高模型精度。

虚拟变量陷阱 dummy variable trap 在同一模型中，可以引入多个虚拟变量。但其设置必须遵循以下原则：如果有 m 种互斥的属性类型，则在模型中引入 $m-1$ 个虚拟变量 $D_i(i =$

$1, 2, \cdots, m-1$)。若不然，定义 m 个虚拟变量，则必有 $D_1 + D_2 + \cdots + D_m = 2$，这就导致了完全多重共线性，使参数估计不能唯一确定。这就是所谓的虚拟变量陷阱（$dummyvariabletrap$）。

虚拟变量的应用

- 测量截距的移动
- 测量斜率的变动
- 调整季节波动
- 作为数值因素和非数值因素的代表
- 作为因变量的代表

求和算子的运用 定义 X 变量的 N 个观测值的均值或平均值为：

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

，那么

$$\sum_{i=1}^{N}(X_i - \bar{X}) = 0$$

定义 X 变量的方差为

$$Var(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

定义 X 与 Y 的协方差是

$$Cov(X, Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

法则:

$$Cov(X, Y) = \frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \bar{X}\bar{Y}$$

双重求和:求和运算用于两个随机变量。令 X_{ij} 为一个随机变量，定义 N^2 个值的双重求和为

$$\sum_{i=1}^{N}\sum_{j=1}^{N} X_{ij} = \sum_{i=1}^{N}(X_{i1} + X_{i2} + \cdots + X_{iN})$$

法则:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} X_i Y_j = (\sum_{i=1}^{N} X_i)(\sum_{j=1}^{N} Y_j)$$

4.4 总结

(1) 如果在不同时段，数据方差已知，则加权是一种好的处理异方差问题。

第五章　经典单方程计量经济学模型的扩展

5.1　前言

经典模型结构的内涵 经典计量经济学模型结构的内涵是：

- 变量的选择的依赖于经济理论；
- 模型所揭示的是变量之间的因果关系；
- 模型关系是明确且线性；
- 模型中参数是不变的。

凡是与这些相背的，就称为模型结构非经典的计量模型。**古典假定松驰带来的问题**

原因 即确定有无不满足古典假定的可能性；

后果 即如果不满足古典假定，用普通最小二乘法进行估计，将会发生什么偏差？

诊断 即用什么方法来诊断古典假定是否成立。

补救 即古典假定若不满足，应采取什么补救措施。

5.2　经典计量经济学模型

经典线性计量经济学模型的形式 经典线性计量经济学模型分为单方程模型和联立方程模型。单方程模型用以研究单一的经济活动中各变量之间的关系，联立方程模型用以研究经济系统中各变量之间的关系。

单方程模型的一般形式：

$$Y = BX + N \tag{5-1}$$

联立方程模型的一般形式：

$$BY + \Gamma X = N \tag{5-2}$$

经典线性计量经济学模型理论模型的设定 对所要研究的经济现象进行深入的分析，根据研究的目的，选择模型中将包含的因素，根据数据的可得性选择适当的变量来表征这些因素，并根据经济行为理论和样本数据显示出的变量间的关系，设定描述这些变量之间关系的数学表达式，即理论模型。例如生产函数

$$Y = Ae^{rt}K^{\alpha}L^{\beta}$$

就是一个理论模型。理论模型的设计主要包含三部分工作，即选择变量、确定变量之间的数学关系、拟定模型中待估计参数的数值范围。

确定模型所包含的变量，主要是指确定解释变量。可以作为解释变量的有下列几类变量：外生经济变量、外生条件变量、外生政策变量和滞后被解释变量。其中有些变量，如政策变量、条件变量经常以虚变量的形式出现。

思考 5.1　怎样才能正确的选择解释变量？

(1) 正确理解和把握所研究的经济现象中所包含的经济理论和经济行为规律，这是基础；

(2) 要考虑数据的可获得性，这需要对经济统计学有一定的了解；

(3) 要考虑入选变量之间的关系，它们之间应是相互独立的，这是建模的技术要求。

经典线性计量经济学模型的样本数据 样本数据的收集与整理，是建立计量经济学模型过程中最为费时费力的工作，也是对模型质量影响极大的一项工作。从工作程序上讲，它是在理论模型建立之后进行，但实际上经常是同时进行的，因为能否收集到合适的样本观测值是决定变量取舍的主要因素之一。**几类常用的样本数据** 常用的样本数据有三类：时间序列数据、截面数据和虚变量数据。

时间序列数据是一批按照时间先后排列的统计数据。一般由统计部门提供，在建立计量经济学模型时应充分加以利用，以减少收集数据的工作量。在利用时间序列数据作样本时，要注意以下几个问题。一是所选择的样本区间内经济行为的一致性问题。例如，我们建立纺织行业生产模型，选择反映市场需求因素的变量，诸如居民收入、出口额等作为解释变量，而没有选择反映生产能力的变量，诸如资本、劳动等，原因是纺织行业属于供大于求的情况。对于这个模型，利用时间序列数据作样本时，只能选择80年代后期以来的数据，因为纺织行业供大于求的局面只出现在这个阶段，而在 80 年代中期以前的一个长时期里，我国纺织品是供不应求的，那时制约行业产出量的主要因素是投入要素。二是样本数据在不同样本点之间的可比性问题。经济变量的时间序列数据往往是以价值形态出现的，包含了价格因素，而同一件实物在不同年份的价格是不同的，这就造成样本数据在不同样本点之间不可比。需要对原始数据进行调整，消除其不可比因素，方可作为模型的样本数据。三是样本观测值过于集中的问题。经济变量在时间序列上的变化往往是缓慢的，例如，居民收入每年的变化幅度只有 5% 左右。如果在一个消费函数模型中，以居民消费作为被解释变量，以居民收入作为解释变量，以它的时间序列数据作为解释变量的样本数据，由于样本数据过于集中，所建立的模型很难反映两个变量之间的长期关系。这也是时间序列数据不适宜于对模型中反映长期变化关系的结构参数的估计的一个主要原因。四是模型随机误差项的序列相关问题。用时间序列数据作样本，容易引起模型随机误差项产生序列相关。

截面数据是一批发生在同一时间截面上的调查数据。例如，工业普查数据、人口普查数据、家计调查数据等。主要由统计部门提供。用截面数据作为计量经济学模型的样本数据，应注意以下几个问题。一是样本与母体的一致性问题。计量经济学模型的参数估计，从数学上讲，是用从母体中随机抽取的个体样本估计母体的参数，那么要求母体与个体必须是一致的。例如，估计煤炭企业的生产函数模型，只能用煤炭企业的数据作为样本，不能用煤炭行业的数据。那么，截面数据就很难用于一些总量模型的估计，例如，建立煤炭行业的生产函数模型，就无法得到合适的截面数据。二是模型随机误差项的异方差问题。用截面数据作样本，容易引起模型随机误差项产生异方差。

虚变量数据也称为二进制数据，一般取 0 或 1。虚变量经常被用在计量经济学模型中，以表征政策、条件等因素。例如，建立我国的粮食生产计量经济学模型，以粮食产量作为被解释变量，解释变量中除了播种面积、化肥使用量、农机总动力、成灾面积等变量外，显然，政策因素是不可忽略的。1980 年前后，由于实行了不同的政策，即使上述变量都没有变化，粮食产量也会发生大的变化。于是必须在解释变量中引入政策变量，用一个虚变量表示，对于以后的年份，该虚变量的样本观测值为 1，对于以前的年份，该虚变量的样本观测值为 0。**样本数据的质量** 样本数据的质量问题大体上可以概括为完整性、准确性、可比性和一致性四个方面。

完整性，即模型中包含的所有变量都必须得到相同容量的样本观测值。这既是模型参数估计的需要，也是经济现象本身应该具有的特征。但是，在实际中，”遗失数据”的现象是经常发生的，尤其在中国，经济体制和核算体系都处于转轨之中。在出现”遗失数据”时，如果样本容量足够大，样本点之间的联系并不紧密的情况下，可以将”遗失数据”所在的样本点整个地去掉；如果样本容量有限，或者样本点之间的联系紧密，去掉某个样本点会影响模型的估计质量，则要采取特定的技术将”遗失数据”补上。

准确性，有两方面含义，一是所得到的数据必须准确反映它所描述的经济因素的状态，即统计数据或调查数据本身是准确的；二是它必须是模型研究中所准确需要的，即满足模型对变

量口径的要求。前一个方面是显而易见的，而后一个方面则容易被忽视。例如，在生产函数模型中，作为解释变量的资本、劳动等必须是投入到生产过程中的、对产出量起作用的那部分生产要素，以劳动为例，应该是投入到生产过程中的、对产出量起作用的那部分劳动者。于是，在收集样本数据时，就应该收集生产性职工人数，而不能以全体职工人数作为样本数据，尽管全体职工人数在统计上是很准确的，但其中有相当一部分与生产过程无关，不是模型所需要的。

可比性，也就是通常所说的数据口径问题，在计量经济学模型研究中可以说无处不在。而人们容易得到的经济统计数据，一般可比性较差，其原因在于统计范围口径的变化和价格口径的变化，必须进行处理后才能用于模型参数的估计。计量经济学方法，是从样本数据中寻找经济活动本身客观存在的规律性，如果数据是不可比的，得到的规律性就难以反映实际。不同的研究者研究同一个经济现象，采用同样的变量和数学形式，选择的样本点也相同，但可能得到相差甚远的模型参数估计结果。为什么？原因在于样本数据的可比性。例如，采用时间序列数据作为生产函数模型的样本数据，产出量用不变价格计算的总产值，在不同年份间是可比的；资本用当年价格计算的固定资产原值，在不同年份间是不可比的。对于统计资料中直接提供的这个用当年价格计算的固定资产原值，有人直接用于模型估计，有人进行处理后再用于模型的估计，结果当然不会相同。

一致性，即母体与样本的一致性。上面在讨论用截面数据作为计量经济学模型的样本数据时已经作了介绍。违反一致性的情况经常会发生，例如，用企业的数据作为行业生产函数模型的样本数据。

经典线性计量经济学模型参数的估计

模型参数的估计是一个纯技术的过程，包括对模型进行识别（对联立方程模型而言）、估计方法的选择、软件的应用等内容。

两类参数估计方法可以用于经典线性计量经济学模型的估计：最下二乘法和最大似然法。例如，在模型满足基本假设时采用的普通最小二乘法，在模型存在异方差性时采用的加权最小二乘法，在模型存在序列相关性时采用的广义最小二乘法和广义差分法，在模型存在随机解释变量时采用的工具变量方法，估计联立方程模型的二阶段最小二乘法、三阶段最小二乘法，等等。在经典线性计量经济学模型估计方法中，存在着一个与最小二乘方法对应的最大似然方法系列，例如与普通最小二乘法对应的最大似然法，与二阶段最小二乘法对应的有限信息最大似然法，与三阶段最小二乘法对应的完全信息最大似然法。但是，由于其方法的数学描述较为复杂，在实际中并不常用。

经典线性计量经济学模型的检验 **方程计量经济学模型检验** 一般讲，单方程计量经济学模型必须通过四级检验，即经济意义检验、统计学检验、计量经济学检验和预测检验。

经济意义检验主要检验模型参数估计量在经济意义上的合理性。主要方法是将模型参数的估计量与预先拟定的理论期望值进行比较，包括参数估计量的符号、大小、相互之间的关系，以判断其合理性。

统计检验是由统计理论决定的，目的在于检验模型的统计学性质。通常最广泛应用的统计检验准则有拟合优度检验、变量和方程的显著性检验等，分别采用 t, F, χ^2 作为检验统计量。有时也称为一级检验。

计量经济学检验是由计量经济学理论决定的，目的在于检验模型的计量经济学性质。通常最主要的检验准则有随机误差项的序列相关检验和异方差性检验，解释变量的多重共线性检验等。有时也称为二级检验。

预测检验主要检验模型参数估计量的稳定性以及相对样本容量变化时的灵敏度，确定所建立的模型是否可以用于样本观测值以外的范围，即模型的所谓超样本特性。具体检验方法为：利用扩大了的样本重新估计模型参数，将新的估计值与原来的估计值进行比较，并检验二者之间差距的显著性；将所建立的模型用于样本以外某一时期的实际预测，并将该预测值与实际观

测值进行比较，并检验二者之间差距的显著性。**联立方程计量经济学模型的检验** 一。拟合效果检验。常用的检验统计量是均方百分比误差 RMS：

$$RMS_i = \sqrt{\sum_{t=1}^{n} e_{it}^2 / n e_{it}} = \frac{y_{it} - \hat{y}_{it}}{y_{it}} \tag{5-3}$$

二。预测性能检验。RE ：

$$RE = \frac{y_{it} - \hat{y}_{it}}{y_{it}}$$

三。方程间误差传递检验。由于联立方程模型系统中变量之间互为解释变量，那么就存在误差的传递，需要对此进行检验。一个总体结构清晰的计量经济学模型系统，应该存在一些明显的关键路径，描述主要经济行为主体的经济活动过程，这是由经济系统的特征所决定的。在关键路径上，方程之间存在明显的递推关系。例如，在一个中国宏观经济模型中，生产方程、收入方程、分配方程、投资方程、固定资产形成方程等，就构成一个关键路径。而且存在着递推关系，由固定资产决定总产值，由总产值决定国民收入，由国民收入决定财政收入，由财政收入决定投资，由投资决定固定资产。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。如果关键路径上方程数目为 T，e_i 为第 i 个方程的随机误差估计值，下面三个统计量可用于检验传送误差：

(1) 误差均值

$$\frac{1}{T}\sum_{i=1}^{T} e_i$$

(2) 均方误差

$$\sqrt{\frac{1}{T}\sum_{i=1}^{T} e_i^2}$$

(3) 冯诺曼比

$$\frac{T}{T-1}\frac{\sum_{i=1}^{T}(e_i - e_{i-1})^2}{\sum_{i=1}^{T} e_i^2}$$

四。样本点误差传递检验。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。所以对模型进行滚动预测检验是必要的。给定第一期所有先决变量的观测值，求解方程组，得到内生变量的预测值。在第二期，只给定外生变量的预测值，滞后内生变量则以第一期的预测值代替，求解方程组，得到预测值。依此类推直至第 N 期，得到预测值 $\hat{Y}_N$。并求出该滚动预测值与实际预测值的相对误差 RMS。

经典计量经济学模型成功的三要素 从上述建立计量经济学模型的步骤中，不难看出，任何一项计量经济学研究、任何一个计量经济学模型赖以成功的要素应该有三个：理论、方法和数据。理论，即经济理论，所研究的经济现象的行为理论，是计量经济学研究的基础。方法，主要包括模型方法和计算方法，是计量经济学研究的工具与手段，是计量经济学不同于其它经济学分支学科的主要特征。数据，反映研究对象的活动水平、相互间联系以及外部环境的数据，或更广义讲是信息，是计量经济学研究的原料。这三方面缺一不可。

误差修正模型 经济变量向量 X_t 的分量之间的协整性，是与经济变量间的均衡 $equilibrium$ 及误差校正机制 $errorcorrectionmechanism$ 联系在一起的。ECM 实质上就是短期均衡与长期均衡矛盾的一种手段，如消费理论中的 $Friedman$ 的“持久收入假说”。

定义 5.1 (**误差修正模型**（$ErrorCorrectionModel$）) 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。

称 X_t 的各分量是均衡的，若存在非零的常数向量 α，使得 $\alpha^T X_t = 0$ 成立（即有一线性约束）。若 X_t 非均衡，则称 $\alpha^T \delta X_t$ 为均衡误差项 $equilibriumerrorterm$。短期非均衡的 X_t 可以是长期均衡的 $long-runequilibrium$，只要存在 β 使得序列 $\beta^T X_t$ 是平稳的。这样，具有协整性的 X_t 一定是长期均衡的:

$$\alpha^T \delta X_t = -\gamma \beta^T X_{t-1} + \varepsilon_t$$

其中 $\beta^T X_{t-1}$ 称为误差校正项 $errorcorrectionterm$， ε_t 是一个平稳的多元随机扰动项。

对于 $(1,1)$ 阶自回归分布滞后模型

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \varepsilon_t$$

移项后得到

$$\begin{aligned}\Delta y_t &= \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1) y_{t-1} + \beta_3 z_{t-1} + \beta_1 z_{t-1} + \varepsilon_t \\ &= \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1)(y - \tfrac{\beta_1+\beta_3}{1-\beta_2} z)_{t-1} + \varepsilon_t\end{aligned}$$

即为误差修正模型。

$$y - \frac{\beta_1 + \beta_3}{1 - \beta_2} z$$

为误差修正项，或均衡误差。

$$\frac{\beta_1 + \beta_3}{1 - \beta_2} z$$

为协整系数。

这个模型反映的是变量的短期均衡关系，即变量 y_t 在短期是如何由 z_t 所决定的，或者说 y_t 短期波动 δy_t 是如何被决定的。如果变量 y_t 和 z_t 之间存在长期均衡关系，即有 $y_t = \alpha z_t$，那么 y_t 的均衡值与 z_t 稳定值之间存在如下关系：

$$y_t = \frac{\beta_1 + \beta_3}{1 - \beta_2} z$$

与误差修正项是一致的，反映长期均衡对短期波动的影响。

模型中的差分项表明，δy_t 的波动分为短期波动和长期均衡两部分。

原模型可以改写为：

$$\delta y_t = \beta_0 + \beta_1 \delta Z_t + \gamma \mathrm{ecm} + \varepsilon_t$$

从中可以看出误差修正项的修正作用。

实际上是先有误差修正模型，然后用协整理论去解释误差修正模型。那么在今天，我们就可以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，求出协整系数，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。

例子

源代码 5.1 ECM

```
data cosuming;
      input cosuming @@;
      datalines;
184 236 437 803 896 1070 1331 1746 2236 2641 2834 2972 3138 33973611
3791 ; data gdp;
      input gdp @@;
      datalines;
3624.1 4038.2 4517.8 4862.4 5294.7 5934.5 7171.0 8964.4 10202.2
11962.5 14928.3 16909.2 18547.9 21617.8 26638.1 34634.4 46759.4
58478.1 67884.6 74462.6 78345.2 82067.5 89468.1 97314.8 ; data temp;
      merge gdp cosuming;
 title  'Error_Correction_Mechanism'; title2 'Estimation_cointegration
vector:Residuals';
 proc reg data=temp;
      c_gdp: model cosuming=gdp;
      output out=abc residual=r;
run; data temp(keep=d1_c d1_g l_r);
      set  temp abc;
      d1_c=dif(cosuming);
      d1_g=dif(gdp);
      l_r=lag(r);
      output;
run;  title2  'ECM''s_Regression'; proc reg data=temp;
      ECM: model d1_c=d1_g l_r /dw;
run;
```

5.3 非线性经济计量检验

经典的线性回归模型的基本假定是关于回归变量、扰动项及参数都是线性的，这种简单化处理是很难找到理论根据的。经济理论常常提出非线性关系，只有定义方程、会计恒等式例外。线性只是一种近似，一种简单化，一种逼近。

如果正确的回归模型是非线性模型，而用线性模型代替，则会产生模型设定误差，估计量是有偏且非一致性的。产生自相关、异方差及非正态误差的原因之一是这种非线性模型的线性设定。

探测非线性的基本方法有二：*Andrews* 检验（$Andrews, 1971$）与 *GodfreyWickens* 检验（$Godfrey\&Wickens, 1981$）。

一般地，非线性回归模型可表示为：

$$\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{\alpha}) + \boldsymbol{\epsilon} \tag{5-4}$$

模型(5-4)的参数估计方法可分为以下三类：

（对数）线性化 直接进行变量替换；通过求导数或**Taylor** 展开间接地将模型转化为可直接线性化的模型。

非线性最小二乘法 使残差平方和

$$\begin{aligned}\boldsymbol{S_T(\alpha)} &= \|\boldsymbol{Y} - f(\boldsymbol{X},\boldsymbol{\alpha})\|^2 \\ &= (\boldsymbol{Y} - f(\boldsymbol{X},\boldsymbol{\alpha})'(\boldsymbol{Y} - f(\boldsymbol{X},\boldsymbol{\alpha})) \end{aligned} \tag{5-5}$$

达到最小值。已有人证明（$Amemiya, 1985, pp.125-135$）：若随机扰动项独立同分布并具有零均值、同方差，而函数 f 满足一些正则条件，则由(5-5)求得的非线性最小二乘估计是一致估计且具有渐近正态性，但不是有效率估计。

极大似然估计 随机扰动项具有零均值，且服从正态分布，即 $\epsilon \sim N(0,\sum)$，则令似然函数

$$\begin{aligned}\mathrm{Lik}(\alpha,\sum|Y) = &\frac{1}{(2\pi)^{T/2}|\sum|^{1/2}} \\ &\exp(-\frac{1}{2}(Y-f(X,\alpha))'(\sum)^{-1}(Y-f(X,\alpha)))\end{aligned} \tag{5-6}$$

达到极大值，便可得到 α 与 $\sum$ 的极大似然估计。可以证明，这种估计是一致、渐近正态且渐近有效率的估计（见 $Bunke, 1980, pp.609-611$）。

若假定3不满足，则 α 的 **OLS** 估计是有偏的。假定3是一个无法检验的假定，这就必须在建军模时保证：

(1) 模型中已包含了所有重要的解释变量。被略去的变量不但不是不重要的，而且其对因变量的影响是正负等可能的，是可以相互抵消的。

(2) 因变量不能存在系统的测量正误差或负误差。

这两项要求难以满足，使得模型在抽象过程中偏离现实。

相应的补救措施是很简单的，只要对 Y 作平移变换：以 $Y-\omega$ 代替 Y 即可。

非正态性误差的原因与后果 在下面几种情形中，正态误差的假定将不成立：

(1) 经济变量的本质属性，使观察数据不服从正态分布。如价格服从特殊的 $Pareto$ 分布。

(2) 因变量理论服从正态分布，但实际中：

- 模型设定不正确，导致模型的误差反映在非正态残差中；
- 观察样本太小或异常值($outliers$)，以致观察数据不能反映真实模型的正态性；
- 实施了数据变换，使原来的正态误差变为非正态误差。

在大样本情形下，非正态误差所导致的后果显得不太严重。

$$H_0: \epsilon_t \quad \text{来自于正态总体} \quad t = 1, 2, \cdots$$

W 检验（Shapiro,S.S. & Wilk, M. B.,1965）

$$\mathrm{W} = \frac{(\sum_{i=1}^{T/2}\alpha_{T-i+1}(\hat{e}_{T-i+1}-\hat{e}_i))^2}{\sum_{i=1}^{T}(\hat{e}_i-\bar{e})^2} \tag{5-7}$$

非线性单方程计量经济学模型概述

解释变量非线性 如需求函数模型：

$$\frac{1}{Q} = \alpha + \beta\frac{1}{P} + \mu \tag{5-8}$$

可线性化的参数非线性 如不变替代弹性 CES 函数：

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} \tag{5-9}$$

不可线性化的参数非线性 如：

$$Y_i = f(X_i, B) + u_i \tag{5-10}$$

非线性普通最小二乘法

普通最小二乘原理

残差平方和是参数估计量的函数

高斯—牛顿迭代法 迭代从

$$S(\hat{\beta}) = \sum_{i=1}^{n} \{y_i - f(x_i, \beta)\}^2 \tag{5-11}$$

开始的。

对 $f(x_i, \beta)$ 展开；取一阶近似值。**牛顿—拉弗森迭代法** 对 $S(\hat{\beta})$ 展开；取二阶近似值。**例子**

例 5.1 (农民收入影响因素分析) 下面分别用线性化后的普通最小二乘法和非线性普通最小二乘法进行一个实际模型的估计。模型的目的是分析农民收入的增长是由哪些因素决定的，以及各个因素的贡献，进行研究提高农民收入的措施。经过反复模拟，剔除从直观上看可能对农民收入产生影响但实际上并不显著的变量后，得到如下结论：改革开放以来，影响我国农民收入总量水平的主要因素是从事非农产业的农村劳动者人数、农副产品收购价格和农业生产的发展规模。用 I 表示农民纯收入总量水平、 Q 表示农业生产的发展规模、 P 表示农副产品收购价格、 L 表示从事非农产业的农村劳动者人数。收入采用当年价格；农业生产的发展规模以按可比价格计算的、包括种植业、林业、牧业、副业和渔业的农业总产值指数为样本数据；农副产品收购价格以价格指数为样本数据。

相关数据如下：

源代码 5.2 农民收入数据集

```
data countrymen;
    input year I Q P L @@;
    cards;
1978    62.45   100.0   100.0   31.52 1979    79.30   107.5   122.1
31.90 1980    96.50   109.0   130.8   35.02 1981    107.65  115.3
138.5   36.92 1982    120.80  128.4   141.5   38.05 1983    142.40
138.4   147.8   43.40 1984    185.85  155.4   153.7   58.88 1985
238.70  160.7   166.9   67.13 1986    285.52  166.1   177.6   75.22
1987    343.80  175.8   198.9   81.30 1988    442.60  182.6   244.6
86.11 1989    495.30  188.3   281.3   84.98 1990    524.66  202.6
274.0   86.74 1991    559.30  210.1   268.4   89.06 1992    613.66
223.5   277.5   97.65 1993    743.49  241.0   314.7   109.98 1994
979.39  261.7   440.3   119.64 1995    1271.16 290.2   527.9 127.07
1996    1567.33 317.5   550.1   130.28 1997    1721.71 333.7 525.3
135.27 ; run;
```

建立如下数理模型：

$$I = AQ^{\alpha_1}P^{\alpha_2}L^{\alpha_3} \tag{5-12}$$

将数据对数化建模:直接应用普通最小二乘法： 从 $DW = 0.895$ 值可以判断模型存在一阶序列相关。将模型线性化用广义差分法（克服序列相关）进行估计：

5.4 变参数线性计量经济学模型

定义 5.2 (**常参数模型**) 在经典计量模型中，参数被认为在样本期内是常数，即经济结构保持不变，解释变量对被解释变量的影响不变，我们称之为常参数模型。

变参数模型一般形式是：

$$y_t = \alpha_t + \beta_t x_t + u_t \tag{5-13}$$

确定性变参数模型 参数是变量，但不是随机变量，而是确定性变量。**参数随某一个变量呈规律性变化** 可将描述式代入原方程，用 OLS 法估计。**参数作间断性变化** 在实际经济问题中，往往用来表示某项政策带来的影响。

$$\alpha_t = \alpha_0 + \alpha_1 p_t$$
$$\beta_t = \beta_0 + \beta_1 p_t$$

源代码 5.3 农民收入估计

```
proc reg data=countrymen;
    model log(I) = log(Q) log(P) log(L) /dw r p;
    output out=temp r=residual p=yhat;
run;

proc autoreg data=countrymen;
    model I = Q P L /nlag=2;
run;
```

当

$$\begin{cases} 1 \le t \le n_0 \ p_t = 0 \\ \quad t > n_0 \quad p_t = 1 \end{cases}$$

第一种情况，n_0 已知。

估计方法 5.1 (Chow **方法**)　由 $Chow$ 1960 提出。分段建立模型，分段估计模型。

改写(5-13)为

$$\begin{aligned} y_t = \alpha_0 + \beta_0 x_t + \mu_{1t} \qquad\quad & t = 1, \cdots, n_0 \\ y_t = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1) x_t + \mu_{2t} \ & t = n_0 + 1, \cdots \end{aligned} \tag{5-14}$$

估计方法 5.2 (Gujarati **方法**)　Gujarati 1970 提出。采用虚变量，建立一个统一的模型。

$$y_t = (\alpha_0 + \alpha_1 D) + (\beta_0 + \beta_1 D) x_t + \mu_t$$

一些实例表明，两种方法得到的参数估计量具有很好的一致性。

例 5.2 (**台湾收入与储蓄估计**)　1964 − 1981 年台湾个人收入和储蓄额的数据，对两种方法进行验证。采用 $Gujarati$ 方法，以 1964 − −1981 年数据为样本，估计一元线性模型:

$$y_t = \alpha_0 + \alpha_1 + (\alpha_2 - \alpha_1) D + \beta_1 X_t + (\beta_2 - \beta_1) D X_t + u_t \tag{5-15}$$

采用 $Chow$ 方法，分别以 1964 − 1972 年和 1973 − 1981 年数据为样本，估计一元线性模型。

第二种情况，n_0 未知，但 $Var(u_{it} = Var(u_{2t})$。此时，可以选择不同的 n_0，依照(5-14)进行试估计，从多次估计选择最优者。选择的标准是两段议程的残差平方和最小。

第三种情况是 n_0 未知，且 $Var(u_{it} \neq Var(u_{2t})$。此时将 n_0 视作待估参数，模型采用(5-14)形式。假设：

$$\begin{aligned} u_{1t} \sim N(0, \sigma_1^2) \\ u_{2t} \sim N(0, \sigma_2^2) \end{aligned}$$

且不存在自相关。$Goldfeld$和 $Quandt$ 于 1973 年研究并提出用最大或然法进行估计。构造关于 n_0 的对数或然函数为：

$$\begin{gathered} \ln L(\beta, \sigma^2 \,|n_0) \\ = -\tfrac{n}{2} \ln(2\pi) - n_0 \ln \sigma_1 - (n - n_0) \ln \sigma_2 \\ - \tfrac{1}{2\sigma_1^2} \sum_{t=1}^{n_0} (y_t - \alpha_0 - \beta_0 x_t)^2 - \tfrac{1}{2\sigma_2^2} \sum_{t=n_0+1}^{n} (y_t - (\alpha_0 + \alpha_1) - (\beta_0 + \beta_1) x_t)^2 \end{gathered}$$

遍取 $n_0 = 1, 2, \cdots, n$ 代入上式，取使对数似然函数值最大的 n_0 作为突变点的估计值。

随机变参数模型 参数是随机变量。**参数在一常数附近随机变化**

$$\begin{aligned} \alpha_t = \alpha_0 + \varepsilon_t \\ \beta_t = \beta_0 + \eta_t \end{aligned} \tag{5-16}$$

其中 ε_t, η_t 为零均值。于是有：

$$y = \alpha + \beta x + \omega$$

源代码 5.4 台湾收入与储蓄数据

```
data twincome;
     input income saving@@;
     datalines;
     8.8     0.36
     9.4     0.21
     10.0     0.08
     10.6     0.20
     11.0     0.10
     11.9     0.12
     12.7     0.41
     13.5     0.50
     14.3     0.43
     15.5     0.59
     16.7     0.90
     17.7      0.95
     18.6     0.82
     19.7     1.04
     21.1     1.53
     22.8     1.94
     23.9     1.75
     25.2     1.99  ;
data twincome;
     retain  year  1963;
     set  twincome;
     year+1;
     output;
run;
```

源代码 5.5 Chow method

```
data chow1;
     set  twincome;
     if  year  le  1972;
run;  proc  print;run;  data chow2;
     set  twincome;
     where year >= 1973;
run;  proc reg  data=chow1;
     mode saving=income;
run;  proc reg  data=chow2;
     mode saving=income;
run;
```

源代码 5.6 Gujarati method

```
data Gujarati;
     set twincome;
     if year le 1972 then d=1;
        else d=0;
     dincome=d*income;
     output;
run; proc reg data=gujarati;
     model saving=income d dincome/covb;
run;
```

其中

$$
\begin{aligned}
&\omega_t = \varepsilon_t + \mu_t + \eta_t x_t \\
&E\omega_t = 0 \\
&E(x_t\omega_t) \\
&= E(\varepsilon_t x_t + \mu_t x_t + \eta_t x_t^2) \\
&= 0 \\
&\mathrm{var}(\omega_t) \\
&= E(\varepsilon_t + \mu_t + \eta_t\omega_t)^2 \\
&= E(\mu_t^2) + E(\varepsilon_t^2) + E(\eta_t x_t)^2 \\
&= (2 + x_t^2)\sigma^2
\end{aligned}
$$

模型具有异方差性，而且已经推导出随机误差项的方差与解释变量之间的函数关系，所以可以采用经典线性计量经济学模型中介绍的估计方法，例如加权最小二乘法等方法很方便地估计参数。

定义 5.3 (Hildreth-Houck **模型**) 由 *Hildreth* 和 *Houck* 于 1968 年提出了如下的变参数模型：

$$
\begin{aligned}
y_t = \beta_{0t} + \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \cdots + \beta_{kt}x_{kt} + \mu_t \quad & t = 0, 1, \cdots, n \\
\beta_{jt} = \beta_j + \varepsilon_{jt}, Var(\varepsilon_{jt}) = \sigma_j^2 \quad & j = 1, \cdots, k
\end{aligned}
$$

参数随某一变量作规律性变化，同时受随机因素影响 参数可以表示为：

$$
\begin{aligned}
\alpha_t &= \alpha + \theta p_t + \varepsilon_t \\
\beta_t &= \beta + \delta p_t + \eta_t
\end{aligned}
$$

模型表示为

$$y_t = \alpha + \theta\, p_t + \beta\, x_t + \delta\, p_t x_t + \mu_t + \varepsilon_t + \eta_t + \eta_t x_t$$

是一具有异方差性的多元线性模型，也可以采用经典线性计量经济学模型中介绍的估计方法，例如加权最小二乘法等方法很方便地估计参数。

自适应回归模型

定义 5.4 (**自适应回归模型**) 如果模型

$$y_t = \alpha_t + \beta_t x_t + \mu_t$$

中的参数 α_t 可以表示为

$$
\begin{aligned}
&\alpha_t = \alpha_{t-1} + \varepsilon_{t-1} \\
&E(\varepsilon_t) = 0 \\
&Var(\varepsilon_t) = \sigma^2
\end{aligned}
$$

且

$$\beta_t = \beta$$

则称该模型为自适应回归模型。

它是由影响 α_t 的变量具有自适应回归模式所形成的。如：

$$\alpha_t = \alpha_0 + \delta p_t$$

$$p_t = \rho p_{t-1} + \varepsilon_t$$

则有

$$\alpha_t = \alpha_0 + \delta\rho p_{t-1} + \delta\varepsilon_t$$

如果 p_t 具有较高程度的自相关 $\rho \approx 1$，则

$$\alpha_t = \alpha_{t-1} + \varepsilon_t$$

就引出了自回归模型。

例如，如果它是一个消费方程，α_t 表示自发性消费（即在收入等于 0 时的消费水平），国家的消费政策（刺激、鼓励、一般或抑制的政策）使得自发性消费是一个随机变量，而国家的消费政策往往具有一阶自相关性，引起自发性消费也具有一阶自相关性。所以自适应回归模型在实际经济研究中是经常出现的。

$$\begin{aligned}
&y_t \\
&= \alpha_t + \beta x_t + \mu_t \\
&= \alpha_{t-1} + \beta x_t + \mu_t + \varepsilon_{t-1} \\
&= \alpha + \beta x_t + \omega_t \\
&\underline{\underline{\alpha_{n+1} = \alpha}} \\
&{\scriptstyle \alpha_n = \alpha - \varepsilon_n} \\
&\cdots\cdots\cdots\cdots \\
&{\scriptstyle \alpha_1 = \alpha - \sum \varepsilon}
\end{aligned}$$

$$\begin{aligned}
&Var(\mu_t) = \sigma_\mu^2 \\
&Var(\varepsilon_t) = \sigma_\varepsilon^2 \\
&Cov(\omega) = \begin{bmatrix}
1+n\lambda & (n-1)\lambda & \cdots & 3\lambda & 2\lambda & \lambda \\
(n-1)\lambda & 1+(n-1)\lambda & \cdots & 3\lambda & 2\lambda & \lambda \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
2\lambda & 2\lambda & \cdots & 2\lambda & 1+2\lambda & \lambda \\
\lambda & \lambda & \cdots & \lambda & \lambda & 1+\lambda
\end{bmatrix} \sigma_\mu^2
\end{aligned}$$

其中

$$\lambda = \sigma_\varepsilon^2 / \sigma_\mu^2$$

当 λ 已知时，容易采用广义最小二乘法（GLS）估计模型参数；如果未知，可以选择不同的值试估计，选择拟合最好者。

增长曲线模型 所谓经典计量经济学模型，专指那些在数学上采用回归分析的方法，在经济意义上揭示因果关系的经济数学模型。增长曲线模型和随机线性时间序列模型，它们的共同特点是采用回归分析的方法估计模型的参数，揭示经济变量的某种变化规律，但并不反映经济活动中的因果关系。**概述**

定义 5.5 (**增长曲线模型**) 增长曲线模型描述经济变量随时间变化的规律性，从已经发生的经济活动中寻找这种规律性，并用于未来的经济预测。但是，时间并不是经济活动变化的原因，所以增长曲线模型不属于因果关系模型。

(1) 多项式增长曲线模型

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_k t^k$$

(2) 简单指数型增长曲线模型

$$y_t = ab^t$$

(3) 修正指数型增长曲线模型

$$y_t = k + ab^t$$

(4) 逻辑($Logistic$)增长曲线模型

(5) 龚珀兹($Gompertz$)增长曲线模型。由 $B.Gompertz$ 于 1825 年提出，其数学形式为：

$$y_t = Ka^{bt}$$

Logistic 增长曲线模型 俗称“S 曲线“，由 $Verhulst$ 1845 年提出，当时提出的目的是模拟人口增长，其一般形式是：

$$y_t = \frac{K}{1 + e^{\phi(t)}}$$

其中

$$\phi(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$$

后经简化，其狭义形式是：

$$y_t = \frac{K}{1 + ae^{-bt}} \tag{5-17}$$

由(5-17)表示的曲线由两个重要特征：

- Y 在 $(0, K)$ 区间取值，K 是其饱和值；
- 曲线存在一个拐点，在此之前，增长速度越来越快；在此之后，越来越慢。

$Logistic$ 模型参数估计方法：一是线性化(5-17)

$$\frac{1}{y} = \frac{1}{K} + \frac{a}{K}e^{-bt} \tag{5-18}$$

再估计。关键在于给定 K 的值。首先根据模型的经济背景给出 K 的上下限。例如一种新产品、新技术的普及率的上限为 100%，下限是已经实现的普及率，比如 60%。然后分别以上、下限作为 K 的给定值，估计模型，计算残差平方和。接下来可以根据优选法提供的思路给定新的 K 值，反复试算，直到得到残差平方和最小时的 K 值和 $a!b$ 的估计值。

二是”三和法”。

估计方法 5.3 (Logistic **的三和法**)　首先将样本分成三段，分别估计线性化后的和；得到三个和之后依次相减，得到两个量；两量相除得到 b 的值；两量相减后再除第一量的平方得到 K, a 的估计值。

将样本分成三段

$$\begin{gathered} t = 1, 2, \cdots, r \\ t = r+1, r+2, \cdots, 2r \\ t = 2r+1, 2r+2, \cdots, n \end{gathered}$$

分别计算(5-18)的和：

$$\begin{gathered} S_1 = \sum_{t=1}^{r} \tfrac{1}{y_t} = \tfrac{r}{K} + \tfrac{a}{K}\tfrac{e^{-b}(1-e^{-rb})}{1-e^{-b}} \\ S_2 = \sum_{t=r+1}^{2r} \tfrac{1}{y_t} = \tfrac{r}{K} + \tfrac{a}{K}\tfrac{e^{-(r+1)b}(1-e^{-rb})}{1-e^{-b}} \\ S_3 = \sum_{t=2r+1}^{3r} \tfrac{1}{y_t} = \tfrac{r}{K} + \tfrac{a}{K}\tfrac{e^{-(2r+1)b}(1-e^{-rb})}{1-e^{-b}} \end{gathered}$$

令

$$D_1 = S_1 - S_2, D_2 = S_2 - S_3$$

有

$$D_1 = \frac{a}{K}\frac{e^{-b}(1-e^{-rb})^2}{1-e^{-b}}$$
$$D_2 = \frac{a}{K}\frac{e^{-(r+1)b}(1-e^{-rb})^2}{1-e^{-b}}$$

可得

$$D_1/D_2 = e^{rb}$$
$$b = \frac{1}{r}(\ln D_1 - \ln D_2)$$

当得到 y 的样本观测值后，可由上式估得 b 的值。又因为

$$D_1 - D_2 = \frac{a}{K}\frac{e^{-b}}{1-e^{-b}}(1-e^{-rb})^3$$
$$D_1^2 = \frac{a^2}{K^2}\frac{e^{-2b}}{(1-e^{-b})^2}(1-e^{-rb})^4$$
$$D_1^2\Big/(D_1 - D_2) = \frac{a}{K}\cdot\frac{e^{-b}}{1-e^{-b}}(1-e^{-rb}) = S_1 - \frac{r}{K}$$

得到

$$K = \frac{r}{S_1 - D_1^2/(D_1 - D_2)}$$
$$a = K\frac{D_1^2}{D_1 - D_2}\cdot\frac{e^b - 1}{1-e^{-rb}}$$

三是“非线性估计”。将(5-17)直接写成：

$$Z_t = \alpha + \beta\gamma^t$$

应用非线性最小二乘法估计。

例 5.3　某城市人均用电量估计。其增长曲线为一逻辑曲线。

5.5 小结

(1) 在二个或二个以上的经济变量（如消费与收入，工资与物价等）之间常常存在一种长期均衡关系。然而，就短期而言，它们可能是不均衡的。但一个时期中不均衡的部分（称为均衡误差）将在下一时期中得到校正。如在一个时期中的价格变化可能依赖于上一时期中超额需求$excessdemand$。

(2) 为了应用通常的渐近理论，我们排除了过程中有单位根的可能情况，因为单位根的出现意味着现在的一个冲击会有长久不衰的影响。

源代码 5.7 Logistic人均用电量

```
data electricity ;
    input perele @@;
    datalines;
    330 359 405 456
    484 522 554 586
    652 764 848 930
    1035 1113 1150
    1227 1341 1376
    1468 1513 1548
    1601 1609 1641
    1672 1695 1710
    ;
data electricity ;
    retain year 1970 t 0;
    set electricity ;
    year+1;
    t+1;
    output;
run;
proc model data=electricity;
    perele=K/(1+a*(exp(-b*t)));
    fit perele;
run;
```

第六章 联立方程计量经济学模型理论与方法

6.1 前言

经济模型表达了所得到的有关正在研究的系统的所有信息，具体在，应包括以下内容：

- 经济变量的分类
- 列入方程的各变量
- 任何可能涉及的滞后
- 有关单一参数或参数组合的非抽样信息
- 应设立多少个方程，应如何闭合这个方程组使之完备

这个系统的方程便是结构方程。**联立方程的概念** 联立方程模型 $simultaneous-equationsmodel$：对于实际经济问题，描述变量间联立依存性的方程体系。

联立方程模型最大的问题是：

$$E(X`u) \neq 0$$

当用 OLS 法估计参数时会产生联立性偏倚，即参数的 OLS 估计量有偏、不一致。

联立方程中变量与模型 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。

定义 6.1 (内生变量) 内生变量（$endogenousvariable$）：由模型内变量所决定的变量。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。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。

定义 6.2 (外生变量) 外生变量（$exogenousvariable$）：由模型外变量所决定的变量。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。

定义 6.3 (前定变量) 前定变量（$predeterminedvariable$）：包括外生变量、外生滞后变量(Lagged Endogenous Variables)、内生滞后变量。它只能作解释变量。

定义 6.4 (结构模型) 结构模型（$structuralmodel$）：把内生变量表述为其他内生变量、前定变量与随机误差项的方程体系。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。

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。

定义 6.5 (简化式模型) 简化型模型（$reduced-formequations$）：把内生变量只表示为前定变量与随机误差项函数的联立模型。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。

定义 6.6 (递归模型) 递归模型（recursive system）：在结构方程体系中每个内生变量只是前定变量和比其序号低的内生变量的函数。

6.2 联立方程组模型Simultaneous Equations Models

在单一方程中，解释变量是“因”，被解释变量是“果”，这种因果关系是单向的“一因一果”或“多因多果”关系。

现实的经济变量的关系是双向的，即一个经济变量影响着另一个（些）经济变量，反过来又受它（们）的影响。如：消费与收入。

联立方程组是由多个互相联系的单一方程构成的，从而所表达的因果关系是双向的，或一果多因，或一因多果。

$I.Walras$ 的经济体制概念，将经济关系描述为一个联立方程组。对这个体制的观察所得的数据一般都是非实验性的，具有不可能完全控制的性质。

联立方程模型的表述 经济模型表达了所得到的有关正在研究的系统的所有信息，具体地，应包括：

- 经济变量的分类
- 列入方程的各变量
- 任何可能涉及的滞后
- 有关单一参数或参数组合的非抽样信息
- 应设立多少个方程，应如何闭合这个方程组或使之完备。

这个系统的方程被称为**结构方程**，结构方程中的参数称为**结构参数**。

若这个方程组中方程的个数= 内生变量的个数，则称为**完备的** （complete)。

结构方程的类型包括：

- 随机、含参数的行为方程与技术方程；
- 非随机、不含参数的制度方程与恒等式。

有些模型真正具有联立的特性，即需要同时求解；而另一些模型并不真正具有联立的特性，即不必同时求解。依此，将联立方程模型区分为：结构式模型($structuralform$)与简约式模型($reducedform$)。递归模型($recursivemodel$)、分段递归模型与似乎不相关模型($seeminglyunrelatedmodel$)是特殊的结构式模型。

结构式模型与简约式模型的参数之间的关系是否有唯一的表示问题，是联立方程模型的最重要的一个问题。

结构式模型就是在一定经济理论基础上建立的，能够反映经济变量之间结构形式的一类联立方程模型。

$$Y_t = BY_t + \Gamma X_t + \varepsilon_t$$

当 B 为一个零矩阵时，对应的结构式模型为似乎不相关模型；当为 B 一个下三角矩阵时，对应的结构式模型为递归模型。

结构式模型具有以下特点：

(1) 每个随机方程的内生变量，都可以表为其它内生变量、前定变量和随机扰动项的函数；

(2) 结构参数表示每个解释变量对因变量的直接影响，一般来说就是模型要素的边际倾向或弹性，应该具有一定的经济意义；

(3) 有些结构参数为零，这是对这一参数所加的零约束条件；

(4) 模型是完备的，即内生变量个数等于方程个数；

(5) 模型的内生变量与随机扰动项相关，OLS估计量有偏。

若结构参数矩阵 $I-B$ 可逆，则可解出 Y_t，得

$$\begin{aligned} Y_t &= (I-B)^{-1}\Gamma X_t + (I-B)^{-1}\varepsilon_t \\ &= \Pi X_t + V_t \end{aligned}$$

这就是简约式模型。简约式模型就是把结构式模型中的内生变量表示为前定变量和随机扰动项的函数的联立方程模型。

联立方程模型的建立

$Phillips$ **型工资——价格确定系统**

货币工资增长率： $\dot{W}_t = \alpha_0 + \alpha_1 UN_t + \alpha_2 \dot{P}_t + u_{1t}$
价格增长率： $\dot{P}_t = \beta_0 + \beta_1 \dot{W}_t + \beta_2 \dot{R}_t + \beta_3 \dot{M}_t + u_{2t}$

$UN =$ 失业率；
$\dot{R} =$ 资本成本增长率；
$\dot{M}$=进口原材料价格增长率。

内生变量： $\dot{W}$，$\dot{P}$
外生变量： UN，$\dot{R}$，$\dot{M}$

该系统的矩阵形式是：

$$\begin{pmatrix} 1 & -\alpha_2 \\ -\beta_1 & 1 \end{pmatrix}\begin{pmatrix} \dot{W}_t \\ \dot{P}_t \end{pmatrix} + \begin{pmatrix} -\alpha_0 & -\alpha_1 & 0 & 0 \\ -\beta_0 & 0 & -\beta_2 & -\beta_3 \end{pmatrix}\begin{pmatrix} 1 \\ UN_t \\ \dot{R}_t \\ \dot{M}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

联立性偏误

定义 6.7 [联立性偏误 simultaneous bias] 在结构式模型中，每个方程所表达的一个内生变量都是其它内生变量的函数。从而，同一内生变量在一个结构方程中可能作为因变量，而在另一结构方程中却可能成为解释变量。这样便造成解释变量与随机扰动项的相关，从而违背了 OLS 法的古典假定。一般地，若此时直接运用 OLS 法来估计参数，则得到的估计将是有偏且非一致估计，我们称这种现象为联立性偏误。

成因：约简型参数是结构型参数线性表出或者非线性表出，所以约简型参数间一般可能不是独立的，甚至可能存在比较复杂的相关关系。

结构式模型的联立性偏误是由内生变量作为解释变量引起的。这种偏误与样本容量无关，并不因样本容量增大而消失。在结构式方程中，只要作为解释变量的都是外生变量，这种偏误即可消失。

与因变量同时决定的解释变量一般都与误差项相关，这就导致 OLS 中存在偏误和不一致性。如两个方程的结构模型：

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$
$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

当 y_2 与 u_1 因联立而相关时，我们就说 OLS 存在联立性偏误。

联立性检验 当存在联立性时，可以用测量误差检验方法来检验联立性。

如果有联立性，普通最小二乘估计将不是一致和有效的，而工具变量将是一致且有效的；如果没有联立性，普通最小二乘法是一致且有效的，但工具变量法（包括二阶段最小二乘法）将是一致但非有效的。

在例(3.3)中讨论的是测量误差。假设我们担心的不是测量误差，而是联立性问题。假设 AID 不仅受 PS 影响，而且也受 EXP 影响。对于这个联立模型，按以下步骤来检验变量 AID 的联立性。第一步，用 AID 对 INC, POP, PS 进行回归，计算残差变量；第二步，把残差变量加到原始回归方程中“修正“联立方程。

源代码 6.1 SAS:联立性检验

```
PROC IMPORT OUT=LI.HAUSMAN
            DATAFILE="D:\MYDOCUMENT\DATASET\PINDYCK\EX73.XLS"
                 REPLACE;
            GETNAMES YES;
RUN; PROC REG DATA=LI.HAUSMAN ;
     MODEL AID=PS ;
     OUTPUT OUT=W R=WHAT;
RUN; DATA TEMP(KEEP=WHAT POP INC AID EXP PS);
     MERGE LI.HAUSMAN W;
RUN; PROC REG DATA=TEMP;
     MODEL EXP=AID INC POP WHAT;
RUN;
```

源代码 6.2 SAS:HAUSMAN

```

 PROC MODEL DATA=TEMP OUT=B;
     ENDOGENOUS EXP AID;
     EXP = A1 + A2*AID + A3*INC + A4*POP;
     AID = B1+B2*EXP+B3* PS;
     FIT EXP AID /OLS SUR 2SLS 3SLS FIML HAUSMAN;
     INSTRUMENTS INC POP PS;
RUN; PROC PRINT DATA=B; RUN;
```

6.3 联立方程模型的识别(identification)

当我们用 *OLS* 估计一个模型时，关键的识别条件是每个解释变量都与误差项无关。这个重要的条件对 *SEM* 而言一般不同成立。

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$
$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

如果 z_1 和 z_2 都包含不同外生变量的事实意味着我们对模型施加了**排除性约束**(*exclusiontestrictions*)。**识别的定义**

- "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。"

从代数意义上讲，当与结构方程参数相与对应的简化型方程参数有一一对应关系时，结构模型是恰好识别的。

当简化型参数多于结构参数时，结构模型是过度识别的。当简化型参数少于结构参数时，结构模型是不可识别的。

识别问题是完整的联立方程模型所特有的问题。

只有行为方程才存在识别问题，对于定义方程或恒等式不存在识别问题。

识别问题不是参数估计问题，但是估计的前提。不可识别的模型则不可估计。

识别依赖于对联立方程模型中每个方程的识别。若有一个方程是不可识别的，则整个联立方程模型是不可识别的。

可识别性分为恰好识别 $JustIdentification$ 和过度识别 $Overidentification$。

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。**联立方程模型的识别** 结构式方程能反映经济系统内各经济变量之间的真实关系，并刻划、描述经济活动主体的具体行为方式；但解释变量包括内生变量，与扰动项相关，从而使 OLS 法失效。

简约式方程不能反映内生变量间的关系，无法说明诸如消费、投资等具体的经济行为方式，但解释变量只是前定变量，与扰动项无关，利用 OLS 法可得到较好的估计。

自然地，确定结构式模型参数的思路是先用估计简约式模型中的参数，然后利用两类参数间的关系，将简约式参数估计还原为结构式参数估计。这种“还原”是有条件的，就是：结构式模型可识别，即每一结构式方程可识别。

设有结构式模型：

$$\boldsymbol{Y_t} = \boldsymbol{B}\boldsymbol{Y_t} + \Gamma\boldsymbol{X_t} + \boldsymbol{\epsilon}_t \tag{6-1}$$

对应的简约式模型为

$$\boldsymbol{Y}_t = \Pi\boldsymbol{X_t} - \boldsymbol{V}_t \tag{6-2}$$

并具有参数关系式体系

$$\boldsymbol{\Pi} = (\boldsymbol{I}-\boldsymbol{B})^{-1}\Gamma \quad 或 \quad (\boldsymbol{I}-\boldsymbol{B})\Pi = \Gamma \tag{6-3}$$

识别方法

阶条件（order condition） 不包含在待识别方程中的变量（被斥变量）个数 $\geq$ 联立方程模型中的方程个数 $-$ 1。阶条件是必要条件但不充分，即不满足阶条件是不可识别的，但满足了阶条件也不一定是可识别的。

秩条件（rank condition） 待识别方程的被斥变量系数矩阵的秩= 联立方程模型中方程个数 $-$ 1。秩条件是充分必要条件。满足秩条件能保证联立方程模型内每个方程都有别于其他方程。

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。

识别两方程联立方程模型中第一个方程的充要条件是，第二个方程中至少包含第一个方程所排除的外生变量中的一个（具有非零系数）。

结构式方程的识别问题 一是从结构参数的求解情况，即（6-1）的解，来定义可识别性：一个结构式方程称为可识别($identifiable$)，是指该方程的全部结构式系数可以从参数关系式体系(6-3) 求出。进一步，若这个解是唯一的，则称该方程恰好识别($exactlyidentification$)，否则，称为过度识别($over-identificaiton$)。

另一是从结构式方程的统计形式是否确定来定义可识别性：若要识别的结构式方程具有确定的统计形式（即：模型中，不能通过部分或全部的线性组合导出一个与被识别方程有完全相同的变量的方程），则称该结构式方程可识别；否则，称为不可识别。

对于恒等式和制度方程，由于不含待估计的结构参数，所以不存在识别问题。

(1) 如果方程组中某方程不是以唯一的统计形式表示的，则该方程不可识别。

(2) 不可识别的方程是不可能用任何经济计量方法估计其系数的。可识别的方程总可以用某种经济计量方法给以估计。

(3) 识别的真实含义是能够由简约式参数确定结构式参数。

(4) 多重共线性的实质，就是模型不可识别的一种特殊表现。可识别性和无多重共线性都是模型估计所必不可少的前提。

可识别性的基本条件 对某一要识别的结构方程，如果将模型中全部变量的集合 $\mathbf{X}$ 分解为两部分：

$$\mathbf{X} = \mathbf{X}_1 \bigcup \mathbf{X}_2$$

其中 $\mathbf{X}_1$ 是出现在待识别方程中的变量的集合，而 $\mathbf{X}_2$ 是其余变量的集合。这样，可识别的第二种定义：若通过模型中其它方程式的线性组合得到一个以 $\mathbf{X}_1$ 为变量集的方程，或同时消去 $\mathbf{X}_2$ 中的所有变量的话，则所讨论的方程式是不可识别的；否则是可识别的。

能否通过模型中其它方程式的线性组合同时消去 $\mathbf{X}_2$ 中的所有变量，可以结合线性方程组中的 $\mathbf{X}_2$ 对应的系数行列式为零的条件来判断。

由这个基本条件，可以较容易地判断一个结构方程是否可识别，但不能确定可识别的类型。因此，需要利用阶条件与秩条件。

可识别性的阶条件与秩条件

两方程情形下，识别联立方程中第一个方程的充要条件是：第二个方程中至少包含第一个方程所排除的外生变量中的一个（具有非零系数）；多方程情形下，对任何一个 SEM 中的一个方程，如果它排除的外生变量数不少于其右端包含的内生变量数，则它就满足识别的阶条件。注意，上述这个条件是必要条件，但非充分条件。

方程识别的阶条件：我们需要排斥的外生变量至少与结构方程中包括的内生解释变量一样多。

识别的阶条件：

- 对于任何一个 SEM 中的一个方程，如果它排除的外生变量数不少于其右端包含的内生变量数，那么它就满足识别的阶条件
- 阶条件只是识别的必要条件而非充分条件
- 过度识别、恰好识别和不能识别

记 $card(A)$ 为集合 A 元素的个数，则对任一待识别的结构式方程，成立：

$$\mathrm{card}(X_2) \geq g-1, \quad \mathrm{g} = \text{联立方程模型中方程的个数} \tag{6-4}$$

假定 $card(X_1) = M$，$card(X_2) = N$，所讨论的方程式以外的 $g-1$ 个方程式可写为：

$$\boldsymbol{RZ_{1t} + SZ_{2t} = \omega_t}$$

不可识别性的充分条件为以下条件之一

- $\mathbf{N} < g-1$。
- $\mathbf{N} \geq g-1$ 且 $rank(\mathbf{S}) = \ g-1$。

可识别性的充分条件为：

$$\boldsymbol{N} \geq g-1, \quad \mathrm{rank}(\boldsymbol{S}) = g-1 \tag{6-5}$$

建模过程中识别性的保证 “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。

不可识别的补救措施

- 对结构参数值施加约束。某个方程的识别是以该方程被斥（不出现的）变量为基础的，可识别方程必须比模型的其它方程式少一个或更多重要的变量。
- 对结构参数之间的关系施加约束。若能够通过经济理论和其它渠道认识参数间的关系，就可以识别方程。
- 利用额外估计量。
- 对方程中随机扰动项方差或协方差施加约束，这些约束通常可根据研究者对经济现象的理解而得出。

非线性模型的识别

$$f_h(\boldsymbol{Y}_i, \boldsymbol{Z}\boldsymbol{X}_i; \boldsymbol{\alpha}) = \boldsymbol{\epsilon}_{ih}, \quad h = 1, 2, \cdots, g; \quad i = 1, 2, \cdots, T \tag{6-6}$$

其中有 g 个内生变量， k 个外生变量。对随机扰动项作零均值、同方差假定。关于参数非线性的模型识别。对线性模型，模型可识别意味着关于方程的先验性及观察性的约束将得到关于结构参数 α 的唯一解。从而，若线性模型是可识别的，则模型将全局可识别。对非线性模型，解不是唯一的，其中的一个解可以对应于真实结构，使得不需要附加信息，就能指导这些解加以区别。故非线性模型的可识别只具有局部意义。

联立方程模型的估计 单一方程法，双称有限信息法，是对联立方程模型中单个结构方程逐个地进行估计的方法，包括：间接最小二乘法，二阶段最小二乘法，工具变量法与有限信息极大似然法。这类方法的特点是：只利用与被估计方程有关的信息，而不涉及其它方程的信息。

系统估计法，双称为方程体系法或完全信息法，是对整个联立方程模型中的所有结构方程同时地进行估计，一次估计出其中全部参数的方法：包括：三段最小二乘法和完全信息极大似然法。

可以由 SAS/ETS 的 $SYSLIN$ 过程实现。

6.4 递归模型与似乎不相关模型

递归模型 或称三角形模型或因果模型，是一种貌似联立方程组，但对内生变量不必同时求解，可以顺序地逐一求解的模型。严格地，若模型的结构式

$$Y_t = BY_t + \Gamma X_t + \varepsilon_t \tag{6-7}$$

满足如下条件：

(1) B 是主对角元全为零的下三角阵

(2) ε_t 协差阵为对角阵

则称该模型是递归模型。

前一方程的内生变量，对后一方程而言是前定变量；而后一方程的内生变量，对前一方程没有影响，显示出一种单向的因果关系，因而可以应用 OLS 法。

递归方程具有如下良好性质：

(1) 恰好识别

(2) 适用 OLS 法，但不再具有线性特性。

一个结构式联立方程模型若其结构参数矩阵能写成分块的主对角元为零下三角阵的形式则称为分块递归模型，是指整个方程组分为若干段，段内方程式具有联立性质，段间具有递归性质。

似乎不相关模型与$Zeller$ 估计方法 似乎不相关模型是递归模型的一种特例：每个内生变量均由不同的结构参数和前定变量表示，即在(6-7)中， $B=0$。

当误差项协差阵为对角矩阵，即各方程的扰动项互不相关时，可以直接应用 OLS 法进行估计，得到一致的、最小方差的、无偏的但不是线性的估计；否则，需要应用 $Zellner$ 估计方法，其要点是：

(1) 将 SUR 模型用矩阵表示为一个单一方程式；

(2) 由于方程组中各个方程式的误差项之间有可能有相关关系，故使用广义最小二乘法进行估计。

(3) 扰动项满足：1）对同一次观测来说，跨方程式的扰动项存在相关；2）对不同次观测来说，扰动项是不相关的。

6.5 联立方程模型的估计方法 Estimation Methods

单方程估计方法与系统估计方法 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。

对于结构模型有两种估计方法。一种为单一方程估计法，即有限信息估计法；另一种为方程组估计法，系统估计法，即完全信息估计法。前者只考虑被估计方程的参数约束问题，而不过多地考虑方程组中其他方程所施加的参数约束，因此称为有限信息估计方法。后者在估计模型中的所有方程的同时，要考虑由于略去或缺少某些变量而对每个方程所施加的参数约束。因此称为完全信息估计法。

间接最小二乘法 ILS 法只适用于恰好识别模型。其基本思想是：由导出的简约式模型，再对每一简约式方程运用 OLS 法进行估计，然后利用“参数关系式体系”还原出结构参数的估计值。具体估计步骤是先写出与结构模型相对应的简化型模型，然后利用 OLS 法估计简化型模型参数。因为简化型模型参数与结构模型参数存在一一对应关系，利用

$$\Pi=A^{-1}B$$

可得到结构参数的唯一估计值。

ILS 估计量是有偏的，但具有一致性和渐近有效性。当结构方程为过度识别时，其相应简化型方程参数的 OLS 估计量是有偏的，不一致的。采用 ILS 法时，简化型模型的随机项必须满足 OLS 法的假定条件。当不满足时，简化型参数的估计误差就会传播到结构参数中去。

对该法的运用，需作下列假定：

(1) 被估计的结构方程必须是恰当识别的；

(2) 每个简约式方程的随机扰动项应满足最小二乘法的古典假定；

(3) 前定变量之间不存在多重共线性。

具体的计算步骤如下：

(1) 识别；

(2) 将结构式模型转换成简约式模型，即把所有内生变量表为前定变量的函数；

(3) 对每个简约式方程应用 OLS，得到简约式系数的估计值；

(4) 由简约式系数的估计值确定结构参数的估计值。

所得到的估计值具有如下的性质：

(1) 简约式参数估计是一致最优线无偏估计

(2) 结构式参数的估计一般是有偏的，但具有一致性。

ILS 法所考虑的结构式必须是恰好可识别的，即所估计的方程中不包括的前定变量个数正好等于该方程中的内生变量（包括被解释变量）的个数减 1。对于过度识别的方程式，由于结构式参数的估计不能唯一地由简约式参数估计提供，该方法失效。

工具变量法 工具变量应满足：

(1) 与被替代的内生变量高度相关；

(2) 是联立方程组中真正的外生变量，因而与结构方程中的扰动项不相关；

(3) 同一方程中使用多于一个工具变量时，要求工具变量之间的相关程度很小；

(4) 应有明确的经济意义。

IV 法的计算步骤是：

(1) 选择适当的工具变量以代替在结构方程左边出现的作为解释变量的内生变量；

(2) 用已选定的工具变量去乘结构方程，并对 n 次观测求和，得到方程个数与未知结构参数个数相同的一个线性方程组；

(3) 解这个线性方程组，求得结构参数的估计值。

IV 法的缺点是：

(1) 工具变量的选择具有任意性，因而使得估计结果也具有任意性；

(2) 只考虑了部分外生变量对相应内生变量的影响；

(3) 实际中，内生变量与许多前定变量相关，工具变量的选择相当困难，只能凭经验加以决断；

(4) 由于扰动项是不可观测的，难以判断所选择的工具变量是否与扰动项相关。

IV 法的基本思想是：以适当的外生变量为工具变量代替作为解释变量的内生变量，以减少解释变量与扰动项的相关性，然后运用 OLS 法进行参数估计。估计一般是有偏的，但具有一致性。**二阶段最小二乘法** $2sLS$ 是 ILS 和 IV 法的理论推广，由 $H.Theil$ 与 $L.Basman$ 于二十世纪五十年代分别独立提出来的，已公认为估计过度识别模型的单方程估计法中最重要的方法。该方法的目的在于：尽可能地清除联立方程的偏误。直观地说，两阶段最小二乘估计法的第一阶段应该是创造一个工具变量，第二阶段则工具变量估计法的某种变形。

应用该方法必须满足以下条件：

(1) 原始结构方程中的误差项为零均值、常数协方差且序列不相关；

(2) 原始结构方程中的所有前定变量同误差序列不相关；

(3) 前定变量之间不存在渐近的多重共线性，即

$$\lim \frac{1}{T}X'X \geqslant 0$$

(4) 为了估计简约式模型的系数，样本容量必须足够大，至少大于等于方程出现的前定变量的个数；

(5) 所考虑的结构方程式应是可识别的。

二阶段最小二乘法的使用步骤是：

(1) 用 OLS 法依次就每一内生变量取关于所有前定变量的回归的得拟合值或预测值

(2) 对作为解释变量的内生变量，以第一步中所得的拟合值代替实际观测值，然后按每一结构方程用 OLS 法进行参数估计，所得的估计值即为 $2SLS$ 估计值。

例 6.1 简单宏观经济模型：

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{t1} \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\ Y_t = I_t + C_t + G_t \end{cases} \tag{6-8}$$

这个模型中消费方程是恰好识别的，投资方程是过度识别的。以我国 1978 – 1996 年的数据进行估计。

源代码 6.3 SAS:2SLS

```
libname li 'd:\Mydocument\dataset\sas'; data temp;
    set li.macroeco;
    CL=lag(C);
    if _N_>1;
    output;
run; proc model data=temp;
    C = a0 + a1*Y + a2*CL ;
    fit C /2SLS FSRSQ;
    instruments Y G CL;
run;
```

三阶段最小二乘法 $3sLS$ 法是由 $H.Theil$ 与 $A.Zeller$ 于 1962 年提出来的一种最重要的系统估计方法，在计算方法上，它是在 $2sLS$ 的基础上，应用广义最小二乘法，如他们所证明，它比 $2sLS$ 估计有更好的渐近有效性。

采用 $3sLS$ 法的出发点在于，$2sLS$ 法只使用了模型的部分信息，忽视了模型结构对其它方程的参数值所施加的全部约束条件。特别是，当联立方程模型各方程的随机扰动项同期相关时，$2sLS$ 不再有效，此时需要引入 GLS 以克服各方程之间的联立性偏误。

应用 $3sLS$ 法需满足以下条件：

(1) 每个结构方程都是可识别的

(2) 全部方程式均已用代换方法消去

(3) 模型中所有结构方程都正确设定。不仅要正确设定各个方程中的变量，而且要正确设定各个方程的数学形式

(4) 每个结构方程的扰动项具有零均值、同方差性及无自相关性

(5) 不同结构方程的扰动项是同期相关的，如不是，则 $3sLS$ 法简化为 $2sLS$ 法

在这些假定下，计算出来的估计量是有偏但一致的。

$3SLS$ 是先用 $2SLS$ 估计每个方程，然后再对整个方程组运用广义最小二乘法。

(1) 在第一阶段，先估计联立模型的简化形式；然后，用内生变量的拟合值得到方程组中所有方程的 $2SLS$ 估计。

(2) 一旦计算出 $2SLS$ 的参数，每个方程的残差值就可以用来估计跨方程的方差和协方差。

(3) 计算广义最小二乘法的估计量。

设所考虑的结构式方程的矩阵形式是

$$\begin{aligned}
\underset{T\times 1}{y_i} &= \underset{T\times(g_i-1)\times 1}{Y_i b_i} + \underset{T\times k_i\times 1}{X_i\gamma_i} + \underset{T\times 1}{\varepsilon_i} \\
&= \begin{pmatrix} Y_i & X_i \end{pmatrix}\begin{pmatrix} b_i \\ \gamma_i \end{pmatrix} + \varepsilon_i \\
&= Z_i\delta_i + \varepsilon_i
\end{aligned}$$

$$X = \begin{pmatrix} X_1 & X_2 & \cdots & X_g \end{pmatrix}$$

$$y = X' \otimes \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_g \end{pmatrix}$$

$$Z = X' \otimes \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_g \end{pmatrix}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_g \end{pmatrix}$$

$3sLS$ 的计算如下：

(1) 运用 $2sLS$ 法估计模型中的参数 δ_i 并计算残差向量 $\hat{\varepsilon}_i$

$$\hat{\delta}_i = (Z_i'X(X'X)^{-1}X'Z_i)^{-1}Z_i'X(X'X)^{-1}X'y_i$$

$$\hat{\varepsilon}_i = y_i - Z_i\hat{\delta}_i$$

(2) 计算 $\hat{\Sigma}$ 和 $\hat{R}$

$$\hat{\Sigma} = \frac{1}{T}\begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_g \end{pmatrix}$$

$$\hat{R} = \hat{\Sigma} \otimes (X'X)$$

(3) 计算模型

$$y = Z\delta + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

的参数的 GLS 估计

$$\hat{\delta} = (Z'\hat{R}^{-1}Z)^{-1}Z'\hat{R}y$$

在应用三阶段最小二乘法时，值得注意的是：

- 它在二阶段最小二乘法的基础上进行的，第三阶段所用广义最小二乘法并不能消除估计量的有偏性，只是由于使用了模型系统的全部信息而比二阶段最小二乘法更为有效，估计的方差较小。
- 三阶段最小二乘法要同时估计模型的全部参数，当模型的变量较多时，需要的样本观测值很多。因此，在样本容量很小时，不宜运用。
- 它是一种系统估计法，要求整个模型的所有方程都必须是正确设定的。如果模型中有一个方程的设定不正确，则由此产生的设定误差将会影响整个模型。由此可见，它对设定误差是极敏感的。

似无关回归估计方法 SUR 又称联合广义最小二乘法或 $ZELLER$ 回归。

定义 6.8 (**似无关回归模型**) 似无关回归模型($seeminglyunrelatedregression$，$SUR$)是一种递归模型，它所包含的一系列内生变量被作为一组处理，因为理论上它们彼此之间有密切的联系。它由一系列方程组成，这些方程因误差项之间的跨方程相关而互相联系。

当方程系统是联立的时，两阶段最小二乘法和工具变量法可以产生一致的参数估计量，但一般地来说，它们都不能得到有效的估计量：

- 一是因为这两种方法都只是应用于联立方程模型中的一个方程，尽管它们考虑到了被除数估计方程省略了一个或多个先决变量的事实，但却没有注意到其他方程可能也省略了先决变量；
- 二是它们没有考虑到误差项之间的跨方程相关。

把 SUR 估计法推而广之则是三阶段最小二乘法 $3SLS$。

SUR 估计法只是广义最小二乘法在一组似无关方程组中的的简单运用。这些方程通过不同方程的误差项在某一时刻的非零协方差而联系在一起：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_G \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_G \end{bmatrix} \tag{6-9}$$

根据似无关模型的假设，方程内没有自相关，而方程间存在着相关，即

$$E(u_i u_i') = \begin{bmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{bmatrix} = \sigma_{ij} I \tag{6-10}$$

这种关系适用于模型里 G 个方程中任意两个方程之间的协方差：

$$\begin{aligned} \Omega &= E(uu') \\ &= \begin{bmatrix} E(u_1u_1') & E(u_1u_2') & \cdots & E(u_1u_G') \\ E(u_2u_1') & E(u_2u_2') & \dots & E(u_2u_G') \\ \vdots & \vdots & \ddots & \vdots \\ E(u_Gu_1') & E(u_Gu_2') & \cdots & E(u_Gu_G') \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & \cdots & \sigma_{1G}I \\ \sigma_{21}I & \sigma_{22}I & \dots & \sigma_{2G}I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1}I & \sigma_{G2}I & \cdots & \sigma_{GG}I \end{bmatrix} \end{aligned} \tag{6-11}$$

所有有关误差协方差的信息全都包含在矩阵 Ω 里。模型最有效的估计是由广义最小二乘法得到的：

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}) \tag{6-12}$$

以及

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1} \tag{6-13}$$

实际上，Ω 的元素也需要估计。通过对 G 个方程中的每一个方程使用 OLS 法得到其残差估计值可以完成这个估计：

$$\hat{\sigma}_{ij} = \frac{\hat{u}_i \hat{u}_j'}{\sqrt{(N-K_i)(N-K_j)}} \tag{6-14}$$

运用简单的矩阵代数可以证明：当对所有的 $i, j i \neq j, \sigma_{ij} = 0$；当每一个方程中出现的自变量完全相同时；$SUR$ 估计法等价于逐个对方程使用 OLS 估计。注意，SUR 方法需要 $\sum$ 阵的估计，这会对小样本情况增加估计的抽样变化性($samplingvariability$)。

源代码 6.4 SAS:SUR

```
libname li 'd:\Mydocument\dataset\sas'; data temp;
        set li.macroeco;
        CL=lag(C);
        if _N_>1;
        output;
run; proc model data=temp;
        C = a0 + a1*Y + a2*CL ;
        I = b0 + b1*Y;
```

```
      Y = C + I + G;
      fit  C I /sur covs;
run;
```

有限信息最大似然法 要点 只利用结构式模型中特定方程式的信息，而不利用其它结构式方程的信息。具体地，就是对特定结构方程式，利用方程中各有关变量的观测值，求出似然函数，继而确定似然函数的最大值及相应参数的最大似然估计。**步骤** 设所考虑的结构式方程的矩阵形式（由观测值所作成）为

$$y_1 = Y_2 b + X_1\gamma + \varepsilon_1 \text{或} (y_1 Y_2)\begin{pmatrix} 1 \\ -b \end{pmatrix} = X_1\gamma + \varepsilon_1 \tag{6-15}$$

(1) 运用 OLS 法，就全部内生变量依次取关于所考虑的结构式方程中的所有的外生变量的线性回归，得 OLS 残差向量，以此为列构成协差矩阵 M_1:

$$M_1 = (y_1Y_2)'(I - X_1(X_1'X_1)^{-1}X_1')(y_1Y_2)$$

(2) 运用 OLS 法，就全部内生变量依次取关于所有外生变量的线性回归，得 OLS 残差向量，以此为列构成协方差矩阵 M：

$$M = (y_1Y_2)'(I - X(X'X)^{-1}X')(y_1Y_2)$$

(3) 求广义特征问题

$$(M_1 - \lambda M)\beta = 0$$

的最小特征值 λ_1，及其对应的特征向量 $\hat{\beta}_1$，满足

$$\hat{\beta}_1 = \begin{pmatrix} 1 \\ -\hat{b} \end{pmatrix} = \hat{\beta}$$

即 $\hat{b}$ 是 b 的 $LIML$ 估计；

(4) 计算 γ 的 $LIML$ 估计，满足

$$\begin{pmatrix} \hat{\gamma} \\ 0 \end{pmatrix} = (X'X)^{-1}X'Y\hat{\beta}$$

由 $LIML$ 法所得的估计具有下面的性质：

- 当结构式为恰好识别时，它等价于 ILS 法。
- 在大样本情形，它是一致估计。
- $E(\hat{b}_{LIML}) = \infty$

例 6.2　设有模型

$$\begin{cases} Y_{1t} = b_{12}Y_{2t} + \gamma_{11}X_{1t} + \varepsilon_{1t} \\ Y_{2t} = b_{21}Y_{1t} + \gamma_{22}X_{2t} + \gamma_{23}X_{3t} + \varepsilon_{2t} \end{cases}$$

$\varepsilon_t = (\varepsilon_{1t}\varepsilon_{2t})'$ 是零均值，具有同协方差矩阵的独立正态误差向量序列。若变量观测值的二阶矩矩阵为

$$Y'Y = \begin{pmatrix} 14 & 6 \\ 6 & 10 \end{pmatrix}$$

$$Y'X = \begin{pmatrix} 2 & 3 & 0 \\ 2 & 1 & 0 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

求第一个方程中的参数 b_{12}, γ_{11} 的有限信息估计与二阶段最小二乘估计与普通最小二乘估计。

运用 OLS 法，就全部内生变量依次取关于所考虑的结构式方程中的所有外生变量的线性回归，得 OLS 残差向量，以此为列构成协差矩阵 M_1。

$$\begin{aligned}
&M_1 \\
&= \begin{pmatrix} y_1 & Y_2 \end{pmatrix}' (I - X_1(X_1'X_1)^{-1}X_1') \begin{pmatrix} y_1 & Y_2 \end{pmatrix} \\
&= Y'Y - Y'x_1(x_1'x_1)^{-1}x_1'Y \\
&= \begin{pmatrix} 14 & 6 \\ 6 & 10 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \end{pmatrix} (I)^{-1} \begin{pmatrix} 2 & 2 \end{pmatrix} \\
&= \begin{pmatrix} 10 & 2 \\ 2 & 6 \end{pmatrix}
\end{aligned}$$

运用 OLS 法，就全部内生变量依次取关于所有外生变量的线性回归，得 OLS 残差向量，以此为列构成协差矩阵 M。

$$\begin{aligned}
&M \\
&= \begin{pmatrix} y_1 & Y_2 \end{pmatrix}' (I - X(X'X)^{-1}X') \begin{pmatrix} y_1 & Y_2 \end{pmatrix} \\
&= Y'(I - X(X'X)^{-1}X')Y \\
&= Y'Y - Y'X(X'X)^{-1}X'Y \\
&= \begin{pmatrix} 14 & 6 \\ 6 & 10 \end{pmatrix} - \begin{pmatrix} 2 & 3 & 0 \\ 2 & 1 & 0 \end{pmatrix} I \begin{pmatrix} 2 & 2 \\ 3 & 1 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix}
\end{aligned}$$

求广义特征问题：

$$(M_1 - \lambda M)\beta = 0$$

$$\begin{pmatrix} 10-\lambda & 2+\lambda \\ 2+\lambda & 6-5\lambda \end{pmatrix} \beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

的最小特征值得 $\lambda_{min} = 1$ 及满足 $\beta = (1 - \hat{b}_{12})'$ 的特征向量可得 $\hat{b}_{12} = 3$。再由

$$\begin{pmatrix} \hat{\gamma}_{11} \\ 0 \end{pmatrix} = (X'X)^{-1}X'Y\hat{\beta}$$

得 $\hat{\gamma}_{11} = -4$。

为计算 $2sLS$ 估计，对第一个方程中作为解释变量的内生变量 Y_2，运用 OLS 法取其关于

外生变量 X_1, X_2, X_3 的回归，得拟合值

$$\hat{y}_2 = X(X'X)^{-1}X'y_2 = X\begin{pmatrix}2\\1\\0\end{pmatrix}$$

以此代替 y_2，再应用 OLS 法便得到 $2sLS$ 估计值

$$\begin{pmatrix}\hat{b}_{12}\\\hat{\gamma}_{11}\end{pmatrix} = \begin{pmatrix}\hat{y}_2'\hat{y}_2' & \hat{y}_2'x_1\\ x_1'\hat{y}_2 & x_1'x_1\end{pmatrix}^{-1}\begin{pmatrix}\hat{y}_2'y_1\\ x_1'y_1\end{pmatrix} = \begin{pmatrix}5 & 2\\2 & 1\end{pmatrix}^{-1}\begin{pmatrix}7\\2\end{pmatrix} = \begin{pmatrix}3\\-4\end{pmatrix}$$

对第一个方程直接应用 OLS 法得到估计值

$$\begin{pmatrix}\hat{b}_{12}\\\hat{\gamma}_{11}\end{pmatrix} = \begin{pmatrix}y_2'y_2 & y_2'x_1\\ x_1'y_2 & x_1'x_1\end{pmatrix}^{-1}\begin{pmatrix}y_2'y_1\\ x_1'y_1\end{pmatrix} = \begin{pmatrix}10 & 2\\ 2 & 1\end{pmatrix}\begin{pmatrix}6\\2\end{pmatrix} = \begin{pmatrix}\frac{1}{3}\\\frac{4}{3}\end{pmatrix}$$

k– 类估计量 $H.Theil$(1958) 提出来了一类很一般地称之为 $k--$ 类估计量的有限信息估计量。$OLS, 2sLS, LIML$ 估计均是其特例。考虑第一个结构式方程：

$$y_1 = Y_1 b + X_1\gamma + \varepsilon_1 \tag{6-16}$$

设 k_1, g_1 分别为该方程中所含有的前定变量的个数、内生变量的个数，k_2 为不在方程中出现的变量（内生、前定）的个数。

$$\begin{pmatrix}\hat{b}(k)\\\hat{\gamma}(k)\end{pmatrix} = \begin{pmatrix}Y_1'Y_1 - k\hat{\varepsilon}_1'\hat{\varepsilon}_1 & Y_1'X_1\\ X_1'Y_1 & X_1'X_1\end{pmatrix}^{-1}\begin{pmatrix}Y_1' - k\hat{\varepsilon}_1'\\ X_1'\end{pmatrix}y_1$$

其中

$$\hat{\varepsilon}_1 \triangleq (I - X(X'X)^{-1}X')Y_1 \tag{6-17}$$

是 Y_1 的简约式残差。

- 当 $k = 0, 1$ 是，所确定的估计量便是 OLS 或 $2sLS$；
- 当 k 取最小特征根时为 $LIML$ 估计；
- 当 k 满足条件

  $$plimk = 1$$

  时，是一致估计。

关于 k 类估计的小样本性质，除了有偏性外，对于估计的矩，$Kinal$(1980) 给出如下研究结果：

- $2sLS$ 估计存在 $L \triangleq k_2 - (g_1 - 1)$ 阶为止的矩，其中 L 称为过度可识别度；
- 当 $0 < k < 1$ 时，k 类估计存在 $T - (k_1 + g_1 - 1)$ 阶为止的矩；
- 当 $k > 1$ 是，k 类估计的均值与各阶矩都不存在。

完全信息最大似然法 在一个联立方程模型中，如果各方程的扰动项构成一个具有零均值的多元正态分布，则结构式参数可用极大似然法求出。由于利用了模型系统的全部信息，称为完全信息极大似然法 $FIML$。

它的应用条件是：

- 模型是过度识别的；
- 研究者了解模型全部结构方程的定式。

- 需要大样本。

设所考虑的结构式模型为：

$$AY_t = \Gamma X_t + \varepsilon_t \qquad t = 1, 2 \cdots, T \tag{6-18}$$

设 ε_t 服从多元正态分布 $N(0, \Sigma)$，即有分布密度函数：

$$f(\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{gt}) = ((2\pi)^{g/2} |\Sigma|^{1/2})^{-1} exp(-\frac{1}{2}\varepsilon_t' \Sigma^{-1} \varepsilon_t) \tag{6-19}$$

得内生变量的似然函数：

$$Lik = \frac{|A|}{(2\pi)^{\frac{Tg}{2}} |\Sigma|^{\frac{T}{2}}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{T}(AY_t - \Gamma X_t)'\Sigma^{-1}(AY_t - \Gamma X_t)\right\}$$

完全信息极大似然估计量在小样本条件下是有偏的，对于大样本，估计是一致且有效的。

6.6 工具变量估计与两阶段最小二乘法

- 遗漏一个重要变量时 OLS 估计量有偏误
- 遗漏变量时，OLS 通常是非一致性的
- 对未被观测到的解释变量给出适宜的代理变量，能消除或减轻遗漏变量偏误。
- 存在不随同时间变化的遗漏变量情况下，对综列数据可用固定效应估计或一阶差分来估计随同时间变化的自变量的影响

定义 6.9　假设模型中不是包含了一个无关变量，而量遗漏了一个实际上应包括在真实（或总体）模型中的变量。这通常被称为排除一个有关变量excluding a relevant variable或对模型设定不足underspecifying the model。

潜在的真实模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

实际估计的方程：

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

$$\tilde{\beta}_1 = \frac{\sum\limits_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{\sum\limits_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

将

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

代入 $\tilde{\beta}_1$ 的表达式中并求期望可得：

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum\limits_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum\limits_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

β_2 的系数有一个简单解释：如果使用我们的自变量样本，将 x_2 对 x_1 进行回归，它就是回归方程中的斜率系数。

定义 6.10 (遗漏变量偏误) omitted variable bias:

$$E(\tilde{\beta}_1)-\beta_1=\beta_2\frac{\sum\limits_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i2}}{\sum\limits_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}=\beta_2\delta$$

一致性涉及的是一个设想的实验，考虑的是样本容量变大时发生的情况，而另一方面我们得到的是每个样本容量的大量随机样本。假定

- 对参数而言是线性的
- 随机抽样性
- 条件均值为零
- 无完全共线性

保证了 OLS 的无偏性和一致性。

$$\hat{\beta}_1=\beta_1+\frac{n^{-1}\sum\limits_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i}{\sum\limits_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}$$

我们可以在分子分母中应用大数定律，则分别在概率上收敛于总体值 $Cov(x_1,u)$ 和 $Var(u)$。利用假定三：

$$p\lim\hat{\beta}_1=\beta_1+\frac{Cov(x_1,u)}{Var(x_1)}=\beta_1$$

假定三意味着零均值和零相关

$$E(u)=0,Cov(x_j,u)=0$$

但反之不一定成立。

6.7 总结

(1) 联立方程模型分为结构式模型与简约式二种。

(2) 结构式方程的识别问题就是能否由所估计的简约式参数得到结构式参数的数值解问题。

(3) 为评价结构方程的识别性，可以应用简约式方程技术，即把内生变量唯一地表示为前定变量的函数；也可以求助于阶条件与秩条件。

(4) 联立性问题可以利用 $Hausman$ 设定检验来探测。此外该检验还可以用来确定一个或一组变量是内生的还是外生的。

(5) 对于一个真正的联立方程模型，普通最小二乘法估计既不一致也是有偏的。

(6) 两阶段最小二乘估计法和工具变量法可以提供一致的单方程参数估计量，但它们同样是有偏的。

(7) 研究表明，二阶段最小二乘法估计的方差比普通最小二乘法的大。

(8) 研究表明，系统估计法将产生比单方程估计法方差更小的估计量。

第七章　时间序列计量经济学模型理论与方法

来自[24]。

7.1　基础知识

定义 7.1 (**动态完整模型**)　设更多的滞后因变量，或设更多的滞后解释变量都无助于解释因变量的均值的时间序列模型。即，无论 x_t 包含了什么，它都包括了足够多的滞后，以至于 y 和解释变量的其他滞后对解释 y_t 都没有任何意义。

定义 7.2 (**动态模型**)　Dynamic models 考虑了变量跨时期的影响关系的模型

定义 7.3 (**分布滞后模型**)　Distributed lag models 考虑了解释变量滞后的模型

在时间序列数据回归分析中，由于不能保持在横截面数据回归分析中随机抽样的假定，我们必须研究能够限制时间序列中在**在时间上相关**$temporalcorrelation$ 的条件，以确保通常的渐近分析有效。

标注有时间的一个随机变量序列被正式地称为**随机过程**$stochasticprocess$ 或者**时间序列过程**$timeseriesprocess$。当我们搜集到一组时间序列数据时，我们得到了该随机过程的一个可能的结果或实现$realization$。

有限分布滞后模型$finitedistributedlagFDLmodel$

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \cdots + \delta_q z_{t-q} + u_t$$

静态模型作为一个特例被包含于其中。

- 即期倾向或即期乘数
- 长期倾向或长期乘数

假定 7.1 (TS.1 **对参数而言是线性的**)　对参数而言是线性的

假定 7.2 (TS.2 **条件均值为零**)　零条件均值。即时间为 t 的误差项 u_t 与任何时期的解释变量都线性无关。

当这个假定被满足时，我们称解释变量是**严格外生**$strictlyexogenous$。严格外生的解释变量无法对过去发生在 y 上的变化作出反应.政策变量，如货币供给的增长、福利开支等经常受结果变量过去的情况的影响。

定义 7.4 (**同期外生**)　$contemporaneouslyexogenous$ 若条件

$$E(u_t|x_t) = 0$$

被满足，即误差项与解释变量之间是同期不相关的。

假定 7.3 (TS.3 **无完全多重共线性**)　在样本中，并因此也在潜在的时间序列过程中，没有任何自变量是恒定不变的，或者是其他自变量的一个完全的线性组合。

定理 7.1 (OLS **的无偏性**)　当假定TS.1-3 成立时，对于给定的 X_t ，OLS 估计量是无偏的。

假定 7.4 (TS.4 **同方差性**)　给定 X_t，u_t 的方差在所有的时间 t 上都相等：$Var(u_t|X_t) = Var(u_t) = \sigma^2$

假定 7.5 (**TS.5 无序列相关**) 给定 X，任意两个不同时期的误差不相关：$Corr(u_t, u_s|X) = 0$

定理 7.2 (**OLS 的样本方差**) 在时间序列的高斯–马尔科夫假定TS.1-5成立时，$\hat{\beta}_j$ 对 X 的条件方差为$Var(\hat{\beta}_j|X) = \sigma^2/[SST_j(1 - R_j^2)]$

定理 7.3 (**$\hat{\sigma}^2$的无偏估计**) 在假定TS.1-5下，估计量$\hat{\sigma}^2 = SSR/(n - k - 1)$ 是 σ^2 的一个无偏估计量。

定理 7.4 (**高斯–马尔科夫**) 在假定TS.1-5下，给定 X 的值，OLS 估计量是最优线性无偏估计量。

假定 7.6 (**TS.6 正态性**) 误差 u_t 独立于 X，且为 $i.i.d.$。

它蕴涵了 $TS.3 - 5$，同时还假定了独立性和正态性。

定理 7.5 (**正态抽样分布**) 在时间序列的 CLM 假定 $TS.1 - 6$ 下，给定 X，OLS 估计量遵循正态分布。而且，在虚拟假设下，每个 t 统计量呈 t 分布，F 统计量呈 F 分布，通常构造置信区间的方法也是有效的。

推断方法的优劣和隐含的假定的真伪是一致的。

定义 7.5 (**事件研究**) 在所谓的事件研究 $eventstudy$ 中，虚拟变量是个关键的部分。事件研究的目标是确定某个特定的事件是否会影响到某项结果。

线性时间趋势

$$y_t = \alpha_0 + \alpha_1 t + e_t \tag{7-1}$$

$$\begin{cases} E(y_t) = \alpha_0 + \alpha_1 t \\ Var(y_t) = Var(e_t) = \sigma^2 \end{cases}$$

即均值随时间变化，而方差不变。

定义 7.6 (**指数趋势**) $exponentialtrend$: 当一个序列从一个时期到另一个时期的平均增长率为恒定时，它就服从指数趋势。

在实践中，时间序列中的指数趋势可以通过建立有线性趋势的自然对数模型得到：

$$log(y_t) = \beta_0 + \beta_1 t + e_t$$

在回归模型中引进时间趋势，相当于在回归分析中在使用原始数据之前将它们的趋势去掉。

如果一个时间序列是由定期如每月或每季度观测而得到的，它就有可能表现出**季节性**。

如果误差项不是来自正态分布，那么我们必须依靠中心极限定理来为常见的 OLS 统计量和置信区间提供依据。

有滞后因变量的模型必定使假定 $TS.2$ 不成立。

平稳性并没有对相邻项的**相关性**做出要求，但平稳性要求所有时期的相邻项之间的相关关系具有相同的性质。

我们很难判断搜集到的数据是否产生于一个平稳的过程。但是，要指出某些序列不是平稳的却很容易。

定义 7.7 (**协方差平稳**) 一个有有限二阶矩的随机过程在满足下列条件时是协方差平稳 $covariancestationary$ 的：

- $E(x_t)$ 为常数
- $Var(x_t)$ 为常数
- 对任何 t，$h \neq 1$，$Corr(x_t, x_{t+h})$ 只取决于 h，不取决于 t

在技术层面上，平稳性简化了大数定律和中心极限定理的表述。在操作层面上，如果我们想通过回归分析掌握两个或更多变量之间在关系，就需要假定某种跨时期的平稳性。

任何时间序列数据都可以看作是由一个随机过程产生的，看作是一个基本的随机过程的现实($realization$)。这个随机过程所反代表的概率结构，就是时间序列的概率结构。作为类比，时间序列数据与随机过程的关系就如同样本数据与总体分布的关系。对随机过程所作的统计推断，正是藉着时间序列数据而实现的。

定义 7.8 (**弱相依**) 弱相依 $weaklydependent$ 对在随机变量 x_t 和 x_{t+h} 之间的时间距离 h 变大时这二者之间有多大程度的关系作出限定。当 h 无限增大时，如果 x_t 和 x_{t+h} 是“近乎独立”的，这个时间序列过程就被称为弱相依。

弱相依暗示了大数定理 LLN 和中心极限定理 CLT 的成立，并因此替代了随机抽样的假定。时间序列数据的中心极限定理，要求平稳性和弱相依性。独立同分布序列、一阶移动平均序列、稳定的自回归序列、自回归和移动平均过程的混合过程是弱相依的。

定义 7.9 (**协方差平稳**) 当 $h \to \infty$ 时$Corr(x_t, x_{t+h}) \to 0$ 的协方差平稳序列即为 $asymptoticallyuncorrelated$ 的。

定义 7.10 (**稳定的** $AR(1)$ **过程**) 如果 $AR(1)$ 过程要具有弱相依性就需要作一个关键假定–**稳定性条件** $stabilitycondition$：$|\rho| < 1$。一旦假定得到满足，我们就称 $y_t = \rho y_{t-1} + e_t$ 是一个 $stableAR(1)process$。

$$Corr(y_t, y_{t+h}) = \rho^h \tag{7-2}$$

趋势序列，虽然必定是非平稳的，但可以是弱相依的。

定义 7.11 (**趋势平稳过程**) 如果一个序列是弱相依的，而且在除掉趋势以后是平稳的，我们称它是 $trend - stationaryprocess$。

定理 7.6 (OLS **的一致性**) 在假定：线性与弱相依、零条件均值和无完全共线性成立时，OLS 估计量是一致的：

$$plim\hat{\beta}_j = \beta_j$$

我们说 OLS 估计量是一致的，但并不一定是无偏的；我们没有要求解释变量是严格外生的，而是要求潜在的时间序列是弱相依的。

定理 7.7 在假定线性与弱相依、零条件均值、无完全共线性、同方差和无序列相关成立下，OLS 估计量是渐近正态分布的。而且，通常的 OLS 标准误、t 统计量、F 统计量和 LM 统计量是渐近生效的。

它说明，即使经典线性模型假定都不成立，OLS 仍然是一致的，通常的推断程序也是生效的。

一个线性的附加预期的**菲利普斯曲线**：

$$inf_t - inf_t^* = \beta_1(unem - \mu_0) + e_t \tag{7-3}$$

定义 7.12 (周期性失业) 实际失业和自然失业之间的差距称为 $cyclicalunemployment$。

定义 7.13 (未被预期的通货膨胀) 实际的通货膨胀与预期的通货膨胀之间的差距称为 $unanticipatedinflation$。

定义 7.14 (供给冲击) 误差项 e_t 称为 $supplyshock$。

如果在未被预期到的通货膨胀和周期性失业之间有相互替代关系的话，应有 $\beta_1 < 0$。在**适应性预期**$adaptiveexpectation$ 的作用下，当前通货膨胀的预期值决定于最近观测到的通货膨胀。

只要使用的时间序列是弱相依的，常见的 OlS 推断程序比在经典线性假定更弱的假定下生效。当数据不是弱相依的时候，常见的推断程序对这些假定是否成立非常敏感，因为我们无法借助大数定理和中心极限定理。

定义 7.15 (强相依) **高度持久或强相依**$highlypersistentorstronglydependent$ 时间序列：随机游走过程–现在的 y 值和过去的 y 值，即使是很远的 y 值都高度相关。很多人认为，利率、通货膨胀率和失业率这些因素是高度持久的。

有趋势的序列不一定是高度持久的。高度持久的序列包含明显的趋势的例子是**带漂移的随机游走**。

可以用单位根过程表示的一类高度持久的时间序列，在不满足 CLM 假定的情况下，一旦用于回归方程中，可能导致误导性结果。弱相依过程被称为**零阶积整**$integratedoforderzeroI(0)$。：在回归分析中使用之前无须对这种序列进行任何处理，这种序列的均值已经满足标准的极限定理。单位根过程，如随机游走（有或无漂移项）被称作一**阶积整**$integratedoforderfirstI(1)$：这个过程的**一阶差分**是弱相依的（而且通常是平稳的）。

$AR(2)$ 序列相关的平稳性条件:

$$\rho_2 > -1, \rho_2 - \rho_1 < 1, \rho_1 + \rho_2 = 1$$

时间序列的数字特征：

- 均值函数$\mu(t) = E\{Y_t\}$
- 自协方差函数$\gamma(t, s) = Cov(Y_t, Y_s)$
- 自相关函数$\rho(t, s) = Corr(Y_t, Y_s) = \gamma(t, s)/\sqrt{(\gamma(t, t)\gamma(s, s)}$
- 偏自相关函数$\phi(t, s) = Corr\{Y_t, Y_s|Y_{s+1}, \cdots, Y_{t-1}$

滞后算子($lagoperator$) L:

$$L^k Y_t = Y_{t-k}$$

差分运算（$differenceoperation$）：

$$\delta^d Y_t = \delta(\delta^{d-1} Y_t) = \sum_0^d (-1)^j C_d^j Y_{t-j}$$

定义 7.16 (时间序列) 时间序列，就是各种社会、经济、自然现象的数量指标按照时间次序排列起来的统计数据。

定义 7.17 (时间序列模型) 时间序列分析模型，就是揭示时间序列自身的变化规律和相互联系的数学表达式。时间序列分析模型分确定性模型和随机模型两大类。

源代码 7.1 产生AR(1)随机过程

```
DATA TEMP(KEEP=Y1−Y50);
      ARRAY Y Y1−Y50;
      Y(1)=RANNOR(−1);
      DO J=2 TO 50;
         Y(J)=SQRT(0.5)∗Y(J−1) + SQRT(0.5)∗RANNOR(−1);
      END;
RUN; PROC TRANSPORT OUT=TEMP; RUN; DATA TEMP;
  SET TEMP;

PROC ARIMA DATA=TEMP;
      IDENTIFY VAR=COL1 NLAG=6;
RUN;

PROC ARIMA DATA=TEMP;
      IDENTIFY VAR=COL1 NLAG=6;
RUN;

      IDENTIFY VAR=COL1(1) NLAG=6;
RUN;
      ESTIMATE P=1;
RUN;
      ESTIMATE P=1 Q=1;
RUN;
      FORECAST LEAD=6 OUT=FORECAST;
RUN;
      IDENTIFY VAR=COL1(1,1) ;/∗second−difference∗/
RUN;
```

$ARIMA$ 语句是交互式语句，可以通过提供一个 $DATA$ 语句、$PROC$ 语句、$ENDSAS$ 语句或 $QUIT$ 语句中止。

$$
\begin{aligned}
W_t &= \mu + \frac{\theta(B)}{\phi(B)}\varepsilon_t \\
W_t &= (1-B)^d(1-B^s)^D Y_t \text{季节差分} s \text{是季节循环长度}
\end{aligned}
$$

模型常数项

$$constant = \phi(B)\mu$$

自相关图表示当前值与过去值之间的相关程度。散点图称为自相关函数，是因为它有函数形式显示出序列过去值之间的相关程度，自相关函数是是时间间隔的函数。通过检查散点图，可以检查序列是平稳还是非平稳的。

定义 7.18 (**拖尾与截尾**) 函数 $f(k)$ 拖尾(tails off)是指 $f(k)$ 按指数速度衰减，即在存在 $\alpha_1 > 0, \alpha_2 > 0$，使得 $|f(k)|\,\alpha_1 e^{-\alpha_2 k} \quad k = 0, 1, \cdots$。

$f(k)$ p 步截尾(cutts off)是指当 $k > p$ 时，成立 $f(k) = 0$。

时间序列模型与标准的经济计量模型在许多方面是有不同的：

- 首先，时间序列模型通常是为预测而建立的，在应用中参数的估计和显著性检验的目的仅是为了进行预测，以求由最佳参数估计产生最佳预测；而一般的经济计量模型除了为进行经济预测之外，还为进行经济结构分析及经济政策评价而建立的，特别是对模型的参数值感兴趣；
- 其次，经济计量模型的建立基于经济理论，每一参数都有明确的经济意义；而时间序列模型的建立几乎不依赖于任何理论，参数估计亦然，模型参数一般都不需要有明确的经济意义。

对时间序列数据进行分析，平稳性($stationary$)这一假定是基本的(平稳性假定指：时间序列分析的基本用途是根据过去预报未来，因此必须假定过去的发展是什么样的，将来的发展也应是什么样的)。否则，基于 t, F, χ^2 检验等常规假设检验方法的结论是可疑的。一个时间序列关于其它时间序列的回归常常会导致无意义或假性($spurious$)回归。防止这种现象的一种方法就是检查这些序列是否有协整性($cointegration$)。

平稳时间序列建模的一种很流行的方法是自回归求和滑动平均($autoregressiveintegratedmovingaverage$)方法，即著名的 $Box-Jenkins$ 方法。作为一种预测方法，$ARIMA$ 方法的重点不在于构造单一方程或联立方程，而是在“让数据为自身说话”的信念下分析序列本身的概率或随机性质。

同时，作为多元线性回归模型的推广，与多个时间序列有关的传递函数模型及干预模型有着广泛的应用。

7.2 时间序列回归中的序列相关和异方差

如果一个模型的动态在适当的意义上被完整地设定了，它的误差应不会序列相关。所以，检验序列相关可以用来探测动态的错误设定。另外，静态模型和有限分布滞后模型经常有序列相关的误差，即使在模型没有被错误设定的时候也是这样。

只要解释变量是严格外生的，无论误差的相关程度如何，实际上，$\hat{\beta}_j$ 都是无偏的。这与误差差的异方差不会造成 $\hat{\beta}_j$ 的偏误这一观察结果很相符。

当把严格外生性假定放松到 $E(u_t|x_t) = 9$，当数据是弱相依时，$\hat{\beta}_j$ 的分布就紧缩成单一个点 β_j:一致但可能不无偏。这个结论不会因有关误差的序列相关的任何假定而改变。

$AR(1)$ 模型的标准假定：

$$E(e_t|u_{t-1}, u_{t-2}, ...) = 0$$

$$Var(e_t|u_{t-1}) = Var(e_t) = \sigma_e^2$$

回归元严格外生时 $AR(1)$ 序列相关的检验

(1) 做 OLS 回归得到残差 $\hat{u}_{t-1}$ $t = 1, 2, \cdots, n$

(2) 运行回归 $\hat{u}_t$ 对 $\hat{u}_{t-1}$ $t = 2, 3, \cdots, n$得到系数 $\hat{\rho}$ 和它的 t 统计量

(3) 检验 $H_0: \rho = 0$ 和 $H_1: \rho \neq 0$

$DurinandWatson, 1950$ 的 DW 是以全套经典线性模型假定为条件的，包括误差的正态性。这个分布取决于自变量的值，样本规模，回归元的数量和回归是否包含截距。

$$DW = \frac{\sum\limits_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum\limits_{t=1}^{n} \hat{u}_t^2}$$

$$DW \approx 2(1 - \hat{\rho})$$

$$H_1 : \rho > 0$$

为拒绝原假设接受备选假设，我们希望得到一个显著地小于 2 的 DW 值。

当解释变量不是严格外生时，以上两个统计量都无效。

一般回归元的序列相关的检验

(1) 得出 OLS 残差 $\hat{u}_t$ $t = 1, 2, \cdots, n$

(2) 做 u_t 对所有自变量、截距及滞后残差 $\hat{u}_{t-1}$ $t = 2, 3, \cdots, n$的回归，得到 $\hat{u}_{t-1}$ 的系数 $\hat{\rho}$ 及它的 t 统计量

(3) 检验

当采用 LM $LM = (n-q)R^2_{\hat{u}}$ 代替 F 统计量以检验更高阶序列相关时，就是**戈弗雷检验**$Breush-Godfreytest$。

假设误差遵循 $AR(1)$ 模型：

$$u_t = \rho u_{t-1} + e_t t = 1, 2, \ldots$$

u_t 的方差是

$$Var(u_t) = \frac{\sigma_e^2}{1-\rho^2}$$

对方程进行变换

$$y_t - \rho y_{t-1} = (1-\rho)\beta_0 + \beta_1(x_t - x_{t-1}) + e_t t \geq 2$$

上式可写为

$$\bar{y}_t = (1-\rho)\beta_0 + \beta_1 \bar{x}_t + e_t t \geq 2$$

被称为**准差分数据**$quasi-differenceddata$。

$AR(1)$ 模型的可行的 GLS 估计:

(1) OLS 回归得出残差 $\hat{u}_t$

(2) $\hat{u}_t$ 对 $\hat{u}_{t-1}$ 回归，求出 $\hat{\rho}$

(3) 用 OLS 估计准差分变换后的方程，常见的统计量都是渐近正确的。

用 $\hat{\rho}$ 来代替 ρ 的代价是，可行的 GLS 估计量失去了易于处理的有限样本性质。特别是当数据是弱相依时，它就不再是无偏的了，尽管仍然是一致的。同时由于存在估计误差，即便 e_t 是正态分布，其他常用统计量也只是渐近分布。

$\hat{\beta}_1$ 的序列相关–稳健性标准误$serialcorrelation-robuststandarderror$：

(1) 用 OLS 估计方程 $y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t, t = 1, 2, \cdots, n$，得到通常的（但不正确的）$OLS$ 标准误 $se(\hat{\beta}_1)$ 和普通的回归标准误 $\hat{\sigma}$ 及 OLS 残差 $\{\hat{u}_t, t = 1, \cdots, n\}$

(2) 通过辅助方程 $x_{t1} = \delta_0 + \delta_2 x_{t2} + \cdots + \delta_k x_{tk} + r_t$计算残差 $\{\hat{r}_t : t = 1, 2, \cdots, n\}$，然后构造 $\hat{\alpha}_t = \hat{r}_t \hat{u}_t$ (对每个 t)

(3) 对于选定的 g，如式 $\hat{v} = \sum\limits_{t=1}^{n} \hat{\alpha}_t^2 + 2\sum\limits_{h=1}^{n} [1 - h/(g+1)] \left(\sum\limits_{t=h+1}^{n} \hat{\alpha}_t \hat{\alpha}_{t-h} \right)$ 计算 $\hat{v}$

(4) $se_{SCR} = [se(\hat{b}eta_1)/\hat{\sigma}]^2 \sqrt{\hat{v}}$

异方差也可能出现在时间序列模型中，它虽不会造成 $\hat{\beta}_j$ 的偏误和不一致，但会使通常的标准误和统计量变得无效。

7.3 确定性时间序列模型

对于一个时间序列:

$$y_1, y_2, \cdots, y_n$$

滑动平均模型

$$\hat{y}_t = \frac{y_t + y_{t-1} + \cdots + y_{t-N+1}}{N}$$

滑动平均模型的主要作用是消除干扰，显示序列的趋势性变化，并用于趋势预测。**加权滑动平均模型**

$$\hat{y}_{tw} = \frac{a_0 y_t + a_1 y_{t-1} + \cdots + a_{N-1} y_{t-N+1}}{N}$$

加权因子满足

$$\frac{\sum\limits_{i=0}^{N-1} a_i}{N} = 1$$

二次滑动平均模型 二次滑动平均是对经过一次滑动平均产生的序列再进行滑动平均。

$$\hat{\hat{y}}_t = \frac{\hat{y}_t + \hat{y}_{t-1} + \cdots + \hat{y}_{t-N+1}}{N}$$

指数平滑模型

$$\hat{y}_t = \hat{y}_{t-1} + \alpha(y_{t-1} - \hat{y}_{t-1})$$

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}$$

即预测值是前期实际值与预测值的加权和。

随机时间序列模型 自回归滑动平均模型（$ARMA$）是随机时间序列分析模型的普遍形式，自回归模型(AR)和滑动平均模型(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。**自回归模型** 时间序列是它的前期值和随机误差项的线性函数：

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mu_t$$

随机项服从零均值有限方差的正态分布，与前期值不相关。引入滞后算子

$$y_t = \varphi_1 B y_t + \varphi_2 B^2 y_t + \cdots + \varphi_p B^p y_t + \mu_t$$

令

$$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p)$$

模型可写成：

$$\varphi(B) y_t = \mu_t$$

滑动平均模型 时间序列是它的当前和前期误差项和的线性函数：

$$y_t = \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q}$$

引入滞后算子，模型可写为：

$$y_t = \theta(B)\mu_t$$

自回归滑动平均模型 时间序列是它的当前和前期误差项与前期值和的线性函数：

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q}$$

引入滞后算子：

$$\varphi(B) y_t = \theta(B)\mu_t$$

7.4 单整的单位根检验

单整的DF检验 对时间序列建立下列方程：

$$x_t = \rho x_{t-1} + \varepsilon_t$$

如果 ρ 显著地不为零，则 x_t 至少是一阶单整。问题是如何判断 ρ 是否显著在不为零。

构造 t 统计量，进行显著性检验。但此时 t 统计量服从由 $Dicky--Fuller$ 于 1976 年提出的 DF 分布。

一般讲，在经济数据中，表示流量的序列，例如以不变价格表示的消费额、收入等经常表现为 1 阶单整；表示存量的序列，例如以不变价格表示的资产总值、储蓄余额等经常表现为 2 阶单整；用当年价格表示的流量的序列，例如以当年价格表示的消费额、收入等，由于价格指数的作用，也经常表现为 2 阶单整；而象利率等序列，经常表现为 0 阶单整。了解这些，对于选择什么变量进入模型是十分重要的。

单整的ADF检验 在 DF 检验中，由于不能保证 ε_t 是白噪声，所以不能保证 $\rho-1$ 的无偏性。为此，在方程右边加上滞后项（$Dicky--Fuller, 1970, 1980$）：

- 模型一：

$$\Delta x_t = (\rho-1)x_{t-1} + \sum_{i=1}^{p}\theta_i \Delta x_{t-i} + \varepsilon_t$$

- 模型二：

$$\Delta x_t = \alpha + (\rho-1)x_{t-1} + \sum_{i=1}^{p}\theta_i \Delta x_{t-i} + \varepsilon_t$$

- 模型三：

$$\Delta x_t = \alpha + \beta t + (\rho-1)x_{t-1} + \sum_{i=1}^{p}\theta_i \Delta x_{t-i} + \varepsilon_t$$

具体步骤：
一。估计和检验模型三

(1) 第一步，检验

$$H_0 : \rho = 1$$

。从 ADF 分布临界值表中查得给定显著性水平下的、用于模型三检验的 τ_ρ 的临界值。如果参数 $\rho-1$ 的 t 统计量小于临界值（因为 τ_ρ 的临界值为负值，参数 $\rho-1$ 的估计量一般为负，其 t 统计量也为负。所以 t 统计量小于临界值等价于其绝对值大于临界值的绝对值，与以前的概念不矛盾），则拒绝零假设，此时即可以得出序列不存在单位根的结论，不再进入下面的步骤；否则，进入下一步骤。

(2) 第二步，给定 $\rho=1$，检验 β 是否等于零。如果拒绝，进入下一步骤；如果接受，说明模型不包含时间趋势项，转入估计和检验模型二。

(3) 用一般的 t 分布检验 $H_0 : \rho = 1$。如果拒绝零假设，则原序列不存在单位根，为平稳序列；否则，说明原序列是不平稳的，必须对其差分后进一步检验其单位根。

例 7.1 (*GDP* **的** *ADF* **检验**)　应用 *ADF* 检验，探测我国 *GDP* 序列的平稳性。

源代码 7.2　*GDP* 的 *ADF* 检验

```
data temp;
        retain year 1977;
        set gdp;

        d1=dif1(gdp);
        d2=dif2(gdp);

        l1=lag(gdp);
        l2=lag(d1);

        d3=dif(l1);

        year+1;

        output;

         title 'Original␣Series␣GDP';

proc reg data=temp;
        m0: model d1=l1 /noint dw;
        m1: model d1=l1 /dw;
        m2: model d1=year l1 /dw;
        mw: model d1=year l1 d3 /dw;
run;

 title 'Difference␣Series␣GDP'; proc reg data=temp;
        md1: model d2=l2 /noint dw;
        md2: model d2=l2 /dw;
run;
```

7.5　协整cointegration

协整的定义及其性质

定义 7.19 (**协整**)　如果序列 $X_{it}, X_{2t}, \cdots, X_{kt}$ 都是 d 阶单整，存在一个向量 $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$ 使得 $Z_t = \alpha X_t \sim I(d-b)$，则认为序列是 (d, b) 阶协整，记为 $X_t \sim CI(d, b)$，α 为协整向量。

如果两个变量是单整变量，只有当它们单整阶相同时才有可能协整；如果两个变量单整阶不同，就不可能协整。对于三个以上的变量，如果具有不同的单整阶数，则有可能经过线性组合转化为协整。协整的经济意义在于，如果两个变量是具有各自的长期波动率，只要它们是协整的，则它们之间存在一个长期的稳定的比例关系。

例 7.2 (**协整的经济意义**)　居民收入 Y_t 和居民消费 C_t，如果它们都是一阶单整的，则它们之间存在一个稳定的比例关系，就是边际消费倾向不变。可以建立计量经济模型如下：

$Y_t = \alpha + \beta C_t + u_t$。变量选择是合理的，模型误差项一定是白噪声，模型参数具有一定的经济解释能力。

反之，居民消费 C_t 和居民储蓄余额 S_t，前者是一阶单整，后者是二阶单整，它们之间就没有一定的稳定的比例关系，建立如下的经济计量模型：$S_t = \alpha + \beta C_t + u_t$ 就不是合理的，模型误差项就一定不是白噪声，模型参数没有合理的经济解释。

> 从这里，我们已经初步认识，检验变量之间的协整关系，在建立计量经济学模型中的重要性。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。从协整理论出发，在建立消费函数模型时，就不会选择居民储蓄余额作为居民消费的解释变量；但是，按照传统的计量经济学建模理论，从已经认识的经济理论出发选择模型的变量，那么选择居民储蓄余额和居民收入共同作为居民消费的解释变量，不仅不感到奇怪，而且被认为是完全合理的，按照"生命周期消费理论"建立的消费函数模型正是这样的[1]。

协整的检验

检验方法 7.1 (协整的检验：EG 检验) 两变量的 $Engle--Granger$ 检验，$Engle$ 和 $Granger$ 于 1987 年提出两步检验法，也称为 EG 检验。第一步，实行协整回归 $Y_t = \alpha X_t + \varepsilon_t$ 得出 $\hat{e} = Y_t - \hat{Y}_t$；第二步，检验 $\hat{e}$ 序列的单整性，如果序列平稳，则 Y_t 和 X_t 是一阶单整的；如果序列是一阶单整的，则 Y_t 和 X_t 是 $I(2,1)$；$\cdots$。注意，此时的临界值要查由 $Engle\&Granger$ （1987）给出的附表，不能用 $(A)DF$ 检验的临界值。

对于多变量的协整检验，采用 $Johansen$ 检验。$Johansen$于 1988 年，以及与 $Juselius$ 于 1990 年提出了一种用向量自回归模型进行检验的方法，

检验方法 7.2 (协整检验：$CRDW$ 检验) $Sargan\&Granger$(1983) 提出的协整回归 $DurbinWatson$ 检验，$CRDW, cointegratingregressionDurbinWatsontest$。对应的零假设是 $H_0 : d = 0$，检验用统计量为通常的 DW 统计量，$CRDW$ 检验临界值另外给出。

例 7.3 (协整检验：$EG\&CRDW$) 用 OLS 法取我国居民消费关于国民生产总值(1978 − 2002)的线性回归，试问这一回归是否为假性回归，分别用 EG 及 $CRDW$ 检验验证之。

源代码 7.3 协整检验：*EG&CRDW*

```
 title  'Calculate_OLS_Residuals_of_Cosuming_Regressing_on_GDP'; proc
reg  data=LI.cosgdp;
        C_GDP: model cosuming=gdp /dw;
        output out=abc residual=r;
run; data temp;
        set  abc;
        d_r=dif(r);
        l_r=lag(r);
        output;
 title  'Engle-Granger_Test'; proc reg data=temp;
        EG: model d_r=l_r /noint;
run;
```

7.6 自回归与移动平均过程

AR(p)过程和AR(1)、AR(2)过程

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{7-4}$$

用滞后算子表示：

$$(1-\phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)x_t = \phi(L)x_t = u_t \tag{7-5}$$

其中 $\phi(L)$ 称为特征多项式或自回归算子。

思考 7.1 为什么若特征方程所有根的绝对值大于 1，则 $AR(p)$ 过程是平稳的随机过程?

对于 $AR(1)$ 过程：

$$x_t = \phi_1 x_{t-1} + u_t \tag{7-6}$$

均值为零，方差是

$$\frac{1}{1-\phi_1^2}\sigma_u^2$$

。

又

$$\begin{cases} x_t \\ = (1-\phi_1 L)^{-1} u_t \\ = \frac{1}{1+\phi_1 L + (\phi_1 L)^2 + \cdots} u_t \end{cases}$$

对于 $AR(2)$ 过程：

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t \tag{7-7}$$

其平稳性条件为：

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ -\phi_1 + \phi_2 < 1 \\ \phi_2 < 1 \end{cases}$$

如图：

思考 7.2 (如何判别 $AR(2)$ 过程的平稳性)

- 特征方程的根是否在单位圆外?
- (ϕ_1, ϕ_2) 是否落在平稳域?
- 能否满足条件(??)

对于一个自回归时间序列的平稳性必要而不充分的条件是：

$$\sum_{i=1}^{p} \phi_i < 1$$

特征方程的根必须全部在单位圆外时，自回归过程才平稳。

一个平稳的 $AR(p)$ 过程可以转换成一个移动平均过程。

移动平均过程AM(q)

$$x_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \tag{7-8}$$

它具有可逆性的条件是特征方程

$$\Theta(L) = (1 + \theta_1 L + \cdots + \theta_q L^q) = 0$$

的全部根的绝对值大于 1。

自回归移动平均过程

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t \tag{7-9}$$

或

$$\Phi(L) x_t = \Theta(L) u_t$$

如图：

单整自回归移动平均过程

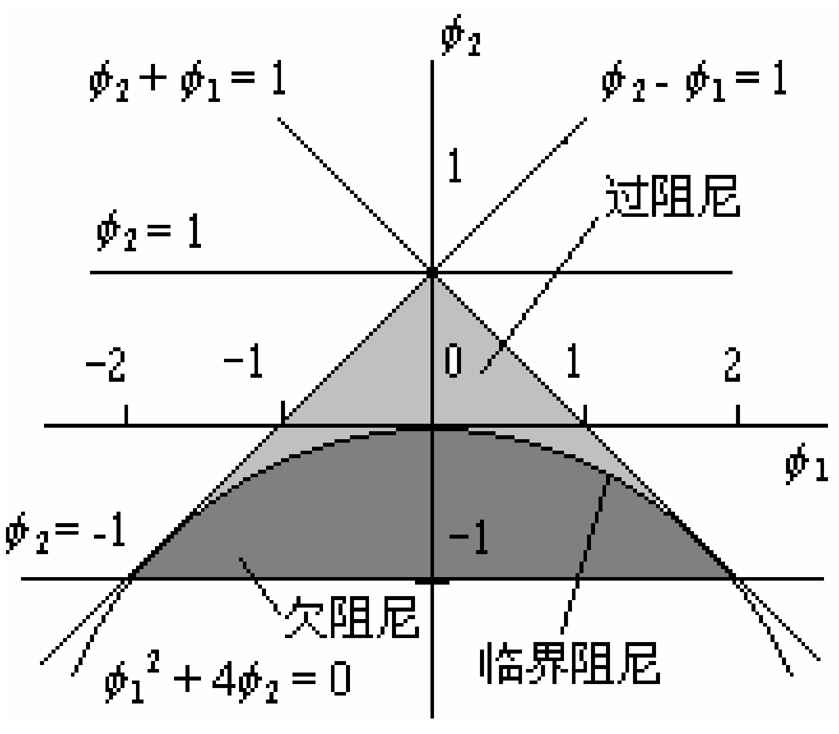


图 7-1 AR(2) 过程的平稳域

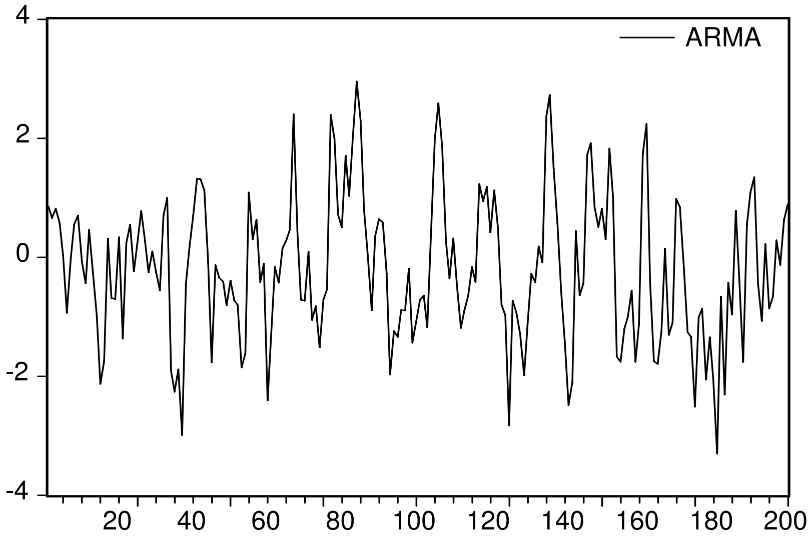


图 7-2 ARMA(1,1) 过程

定义 7.20 (**单位根**) 特征方程的若干根取值恰好在单位圆上。这种根称为单位根，这种过程也是非平稳的。

定义 7.21 (**单整自回归移动平均过程**) 假设一个随机过程含有 d 个单位根，其经过 d 次差分之后可以变换为一个平稳的自回归移动平均过程。则该随机过程称为单整自回归移动平均过程。

$$\Phi(L)\Delta^d y_t = \Theta(L)u_t \tag{7-10}$$

做 $\Delta^d y_t = x_t$ 的逆运算：

$$y_t = S^d x_t$$

，其中 S 是无限累加（积分）算子。

7.7 自相关函数与偏自相关函数

对于 $ARMA$ 模型，在进行参数估计之前，需要进行模型的识别。识别的基本任务是找出 $ARMA(p,q)!AR(p)!MA(q)$ 模型的具体特征，最主要的，是确定模型的阶，即 $ARMA(p,q)$ 中的 p 和 $q!AR(p)$ 中的 $p!MA(q)$ 中的 q。识别的方法是利用时间序列样本的自相关函数和偏自相关函数。

思考 7.3 (**时间序列的识别**) 如何判别其是自回归过程还是移动平均过程？

如何判别其过程的阶数呢？

如何通过一个时间序列研究其过程的平稳性呢？

定理 7.8 (Wald **分解定理**) 若时间序列可以用一个白噪声序列线性表示，则该序列是（宽）平稳的。

$$x_t - \mu - d_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \cdots = \sum_{j=0}^{\omega} \psi_j u_{t-j} \tag{7-11}$$

d_t 是线性确定性成分，μ 表示 x_t 的期望，$\sum \phi_j u_(t-j)$ 是线性非确定性成分，当 $d_j = 0$ 时为纯线性非确定性成分。

$Wald$ 分解定理是 $Wald$ 1937 年提出，它只要求过程 2 阶平稳。要得到过程的 $Wald$ 分解，必须得到无限个 ϕ_j，但这对于有限样本是不可能的，一般如下处理：把 $\Phi(L)$ 看作是两个多项式之比

$$\Psi(L) = \sum \psi_j u_{t-j} = \frac{\Theta(L)}{\Phi(L)} \tag{7-12}$$

定义 7.22 (自协方差函数) 随机过程的每一个元素都是随机变量，相隔 k 期的两个随机变量的协方差即滞后 k 期的自协方差：

$$\gamma_k = Cov(x_t, x_{t-k}) \tag{7-13}$$

自协方差序列

$$\gamma_k, k = 0, 1, 2, \cdots \tag{7-14}$$

称为随机过程 x_t 的自协方差函数。

定义 7.23 (自相关函数) 自相关系数序列

$$\rho_k = \frac{Cov(x_t, x_{t-k})}{\sqrt{Var(x_t)}\sqrt{Var(x_{t-k})}} \tag{7-15}$$

称为随机过程的自相关函数，它是零对称函数。

对于一个平稳过程有：

$$var(x_t) = Var(x_{t-k}) = \sigma_x^2 \tag{7-16}$$

所以(7-15)可以改写为

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{7-17}$$

AR(p)过程的自相关函数 自协方差为：

$$\begin{aligned}
&r_k \\
&= E(y_{t+k}y_t) \\
&= E(\varphi_1 y_{t+k-1} + \varphi_2 y_{t+k-2} + \cdots + \varphi_p y_{t+k-p} + \mu_{t+k})y_t \\
&= \varphi_1 r_{k-1} + \varphi_2 r_{k-2} + \cdots + \varphi_p r_{k-p}
\end{aligned}$$

从而有自相关函数：

$$\begin{aligned}
&\rho_k \\
&= \tfrac{r_k}{r_0} \\
&= \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p}
\end{aligned}$$

此种情形称为拖尾。

可通过解 $YuleWalker$ 方程组求得参数。模型方差可由下式估计：

$$\sigma_\mu^2 = E\mu_t^2 = \cdots = r_0 - \sum_{i,j=1}^{p} \varphi_i \varphi_j r_{j-i}$$

对(**??**)两边同乘 x_{t-k} 再求期望可得

$$\gamma_k = \phi_1 \gamma_{k-1} = \phi_1^k \gamma_0 = \phi_1^k \tag{7-18}$$

对于平稳序列有 $|\phi_1| < 1$，所以当 ϕ_1 为正时，自相关函数衰减至零；当 ϕ_1 为负时，正负交错地衰减至零。

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \phi_3 \gamma_{k-3} + \cdots + \phi_p \gamma_{k-p} \tag{7-19}$$

易得

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p} \tag{7-20}$$

可改写为

$$(1-\phi_1 L-\cdots-\phi_p L^p)\rho_k=\Phi(L)\rho_k=0 \tag{7-21}$$

因式分解

$$\Phi(L)=\prod_{i=1}^{p}(1-G_iL) \tag{7-22}$$

则(7-20)的通解是

$$\rho_k=A_1G_1^k+A_2G_2^k+\cdots+A_pG_p^k \tag{7-23}$$

移动平均过程的自相关函数 移动平均过程的自协方差函数为

$$r_k=E(y_{t+k}y_t)=\begin{cases}\sigma_\mu^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2) & k=0\\ \sigma_\mu^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q) & 1\le k\le q\\ 0 & k>q\end{cases}$$

其中利用了

$$E\mu_t=0$$
$$E(\mu_{t+j}\mu_t)=\begin{cases}0 & j\neq 0\\ \sigma_\mu^2 & j=0\end{cases}$$

依据自相关函数定义：

$$\rho_k=\frac{r_k}{r_0}=\begin{cases}1 & k=0\\ (-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)/(1+\theta_1^2+\cdots+\theta_q^2) & 1\le k\le q\\ 0 & k>q\end{cases}$$

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。

对于 $MA(1)$ 过程而言

$$\rho_k=\frac{\gamma_k}{\gamma_0}=\begin{cases}\frac{\theta_1}{1+\theta_1^2} & k=1\\ 0 & k>1\end{cases} \tag{7-24}$$

其自相关函数具有截尾特征。

$$\rho_k=\begin{cases}\frac{\theta_k+\theta_1\theta_{k+1}+\theta_2\theta_{k+2}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\cdots+\theta_q^2} & k<q\\ 0 & k>q\end{cases} \tag{7-25}$$

ρ_k 从 ρ_1 开始指数衰减。

rho_1 的大小取决于 ϕ_1 和 θ_1，符号取决于 $(\phi_1-\theta_1)$。

自相关图平稳性检验

定义 7.24 (**相关图**correlogram)　由 $ARMA$ 模型性质，一个时间序列若是平稳的，则其 ACF 就是拖尾或截尾的。自相关图是由点(k,ρ_k) 作成的点图。基于样本 $Y_1,Y_2,\cdots,Y_T$ 取 $\hat{\gamma}_k$ 为 γ_k 的一个矩估计：

$$\hat{\gamma}_k=\frac{\sum(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{T}\quad (k=0,1,\cdots)$$

从而得到 ρ_k 的一个估计 $\hat{\rho}_k=\hat{\gamma}_k/\hat{\gamma}_0$。

当用样本矩估计随机过程的自相关函数，则称其为相关图或估计的自相关函数，记为

$$\gamma_k=\frac{C_k}{C_0}\quad k=0,1,\cdots,K\quad k<T \tag{7-26}$$

其中

$$C_k = \frac{1}{T}\sum_{t-1}^{T-k}(x_t - \bar{x})(x_{t-k} - \bar{(}x))$$

相关图是对自相关函数的估计。由于 MA 和 $ARMA$ 过程中的 MA 分量的自相关函数具有截尾特征，所以通过相关图可以估计 MA 过程的阶数 q。实际应用中相关图一般取 $k = 15$ 就足够了。

为检验时间序列的平稳性，首先考虑序列的纯随机性，即白噪声检验，即检验所有滞后超过 1 的 ACF 同时为零：

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$$

利用 $Box - PierceQ$ 统计量：

统计量 7.1 ($Box - PierceQ$ **统计量**)

$$Q(m) = T\sum_{k=1}^{m}\hat{\rho}_k^2$$

它具有大样本分布 χ_m^2。

利用 χ^2 的可分割性，可利用 Q 统计量来验证自相关函数的”拖尾性“，即自某一滞后开始的零假设是否能被拒绝。若自相关函数不具有拖尾性，尤其介呈随滞后值 k 增加的趋势，则时间序列是非平稳的。

例 7.4 (**自相关图检验**)　用 Q 统计量对我国 1987 – 2004 年国民生产总值进行平稳性检验

例 7.5 (**差分平稳和** d **求和阶**)　对我国国民生产总值进行差分平稳 $differencestationary$，确定其 d 求和阶 $integratedoforderd$。

源代码 7.5　差分平稳和 d 求和阶

```
proc arima data=gdp;
            identify  var=gdp(1);
run;
```

源代码 7.4　自相关图检验

```
data gdp;
     input gdp;
     datalines;
3624.1 4038.2 4517.8 4862.4 5294.7 5934.5 7171.0 8964.4 10202.2
11962.5 14928.3 16909.2 18547.9 21617.8 26638.1 34634.4 46759.4
58478.1 67884.6 74462.6 78345.2 82067.5 89468.1 97314.8 105172.3
117390.2 136875.9 ;

proc arima data=gdp;
     identify  var=gdp;
run;
```

ARMA(p,q)的自相关函数 $ARMA(p,q)$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。经过推导得到，$ARMA(p,q)$ 的自协方差函数为：

$$\begin{aligned}&r_k\\&=E(y_{t+k}y_t)\\&=\varphi_1 r_{k-1}+\varphi_2 r_{k-2}+\cdots+\varphi_p r_{k-p}+r_{y\mu}(k)-\theta_1 r_{y\mu}(k-1)-\cdots\theta_q r_{y\mu}(k-q)\end{aligned}$$

其中

$$r_{y\mu}(k)=E(y_t\mu_{t+k})=\begin{cases}0 & k>0\\ \sigma_\mu^2\psi_{-k} & k<0\end{cases}$$

$ARMA(p,q)$ 的自相关函数为：

$$\rho_k=\frac{r_k}{r_0}=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}+\cdots+\varphi_p\rho_{k-p}$$

偏自相关函数 γ_k 的方差近似地为 T^{-1}。当 T 充分大时：

$$\frac{\gamma_k-0}{T^{-\frac{1}{2}}}=\gamma_k\sqrt{T}\sim N(0,1) \tag{7-27}$$

偏自相关函数在滞后期 p以后有截尾特性，因此可用此特征识别$AR(p)$过程的阶数。

偏相关图

对于被估的位于时间间隔 K 的偏自相关函数的标准误差的近似值的计算基于如下零假设：阶数为 $K-1$ 的纯自回归高斯过程生成所考虑的序列，其标准误差为 $\frac{1}{\sqrt{n}}$，置信区间为 95。

时期差为 k 的协方差为

$$\gamma_k=E[y_{t-k}(\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+\epsilon_t)]$$

令 $k=0,1,\cdots,p$，我们得到 $p+1$ 个差分方程，从它们可以解出 $\gamma_0,\gamma_1,\cdots,\gamma_p$:

$$\begin{gathered}\gamma_0=\phi_1\gamma_1+\phi_2\gamma_2+\cdots+\phi_p\gamma_p+\sigma_\epsilon^2\\ \gamma_1=\phi_1\gamma_0+\phi_2\gamma_1+\cdots+\phi_p\gamma_{p-1}\\ \cdots\cdots\cdots\cdots\\ \gamma_p=\phi_1\gamma_{p-1}+\phi_2\gamma_{p-2}+\cdots+\phi_p\gamma_0\end{gathered}$$

在上式每个方程两边同时除以 γ_0，我们就可以解出 $\rho_1,\rho_2,\cdots,\rho_p$ 的 p 个方程。这个方程称为

定义 7.25 (Yule-Walker **方程组**)

$$\begin{aligned}1=&\ \phi_1\rho_1+\phi_2\rho_2+\cdots+\phi_p\rho_p+\sigma_\epsilon^2/\gamma_0\\ \rho_1=&\ \phi_1+\phi_2\rho_1+\cdots+\phi_p\rho_{p-1}\\ \cdots&\ \cdots\cdots\cdots\\ \rho_p=&\ \phi_1\rho_{p-1}+\phi_2\rho_{p-2}+\cdots+\phi_p\end{aligned} \tag{7-28}$$

当 $\rho_1,\rho_2,\cdots,\rho_p$ 已知时，由这些方程组就可以解出 $\phi_1,\phi_2,\cdots,\phi_p$。

$Yule-Walker$ 方程组求解需要知道自回归过程的阶数。因此，我们可以（利用样本自相关函数）对连续的 p 值求解 $Yule-Walker$ 公式，得到偏自相关函数。偏自相关函数近似地服从均值为零，方差为 1/4 的正态分布。因此，我们可以在显著性水平 5% 下通过考察偏自相关函数序列的绝对值是否大于 $2/\sqrt{T}$ 来检验某个值是否显著在不为 0。

7.8 随机时间序列分析模型（AR、MA、ARMA）的估计

自回归模型：AR 模型

$$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\cdots+\phi_p y_{t-p}+\delta+\varepsilon_t \tag{7-29}$$

如果自回归过程是平稳的，则它的均值关于时间一定是常数：

$$\mu=\frac{\delta}{1-\phi_1-\phi_2-\cdots-\phi_p}$$

时期差为 k 的协方差为：

$$\rho_k=\phi_1^k\gamma_0=\frac{\phi_1^k\sigma_\varepsilon^2}{1-\phi_1^2}$$

齐次非平稳过程：ARIMA　模型 如果

$$w_t=\Delta^d y_t$$

平稳序列，我们就称 y_t 是 d 阶齐次非平稳序列。

对序列 y_t 差分得到平稳序列 w_t 后，我们可以用 $ARMA$ 过程对 w_t 建立模型。如果 $w_t=\Delta^d y_t$ 且 w_t 是一个 $ARMA(p,q)$ 过程，则我们称 y_t 是 (p,d,q) 阶综合自回归——移动平均过程（$anintegratedautoregressive--movingaverageprocessoforder$）或 $ARMA(p,d,q)$。

我们可以利用向后位移算子将其写为：

$$\phi(B)\Delta^d y_t=\delta+\theta(B)\epsilon_t$$

其中：
自回归算子

$$\phi(B)=1-\phi_1B-\phi_2B^2-\cdots-\phi_pB^p$$

移到平均算子

$$\theta(B)=1-\theta_1B-\theta_2B^2-\cdots-\theta_qB^q$$

且有 $w_t=\Delta^d y_t$ 的均值为：

$$\mu_w=\frac{\delta}{1-\phi_1-\phi_2-\cdots-\phi_p}$$

如果 θ 不为0，综合序列 y_t 将有一个确定性的趋势。

注意到如果有序列 w_t，我们可以通过对其求和 d 次而回到 y_t：$y_t=\sum^d w_t$ 。**AR(p)模型的Yule Walker方程估计**

$$\begin{bmatrix}\hat\varphi_1\\ \hat\varphi_2\\ \vdots\\ \hat\varphi_p\end{bmatrix}=\begin{bmatrix}\hat\rho_0 & \hat\rho_1 & \cdots & \hat\rho_{p-1}\\ \hat\rho_1 & \hat\rho_0 & \cdots & \hat\rho_{p-2}\\ \vdots & \vdots & \ddots & \vdots\\ \hat\rho_{p-1} & \hat\rho_{p-2} & \cdots & \hat\rho_0\end{bmatrix}^{-1}\begin{bmatrix}\hat\rho_1\\ \hat\rho_2\\ \vdots\\ \hat\rho_p\end{bmatrix}$$

$$\hat\sigma_\mu^2=\hat r_0-\sum_{j=1}^{p}\hat\varphi_j\hat r_j=\hat r_0-\sum_{i,j=1}^{p}\hat\varphi_i\hat\varphi_i\hat r_{j-i}$$

MA(q)模型的矩估计

$$\hat r_k=\begin{cases}\hat\sigma_\mu^2(1+\hat\theta_1^2+\hat\theta_2^2+\cdots+\hat\theta_q^2) & k=0\\ \hat\sigma_\mu^2(-\hat\theta_k+\hat\theta_1\hat\theta_{k+1}+\cdots+\hat\theta_{q-k}\hat\theta_q) & 1\le k\le q\\ 0 & k>q\end{cases}$$

ARMA(p,q)模型的矩估计

$$\begin{bmatrix}\hat\varphi_1\\ \hat\varphi_2\\ \vdots\\ \hat\varphi_p\end{bmatrix}=\begin{bmatrix}\hat\rho_q & \hat\rho_{q-1} & \cdots & \hat\rho_{q-p+1}\\ \hat\rho_{q+1} & \hat\rho_q & \cdots & \hat\rho_{q-p}\\ \vdots & & & \\ \hat\rho_{q+p-1} & \hat\rho_{q+p-2} & \cdots & \hat\rho_q\end{bmatrix}^{-1}\begin{bmatrix}\hat\rho_{q+1}\\ \hat\rho_{q+2}\\ \vdots\\ \hat\rho_{q+p}\end{bmatrix}$$

7.9 传递函数模型

定义 7.26 (传递函数模型) 如果时间序列模型描述某一时间序列的性态是由另一些时间序列来解释的，则称这一模型为多变量时间序列模型 $multipletimeseriesmodel$。描述这些变量之间的动态关系的模型，称为传递函数模型 $transferfunctionmodel$。传递函数模型一词与多变量时间序列模型一词是可以交换使用的。

输出 Y 与输入 X 间的动态关系可以用一个线性模型或线性系统表示：

$$Y_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots = \alpha(L) X_t$$

其中 $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$ 称为滤波器 $linearfilter$，权重 $\alpha_0, \alpha_1, \cdots$ 称为脉冲响应函数 $impulseresponsefunction$。称 $\alpha(L)$ 是稳定的 $stable$，如果对于 $|L| \leq 1$，级数 $\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$ 收敛。

一般地，传递函数模型具有如下的形式：

$$Y_t = \frac{\beta(L)}{\delta(L)} X_t + \frac{\theta(L)}{\alpha(L)} \varepsilon_t$$

其中 $varepsilon_t$ 是白噪声序列，$\beta(\cdot), \delta(\cdot), \theta(\cdot), \alpha(\cdot)$ 为多项式，并要求 Y_t 与 X_t 都是平稳的。

7.10 时间序列ARIMA建模:BOX-Jenkiens

建模步骤

$$\Phi(L)\Delta^d y_t = \theta_0 + \Theta(L) u_t \tag{7-30}$$

- $\phi(L)$ 和$\Theta(L)$ 是 L 的函数，其根在单位圆外
- θ_0 是位移项
- y_t 经过 d 次差分后可表达为一个平稳可逆的 $!ARMA$ 过程

建模过程见(7-3)：

模型的识别model identification 通过对相关图的分析，初步确定 $ARIMA$ 模型的形式，即 p, d, q 的值。第一步：

(1) 模型的识别主要依赖于自相关和偏相关图

(2) 对经济数据取对数以消除异方差

(3) 判断随机过程是否平稳：如果一个随机过程是平稳的，其特征方程的根都应在单位圆之外：如果 $\Phi(L) = 0$ 的根接近单位圆，自相关函数将衰减的很慢。

(4) 在分析相关图时，如果发现其衰减很慢，即可认为该时间序列是非平稳的。这时应对该时间序列进行差分，同时分析差分序列的相关图以判断差分序列的平稳性，直至得到一个平稳的序列。

(5) 参数 d 通常只取 0，1或 2。

(6) 实际中也要防止过度差分。一般来说平稳序列差分得到的仍然是平稳序列，但当差分次数过多时存在两个缺点，（1）序列的样本容量减小；（2）方差变大；所以建模过程中要防止差分过度。对于一个序列，差分后若数据的极差变大，说明差分过度。

第二步：

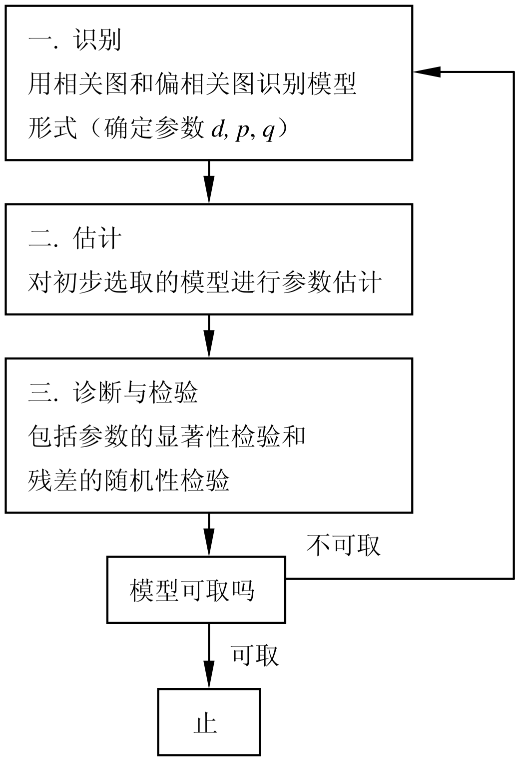


图 7-3 ARIMA 建模过程图

(1) 在平稳时间序列基础上识别模型的阶数。

(2) 时间序列的相关图与偏相关图可为识别模型参数 p, q 提供信息。相关图和偏相关图（估计的自相关系数和偏自相关系数）通常比真实的自相关系数和偏自相关系数的方差要大，并表现为更高的自相关。

(3) 估计的模型形式不是唯一的，所以在模型识别阶段应多选择几种模型形式，以供进一步选择。

模型的估计model estimation 在实际应用中，最小二乘估计很难被应用，因为存在严重的共线性。一般采用极大似然法对参数进行估计。对一组相互独立的随机变量 x_t，$t = 1, 2, \cdots, T$，获取一个样本 $x_1, x_2, \cdots x_T$，则有似然函数

$$\begin{aligned}&L(\gamma|x_1, x_2, \cdots, x_T)\\&= f(x_1|\gamma)f(x_2|\gamma)\cdots f(x_T|\gamma)\\&= \prod_{t=1}^{T} f(x_t|\gamma)\end{aligned}$$

其中 $\gamma = \gamma_1, \cdots, \gamma_k$ 是一组未知参数。取对数构成对数似然函数，然后采用拉格朗日方法对对数似然函数求极大值，得出 γ。极大似然估计量具有一致性的渐近有效性。

思考 7.4　用极大似然法推导

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_{k-1} x_{t(k-1)} + u_t, u_t \sim N(0, \sigma^2)$$

$$LogL = \sum_{t=1}^{T} logf(y_t) = -\frac{T}{2}log2\pi - \frac{T}{2}log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t-1}^{T}[y_t - E(y_t)]^2 \tag{7-31}$$

对时间序列模型参数进行估计 对于非平稳过程 y_t，经过 d 次差分后可以化为平稳、可逆的综合自回归移动平均过程 x_t:

$$\Phi(L)\Delta^d y_t = \Theta(L)u_t \tag{7-32}$$

它可改写为

$$u_t = \frac{\Phi(L)}{\Theta(L)}x_t \tag{7-33}$$

对于 y_t 假定可以观测到 $T+d$ 个观测值 $y_{-d+1}, y_{-d+2}, \cdots, y_0, y_1, \cdots, y_T$。经过 d 次差分后，x_t 的容量为 T。参数估计的标准的是使观测值与拟合值平方和最小。

根据(7-31)可得条件似然函数

$$LogL = -Tlog\sigma_u - \frac{\sum_t \hat{u}_t^2}{2\sigma_u^2} \tag{7-34}$$

称之为条件似然函数是因为 $\hat{u}_t^2$ 依赖于过去的不可知观测值 $x_0, x_1, \cdots, x_{-p+1}$和 $u_0, u_{-1}, \cdots, u_{-1+1}$。此问题的一般处理方法是取它们的无条件期望值。$u_0, u_{-1}, \cdots, u_{-1+1}$ 的无条件期望值为零。

(1) 若模型(7-32)不包含漂移项，则 $x_0, x_1, \cdots, x_{-p+1}$ 无条件期望值为零；

(2) 若模型(7-32)不包含移动平均项，则(7-33)是一个线性函数，可以用 OLS 法估计参数；

(3) 若模型(7-32)包含移动平均项，则(7-33)是一个非线性函数，要采用非线性方法估计参数。

假定模型是纯自回归形式：

$$\Phi(L)x_t = u_t \tag{7-35}$$

或

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{7-36}$$

对于参数而言，这是一个线性回归模型，用以直接用 OLS 法估计，并且和 ML 法结果近似。

估计量的评价问题 用于线性回归模型的 F, t 统计量不能直接用于评价非线性模型。原因是：虽然 u_t 是正态分布且均值为零，但残差：

$$\hat{u}_t = x_t - \hat{x}_t = x_t - f(x_{t-1}, x_{t-2}, \cdots, x_{t-p}, \gamma_{1.k}, \cdots, \gamma_{p+q.k})$$

并不服从正态分布，则 $\sum \hat{u}_t$ 并不服从 χ^2 分布，参数估计量不服从正态分布。

模型的诊断与检验 model diagnostic checking 以样本为基础检验模型。如果模型某些参数值不能通过显著性检验，或残差序列不能近似白噪声，则要返回第一步。

检验拟合的模型是否合理：

(1) 检验模型的参数估计值是否有显著性，用 t 检验完成；

(2) 检验模型的残差序列是否为白噪声，用 $Box-Pierce$ (1970) 提出的 Q统计量完成。

检验方法 7.3 (Q **检验**) 零假设: $H_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0$ 即假设模型的误差项是一个白噪声。

统计量: Q 统计量$Q = T\sum_{k=1}^{K}\gamma_k^2$ 它近似服从 $\chi^2(k-p-q)$ 分布。γ_k 表示自相关系数值。

判别规则: 用残差序列计算 Q 统计量的值。显然若残差序列不是白噪声，残差序列中必含有其他成份，自相关系数不等于零。则 Q值将很大，反之 Q 值将很小。

修正的统计量:上式定义的统计量与 χ^2_{k-p-q} 分布存在差异，相应值可能偏小，$Ljung$和 Box 提出修正的 Q 统计量

$$Q = T(T+2)\sum_{k=1}^{K}\frac{\gamma_k}{T-k}$$

非线性模型的估计 如果模型包含移动平均项，则模型对于参数而言是非线性的。一般采用如下三种方法：**直接搜索法** 通过改变参数的取值，反复计算残差平方和 $\sum \hat{u}_t^2$ 的值。然后从中选择最小的那个值所对应的参数值作为对参数的估计值。这种方法只有在参数个数较少时才是可行的。当参数个数较多时，计算量将非常大。

直接优化法 求误差平方和函数对每一个参数的偏导数并令其为零，从而求得正规方程

$$\frac{\partial\left(\sum\limits_t \hat{u}_t^2\right)}{\partial\gamma_i} = 0, i = 1, 2, \cdots, p+q$$

其中

$$\begin{aligned}&\gamma\\&= (\gamma_1, \gamma_2, \cdots, \gamma_{p+q})\\&= (\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q)\end{aligned}$$

因为 $p+q$ 个方程中都含有 $p+q$ 个参数，所以必须联立求解。由于计算上的困难，这种方法很少直接采用。

线性迭代法 对任何非线性函数通常都可以按泰勒级数展开。

首先为参数选一组初始值，将

$$x_t = f(x_{t-1}, x_{t-2}, \cdots, x_{t-p}$$

在选定的初始点

$$(\gamma_{1.0}, \cdots, \gamma_{p+q.0})$$

按 $Taylor$ 展开。

取展开式前两项，用 OLS 法估计得出第二个点。依此类推。

中止标准：

$$\frac{|\gamma_{i,j+1} - \gamma_{i,j}|}{\gamma_{i,j}} < \delta$$

7.11 ARCH 模型

若一个平稳随机变量 x_t 可以表示为 $AR(p)$ 形式，其随机误差项的方差可用误差项平方的 q 阶分布滞后模型描述

$$\begin{aligned}&\text{均值方程}\quad x_t = \beta_0 + \beta_1 x_{t-1} + \beta x_{t-2} + \cdots + \beta_p x_{t-p} + u_t\\&ARCH\text{方程}\quad \sigma_t^2 = E(u_t^2) = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2\end{aligned} \tag{7-37}$$

则称 u_t 服从 q 阶的 $ARCH$ 过程。它的无条件方差是

$$\sigma^2 = \frac{\alpha_0}{1-\sum_{i=1}^{q}\alpha_i} \tag{7-38}$$

$ARCH(q)$ 模型是关于 σ_t^2 的分布滞后模型。为避免 u_t^2 的滞后项过多，可采用加入 σ_t^2 的滞后项的方法：

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \lambda_1\sigma_{t-1}^2 \tag{7-39}$$

此模型称为广义自回归条件异方差模型，用 $GARCH(1,1)$ 表示。其中 u_{t-1} 称为 $ARCH$ 荐，σ_{t-1} 称为 $GARCH$ 项。它的无条件方差是

$$\sigma^2 = \frac{\alpha_0}{1-\sum_{i=1}^{q}\alpha_i - \sum_{i=1}^{p}\lambda_i} \tag{7-40}$$

对于 $ARCH(p)$ 模型和 $GARCH(p,q)$ 模型，在实际应用中，条件

$$\begin{aligned} 0 &\le \sum_{i=1}^{q} \alpha_i \le 1 \\ 0 &\le (\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \lambda_i) \le 1 \end{aligned} \tag{7-41}$$

有时不能得到满足。

为了保证方差为正，又提出了如下两种模型形式。一种是绝对值模型 $GARCH/ARCH$ 模型：

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left| u_{t-i} \right| + \sum_{j=1}^{p} \lambda_j \sigma_{t-j}^2 \tag{7-42}$$

另一是指数 $GARCH$(Nelson 1991):

7.12 多项式分布滞后模型Polynomial distributed lags PDLs

对于时间序列数据，由于经济系统中的经济政策的传导、经济行为的相互影响和渗透都是需要一定时间的。它们的数值是由自身的滞后量或其他变量的滞后量所决定的，因此需要考虑变量之间的滞后关系。表现在计量经济模型中，即应当包括某些滞后变量于解释变量中。比如分析货币政策效应时，经常分析货币供给对产出的影响，要在模型中加入货币供给的多期滞后，以反映出货币政策的时滞性。

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} + u_t \quad t = 1, 2 \cdots, T$$

系数 β 反映了 x 对 y 的乘数作用。

7.13 无限分布滞后模型

Infinite Distributed Lag Model

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \cdots + u_t$$

- IDL模型不要求在某个特定时刻截断滞后
- 很远的过去的z 对 y 的解释能力不如较近的过去的 z
- 短期倾向是 δ_0，这意味着，z 的一个暂时变化对 y 的期望没有长期影响
$$E(y_h) = \alpha + \delta_h$$
- 长期倾向是所有系数之后，这意味着 z 的一个单位的永久增长，y 的期望值的长期变化
$$E(y_h) = \alpha + \delta_0 + \cdots + \delta_h$$

这隐含着**严格外生性假定：**z 的任何变化对 y 的期望值没有影响。它不允许 y_t 对将来的 z 的反作用。

$$E(u_t | \cdots, z_{t-2}, z_{t-1}, z_t, z_{t+1}, \cdots) = 0$$

更弱一点的假定是:

$$E(u_t | z_t, z_{t-1}, \cdots) = 0$$

这个假定使得 z 服从一种政策规则。

7.14 几何（或考依克）分布滞后

IDL 不能无约束地估计出一切系数。一种简单形式是$geometricorKoyckdistributedlag, GDL$。在这个模型中，每个 δ_i取决于

$$\delta_j = \gamma\rho^j, |\rho| < 1, j = 0, 1, 2, \cdots$$

- GDL 的即期倾向 IP 是 $\delta_0 = \gamma$
- 长期倾向 $LRP = \gamma(1 + \rho + \rho^2 + \cdots) = \gamma/(1 - \rho)$

7.15 有理分布滞后模型

$rationaldistributedlagRDLmodel$

$$y_t = \alpha_0 + \gamma_0 z_t + \rho y_{t-1} + \gamma_1 z_{t-1} + v_t$$

它与无限分布滞后模型是等价的。

- 即期倾向是 γ_0
- z_{t-h} 的系数是 $\rho^{h-1}(\rho\gamma_0 + \gamma_1)$
- LRP 是 $(\gamma_0 + \gamma_1)/(1 - \rho)$

7.16 趋势和季节性

忽略两个序列按相同或相反趋势发展这一事实会导致错误的结论，认为一个变量的变化由另一个变量的变化引起。什么样的模型能恰当地描述有趋势的行为？

$$y_t = \alpha_0 + \alpha_1 t + e_t \quad e_t \sim i.i.d.$$

参数 α_1 与时间 t 相乘的形式形成了一个线性时间趋势 $lineartimetrend$。α_1 度量了由于时间的流逝，y_t 从一个时期到下一个时期的变化。

- $E(y_t) = \alpha_0 + \alpha_1 t$
- $Var(y_t) = Var(e_t) = \sigma_e^2$
- y_t 是独立但非同一分布的序列

定义 7.27 (指数趋势) 经济序列通用指数趋势 $exponentialtrend$ 来追近。当一个序列从一个时期到另一个时期的平均增长率为恒定时，它就服从指数趋势。

在实践中，时间序列中的指数趋势可以通过建立有线性趋势的自然对数模型得到：

$$log(y_t) = \beta_0 + \beta_1 t + e_t, t = 1, 2, \cdots$$

$beta_1$ 近似等于 y_t 各项增长率的平均值：

$$\beta_1 = \frac{y_t - y_{t-1}}{y_{t-1}} = \Delta\log(y_t)$$

趋势变量并不一定违背经典线性模型的假定。

仅仅因为每个变量都随时间增长，而在两个或更多的趋势变量之间找到某种关系的现象是**谬误回归**$spuriousregression$ 的一个例子。

在回归模型中引进时间趋势，相当于在回归分析中在使用原始数据之前将它们的趋势去掉。

时间序列回归中的 R^2 通常很大，与典型的横截面数据的 R^2 相比尤其大。这是因为：

- 时间序列数据通常是总量数据，这相比个人、家庭或企业的数据容易解释一些
- 因变量有趋势时，时间序列回归中的普通的或较正的 R^2 可能会人为变大

7.17 总结

(1) 基于时间序列数据的回归分析都默许地假定时间序列是平稳的。经典的 t, F 检验等都基于此一假定。在经济实践中，大多数的时间序列都是非平稳的。

(2) 一个平稳的时间序列，从直观上可以看作一条围绕其平均值上下波动的曲线；从理论上，有二种意义的平稳性：一是严平稳($stationaryinthestrictsense$)，另一是宽平稳($stationaryinthewidesense$)。概率分布随同时间的平移而不变的序列是严平稳序列；而均值、方差与协方差函数不随同时间变化的序列是宽平稳序列。

(3) 平稳性的检验有二种方法。较为粗略的方法是周期图性：若周期图衰减得很快，则可认为该序列是平稳的；否则是非平稳的。对纯随机序列，即白噪声序列，在所有大于或等于 1 的滞后处，自相关函数全为零。较为精细的方法是单位根法：若在统计意义上认为单位根存在，则序列不平稳。单位根检验主要有 DF, ADF 检验方法。

(4) 对非平稳的时间序列 Y_t，有二种途径可将其转化为平稳序列。一是差分平稳(DS)，即经过某一阶差分（如 d 阶 $\delta^d Y_t$）将达到平稳，此时，我们就称原序列具有 d 求和阶，记为 $I(d)$；另一是趋势平稳，即 $Y_t - f(t)$ 为一平稳序列，其中 $f(t)$ 是一确定性的趋势曲线。DF, ADF 检验可用来检验一个时间序列是 DS 还是 TS。

(5) 对于非平稳时间序列，一个时间序列变量关于另一个或多个时间序列变量的回归，很有可能会导致假性回归，即拟合效果很好（如有较高的可决系数 R^2），但作深入研究会发现参数估计与常理不符。假性回归检验，经验性地有一经纶准则：若 $R^2 > DW$，则所估计出的样本回归可能是假性回归；严格地，需从协整性的角度来测定。

(6) 一组时间序列 $X_{it}, \cdots, X_{pt} \quad (p \geq 2)$ 称为具有 (d, b) 阶协整性（$0 < b \leq d$），是指：该序列均为 $I(d)$ 序列，且它们的线性组合为一个 $I(d, b)$ 序列。通常所指的协整性是指非平稳的 DS 序列的线性组合达到平稳。协整性的探测方法有 EG 检验、AEG 检验和 $CRDW$ 检验。协整性的存在意味着变量间存在着一种长期的或均衡的关系。

(7) 为了协调经济变量短期性态与长期性态之间的矛盾，可以应用 ECM。

(8) 时间序列经济计量学建模在某些场合是暂时的，需要作进一步的改进。在这一领域最为关心的问题是：时间序列何时、为何平稳或不平稳？

(9) 作为一种预测方法，BJ 方法包括以下四步骤：模型识别；模型估计；模型诊断校验和模型预测。在许多情形，此方法所进行的预测会比传统的经济计量建模所得的预测更可靠，尤其是短期预测。

(10) 经济时间序列的季节性，可用虚拟变量和 $ARIMA$ 模型来调整。

(11) $Granger$ 指出：“协整性检验 $testforcointegration$”可以看作避免假性回归 $spuriousregression$ 的预检验“。

(12) 在小样本情形，似然法倾向于低估误差方差；二乘法的 MSE 高估 $\hat{\sigma}^2$。

第八章 横截面时间序列模型

8.1 基本原理

定义 8.1 (**独立混合横截面**) independently pooled cross section它是在不同时点（经常但不一定是不同年份）从一个大总体里进行随机抽样的结果。如，每隔一年对某城市出售的住房抽取一个关于售价、面积、间数的随机样本。从统计学观点看，这些数据集有一个重要特点：它们都是由独立抽取的观测值构成的，这排除了在不同观测值中误差项有相关关系。如果每个时期都抽取一个随机样本，那么将所得到的随机样本合并起来就给出一个独立混合横截面。

一个独立混合横截面和单一个随机样本的差异在于，在不同时点上对总体进行抽样很可能导致观测点不是同分布的。如，随着时间的流逝，工资和学历的分布已经改变。这可以通过在多元回归模型中容许截距或（和）斜率的改变来处理。

使用的理由是要加大样本容量，因为它是在不同时点从同一总体中抽取的多个随机样本混合而成，从而可以得到更好的估计量和更具功效的检验统计量。但是，**这仅当因变量和某些自变量保持着不随时间而变的关系时，混合才起作用**。使用混合横截面的一个问题是**总体在不同时期有不同分布**，对此，我们允许截距或斜率在不同年份或时点有不同的值。

此外，误差方差还可能随时间而改变。异方差－稳健标准误及其检验统计量可以用于此。通过平方 OLS 残差对所有自变量回归，就能作出 $Breush-Pagan$ 检验；或作残差平方对所有原来模型中包含的解释变量及其平方和交叉乘积项的回归进行 $White$ 检验。$White$ 检验的原假设是不存在异方差，即检验辅助方程中除截距项外的所有自变量的系数为零。如果辅助回归方程的拟合优度 R^2 越大，说明残差平方受到解释变量影响越显著，也就越倾向于认为存在异方差；或者计算 $N\times R^2\sim\chi_k^2$ 来判断；或者用辅助回归方程的 F 统计量来判断模型显著性。

在存在异方差时，普通最小二乘估计量仍然是线性和无偏的，但不有效的，即不再具有方差最小性，所以通常的假设检验不可靠。加权最小二乘程序能解决误差可能随时间而变的问题。WLS 估计或获得有效估计，它是将权重序列分别与每个变量的观测值相乘，然后对变化后的新模型利用 OLS 估计。

如果在方程中考虑所有自变量与某个年份虚拟变量乘积项的交互作用，等价于估计两个不同的方程，一个对基年，一个对年份虚拟变量所代表的年。

邹至庄检验能用来决定两组数据之间的多元回归函数有无差别，我们同样可把这种检验用于两个不同时期。

定义 8.2 (**邹至庄统计量**)

$$F=\frac{SSR-(SSR_1+SSR_2)}{SSR_1+SSR_2}\times\frac{n-2(k+1)}{k+1}$$

它有一个重要的局限，即虚拟假设要求各组之间不存在任何差异。

定义 8.3 (**综列数据**) 兼有横截面和时间序列的维度，但在一些方面不同于独立混合横截面，收集综列数据时要求对相同的观测对象在不同时点进行跟踪而获得数据。

利用综列数据的一个方法是把影响因变量的观测不到的因素分为两类:一类是恒常不变的；一类则随同时间而变。

$$y_{it}=\beta_0+\delta_0d_t+\beta_1x_{it}+\alpha_i+u_{it}$$

在时间恒定的因素称为**非观测效应**unobserved effect，**固定效应**fixed effect，**非观测差异性**unobserved heterogeneity。误差常称为**特异性误差**idiosyncratic erroe或**时变**time-varying。上式还可写成：

$$y_{it}=\beta_0+\delta_0d_t+\beta_1x_{it}+v_{it}\ \ v_{it}=\alpha_i+u_{it}$$

式中，v_it 称为**复合误差**composite error。

定义 8.4 (**差异性偏误**) 在综列数据建模中，即使假定特异误差与解释变量不相关，如果非观测效应与解释变量相关，混合 OLS 估计就是有偏误且不一致的。由此所造成的偏误称做heterogeneity bias。然而，它的确是由于遗漏了一个时间上恒常的变量而引起的。

在大多数应用中，收集综列数据的主要理由是为了考虑非观测效应与解释变量相关。将数据差分之后，非观测效应不再出现于模型中，

定义 8.5 (**平衡综列数据**) 如果对 N 个横截面单元的每一个都有同样的 T 期数据

$AR1$ 模型的 $FGLS$ 估计根据估计 ρ 的方法不同和处理第一次观测的办法不同而不同。$Cochrane-Orcuttestimation$ 省略了第一次观测，$Prais-Winstenestimation$ 则使用了第一次观测：同乘以 $(1-\rho)^{1/2}$。

取一阶差分是消除固定效应 α_i 一种方法，另一种方法是**固定效应变换**$fixedeffectstransformation$，凡是在时间上恒定的解释变量都必定随机之而消失。以一元模型为例：

$$y_{it}=\beta_1x_{it}+\alpha_i+u_{it}$$

对每个 i 求取时间上的平均：

$$\bar{y}_i=\beta_1\bar{x}_i+\alpha_i+\bar{u}_i \tag{8-1}$$

获取**除去时间均值后的数据**$time-demeaneddata$: 、

$$y_{it}-\bar{y}_i=\beta_1(x_{it}-\bar{x}_i)+u_{it}-\bar{u}_i \tag{8-2}$$

或

$$\ddot{y}_{it}=\beta_1\ddot{x}_{it}+\ddot{u}_{it} \tag{8-3}$$

固定效应变换又称**组内变换**$withintransformaiton$。基于除时间均值变量的混合 OLS 估计量被称为**固定效应估计量**$fixedeffectsestimator$ 或**组内估计量**$withinestimator$。(8-3)的 OLS 使用了变量在每一横截面观测之内的**时间变异**$timevariation$。

对横截面方程(8-1)使用 OLS 估计量时，就给出了**组间估计量**$betweenestimator$：同时用 y 和 x 的时间平均代替 y 和 x，然后做一个横截面回归。它忽视了变量怎样在时间上变化所提供的信息。

虽然不能把时间上恒常的那些变量本身包括到固定效应模型中来，却能把它们同随时间而变的变量，特别是年虚拟变量交互起来分析。

定义 8.6 (**虚拟变量回归**) $dummyvariableregression$:对每个 i 估计一个截距，即连同诸解释变量在一起，给每一个横截面观测单元安排一个虚拟变量（也许还给每个时期安排有虚拟变量）。

- 它所给出的 β_j 的估计值，与除均值数据回归得到的相同，标准误与主要统计量雷同
- 拟合优度较高，因为每一横截面都有一个虚拟变量，这能解释数据中的变异的大部分

非观测效应模型：

$$y_{it}=\beta_0+\beta_1x_{it}+\alpha_i+u_{it}$$

不管使用 FE 还是 FD，目的都是要把 α_i 消去。这是因为，α_i 被认为是与解释变量中之一或全部相关。**但若是不相关，则变换会导致非有效的估计量**。如果假定非观测效应 α_i 与每一个解释变量都不相关：

$$Cov(x_itj,\alpha_i)=0$$

则方程为**随机效应模型**$randomeffectsmodel$。

如果我们相信 α_i 与解释变量不相关，则可用单一个横截面对 β_j 做一致性估计：根本不需要什么综列数据。

定义**复合误差项**为

$$v_i t = \alpha_i + u_i t$$

在**随机效应假定**下有

$$Cov(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \ t \neq s$$

由于通常的混合 OLS 标准误忽视了这种相关，以致标准误是不正确的，从而通常的检验统计量也不正确。

$$y_{it} = \alpha_{it} + x_{it}\beta_{it} + u_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T \tag{8-4}$$

对于截面成员方程，待估计参数的个数为 $NT(k+1)+N$，对于时间截面成员方程，待估计参数的个数为 $(NT(k+1)+T)$。所以自由度 NT 远远小于参数个数，这使得模型无法估计。为了实现模型的估计，可分别建立两类模型：从成员角度考虑的建立含有N 个个体成员方程的模型和从时间点上截面建立含有 T 个时间点截面方程的模型。

$$y_t = \alpha + X_t b + u_t \tag{8-5}$$

$$y_i = \alpha + x_i b + u_i \tag{8-6}$$

这个两类估计方法类似，一般讨论含有 N 个个体成员方程的模型。

根据截距项向量 α 和系数向量 β 中各分量的不同限制要求，可以将(8-5)分为三类：

- 无个体影响的不变系数模型，或联合回归模型 $pooledregressionmodel$(8-7)
- 变截距模型，或个体均值修正模型 $individual - meancorrectedregressionmode$(8-8)
- 变系数模型，或无约束模型 $unrestrictedmodel$(8-9)

$$y_{it} = \alpha + x_{it}\beta + u_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T \tag{8-7}$$

$$y_{it} = \alpha_i + x_{it}\beta + u_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T \tag{8-8}$$

$$y_{it} = \alpha_i + x_{it}\beta_i + u_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T \tag{8-9}$$

8.2 模型的设定检验

建立时间序列截面数据模型的第一步是检验被解释变量 y_{it} 的参数 α_i 和β_i 是否对所有个体样本和时间点都是常数。经常使用的检验是协方差分析检验。

首先计算(8-9)的残差平方和：

$$W_{xx,i} = \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i)$$
$$W_{xy,i} = \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(y_{it} - \bar{y}_i)$$
$$W_{yy,i} = \sum_{t=1}^{T} (y_{it} - \bar{y}_i)'(y_{it} - \bar{y}_i)$$

其中

$$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$
$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

则(8-9)的残差平方和记为

$$S_1 = \sum_{i=1}^{N} RSS_i = \sum_{i=1}^{N} (W_{yy,i} - W'_{xy,i}W_{xx,i}^{-1}W_{xy,i}) \tag{8-10}$$

其次计算(8-8)的残差平方和：记

$$W_{xx} = \sum_{i=1}^{N} W_{xx,i} \ W_{xy} = \sum_{i=1}^{N} W_{xy,i} \ W_{yy} = \sum_{i=1}^{N} W_{yy,i}$$

则(8-8)的残差平方和记为

$$S_2 = W_{yy} - W'_{xy}W_{xx}^{-1}W_{xy} \tag{8-11}$$

最后计算(8-5)的残差平方和：记

$$T_{xx} = \sum_{i=1}^{N}\sum_{t=1}^{T} (x_{it} - \bar{x})'(x_{it} - \bar{x})$$
$$t_{xy} = \sum_{i=1}^{N}\sum_{t=1}^{T} (x_{it} - \bar{x})'(y_{it} - \bar{y})$$
$$t_{yy} = \sum_{i=1}^{N}\sum_{t=1}^{T} (y_{it} - \bar{y})^2$$

其中

$$\bar{x} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it} \ \bar{y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$$

则(8-5)的残差平方和记为

$$S_3 = t_{yy} - t'_{xy}T_{xx}^{-1}t_{xy} \tag{8-12}$$

结论：

(1) $S_1/\sigma_u^2 \sim \chi^2[N(T-k-1)]$

(2) $S_3/\sigma_u^2 \sim \chi^2[NT-(k+1)]$ $(S_3-S_1)/\sigma_u^2 \sim \chi^2[(N-1)(k+1)]$

(3) $\chi^2[NT-(k+1)]$ 与 S_1/σ_u^2 独立。

$$F_2 = \frac{(S_3-S_1)/[(N-1)(k+1)]}{S_1/[NT-N(k+1)]}$$
$$F_1 = \frac{(S_2-S_1)/[(N-1)k]}{S_1/[NT-N(k+1)]}$$

8.3 数据类型非经典的计量经济学模型：面板数据模型

定义

定义 8.7 (**面板数据**) 面板数据（$PanelDate$），又称平行数据，是指在时间序列上取多个截面，在这些截面上取多个样本观测值构成的数据，以 Y_{it} 表示。其中，下标 i 表示不同的时间点，下标 t 表示不同的观测对象。

最早使用面板数据进行单位根检验的是 $Bhargava$ 等（$Bhargavaetal, 1982$）。他们利用修正的 DW 统计量提出了一种可以检验固定效应动态模型的残差是否为随机游走的方法。$Abuaf$ 和 $Jorion$ （1990）基于 SUR 回归($seeminglyunrelatedregression$)模型，采用 GLS 估计方法提出了面板单位根检验方法 $SUR-DF$ 检验。

平行数据的使用可以：

(1) 使研究人员区分出单用截面数据或时间序列数据都不能得到的经济作用；（规模经济与技术进步）

(2) 平行数据集通常含有许多数据点，会带来较大的自由度；

(3) 截面变量和时间变量的结合信息能够显著地减少缺省变量所带来的问题。

将截面数据和时间序列数据相混合的过程称为融合。

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \quad i = 1, 2, \cdots, N \quad t = 1, 2, \cdots, T \tag{8-13}$$

固定效应模型

问题 8.1　最小二乘融合方法的问题在于常数截距和常数斜率的假设可能性不合理。最显然的推广方法是引进允许截距项随时间和截面单位变化的虚拟变量。

$$Y_{it} = \alpha + \beta X_{it} + \gamma_2 W_{2t} + \cdots + \gamma_N W_{Nt} + \delta_2 Z_{i2} + \cdots + \delta_T Z_{iT} + \varepsilon_{it} \tag{8-14}$$

其中，如果是第 i 个个体 $i = 2, \cdots, N$，$W_{it} = 1$；如果是第 i 个时段 $i = 2, \cdots, N$，$Z_{it} = 1$。

在模型中添加了 $(N-1)+(T-1)$ 个虚拟变量。如果用普通最小二乘法估计模型，所有的参数估计都会是无偏和一致的。这里的自由度是 $NT-2-(N-1)-(T-1)$ 或 $NT-N-T$。

是否添加虚拟变量可以由统计假设检验决定：

$$F_{N+T-2, NT-N-T} = \frac{(ESS_1 - ESS_2)/(N+T-2)}{ESS_2/(N-T-2)}$$

注意：虚拟变量的使用不直接确认回归直线随机时间和个体变化的原因；虚拟变量的使用用掉较多的自由度。

相对于混合估计模型来说，是否有必要建立个体固定效应模型可以通过 F 检验来完成。原假设是 H_0 : 不同个体的模型截距项相同（建立混合估计模型）。备择假设是 H_1 : 不同个体的模型截距项不同（建立个体固定效应模型）。F 统计量定义为：

$$F = \frac{(SSE_r - SSE_u)/(N-1)}{SSE_u/(NT-N-1)} \tag{8-15}$$

SSE_r 表示约束模型混合估计模型，SSE_u 表示非约束模型个体固定效应模型的残差平方和。非约束模型比约束模型多了 $N-1$ 个被估参数。混合估计模型给出公共截距项。当模型中含有 k 个解释变量时，统计量的分母自由度是 $NT-N-k$。

随机效应模型 随机效应模型（或误差成分模型）：

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it} \tag{8-16}$$

$$\varepsilon_{it} = u_i + v_i + w_{it}$$

其中假设每一个误差成分相互之间不相关，也不存在自相关，即不横向相关，也不纵向相关。

把固定效应模型中的截距项看成两个随机变量，一个时间序列变量和一个截面变量，通过这样的分解就能看出随机效应模型和固定效应模型之间的关系。在固定效应模型中假设截距项

包含了随机时间序列变量和截面变量的平均效应，并且关于这个均值的离差分别就是误差成分 u_i 和 v_i，这就是随机效应模型。

$$Var(\varepsilon_{it}) = \sigma_u^2 + \sigma_v^2 + \sigma_w^2 \tag{8-17}$$

随机效应模型可以看作是介于零混合误差成分（固定效应模型）和无限大混合误差成分（普通最小二乘融合数据模型）两个极端之间的中间模型。

误差项与解释变量不相关的假设可以用 $Hausman$ 确认检验法进行检验，该检验比较固定效应模型的参数估计和随机效应模型的广义最小二乘估计。

可以把随机效应模型作为广义最小二乘回归来估计。用观测值方差的倒数作为观测值的权重。为了加权，采用两阶段估计法。第一阶段用普通最小二乘法和所有的混合数据进行估计，或采用固定效应模型，用估计的残差计算各误差成分的样本方差。第二阶段用这些估计方差得到参数的广义最小二乘估计。

相对于混合估计模型，是否有必要建立个体随机效应模型的检验是：

$$\begin{aligned} &H_0 : \sigma_u^2 = 0 \text{混合估计模型} \\ &H_1 : \sigma_u^2 \neq 0 \text{个体随机效应模型} \end{aligned} \tag{8-18}$$

统计量是:

$$LM = \frac{NT}{2(T-1)} \left(\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \hat{u}_{it} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1 \right) = \frac{NT}{2(T-1)} \left(\frac{T^2\, \bar{\hat{u}}'\bar{\hat{u}}}{\hat{u}'\hat{u}} - 1 \right)^2 \tag{8-19}$$

其中 $\bar{\hat{u}}'\bar{\hat{u}}$ 表示由个体随机效应模型计算的残差平方和，$\hat{u}'\hat{u}$ 表示由混合估计模型计算的残差平方和。统计量服从自由度为 1 的卡方分布。

例子

例 8.1　专利申请和研究开发支出。用 45 个公司 7 年间的平行数据估计专利申请的对数与研究开发支出之间的关系，其中研究开发数据有 5 年滞后。基本的回归模型为

$$P_{it} = \beta_0 + \beta_1 RND_{i,t-5} + \varepsilon_{it}$$

以混合数据集进行普通最小二乘估计。

源代码 8.1 SAS:混合回归

```
DATA TEMP;
      SET LI.RND;
      P=LOG(P);
RUN;
 /*OLS Estimation*/
PROC REG DATA=TEMP;
      MODEL P=RND;
RUN;
```

这个回归结果说明在研究开发支出与专利申请之间存在很强的正相关关系。

不同公司专利行为的表现之间有很大不同，对专利平均数（以时间平均）对平均研究开发支出进行回归。

源代码 8.2 SAS:截面回归

```
DATA RND;
      SET LI.RND;
      P=LOG(P);
RUN; PROC MEANS DATA=RND NORPRINT;
      CLASS CODE;
      VAR P RND;
      OUTPUT OUT=A SUM=TOTAL MEAN=MP MR;
RUN; DATA TEM;
      SET A;
      IF _N_=1 THEN DELETE;
RUN; PROC REG DATA=TEM;
      MODEL MP=MR;
RUN;
```

这个模型只能表现研究开发支出差异所引起的公司专利申请上的差异，无法表现引起专利申请差异的其他原因。为判明这个问题，向模型添加一组截面虚拟变量，即估计一个固定效应模型：

源代码 8.3 SAS:固定效应模型

```
DATA TEMP;
     SET LI.RND;
     P=LOG(P);
RUN;

%MACRO A;
     %DO I=2 %TO 45;
     DATA TEMP;
        SET TEMP;
        B=7*%EVAL(&I);
        A=B-6;
        IF A <=_N_ <= B THEN W&I=1; ELSE W&I=0;
        OUTPUT;
     RUN;
     %END;
%MEND A;

%A
PROC REG DATA=TEMP;
     MODEL P=RND W2-W45;
RUN;

/*F test*/ PROC REG DATA=TEMP NOPRINT OUTEST=A;
     M1:MODEL P=RND W2-W45 /SSE;
     M2:MODEL P=RND /SSE;
RUN; DATA _NULL_;
     SET A;
     IF FIRST._SSE_ THEN CALL SYMPUT('ESS2',_SSE_);
                    ELSE CALL SYMPUT('ESS1',_SSE_);
RUN; DATA _NULL_;
     Q1=FINV(0.95,50,269);
     Q2=((&ESS1-&ESS2)/50)/(&ESS2/269);
     PUT Q1= Q2=;
RUN;
```

估计含有截面误差成分和混合误差成分的随机效应模型：

源代码 8.4 SAS：随机效应模型

为什么使用综列数据？

(1) 综列数据与一定时期内的个体等有关，那么这些个体中一定存在差异性。综列数据估计的技巧能够通过对特定个体变量的考虑而将这种差异性非常清晰地加以研究。

(2) 综列数据提供更加有价值的数据，变量之间增加了多变性和减少了共线性，并且提高了自由度和有效性

(3) 综列数据更适用于变化中的动态的研究

(4) 综列数据能够更好地检测和度量单纯使用横截面数据或时间序列数据无法观测到的影响

(5) 综列数据能够使我们对更加复杂的行为模型进行研究。

(6) 综列数据适用于大量个体单位，将偏差降到最低。

第九章　非参数回归计量经济学模型理论与方法

9.1　前言

一组数据的最基本的信息就是次序．如果可以把数据点按大小次序排队，每一个具体数目都有它的在整个数据中(从最小的数起)的位置或次序，称为该数据的**秩**($rank$)。数据有多少个观察值，就有多少个秩．在一定的假定下，这些秩和它们的统计量的分布是求得出来的，而且和原来的总体分布无关．这样就可以进行所需要的统计推断。

定义 9.1 (**顺序统计量**)　对于样本 $X_1, X_2, \cdots, X_n$，如果按照升幂排列，得到

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

则称 $X_{(k)}$ 是顺序统计量。

定义 9.2 (**顺序统计量分布函数**)　设总体的分布函数是 $F(x)$，则第 r 个顺序统计量的分布函数是：

$$\begin{aligned}
&F_r(x) \\
&= P(X_{(r)} < x) \\
&= P(\text{至少}r\text{个}X_i\text{小于或等于}x) \\
&= \sum_{i=r}^{n} C_n^i F^i(x)[1-F(x)]^{n-i}
\end{aligned}$$

$$=$$

顺序统计量分布函数与密度函数：

$$F_r(x) = \frac{n!}{(r-1)!(n-r)!} \int_0^{F(x)} t^{r-1}(1-t)^{n-r} dt$$

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!} F^{r-1}(x)\left[1-F(x)\right]^{n-r} f(x)$$

9.2　非参数统计讲义

[21]. **计数统计量和秩统计量**

定义 9.3　设随机变量 $X_1, X_2, \cdots, X_n$ 是来自总体 $F(x)$ 的样本，一切可能的 $F(x)$ 组成分布类 $\mathscr{F}$。如果统计量 $T(X_1, X_2, \cdots, X_n$ 对任意的 $F \in \mathscr{F}$ 均有相同的分布，则称 T 关于 $\mathscr{F}$ 是**适应任意分布的**(distribution-free)。

在做假设检验时，我们总要找一个对原假设 H_0 包含的分布类是适应任意分布的统计量。

当考虑大样本统计推断时，还会考虑渐近适应任意分布的统计量。

定义 9.4　设随机变量 $X_1, X_2, \cdots, X_n$ 是来自总体 $F(x)$ 的样本，一切可能的 $F(x)$ 组成分布类 $\mathscr{F}$。如果统计量 $T(X_1, X_2, \cdots, X_n$ 对任意的 $F \in \mathscr{F}$ 均有相同的极限分布，则称 T 关于 $\mathscr{F}$ 是**渐近适应任意分布的**(distribution-free)。

定义 9.5　设 X 是一个随机变量，对一给定的实数 θ_0，定义随机变量

$$\Psi = \psi(X - \theta_0)$$

其中

$$\psi(t) = \begin{cases} 1 & t > 0 \\ 0 & t < 0 \end{cases}$$

随机变量 Ψ 称为 X 按 θ_0 分段的**计数统计量**。

定理 9.1 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立。X_i 有分布 $F_i(x)$。又对任意 i 均有 $F_i(\theta_0)$ 均有 $F_i(\theta_0) = p_0$，则 $\Psi = \psi(X_i - \theta_0$ 是相互独立同分布的随机变量，其共同分布是参数为 $(1-p_0)$ 两点分布。

符号检验在实际工作中常用于对比检验。检验用计数统计量是

$$B = \sum_{i=1}^{n} \Psi_i = \sum_{i=1}^{n} \psi(X_i)$$
$$H_0 : F(0) = \frac{1}{2} \qquad H_1 : F(0) \neq \frac{1}{2}$$

利用符号统计量可以构造总体中位数的置信区间。

定义 9.6 设 $Z_1, Z_2, \cdots, Z_n$ 是来自连续分布 $F(z)$ 的简单随机样本，$Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(N)}$ 为次序统计量，定义随机变量

$$R_i = r, \qquad 当 Z_i = Z_{(r)} \quad i = 1, 2, \cdots, N$$

当 R_i 是唯一确定时，称样本观测值 Z_i 有**秩** R_i。它是第 i 个样本单元 Z_i 在样本次序统计量 $(Z_{(1)}, Z_{(2)}, \cdots, Z_{(N)}$ 中的位置。

定义 9.7 (秩统计量) 设 $x_1, \cdots, x_n$ 为两两互不相等的实数。若在 $x_1, \cdots, x_n$ 中恰好有 r_i 个的值不超过 x_i，则称 r_i 为 x_i 在 $x_i, \cdots, x_n$ 中的秩。若 $X_1, \cdots, X_n$ 为来自单个连续型总体的样本，或来自多个连续型的合样本，R_i 国 X_i 在 $(X_1, \cdots, X_n)$ 中的秩，则称 $R = (R_1, \cdots, R_n)$ 为 $(X_1, \cdots, X_n)$ 的**秩统计量**。

定理 9.2 记 $R = (R_1, R_2, \cdots, R_N)$，集合

$$\mathscr{R} = \{(r_1, \cdots, r_N) | (r_1, \cdots, r_N) 是 (1, \cdots, N) 的一个排列\}$$

则 R 在 $\mathscr{R}$ 上均匀分布。

仅依赖 R 的统计量 $S(R)$ 关于连续分布构成的分布类是适应任意分布的。

$$P\{R = r\} = \frac{1}{N!}$$
$$P\{R_i = r\} = \frac{1}{N}$$
$$P\{R_i = r, R_j = s\} = \frac{1}{N(N-1)}$$

$t_{m,n}(d)$ 表示从 $1, 2, \cdots, m+n$ 中取 n 个数其和恰好为 d 的可能取法的个数。计算 $t_{m,n}(d)$ 可利用下列递推公式：

$$t_{m,n}(d) = t_{m,(n-1)}(d - m - n) + t_{(m-1),n}(d)$$
$$t_{i,0}(0) = 1; t_{i,0}(d) = 0$$
$$t_{0,j}(\frac{j(j+1)}{2}) = 1; t_{0,j}(d) = 0$$

定义 9.8 随机变量 $|X_1|, \cdots, |X_n|$ 对应的秩向量记为 $(R^+, \cdots, R_n^+)$，R_i^+ 称为 X_i 的**绝对秩**，$\Psi_i R_i^+$ 称为 X_i 的**符号秩**。

只依赖于 $\Psi_1, \cdots, \Psi_n, R^+$ 的统计量 $S(\Psi_1, \cdots, \Psi_n, R^+)$，对于关于 0 点对称的连续分布构成的分布类是适应任意分布的。

任一由 $\Psi_1, \cdots, \Psi_n$ 构成的统计量 $S(\Psi_1, \cdots, \Psi_n)$ 关于一切 p_0 分位点为 θ_0 (θ_0是连续点，$F_i(\theta_0) = p_0$)的分布类是适应任意分布的。

9.3 Histogram

Statistical Properties the histogram as an estimator of the unknown pdf $f(x)$

the origin $x_0 = 0$

some point $x \in B_j = [(j-1)h, jh)$

the density estimate assigned by the histogram to x is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} I(X_i \in B_j)$$

Bias Is the histogram an unbiased estimator?

$$E\{\hat{f}_h(x)\} = \frac{1}{nh} \sum_{i=1}^{n} E\{I(X_i \in B_j)\}$$

X_i is $i.i.d \Longrightarrow$ the indicator function I is $i.i.d$

$$E\{\hat{f}_h(x)\} = \frac{1}{nh} nE(I(X_i \in B_j))$$

$$I(X_i \in B_j) = \begin{cases} 0 \ p \ \textbf{is} \ 1 - \int_{(j-1)h}^{jh} f(u)du \\ \quad 1 \ p \ \textbf{is} \ \int_{(j-1)h}^{jh} f(u)du \end{cases}$$

hence it is Bernoulli distributed:

$$\begin{aligned} E\{I(X_i \in B_i)\} &= P\{I(X_i \in B_i\} \\ &= \int_{(j-1)h}^{h} f(u)du \end{aligned}$$

then

$$E(\hat{f}_h(x)) = \frac{1}{h} \int_{(j-1)h}^{h} f(u)du$$

the last term is not in general equal to the $f(x)$ hence, the histogram is in general not an unbiased estimator of the $f(x)$:

$$\begin{aligned} Bias(\hat{f}_h(x)) &= E(\hat{f}_h(x) - f(x)) \\ &= \frac{1}{h} \int_{(j-1)h}^{h} f(u)du - f(x) \\ &= \frac{1}{h} \int_{(j-1)h}^{h} [f(u) - f(x)]du \end{aligned}$$

a first-order Taylor approximation of $f(x) - f(u)$ around the center $m_j = (j - \frac{1}{2})h$ yields:

$$Bias(\hat{f}_h(x)) \approx f'(m_j)(m_j - x)$$

if m_j is equal to the x?

Variance

$$Var\{\hat{f}_h(x)\} = \frac{1}{n^2h^2} n \int_{B_j} f(u)du \left(1 - \int_{B_j} f(u)du\right)$$
$$\approx \frac{1}{nh} f(x)$$

integral:

$$\int_a^b f(x)dx = \lim_{\lambda \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i$$

Mean squared error Variance and bias vary opposite directions with h,we have to settle for finding the value of h that yields the optimal compromise between variance and bias reduction.

the mean squared error($\boldsymbol{MSE}$) of the histogram:

$$MSE\{\hat{f}_h(x)\} = E\{[\hat{f}_h(x) - f(x)]^2\}$$

which can be written as the sum of the variance and the squared bias:

$$MSE\{\hat{f}_h(x)\} = Var(\hat{f}_h(x)) + [Bias(\hat{f}_h(x))]^2$$

Mean integrated squared error the application of the !MSE is difficult in practice since the derived formula for the MSE depends on the unknown density function f both in the variance and the squared bias term. the most widely used global measure of estimation accuracy is the $meanintergratedsquarederror$:

$$\begin{aligned} MISE(\hat{f}_h(x)) &= E\left[\int_{-\infty}^{+\infty} \{\hat{f}_h(x) - f(x)\}^2 dx\right] \\ &= \int_{-\infty}^{+\infty} E[\{\hat{f}_h(x) - f(x)\}^2]dx \\ &= \int_{-\infty}^{+\infty} MSEdx \end{aligned}$$

$$\begin{gathered} MISE(\hat{f}_h(x)) \approx \int \tfrac{1}{nh} f(x)dx \\ + \int \sum_j I(x \in B_j)\{(j - \tfrac{1}{2})h - x\}^2 [f'\{(j - \tfrac{1}{2})h\}]^2 dx \\ = \tfrac{1}{nh} + \sum_j \int_{B_j} \{(j - \tfrac{1}{2})h - x\}^2 [f'\{(j - \tfrac{1}{2})h\}]^2 dx \\ \approx \tfrac{1}{nh} + \sum_j [f'\{(j - \tfrac{1}{2})h\}]^2 \bullet \int_{B_j} \{x - (j - \tfrac{1}{2})h\}^2 dx \\ \approx \tfrac{1}{nh} + \tfrac{h^2}{12} \int \{f'(x)\}^2 dx \\ \underline{\underline{h \to 0}} \tfrac{1}{nh} + \tfrac{h^2}{12} \left\| f' \right\|_2^2 \end{gathered}$$

the asymptotic of MSE :

$$AMSE(\hat{f}_h(x)) = \frac{1}{nh} + \frac{h^2}{12} \left\| f' \right\|_2^2$$

optimal $Binwidth$ differentiating $AMSE$ with respect to h gives:

$$\begin{aligned} \tfrac{\partial AMSE(\hat{f}_h(x))}{\partial h} &= -\tfrac{1}{nh^2} + \tfrac{1}{6} h \left\| f' \right\|_2^2 \\ &= 0 \end{aligned}$$

hence

$$h_0 = \left(\frac{6}{n \left\| f' \right\|_2^2}\right)^{\frac{1}{3}} \sim n^{-\frac{1}{3}}$$

assuming a standard norm distribution:

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

and

$$f'(x) = (-x)f(x)$$

$$\begin{gathered} \|f'\|_2^2 = \|(-x)f(x)\|_2^2 \\ = \{\int x^2 f^2(x)dx\}^{\frac{1}{2}\times 2} \\ = \frac{1}{\sqrt{2\pi}} \int x^2 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2} \times 2)dx \\ = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{2}} \times \int x^2 \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{\frac{1}{2}}} \times \exp(-\frac{x^2}{2\times\frac{1}{2}})dx \\ = \frac{1}{4\sqrt{\pi}} \end{gathered}$$

Dependence of the histogram on the origin this property strongly conflicts with the goal of nonparametric statistics to "let the data speak for themselves"

ASH**:averaged shifted histogram**

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\{\frac{1}{Mh}\sum_{l=0}^{M-1}\sum_{j} I(X_i \in B_{jl})I(x \in B_{jl})\}$$

9.4 Nonparametric density estimation

motivation and derivation

even if the ASH seemed to solve the choice-of-an-origin problem the histogram retains some undesirable properties:

(1) the histogram assigns each x in $[m_j - \frac{h}{2}, m_j + \frac{h}{2})$ the same estimator of the pdf,namely $\hat{f}_h(m_j)$.This seems to be overly restrictive.

(2) The histogram is not a continuous function but has jumps at the boundaries of the bins.

{observations that fall into a small interval containing x }

{observations that fall into a small interval around x }

we consider intervals of the form $[x-h, x+h)$:

$$\hat{f}_h(x) = \frac{1}{2hn}\#\{X_i \in [x-h, x+h)\}$$

the formula can be rewritten if we use a weighting function ,called the uniform kernel function:

$$K(u) = \frac{1}{2}I(|u| \le 1)$$

and let

$$u = \frac{x - X_i}{h}$$

we can write:

$$\begin{gathered} \hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h}) \\ = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{2}I(\left|\frac{x-X_i}{h}\right| \le 1) \end{gathered}$$

maybe we should give more weight to contributions from observations very close to x than to those coming from observations that are more distant.

for instant,Epanechniko kernel:

$$K(\bullet) = \frac{3}{4}(1-u^2)I(|u| \le 1)$$

based on a random sample

$$\{X_1, X_2, \cdots, X_n\}$$

from f

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

where

$$K_h(\bullet) = \frac{1}{h}K(\frac{\bullet}{h})$$

h controls the smoothness of the estimate and the choice of the h is a crucial problem.

the rescaled kernel function is simply:

$$\frac{1}{nh}K(\frac{x - X_i}{h}) = \frac{1}{n}K(x - X_i)$$

$$\int \frac{1}{nh}K(\frac{x - X_i}{h})dx = \frac{1}{nh}\int hK(u)du = \frac{1}{n}$$

Statistic Properties

Bias

$$\begin{aligned} & Bias\{\hat{f}_h(x)\} = E\{\hat{f}_h(x)\} - f(x) \\ & = \tfrac{1}{n}\sum_{i=1}^{n} E\{K_h(x - X_i)\} - f(x) \\ & = E\{K_h(x - X)\} - f(x) \\ & = \int \tfrac{1}{h}K(\tfrac{x-u}{h})f(u)du - f(x) \end{aligned}$$

a second-Taylor expansion of $f(u)$ around x:

$$Bias\{\hat{f}_h(x)\} = \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2)$$

$$\mu_2(K) = \int s^2K(s)ds$$

the bias is proportional to h^2.moreover,the bias is depend on the curvature of the density at x.

Variance

$$\begin{aligned} & Var\{\hat{f}_h(x)\} \\ & = Var\{\tfrac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\} \\ & = \tfrac{1}{n}Var\{K_h\} \\ & = \tfrac{1}{n}\{EK_h^2 - (EK_h)^2\} \end{aligned}$$

using

$$EK_h = f(x) + o(h)$$

and

$$EK_h^2 = \int K_h^2 f(u)du$$

and variable substitution and Taylor expansion arguments in the derivation of the bias:

$$Var\{\hat{f}_h(x)\} = \frac{1}{nh}\|K\|_2^2 f(x) + o(\frac{1}{nh})$$

$\|K\|_2^2$ is shorthand for $\int K^2(s)ds$,the squared L_2 norm of K.

MSE

$$\begin{aligned} & MSE\{\hat{f}_h(x)\} \\ & = \tfrac{h^4}{4}f''(x)^2\mu_2(K)^2 + \tfrac{1}{nh}\|K\|_2^2 f(x) + o(h^4) + o(\tfrac{1}{nh}) \end{aligned}$$

$h_opt(x)$ depends on x and is thus a local bandwidth.

AMSE

$$AMSE = \int MSE dx$$

$$\begin{aligned}&AMSE\\&= \int MSEdx\\&= \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\{\mu_2(K)\}^2\|f''\|_2^2\end{aligned}$$

differentiating $AMSE$ with respect to h and solving the first-order condition for h yields the $AMSE$ optimal bandwidth:

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2\{\mu_2(K)\}^2 n}\right)^{1/5} \sim n^{-\frac{1}{5}}$$

$$\begin{aligned}&AMSE(\hat{f}_{hopt})\\&= \tfrac{5}{4}(\|K\|_2^2)^{4/5}(\mu_2(K)\|f''\|_2)^{2/5}\end{aligned}$$

$AMSE$ is converging at the rate of $n^{-4/5}$;$histogram$ is converging at the rate of $n^{-2/3}$

9.5 无参数回归模型

经典计量经济学模型首先根据经济理论和样本数据设定模型的函数关系，然后估计关系参数并检验所设定的关系。如果模型的函数关系通过检验被证明是成立的，那么回归结果可以外延，其推断有较高的精度，模型的参数一般具有明确的经济意义，可以方便于各方面的应用。但关于模型及参数的一些假定在现实中未必成立。例如$C-D$生产函数模型假定技术进步是中性的、技术进步独立于要素投入量的变化、要素替代弹性为1、具有一次齐次性即不变规模报酬等，这些假定在现实中很难同时成立。因而当模型及参数的假定与实际背离（也包括模型的随机干扰项的正态性假定与实际背离）时，就容易造成模型设定误差。此时，基于经典假设模型所作出的推断的表现可能很差，这就促使人们寻找别的出路，而无参数回归模型则是朝着这个方向的一种努力。

无参数回归模型的估计方法有三大类，一是权函数方法，二是最小二乘估计，三是稳健估计。无参数回归模型的特点是：回归函数的形式可以任意,没有任何约束，解释变量和被解释变量的分布也很少限制，因而有较大的适应性，但无参数回归结果外延困难。无参数技术的目的是放松回归函数形式的限制，为确定或建议回归函数的参数表达式提供有用的工具。无参数技术并不能取代参数技术，两者相结合将会得到用单一方法无法获得的结论。

定义 9.9 (**无参数回归模型**) 无参数回归模型指对 Y 对 X 条件回归函数 $m(x) = E(Y|X = x)$ $x \in R^q$ 依据样本 $\{(X_i, Y_i)_{i=1}^n\}$ 进行估计，得出 $\hat{m}_n(x)$。如果 X 是随机变量，且联合密度函数 $f(X, Y)$ 存在，则模型可表示为

$$m(x) = \int yf(x, y)dy/f(x)$$

$f(x)$ 是 X 的边缘密度函数；如果 X 是确定性变量，模型可表示为

$$Y_i = m(x_i) + \varepsilon_i$$

权函数方法 非参数回归的基本方法有核函数法，最近邻函数法，样条函数法，小波函数法。这些方法尽管起源不一样，数学形式相距甚远，但都可以视为关于 Y_i 的线性组合的某种权函数。也就是说，回归函数 $g(X)$ 的估计 $g_n(X)$ 总可以表为下述形式：

$$g_n(X) = \sum_{i=1}^{n} W_i(X)Y_i$$

其中 $\{W_i(X)\}$ 称为权函数。在一般实际问题中，权函数都满足下述条件：

$$W_i(X;X_1,\cdots,X_n)\geqslant 0,\quad \sum_{i=1}^{n}W_i(X;X_1,\cdots,X_n)=1$$

权函数方法的系统研究始于 $Stone(1977)$。

核函数法 $Nadaraya(1964)$及 $Watson(1964)$ 提出了一种既适合解释变量是确定性变量，也适合解释变量是随机变量的核估计：

选定概率密度

$$\int K(u)du=1$$

为核函数及窗宽 $h>0$。定义核权函数

$$W_{ni}(x)=K_h(x-X_i)/\sum_{j=1}^{n}K_h(x-X_j)$$

其中

$$K_h(u)=\frac{1}{h}K(\frac{u}{h})$$

是一个概率密度函数。

$Nadaraya-Watson$ 核估计是

$$\hat{m}_n(x)=\frac{\sum_{i=1}^{n}K_h(x-X_i)Y_i}{\sum_{j=1}^{n}K_h(x-X_j)}$$

核估计等价于局部加权最小二乘估计。

$$W_i(X;X_1,\cdots,X_n)=K\left(\frac{X-X_i}{a_n}\right)/\sum_{i=1}^{n}\left(\frac{X-X_i}{a_n}\right)$$

$$Y=g(X)=\sum_{i=1}^{n}W_i(X)Y_i=\sum_{i=1}^{n}\frac{K\left(\frac{X-X_i}{a_n}\right)}{\sum\limits_{j=1}^{n}K\left(\frac{X-X_i}{a_n}\right)}Y_i$$

核函数 直方图估计密度 密度估计最基本的方法是直方图估计。先用点 $\{a_i\}_{i=1}^{k}$ 把直线划分为 k 个小区间，计数区间的端点和宽度都是固定的。记 N_i 为样本点 $X_1,X_2,\cdots,X_n$ 落在第 i 个计数区间 $[a_i,a_{i+1})$ 的个数。则密度函数 $f(x)$ 在区间 $[a_i,a_{i+1})$ 的函数估计值是：

$$f_n(x)=\frac{N_i}{n(a_{i+1}-a_i)},a_i\leqslant x<a_{i+1},i=1,\cdots,k$$

计数区间端点可变 这样的直方图估计当然是阶梯函数，于是人们想法改进它。这种直方图估计方法在估计中心比较精确而在估计区间的端点处精度较差。有人提出，对每个 x，各作一个以 x 为中点的小计数区间 $[x-h,x+h]$，再对落在该区间的样本点进行计数，设为 $N(x,h)$，则密度估计为：

$$f_n(x)=\frac{N(x,h)}{2nh}$$

这个想法与直方图不同在于它的计数区间端点划分不是固定的，而是随 x 而变，可以自始至终保持 x 点在计数区间中间。**计数区间宽度可变** 不过此时计数区间宽度 h_n 一般是固定的。如果引进函数：

$$K(x)=\begin{cases}\frac{1}{2} & -1\leqslant x<1\\ 0 & else\end{cases}$$

则上述变端点计数区间的密度估计可写为

$$f_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_n}{h}\right)$$

后来 $Parzen(1962)$ 提出，可以将这种矩形核函数形式放宽限制，只须积分为 1 (最好还为恒正)即可。这就导出了密度的核估计。**核函数性质** 一般来说，要求核函数满足：

$$\begin{cases} K(x) \geqslant 0, \int_{-\infty}^{+\infty} K(x)dx = 1 \\ \sup K(x) < +\infty, \int_{-\infty}^{+\infty} K^2(x)dx < +\infty \\ \lim\limits_{x\to\infty} K(x)\cdot x = 0 \end{cases}$$

对于窗宽，显然样本点越多，窗宽越小，即窗宽是样本点的函数：

$$\lim_{x\to\infty} h(n) = 0, \lim_{n\to\infty} nh(n) = n \to \infty$$

在上述要求的核函数及窗宽条件下，密度 $f(x)$ 的核估计 $f_n(x)$ 是 $f(x)$ 的渐近无偏估计与相合估计。

密度核估计 条件回归函数的导数期望估计 $ADE(Averagederivativeestimation)$

经验分布函数为：

$$F_n(x) = \frac{1}{n}(X_1, X_2, \cdots, X_n\text{中小于}x\text{的个数})$$

取核函数为均匀核：

$$K_0(x) = \begin{cases} \frac{1}{2} & if -1 \leqslant x < 1 \\ 0 & else \end{cases}$$

密度函数估计为

$$\begin{aligned} & \hat{f}_n(x) \\ & = [F_n(x+h) - F_n(x-h)]/2h \\ & = \frac{1}{2h}\int_{x-h}^{x+h} dF_n(t) \\ & = \int_{-\infty}^{+\infty} \frac{1}{h}K_0(\frac{x-t}{h})dF_n(t) \\ & = \frac{1}{nh}\sum_{i=1}^{n} K_0(\frac{x-X_i}{h}) \end{aligned}$$

一般密度核估计：

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h})$$

回归函数的导数的核估计采用 $Priestley-Chao$ 核权定义为

$$\hat{m}_n^{(k)}(x) = h^{-(k+1)}\sum_{i=1}^{n}(X_i - X_{i-1})K^{(k)}(\frac{x-X_i}{h})Y_i$$

P 维随机向量的核函数估计为

$$\hat{f}_n(x) = \frac{1}{nh^p}\sum_{i=1}^{n} K(\frac{x-X_i}{h})$$

最近邻函数法

(1) 首先引进一个距离函数:

$$||u-\upsilon||^2=\sum_{i=1}^{n}(u_i-\upsilon_i)^2$$

或者

$$||u-\upsilon||=\max_{1\leqslant i\leqslant n}||u_i-\upsilon_i||$$

(2) 为了反映各分量的重要程度，引进权因子 $\{C_i\}$ 并满足配方条件

(3) 距离函数改进

$$||u-\upsilon||^2=\sum_{i=1}^{n}C_i(u_i-\upsilon_i)^2$$

或者

$$|u-\upsilon||^2=\max_{1\leqslant i\leqslant n}C_i|u_i-\upsilon_i|$$

假设有样本 (Y_i,X_i)，并指定空间中作任一点 X，估计回归函数在该点的值:

(1) 排序

$$||X_{k_1}-X||<||X_{k_2}-X||<\cdots<||X_{k_n}-X||$$

表示点 X_{k_1} 与 X 的距离最近，依次类推。

(2) 这里的 n 个权函数也满足配方条件，并按从大到小排序

$$k_1\geqslant k_2\geqslant\cdots\geqslant k_n>0,\ \sum_{i=1}^{n}k_i=1$$

即

$$W_{k_i}(X;X_1,\cdots,X_n)=k_i,\ i=1,\cdots,n$$

(3) 最近邻回归函数

$$Y=g(X)=\sum_{i=1}^{n}W_i(X;X_1,\cdots,X_n)Y_i=\sum_{i=1}^{n}k_iY_i=\sum_{i=1}^{n}k_i(X)Y_i$$

权函数估计的矩相合性 如果对样本 $(Y_i,X_i),i=1,2,\cdots,n$，构造了权函数 $W_i=W_i(X)=W_i(X,X_1,\cdots,X_n)$，有了回归函数 $g(X)$ 的权函数估计

$$g_n(X)=\sum_{i=1}^{n}W_iY_i$$

当 Y 的 r 阶矩存在时，若

$$\lim_{n\to\infty}E|g_n(X)-g(X)|^r=0$$

则称这样的权函数为矩相合的权函数。

在什么样的条件下构造的权函数是矩相合的呢? Stone(1977)提出了很一般的，几乎是充分必要的条件。

定理 9.3 设概率权 $\{W_i\}$ 满足下列条件: (1)存在有限常数 C，使对 R^m 上任何非负可测函数(连续函数与分段连续函数是最常见的可测函数) f，必有

$$E\left(\sum_{i=1}^{n}W_if(X_i)\right)\leqslant CEf(X)$$

因为权函数是概率权，必有 $|W_i| < 1, i = 1, 2, \cdots, n$，于是

$$\left| E\left(\sum_{i=1}^{n} W_i f(X_i)\right) \right| \leqslant E\sum_{i=1}^{n} |W_i|\, |f(X_i)| \leqslant E\sum_{i=1}^{n} |f(X_i)| = E\sum_{i=1}^{n} f(X_i)$$

这里常数 C 取值为1。(2) $\forall \varepsilon > 0$，当 $n \to \infty$ 时，

$$\sum_{i=1}^{n} W_i I_{(\|X_i - X\| \geqslant \varepsilon)} \xrightarrow{P} 0$$

(3) 当 $n \to \infty$时

$$\max_{1 \leqslant i \leqslant n} W_i \xrightarrow{P} 0$$

则 $\{W_i\}$ 是矩相合的权函数。

引理 9.1 设概率权函数$\{W_i\}$ 适合定理条件（1）和（2），又对某个 r，有 $E|f(x)|^r < \infty$，则

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i(X)\, |f(X_i) - f(X)|^r\right) = 0$$

$$M = \sup_X |f(X)|$$

由条件(2)，上式右边第二项依概率收敛于0且不大于1。依控制收敛定理有

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i(X) I_{(\|X_i - X) > \varepsilon}\right) = 0$$

故存在 n_0，当 $n > n_0$，有：

$$E\left(\sum_{i=1}^{n} W_i(X) I_{(\|X_i - X) > \varepsilon}\right) \leqslant \frac{\eta}{2}$$

有：

$$E\left(\sum_{i=1}^{n} W_i(X)|f(X_i) - f(X)|^r\right) \leqslant \eta$$

对一般的函数 f，取一个在 R_m 上连续且在一有界域外为零的函数 $\tilde{f}$，使 $E|\tilde{f}(X)|^2 < \infty$，且 $E|f(x) - \tilde{f}(x)|^r < \eta$。因为：

$$\begin{aligned} E\left(\sum_{i=1}^{n} W_i(X)\, |f(X_i) - f(X)|^r\right) \leqslant 3^{r-1} &\left\{ \left(\sum_{i=1}^{n} W_i(X)|f(X_i) - \tilde{f}(X)|^r\right) \right. \\ &\left. + E\left(\sum_{i=1}^{n} W_i(X)|\tilde{f}(X_i) - f(X_i)|^r\right) + E\left(\sum_{i=1}^{n} W_i(X)|\tilde{f}(X) - f(X)|^r\right) \right\} \end{aligned}$$

第一项根据条件(1)不超过

$$CE\left|\tilde{f}(X) - f(X)\right|^r < C_\eta$$

当$n \to \infty$，第二项将趋近于零。第三项等于

$$E\left|\tilde{f}(X) - f(X)\right|^r < \eta$$

。因此，

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i(X)|f(X) - f(X_i)|^r\right) \leqslant 3^{r-1}(C+1)\eta$$

。

定理 9.4 (Lebesgue控制收敛定理)　设 $f_n(x)$ 为 E 上可测函数列，

$$\lim_{n\to\infty} f_n(x) = f(x)\ a.e\ E$$

且存在非负可积函数 $F(x)$，使得

$$f_n(x) \leqslant F(x)\ a.e\ E$$

则 $f(x)$ 在 E 上可积，且

$$\lim_{n\to\infty}\int_E f_n(x)dx = \int_E \lim_{n\to\infty} f_n(x)dx$$

引理 9.2　设 $\{W_i\}$ 为满足定理三个条件的概率权，函数 f 非负且 $Ef(x) < \infty$，则有：

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i^2(X) f(x_i)\right) = 0$$

证明：定义一组新的概率权函数 $W_i' = W_i^2$，因为 $0 \leq W_i' \leq 1$，所以 $0 \leq W_i^2 \leq 1$。由上一个引理我们有：

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i^2(X)\left|f(x) - f(x_i)\right|\right) = 0$$

由定理中的条件三和 $0 \leq W_i^2 \leq 1$ 我们有：

$$\sum_{i=1}^{n} W_i^2 \leqslant (\max_{1\leqslant i\leqslant n} W_i)\sum_{i=1}^{n} W_i = \max_{1\leqslant i\leqslant n} W_i \xrightarrow{P} 0$$

由控制收敛定理有：

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} W_i^2(X) f(X)\right) = 0$$

窗宽 选择窗宽的交错鉴定方法、惩罚函数法和插入法。

平方拟合误差

$$p(h) = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{m}_n(X_i))^2 w(X_i)$$

使平均平方误差最小：

$$d_A(h) = n^{-1}\sum_{i=1}^{n}(\hat{m}_n(X_i) - m(X_i))^2 w(X_i)$$

$$p(h) = n^{-1}\sum_{j=1}^{n}\varepsilon_j^2 w(X_j) + d_A(h) - 2n^{-1}\sum_{j=1}^{n}\varepsilon_j(\hat{m}_n(X_j) - m(X_j))w(X_j)$$

最后一项可重写为：

$$C_{1n}(h) = -2n^{-1}\sum_{j=1}^{n}\varepsilon_j\left[n^{-1}\sum_{i=1}^{n} W_{ni}(X_j)Y_i - m(X_j)\right]w(X_j)$$

其在给定 X 下的条件期望为

$$-2n^{-1}\sum_{j=1}^{n}[n^{-1}W_{nj}(X_j)\sigma^2(X_j)]w(X_j)$$

稳健估计

最小二乘估计 最小二乘法是确定性解释变量无参数回归模型估计的一种主要方法，主要有正交序列估计、样条估计和小波估计。

正交序列估计

样条估计

小波估计

局部多项式回归估计 正交序列估计和样条估计都需要估计出一系列的参数。为了减少参数估计的个数，可以采用局部线性回归估计，即对于给定的 x，认为 $m(x)$ 在 x 的局部邻域近似于线性的。

一元局部多项式回归 假设 $m(x)$ 在 x_0 处 $p+1$ 阶导数存在，进行台劳展开：

$$m(x) \approx m(x_0) + m'(x_0)(x-x_0) + \cdots + \frac{m^{(p)}(x_0)}{p!}(x-x_0)^p$$

从而

$$Y_i = m(x_0) + m'(x_0)(X_i - x_0) + \cdots + \frac{m^{(p)}(x_0)}{p!}(X_i - x_0)^p + \varepsilon_i$$

用加权最小二乘法进行局部拟合，即最小化

$$\sum_{i=1}^{n}\{Y_i - \sum_{j=0}^{p}\beta_j(X_i - x_0)^j\}^2 K_h(X_i - x_0) I(\frac{|X_i - x_0|}{h} \leqslant 1)$$

若 $K(\cdot)$ 是 $[-1,1]$ 上的核函数，上式可简写为

$$\sum_{i=1}^{n}\{Y_i - \sum_{j=0}^{p}\beta_j(X_i - x_0)^j\}^2 K_h(X_i - x_0)$$

矩阵形式是

$$Y = X\beta + \varepsilon$$

其中

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_1 - x_0 & \cdots & (X_1 - x_0)^p \\ 1 & X_2 - x_0 & \cdots & (X_2 - x_0)^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n - x_0 & \cdots & (X_n - x_0)^p \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

参数 β 的加权最小二乘估计是：

$$\min_{\beta}(Y - X\beta)'W(Y - X\beta)$$

$$\hat{\beta} = (X'WX)^{-1}X'Wy$$

其中

$$W = diag\{K_h(X_i - x_0)\}$$

局部回归方法的优点 局部多项式拟合从理论和实践上都是吸引人的。第一，传统回归方法将经济变量局部上的变化差异掩盖了，因而无法反映经济现象的结构变化。传统回归方法处理有结构变化的经济变量之间关系时常用的手法是引入虚拟变量，但虚拟变量只能反映两个时期或若干已知时期经济现象的结构变化。而局部回归的结果能够动态地反映经济现象的结构变化。

第二，以往根据经济理论建立的线性或非线性计量模型，隐含着若干假设条件。例如，$C-D$ 生产函数，隐含着技术中性、规模报酬不变等假设。而这些隐含条件对于具体问题

不一定符合。局部回归假定变量之间的关系未知，因而没有隐含任何假设条件，所以更加符合实际。

第三，其它普遍使用的核估计，如 $Nadaraya-Watson$ 估计导致不必要的偏差，而 $Gasser-Mjller$ 估计在处理解释变量为随机性变量时有较大的方差。

第四，局部回归方法既适合于解释变量为确定性变量的固定设定模型，也适合于解释变量为随机性变量的随机设定模型。

第五，局部回归方法适合于随机设定模型解释变量分布均匀情形，也适合于分布不均匀的情形。

第六，局部多项式拟合不必进行边界修正，它在边界的偏差自动与内部的偏差有相同的阶。在处理多维情况时，这是一个重要的优点。因为多维情况的边界相当大，其它估计方法在进行边界修正非常困难。

第七，局部多项式回归估计在所有线性估计中，在极小极大效率($Minimaxefficiency$)意义上接近于最优，它的有效性为 100%。

偏和方差 其中

表 9-1 $m(x)$ 的核回归估计的逐点渐近偏和方差（$Fan1992$）

方法	偏	方差
$Nadaraya-Watson$	$(m''(x)+\frac{2m'(x)f'(x)}{f(x)})b_n$	V_n
$Gasser-Mjller$	$m^n(x)$	$1.5V_n$
局部线性回归	$m^n(x)$	V_n

$$b_n=\frac{h^2}{2}\int_{-\infty}^{\infty}u^2K(u)du\ V_n=\frac{\sigma^2(x)}{f(x)nh}\int_{-\infty}^{\infty}K^2(u)du$$

记

$$\begin{array}{ll}\mu_j=\int u^jK(u)du & v_j=\int u^jK^2(u)du\\ S=(\mu_{j+l})_{0\leqslant j,l\leqslant p} & c_p=(\mu_{p+1},\cdots,\mu_{2p+1})'\\ \tilde{S}=(\mu_{j+l+1})_{0\leqslant j,l\leqslant p} & \tilde{c}_p=(\mu_{p+2},\cdots,\mu_{2p+2})'\\ S^*=(v_{j+l})_{0\leqslant j,l\leqslant p} & S^{-1}=(S^{jl})_{0\leqslant j,l\leqslant p}\end{array}$$

并且

$$e_{\nu+1}=(0,\cdots,0,1,0,\cdots,0)'$$

$K(t)$ 的同等核定义为

$$K^*_{v,p}(t)=(\sum_{l=0}^{p}S^{vl}t^l)K(t)$$

假设 $f(x_0)>0$，且 $f(\cdot),\sigma^2(\cdot),m^{p+1}$ 在 x_0 的领域内连续，则当 $h\to 0,nh\to\infty$ 时，

$$\hat{m}_v(x_0)=v!\hat{\beta}_v$$

表 9-2 核函数 $K(t)$ 的同等核 $K^*_{v,p}(t)$

v	p	$K^*_{v,p}(t)$
0	1	$K(t)$
0	3	$(\mu_4-\mu_2^2)^{-1}(\mu_4-\mu_2t^2)K(t)$
1	2	$\mu_2^{-1}tK(t)$
2	3	$(\mu_4-\mu_2^2)^{-1}(t^2-\mu_2)K(t)$

的渐近条件方差为

$$\begin{aligned}&Var\{\hat{m}_v(x_0)|X\}\\&=e'_{v+1}S^{-1}S^*S^{-1}e_{v+1}\frac{(v!)^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}+op(\frac{1}{nh^{1+2v}})\\&\approx\int(K^*_{v,p})^2(t)dt\frac{(v!)^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}+op(\frac{1}{nh^{1+2v}})\end{aligned}$$

当 $p-v$ 为奇数时，渐近条件偏为

$$\begin{aligned}&Bias\{\hat{m}_v(x_0)|X\}\\&=e'_{v+1}S^{-1}c_p\frac{v!}{(p+1)!}m^{(p+1)}(x_0)h^{p+1-v}+op(h^{p+1-v})\\&\approx\int t^{p+1}K^*_{v,p}(t)dt\frac{v!}{(p+1)!}m^{(p+1)}(x_0)h^{p+1-v}+op(h^{p+1-v})\end{aligned}$$

当 $p-v$ 为偶数，若 $f^{'}(\cdot),m^{p+2}$ 在 x_0 的领域内连续，渐近条件偏为

$$\begin{aligned}&Bias\{\hat{m}_v(x_0)|X\}\\&=e'_{v+1}S^{-1}\tilde{c}_p\frac{v!}{(p+2)!}\{m^{(p+2)}(x_0)+(p+2)m^{(p+1)}(x_0)\frac{f'(x_0)}{f(x_0)}\}h^{p+1-v}+op(h^{p+2-v})\end{aligned}$$

最佳窗宽的选择

9.6 半参数非回归模型

Stone 于 1977 年提出如下回归模型：

$$\boldsymbol{y}=\boldsymbol{X}\beta+g(\boldsymbol{Z})+\varepsilon \tag{9-1}$$

偏残差估计

估计方法 9.1 偏残差估计方法是由 $Denby(1984,1986)$ 提出的，分两步进行估计。

光滑样条估计 二阶段最小二乘估计

9.7 小结

(1) 核估计的核心问题就是核权函数的选择和窗宽的选择。核权函数在核估计中起光滑的作用，即消除扰动的随机因素，使所得曲线反映变量之间的实际经济关系。

第十章 soft

10.1 Eviews

输入数据

(1) $File \rightarrow New \rightarrow Workfile$

(2) 单击右键 $\rightarrow NewObject \rightarrow Series$

(3) $File \rightarrow Import \rightarrow Excel$

绘制散点图 选中要输入的数据 $\rightarrow$ 右击 $\rightarrow Open \rightarrow asgroup \rightarrow VIEW \rightarrow Graph \rightarrow Scatter \rightarrow SimpleScatter$ **线性回归** 选中要输入的数据 $\rightarrow$ 右击 $\rightarrow Open \rightarrow asequation$

$view \rightarrow representation$

残差图 $view \rightarrow Actual, Fitted, Residual \rightarrow ResidualGraph$ **怎样用** $EViews$ **通过键盘输入数据** 建立新工作文件的方法是从 $EViews$ 主菜单中单击 $File$ 键，选择 New，$Workfile$。则打开一个数据范围选择框（$WorkfileRange$）。需要做出 3 项选择。①选择数据性质。②启始期（$Startdate$）。③终止期（$Enddate$）。3 项选择完毕后，点击"OK" 键。这时，会建立起一个尚未命名的工作文件（$Workfile$），且处于打开状态。当打开新工作文件或现有工作文件后，可以通过键盘输入数据和追加数据。具体操作如下：

从 $EViews$ 主菜单中点击 $Quick$ 键，选择 $EmptyGroup$ 功能。这时会打开一个空白表格数据窗口（$Group$）。每一个空格代表一个观测值位置。按列依次输入每一个变量（或序列）的观测值。键入每一个观测值后，可通过按回车键（$Enter$ 键）或方向指示键进行确认。按方向指示键的好处是在确认了当前输入的观测值的同时，还把光标移到了下一个待输入位置。每一列数据上方的灰色空格是用于输入变量名的。给变量命名时，字符不得超过 16 个。注意：下列名字具有特殊意义，给变量命名时，应避免使用。它们是：$ABSACOSARASINCCONCNORM$，$COEFCOSDDLOGDNORMELSEENDIF$，$EXPLOGLOGI$

回归输出结果 Estimation output 分三部分。

第一部分五行，说明因变量；估计方法；运行时间；样本数据时段；样本个数。

第二部分三行，说明解释变量和常数项的系数估计，系数估计的标准差，t 值和 p 值。

第三部分六行二列：

(1) $R-squared$ 样本可决系数，说明模型按拟合程度

(2) $AjustedR-squared$ 调整的样本可决系数

(3) $S.E.ofregression$ 回归标准误：

$$s = \sqrt{\frac{\hat{e}'\hat{e}}{T-k}} \tag{10-1}$$

(4) $Sumsquaredresid$ 残差平方和：$\sum(y_i - \hat{y}_i)^2$

(5) $Loglikelihood$ 对数似然比检验：

$$L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2} \tag{10-2}$$

(6) $DurbinWatsonstat\ DW$ 统计量，它的值越接近于 2 说明越不存在自相关；

(7) $MeanDependentvar$ 因变量的均值；

(8) $S.D.Dependentvar$ 因变量的标准差；

(9) $Schwarzcriterion$ 施瓦茨准则

(10) $Akaikeinfocriterion$ 赤池信息准则

(11) $F-statistic$ F 统计量

(12) $Prob(F-statistic)$ F 的 p 值

样本期内预测 $Equation$ 界面选择 $Proc \rightarrow Forecast$。

10.2 点预测 Point Forecasts

调整样本范围：在显示各序列的界面下右击界面的 $Range$ 和 $Sample$ 列所在部分，调整样本结构。

单击解释变量序列，输入数据。

重复方程估计与预测。

在序列界面中单击被解释变量序列名，可以看到预测值。**预测结果的评价指标**

(1) 均方根误差 $RMSE$:

$$\sqrt{\frac{1}{h+1}\sum_{t=s}^{s+h}(\hat{y}_t-y_t)^2} \tag{10-3}$$

(2) 平均绝对误差 MAE:

$$\frac{1}{h+1}\sum_{t=s}^{s+h}|\hat{y}_t-y_t| \tag{10-4}$$

(3) 平均相对误差 $MAPE$:

$$\frac{1}{h+1}\sum_{t=s}^{s+h}\left|\frac{\hat{y}_t-y_t}{y_t}\right| \tag{10-5}$$

(4) 泰尔不等系数 $TheilInequalityCoefficient$:

$$\frac{\sqrt{\frac{1}{h+1}\sum_{t=s}s+h(\hat{y}_t-y_t)^2}}{\sqrt{\frac{1}{h+1}\sum_{t=s}^{s+h}\hat{y}_t^2}+\sqrt{\frac{1}{h+1}\sum_{t=s}s+hy_t^2}} \tag{10-6}$$

(5) 偏差比 $BiasProportion$:

$$\frac{(\bar{\hat{y}}-bary)^2}{\sum(\hat{y}_t-y_t)^2/h} \tag{10-7}$$

(6) 方差比 $VarianceProportion$:

$$\frac{(s_{\hat{y}}-s_y)^2}{\sum(\hat{y}_t-y_t)^2/h} \tag{10-8}$$

(7) 协方差比 $CovarianceProportion$:

$$\frac{2(1-r)s_{\hat{y}}s_y}{\sum(\hat{y}_t-y_t)^2/h} \tag{10-9}$$

求相关系数矩阵 (1)点击!*Quick* 键并依次选择 *GroupStatistics*, *Correlations*，将出现一个要求填写序列名的对话框（*SeriesList*），填好序列名后按 *OK*。(2) 在 *Workfile* 窗口中用鼠标选中序列名, 点击 *Show* 键， *OK* 键，从而打开数据组(*Group*) 窗口。在数据组窗口点击 *View* 键选择 *Correlations*。

montecalor EVIEWS:

```
workfile mc u 1 10

!b1=25
!b2=0.5

matrix(100,2)f

vector(10)v1
v1.fill 80,100,120,140,160,180,200,220,240,260
mtos(v1,x)

for !k=1 to 100
series u=3*nrnd
series y=!b1+!b2*x+u
equation eql.ls y=c(1)+c(2)*x
f(!k,1)=c(1)
f(!k,2)=c(2)
next

show f

expand 1 100
smpl 1 100
mtos(f,gr)
freeze ser01.qqplot
freeze ser01.hist
freeze ser02.qqplot
freeze ser02.hist

matrix(1,2)m
m(1,1)=@mean(ser01)
m(1,2)=@mean(ser02)
show m
```

10.3 Maple

10.4 matlab

统计工具箱 自然界和社会上会发生各种各样的现象，其中有的现象在一定条件下是一定要发生的，有的则表现出一定的随机性，但总体上又有一定的规律可循。一般称前者为确定性事件，后者为不确定性事件（或称随机事件）。概率论和数理统计就是研究和揭示不确定事件统计规律性的一门数学学科。作为一门实用性很强的数学分支，概率论和数理统计的理论和方法已经广泛应用于管理、经济、心理、教育、体育、医学、生物、化学、机械、水文、地质、林

表 10-1 微分方程

命令	用途	用法
dsolve	解微分方程	dsolve(DE,v)
rsolve	解差分方程	rsolve(eqn,v)
pdsolve	求解偏微分方程	pdsolve(PDE,v)
PDEplot	画一阶偏微分方程	PDEplot(PDE,int,strange,options)
pdetest	验证偏微分方程解	pdetest(sol,PDE)
odetest	检验常微分方程解	odetest(sol,ode)
mapde	转移偏方程形式	mapde(PDE,into,f)

表 10-2 微积分计算

命令	用途	用法
int	求解定积分	int(f(x),x=a..b)
Doubleint	二重积分	Doubleint(f(x,y),x,y)
		在student 库中；需要附加value() 函数给出结果
Tripleint	三重积分	Tripleint(f(x,y,z),x,y,z)

业、气象、工业生产、建筑、通讯、自动控制等几乎所有社会和科学技术领域。$Matlab$6.0 的统计工具箱相对于前面一些版本，改进较大。目前已经可以与 $SPSS$!SAS 等软件的统计功能相媲美。具体而言，它包括下面几个方面的内容：

- 概率分布:给出了常见的 20 种概率分布类型的概率密度函数、累加分布函数（分布函数）、逆累加分布函数、参数估计函数、随机数生成函数和统计量计算函数。
- 参数估计: 提供了多种分布类型分布参数及其置信区间的估计方法。
- 样本描述: 提供了描述中心趋势和离中趋势的统计量函数，缺失数据条件下的样本描述方法以及其它一些统计量计算函数。
- 方差分析:包括单因子方差分析、双因子方差分析和多因子方差分析。
- 多元方差分析: 包括单因素多元方差分析、分组聚类和多元比较等。
- 回归分析:包括多元线性回归（包括逐步回归）、岭回归、一般线性模型拟合、多项式拟合、稳健回归、响应面分析（包括二维响应面分析和多维响应面分析）、非线性回归。
- 假设检验包括单样本 t 检验、双样本 t 检验和 z 检验。
- 分布的检验:包括 $Jarque-Bera$ 正态性检验、$Kolmogorov-smirnov$ 单样本检验、$Kolmogorov-smirnov$ 双样本检验和 $Lilliefors$ 正态性检验。
- 非参数检验: 包括 $friedman$ 检验、 $Kruskalwallis$ 检验、秩和检验、符号秩检验和符号检验。
- 判别分析
- 聚类分析
- 因子分析
- 统计过程控制:提供了常用的过程管理图和过程性能图。
- 试验设计:包括完全析因设计、不完全析因设计和 $D-$ 优化设计。

- 统计图:包括箱形图、经验累加分布函数图、误差条图、函数交互等值线图、交互画线、交互点标注、散点矩阵图、散点图、添加最小二乘拟合线、正态概率图、帕累托图、q-q图、回归个案次序图、参考多项式曲线、添加参考线、交互插值等值线图和威布尔图。

矩阵函数

det(A)	行列式
rank(A)	秩
inv(A)	逆
pinv(A)	伪逆
trace(A)	迹

值域、零空间和子空间的夹角

orth(A)	正交基,它的列数等于 A 的秩
null(A)	零空间的正交基
subspace(x,y)	夹角
subspace(A,B)	矩阵的列划分的子空间的夹角

A 的向量空间就是由 A 的列划分的线性空间，这个空间的维数是 r，如果 $r = n$，则列线性无关。

A 的零空间是由满足条件 $AX = 0$ 的所有向量 X 组成的线性子空间。

线性系统的求解和 LU 因式分解

$A\backslash b$	求解线性系统
$[L,U] = lu(A)$	求上三角矩阵 U 和交换下三角矩阵 L。
$[L,U,P] = lu(A)$	满足 $LU = PA$。

$MATLAB$ 依据系数矩阵的不同而相应地使用不同的方法求解纯属系统。如果可能，先分析矩阵的结构。如果系数矩阵是对称且正定的，则使用 $Cholesky$ 分解。

如果没有找到可以替代的方法，则使用高斯消元法和部分主元法。主要是对矩阵进行 LU 因式分解。这种方法就是令 $A = LU$， L 是一个带有单位对角线的下三角矩阵， U 是一个上三角矩阵。

L 通常是一个改变序列的下三角矩阵，即有时进行互换，可能显得结构不完整。交换矩阵 P 是一个单位矩阵，按照交换的顺序来交换列向量。交换矩阵的逆等于它本身的转置。

解线性方程组 $AX = b$ 的方法

$x = bicg(A,b,tol$	用双共轭梯度法解方程组
$,maxit,M)$	M 是预处理因子
$bicg(A,b,\cdots,$	$M = M1 * M2$
$M1,M2,x0)$	$x0$ 是初始向量
$[x,flag,relres,iter,$	求 x 中的问题解
$resvect] = bicg(\cdots)$	收敛信息放在 $flag$ 中
$bicgstab(\cdots)$	稳定双共轭梯度法
$cgs(\cdots)$	复共轭梯度平方法
$gmres(A,b,restart)$	广义最小残量法
$pcg(\cdots)$	预处理共轭梯度法
$qmr(\cdots)$	准最小残量法

行梯形矩阵

rref(A)	阶梯矩阵
rref(A,tol)	
rrefmovie(A)	同时给出求解过程

通常使用的精度是 $tol = max(size(A)) * eps * norm(A, inf)$。

Cholesky 因式分解

chol(A)	A的Cholesky因子
$[G, err]$=chol(A)	如果A不是一个正定矩阵 将 err 设为非零值
R1=cholupdate(R,x)	$R = chol(A)$ $A + xx'$ 的上三角cholesky因子
R1=cholupdate(R,x,'-')	$A - xx'$ 的上三角cholesky因子

对称正定矩阵： $A = A'$；对于每个 $x \neq 0$ 都有 $x'Ax > 0$，则存在一个上三角矩阵 G，它的对角线元素都是正数，且满足 $G'G = A$。

比较LU和Cholesky:

flops(0),lu(A);
flops,flops(0),chol(A);
flops

QR 因式分解

$[Q, R] = qr(A)$	
$[Q, R, P] = qr(A)$	$AP = QR$ R 的对角线元素按大小降序排列
$[Q, R] = qr(A, 0)$	
$[Q1, R1] =$ $qrdelete(Q, R, j)$	去掉A中第j列后形成的 矩阵的QR因式分解
$[Q1, R1] =$ $qrinsert(Q, R, b, j)$	在第j列前插入列向量b

假如 A 是 $n \times n$矩阵，那么可以分解为：$A = QR$。其中 Q 是一个正交矩阵， R 是一个大小和 A 相同的上三角矩阵。

QR因式分解能用来解超定方程组，即方程的个数比未知变量的个数多，还能用来求特征值和特征向量。

Givens 和Jacobi 旋转

planerot(x) 求去掉有两个元素的向量X中第1个元素的 2×2 矩阵的Givens 旋转

[G,y]=planerot(x) $y = Gx$

rjr(A) Jacobi 旋转，旋转的角度和平面都是随机数，保存有特征向量、奇异值和对称性

向量范数

norm(x) 欧几里得范数

$$\|x\|_2 = \sqrt{\sum |x_k|^2}$$

norm(x,inf) $\infty-$ 范数

$$\|x\| = \max(abs(x))$$

norm(x,1) $1-$ 范数

$$\|x\|_1 = \sum\nolimits_k |x_k|$$

norm(x,p) $p-$ 范数

$$\|x\|_1 = \sqrt{\sum\nolimits_k |x_k|^p}$$

norm(x,-inf) 求向量X的元素的绝对值的最小值

向量的范数是一个标量，用来衡量向量的长度。

矩阵范数 用向量的范数来定义方阵的 $p-$ 范数：

$$\|A\|_p = \sup_{X\neq 0} \frac{\|AX\|_p}{\|X\|_p}$$

norm(A) 欧几里德范数，等于矩阵的最大奇异值

norm(A,1) 列范数，等于A的列向量的1-范数的最大值

norm(A,2) 欧几里德范数

norm(A,inf) 行范数，等于A的行向量的1-范数的最大值

norm(A,'fro') Frobenius范数

normest(A) 欧几里德范数的估计值，误差小于 10^6

normest(A,tol) 欧几里德范数的估计值，误差小于tol

Frobenius范数：

$$\|A\|_F = \sqrt{\sum\nolimits_i \sum\nolimits_j |a_{ij}|^2}$$

条件数 令 $Ax = b$表示线性方程组。方程组的条件数是一个大于或者等于1的实数，用来衡量关于数据中的扰动，也就是A和（或）b，对解 x 的灵敏度。一个差条件的方程组的条件很大。

条件数的定义：

$$cond(A) = \|A\| \left\|A^{-1}\right\|$$

欧几里德范数的条件数就是A的最大奇异数与最小奇异数的商。

cond(A) 欧几里德范数的条件数

cond(A,p) p范数的条件数

condest(A) A条件数的1范数中的下界估计值

[c,v]=condest(A) A条件数的1范数中的下界估计值，同时还计算向量 v:

$$\|Av\| = \frac{\|A\| \|v\|}{c}$$

[c,v]=condest(A,tr) $tr = 1$ 则显示计算每一步；如果 $tr = -1$，则给出商 $c/rcond(A)$。

rcond(A) 求矩阵定义的方程组的敏感度的另一个估计值。对于差条件矩阵来说，给出一个接近于零的数；对于好条件矩阵则给出一个接近于1的数。

例 10.1 求 *Jilbert* 矩阵的条件数bad $= \mathrm{cond}(\mathrm{hilb}(5))$

表明在最坏情况下右边或系数矩阵的一个扰动和数 *bad* 相乘以后，可能会丢失五位小数。

最小二乘解 对于 $Ax = b$，如果 $m > n$ ，即方程个数多于未知数，则称为超定方程组。通常这些方程组是矛盾的，没有精确解。

关键是要找到一个向量 X，使它对 m 个方程的总误差最小。

最小二乘法：

$$e = \sum_{i=1}^{m}(b_i - \sum_{j=1}^{n} a_{ij}x_j)^2 = \|b - Ax\|_2^2$$

$A \backslash b$ 最小二乘解

$spaugment(A, c)$ 生成一个对称的希疏矩阵 $T = [c * IA; A'0]$。

$nnls(A, b)$ 求非负最小二乘解

$lscov(A, b, v)$ 已知协方差 V 情况下的最小二乘解

泊松分布Poisson Distribution

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda x}, \quad x = 0, 1, \cdots$$

例 10.2 (Poisson Distribution Curve) 绘制$\lambda = 1, 2, 5, 10$ 时 *Poisson* 分布的概率密度函数与概率分布函数曲线。

正态分布normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

例 10.3 (Normal Distribution Curve) 绘制 $\mu = 0, 1, 2$，$\sigma^2 = 1, 2, 3$ 时 *Normal* 分布的概率密度函数与分布函数曲线。

伽马分布Gamma Distribution

$$p_\Gamma(x) = \begin{cases} \frac{\lambda^a x^{a-1}}{\Gamma(a)} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$$

源代码 10.1 分布函数:泊松分布

```
x=[0:15]'; y1=[ ]; y2=[ ]; lambda=[1,2,5,10]; for i=1:length(lambda)
    y1=[y1,poisspdf(x,lambda(i))];
    y2=[y2,poisscdf(x,lambda(i))];
end plot(x,y1), figure; plot(x,y2)
```

源代码 10.2 分布函数:正态分布

```
x=[-15:.05:15]'; y1=[ ]; y2=[ ]; mu=[0,1,2]; sig =[1,2,3]; for
i=1:length(mu)
    y1=[y1,normpdf(x,mu(i),sig(i))];
    y2=[y2,normcdf(x,mu(i),sig(i))];
end plot(x,y1), figure; plot(x,y2)
```

$$\Gamma(a) = a\Gamma(a-1)$$
$$\Gamma(1) = 1$$
$$\Gamma(\frac{1}{2}) = \pi$$

伽马函数的值可通过 $gamma()$ 函数求得。

卡方分布chi distribution

$$p(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

T 分布

$$p(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \left[1 + \frac{x^2}{k}\right]^{-\frac{k+1}{2}}$$

F 分布

$$p(x) = \begin{cases} \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})} p^{\frac{p}{2}} q^{\frac{q}{2}} x^{\frac{p}{2}-1} (p+qx)^{-\frac{p+q}{2}} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

Rayleigh 分布

$$p(x) = \begin{cases} \frac{x}{b^2} e^{-\frac{x^2}{2b^2}} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

概率问题的求解

例 10.4 (条件概率) 已知两维随机变量 ε, η 联合概率密度求条件概率。

$$p(x,y) = \begin{cases} x^2 + \frac{1}{3}xy & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 2 \\ 0 & else \end{cases}$$

求:

$$P(\varepsilon < 1/2 | \eta < 1/2)$$

源代码 10.3 Gamma Distribution Curve

```
  x=[-0.5:.02:15]';
y1=[]; y2=[]; a=[0,-1,1]; lambda=[0.1,1,1]; for i=1:length(a)
    y1=[y1,gampdf(x,a(i),lambda(i))];
    y2=[y2,gamcdf(x,a(i),lambda(i))];
end plot(x,y1), figure; plot(x,y2)
```

源代码 10.4 CHI Distribution Curve

```
x=[-eps:-0.02:-0.05,0:0.02:2]; x=sort(x'); k =[1,2,3,4,5]; y1=[];
y2=[];

for i=1:length(k)
    y1=[y1,chi2pdf(x,k(i))];
    y2=[y2,chi2cdf(x,k(i))];
end plot(x,y1), figure; plot(x,y2)
```

源代码 10.5 T Distribution Curve

```
x=[-5:0.05:5]'; k =[1,2,3,4,5]; y1=[]; y2=[];

for i=1:length(k)
    y1=[y1,tpdf(x,k(i))];
    y2=[y2,tcdf(x,k(i))];
end plot(x,y1), figure; plot(x,y2)
```

源代码 10.6 F Distribution Curve

```
x=[-eps:-0.02:-0.05,0:0.02:2]; =sort(x');
 p =[1,2,3,4,5]';
q =[2,3,4,5,6]'; y1=[]; y2=[];

for i=1:length(p)
    y1=[y1,fpdf(x,p(i),q(i))];
    y2=[y2,fcdf(x,p(i),q(i))];
end plot(x,y1), figure; plot(x,y2)
```

源代码 10.7 Rayleigh Distribution Curve

```
x=[-eps:-0.02:-0.05,0:0.02:2]; x=sort(x');

b =[.2,1,2,3,5]'; y1=[]; y2=[];

for i=1:length(b)
    y1=[y1,raylpdf(x,b(i))];
    y2=[y2,raylcdf(x,b(i))];
end

plot(x,y1), figure; plot(x,y2)
```

例 10.5 (区间概率) 已知随机变量 $X \sim N(1,2)$，求 X 落入区间 $(5,7)$ 的概率。

例 10.6 (随机数生成：Rayleigh) 令 $b=1$，生成 30000×1 个 *Rayleigh* 分布的随机数，并用直方图检验生成数据的概率分布情况。

源代码 10.10

```
b=1; p=raylrnd(1,30000,1);

x1=0:1:6; x2=0:.1:4;

figure(1); hist(p) figure(2); hist(p,x2)

yy=hist(p,x2); yy=yy/30000;

figure(3); bar(x2,yy),y=raylpdf(x2,1); line(x2,y)
```

例 10.7 (求数字特征) 用积分法求伽马分布的均值与方差。

例 10.8 (数字特征：样本数) 生成一组 30000 个正态分布随机数，使其均值为 0.5，标准差为 1.5，分析数据实际的均值、方差和标准差，如果减小随机变量个数，会有什么结果？

例 10.9 (矩) 求随机数的各阶原点矩和中心矩

例 10.10 (联合分布密度图) 绘制二维联合分布正态分布图，如果协方差矩阵非对角线元素为零，绘制新图。

置信度估计

例 10.11 (置信度估计) 生成服从伽马分布，参数为 $a=1.5, \lambda=3$ 的 30000 个伪随机数，用不同的置信度估计:$90\%, 95\%, 98\%$。

例 10.12 线性回归方程 $Y=X_1-1.2X_2+2.3X_3+3.4X_4+5X_5$，生成 120 组随机输入值 X_i，计算输出向量 Y_i。以这些信息为已知，由最小二乘得出待定系数的估计值，并得出 95% 置信求解。叠加方差为 0.5 的白噪声，画出误差差图。

例 10.13 (非线性函数最小二乘拟合) 用参数估计的方法求解最小二乘拟合问题，并得出 95% 的置信度的置信区间，并在实测信号的基础上叠加均匀分布的噪声信号再进行参数与区间估计。原函数为 $y(x)=a_1e^{-a_2x}+a_3e^{-a_4x}\sin(a_5x)$。

源代码 10.8 解条件概率

```
syms x y; f=x^2+x*y/3; p=int(int(f,x,0,1/2),y,0,1/2)
```

源代码 10.9 解区间概率

```
p7=normcdf(7,1,2);
p5=normcdf(5,1,2);
p=p7-p5
```

源代码 10.11 积分法求均值与方差

```
syms x; syms a lambda positive
p=(lambda^a*x^(a-1)/gamma(a))*exp(-lambda*x);
m=int(x*p,x,0,inf) s1=int((x-mu)^2*p,x,0,inf)
s2=simple(int((x-1/lambda*a)^2*p,x,0,inf))
```

源代码 10.12 数字特征：变化的样本容量

```
p=normrnd(0.5,1.5,10000,1); [mean(p) var(p) std(p)]
p=normrnd(0.5,1.5,100,1); [mean(p) var(p) std(p)]
```

源代码 10.13 求矩:随机抽样值与理论值

```
A=[]; B=[];
p=normrnd(0.5,1.5,100000,1);
n=1:5; for r=n;
    A=[A,sum(p.^r)/length(p)];
    B=[B,moment(p,r)];
end
A,B
syms x; A1=[];B1=[];
p=(1/(sqrt(2*pi)*1.5))*exp(-(x-0.5)^2/(2*1.5^2));
for i=1:5
    A1=[A1,vpa(int(x^i*p,x,-inf,inf),12)];
    B1=[B1,vpa(int((x-0.5)^i*p,x,-inf,inf),12)];
end
A1,B1
```

源代码 10.14 生成协方差矩阵

```
p=randn(30000,4); cov(p)
```

源代码 10.15 绘制二维正态分布图

```
mu1=[−1,2]; sig1=[1 1;1 3];

[X,Y]=meshgrid(−3:0.1:1,−2:0.1:4); xy=[X(:) Y(:)];

p=mvnpdf(xy,mu1,sig1); P=reshape(p,size(X)); surf(X,Y,P)

figure (2); sig2=diag(diag(sig1)); p=mvnpdf(xy,mu1,sig2);
P=reshape(p,size(X)); surf(X,Y,P)

figure (3); R=mvnrnd(mu1,sig1,2000); plot(R(:,1),R(:,2),'o')

figure (4); R=mvnrnd(mu1,sig2,2000); plot(R(:,1),R(:,2),'o')
```

源代码 10.16 置信度

```
p=gamrnd(1.5,3,30000,1); Pv=[0.9,0.95,0.98];
num=[30,300,3000,30000,300000]; A=[]; B=[];

for i=1:length(Pv)
    [a,b]=gamfit(p,Pv(i));
    A=[A;Pv(i),a(1),b(:,1)',a(2),b(:,2)'];
end

for i=1:length(num)
    p=gamrnd(1.5,3,num(i),1);
    [a,b]=gamfit(p,0.95);
    B=[B;num(i),a(1),b(:,1)',a(2),b(:,2)'];
end

A,B
```

源代码 10.17 最小二乘估计

```
a=[1 -1.2 2.3 3.4  5]';  X=randn(120,5); Y=X*a;

hat_a=inv(X'*X)*X'*Y

[hat_a1,aci]=regress(Y,X,0.05)

yhat=Y+sqrt(0.5)*randn(120,1);

[a,aint]=regress(yhat,X,0.05)

errorbar (1:5,a,aint (:,1) -a,aint (:,2) -a)
```

源代码 10.18 非线性最小二乘拟合

```
f=inline('a(1)*exp(-a(2)*x)+a(3)*exp(-a(4)*x).*sin(a(5)*x)','a','x');
x=0:0.1:10;  y=f  ([0.12,0.213,0.54,0.17,1.23],  x);

[a,r,j]= nlinfit (x,y,f  ,[1;1;1;1;1])  ;  a

ci=nlparci(a,r,j)

y1=y+0.02*rand(size(x)); [a1,r1,j1]= nlinfit (x,y,f  ,[1;1;1;1;1])  ;  a1
ci1=nlparci(a1,r1,j1)  errorbar (1:5,a1,ci1 (:,1) -a1,ci1 (:,2) -a1)
```

10.5 SAS

来自[25]、[26]。**统计图** 展示各种与数据相关联的图形是统计中呈现数据的重要方法，当数据经过审核、筛选、排序等预处理后，可以把它做成统计图。统计图一般是用点、线、面或立体图像鲜明地表达数量或变化动态。常用的统计图有线图、直方图、条形图和统计地图等。一个好的统计图能够把数据在图中清楚、准确、有效地表达出来。统计软件 *SAS*可以帮助我们把存贮在 *SAS* 数据集中的数据以图形的方式形象直观地显示出来。**常用数据类型及适用的统计图 定类数据—条形图**

表 10-3 常用数据类型及适用的统计图

数据类型	例	统计图
品质数据定类数据	心血管病的死亡原因	条形图；饼图
定序数据	碳氧血红蛋白的含量	累积分布图
数值型数据分组数据	食物中毒患者的潜伏期	折线图
未分组数据	110名男大学生身高	方图；茎叶图；箱线图；散点图
时间序列数据	10年某小学的麻疹曾患率(%)	间隔图形
地域性数据	某省痢疾流行时的患者分布	统计地图

条形图是用宽度相同的条形的高度或长短来表示数据变动的图形。一般用条形图的高度来表示各类别数据的频数。它常用于两个或多个组某指标大小的比较。

条形图的类型一般有三种：

- 简单条形图，用于表现单个指标的大小；
- 分组条形图，用于表现两个或多个指标，比如同时进行男、女性血红细胞和血红蛋白值的比较；
- 分段条形图，用于表现每个直条中某个因素各水平的构成情况，比如男女两种性别的某指标构成的比较。

未分组数据：直方图，茎叶图，箱线图，散点图 直方图，未分组的原始数据，数值型数据表现为数字，在整理时通常进行数据分组。分组是根据统计研究的需要，将数据按照某种标准分成不同的组别。直方图是用矩形的宽度和高度来表示频数分布的图形。用横轴表示数据分组，纵轴示频数或频率。 对于未分组的原始数据，我们可以用茎叶图来显示其分布的特征。茎叶图由“茎”和“叶”两部分构成，其图形是由数字组成的。通过茎叶图，可以看出数据的分布形状及数据的离散状况，比如分布是否对称，数据是否集中，是否有极端值等。 箱线图是由一组数据的5个特征值绘制而成的，它由一个箱子和两条线段组成。5个特征值依次是最大值、上四分位数、中位数、下四分位数和最小值。通过箱线图，可以反映出数据分布的特征。箱线图一般有单批数据箱线图和多批数据箱线图两种。对于多批数据，我们可以将各批数据的箱线图并列起来，从而进行分布特征的比较。 未分组数据－散点图: 表示两种事物变量的相关性和趋势。 时间序列数据－间隔图形: 当观测按日期或时间排列时，数据就构成时间序列数据。我们一般采用“$Timeplot$”过程对一个或几个变量绘制时间间隔的散点图。间隔图形的类型一般有单个变量的间隔图形和多个变量的间隔图形两种。

状态空间模型 PROC STATESPACE 适用于预测几个相关的有动态相互影响的时间序列，并考虑到整个变量集之内的自相关。

其前提是假定时间序列为联合平稳，非平稳序列需要通过初始变换，通常是差分变换，成为平稳序列。

定义 10.1 (**状态空间模型**)　通过辅助变量表示一个多元时间序列，这些辅助变量的某些部分或许不能直接观测到，观测的时间序列被表示成状态变量的线性组合。状态空间模型也称为多元时间序列的马尔可夫表示，或典型表示。它是由 $Akaike$(1974) 概括的。

定义 10.2 (**状态向量**)　状态空间模型中的辅助变量，它概括了所有的来自与时间序列将来值的预测值相关的时间序列的当前及过去值的信息。

定义 10.3 (**状态转移方程**)

$$Z_{t+1} = FZ_t + Ge_{t+1}$$

- $s \times s$ 系数矩阵 F 被称为转移矩阵，它决定了模型的动态性质

源代码 10.19 SAS Code:简单条形图

```
PROC GCHART;
     VBAR X/SUMVAR=Y;
     PATTERN V=X5 C=BLUE;
RUN;
```

源代码 10.20 SAS Code:分组条形图

```
PROC GCHART;
     VBAR GENDER/GROUP=CASE SUMVAR=RATIO PATTERNID=MIDPOINT;
     PATTERN V=X5 C=BLUE;
RUN;
```

源代码 10.21 SAS Code:分段条形图

```
PROC GCHART;
     WHERE GENDER IN ('M', 'F');
     VBAR CASE/SUBGROUP=GENDER SUMVAR=RATIO ;
     PATTERN V=X5 C=BLUE;
RUN;
```

源代码 10.22 SAS Code:直方图

```
data beishu; input beishu@@; cards; 100 200 400 400 400 400 800 1600
1600 1600 3200 ; proc capability ; histogram beishu/ cfill =gray; run;
```

源代码 10.23 SAS Code:茎叶图

```
data beishu; input beishu@@; cards; 100 200 400 400
 400 400 800 1600
1600 1600 3200 ; proc univariate plot;
 var beishu;
 run;
```

源代码 10.24 SAS Code:箱线图

```
data beishu;
 input type beishu@@;
 1 200 1 400 2 400 2 400
 1 340 2 350 1 345 2 350
 1 380 2 390 2 400 1 385
 ;
 proc boxplot;
     plot beishu*type;
 run;
```

源代码 10.25 SAS Code:散点图

```
data beishu;
 input type beishu@@;
 1 200 1 400 2 400 2 400
 1 340 2 350 1 345 2 350
 1 380 2 390 2 400 1 385
 ;
 proc gplot;
     plot beishu*type;
 run;
```

源代码 10.26 SAS Code:多变量时序间隔图

```
data cancer; input year us world@@; cards; 1940 12.6 31.2 1950 10.0
30.6 1960 14.2 46.2 1965 15.0
 55.0 1970 10.2 53.6 1972 13.7 62.9
1973 13.0 63.3 1974 11.5 64.3 ; proc sort; by year; proc timeplot;
plot us world/overlay; id year; run;
```

- $s \times r$ 系数矩阵 G 称为输入矩阵，它决定了转移方程的方差结构，
- 输入向量 e_t 是一个独立的 r 维正态分布随机向量，或称新息向量，震动向量

测量方程

$$X_t = [I_r 0]Z_t$$

X_t 是一个 r 维观测向量，它是经过差分（如果指定了差分）并减去样本均值所组成的。

$Akaike$(1976) 的建模策略，使用典型相关分析作为建模标识。

产生 X 和 Y 序列：

源代码 10.27 产生X 和Y 序列

```
DATA LI.XYIN;
    X=10;Y=40;
    X1=0;Y1=0;
    A1=0;B1=0;
    ISEED=123;
    DO T=-100 TO 200;
      A=RANNOR(ISEED);
      B=RANNOR(ISEED);
      DX=0.5*X1+0.3*Y1+A-0.2*A1-0.1*B1;
      DY=0.5*Y1+0.3*X1+B;
      X=X+DX+0.5;
      Y=Y+DY+0.5;
      IF T>=0 THEN OUTPUT;
      X1=DX;Y1=DY;
      A1=A;B1=B;
    END;
    KEEP T X Y;
RUN; PROC GPLOT DATA=XYIN;
    SYMBOL I=JOIN;
    PLOT Y*T X*T
       /OVERLAY
        HAXIS= 0 TO 200 BY 10
       ;
RUN;
```

对序列进行一阶差分并向前预测 10 步。序列建模前先减去均值，产生预测值后再加回来。可指定 $NOCENTER$ 选项不减去样本均值。建模方式是：

$$x_t = (1 - B)X_t - \bar{x}_t$$
$$y_t = (1 - B)Y_t - \bar{y}_t$$

根据 AIC 定阶。

源代码 10.28 SAS:差分、预测、作图、修正

```
PROC STATESPACE DATA=XYIN OUT=OUT LEAD=10;
        VAR X(1) Y(1);
        ID T;
RUN;

PROC PRINT DATA=OUT;
        ID T;
        WHERE T > 190;
RUN;

PROC GPLOT DATA=OUT;
        SYMBOL1 V=PLUS I=JOIN;
        SYMBOL2 V=STAR I=NONE;
        PLOT FOR1*T=1 FOR2*T=1
                X*T=2 Y*T=2
                /OVERLAY HREF=200.5;
RUN;

PROC STATESPACE DATA=XYIN OUT=OUT LEAD=10 OUTAR=B;
        VAR X(1) Y(1);
        ID T;
        FORM X 2 Y 1;
        RESTRICT F(2,3)=0;
RUN;
```

除差分外，一种对时间趋势去势的方法是对时间的回归。如果时间序列中具有决定性的线性趋势，序列对时间的回归残差应是平稳的。

源代码 10.29 SAS：时间回归去势

```
PROC REG DATA=LI.XYIN;
     DETREDNING:MODEL X Y =T;
     OUTPUT OUT=DETREND R=RX RY;
RUN;

PROC STATESPACE DATA=DETREND OUT=DETREND LEAD=10
    PRINTOUT=LONG;
     VAR RX RY;
     ID T;
RUN;
```

取子数据集 用SET 语句引入全数据集，IF 和WHERE 语句指定取子集的条件，DROP、KEEP 选项或语句取变量的子集，firstobs= 和obs= 选项选择子集。**时间序列横断面形式存储数据** 所有横断面的序列都存入一个ID变量，另一个横断面ID变量用来区别不同序列的观测。观测按横断面ID变量排序，内部按时间排序。**指定时间间隔** 使用诸如 $YEAR, QTR, MONTH, DAY$ 之类指定时间间隔。

- $YEAR2$ 表示时间间隔为两年
- $YEAR.7$ 表示从七月开始的财政年度

纵向绘时间刻度 使用 $TIMEPLOT$ 过程。

A 到B 的转换

源代码 10.30 A 到B 的转换1

```
data a; input id time sbps @@; datalines; 1 1 123 1 2 126 1 3 128 2
1 112 2 2 110 2 3 115 ; run;

proc transpose data=a out=b(drop=_NAME_) prefix=sbps; by id; var
sbps; run;
```

源代码 10.31 A 到B 的转换2

```
data b;
        set a;
        by id;
        retain sbp1-sbp3;
        array sbp[3];
        if first.id then i=1;
        sbp[i]=sbps;
        i+1;
        if last.id then output;
        drop sbps time i;
run;

proc print; run;
```

源代码 10.32 A 到B 的转换3

```
data b;
     retain  id;
     array  u[*]  sbp1-sbp3;
     do i=1 to dim(u);
     set  a  ;
    u[i]=sbps;
    end;
    drop sbps time i;
run;

proc print; run;
```

源代码 10.33 A 到B 的转换4

```
data one; label  x1='601' x2='602' x3='603'; input date x1-x3 @@;
cards;  2001 2.1  1.2  3.5  2002 2.2  1.3  3.6  2003 2.3  1.4  3.7  ;  run;

proc transpose out=two; by date; var x1-x3; proc sort; by _Name_;
proc print;  run;  data three  (drop= _name_); set two; rename
_label_=code col1=pro; proc print; run;
```

10.6 宏

源代码 10.34 SAS：显示自动宏变量的值

```
%PUT _AUTOMATIC_;
%PUT &SYSDAY;
%PUT THIS VERSION IS &SYSVER;
%PUT _AUTOMATIC_;
%PUT _USER_;
%PUT _ALL_;
```

源代码 10.35 SAS：脚注提供报告日期

```
OPTIONS SYMBOLGEN;
%LET A=%SYSFUNC(TODAY(),WORDDATE.);
PROC FREQ DATA=LI.RND;
      TABLES P;
      TITLE "Report Produce on %SYSFUNC(LEFT(%QSYSFUNC(TODAY(),
           WORDDATE.)))";
      FOOTNOTE1 "Created &systime &sysday,&sysdate9";
      FOOTNOTE2 "On the &sysscp system using Release &sysver";
      FOOTNOTE3 "&A";
RUN;
```

如何在同一页上绘制多个散点图 proc greplay ; proc iml; overlay **本记录与上次记录的和与累加和**

源代码 10.36 本记录与上次记录的和

```
data test;
      input a b;
      cards;
2 1 2 1 2 1 2 1 2 3 2 3 2 3 ; run;

data result;
      set test;
      if _n_>1 then do;
      n=_n_-1;
/*find last b*/
      set test(keep=b rename=(b=pre)) point=n;
      b=b+pre;
      drop pre;
      end;
run;
```

[style=numbers,caption=累加和]

源代码 10.37

```
data result;
      set test;
      retain pre;
      b=sum(b,pre);
      pre=b;
run;
```

把data set里observation的数目保存在一个变量中

源代码 10.38

```
data raw; input x; cards; 1 2 3 ; data _null_;
 set raw nobs=n;
  call symput('z',n);
 stop;
run;
%put z=&z;

data _null_; if 0 then set raw nobs=obs_num; /*obs_num is a variable
you want to define*/ put "obs_num=" obs_num; run;

ods listing close; ods html close; proc datasets ; contents data=raw
out=count(keep=nobs) SHORT ; quit; ods listing;

proc print data=count;run;

proc sql; create table b as select count(*) as obs from raw; quit;
```

第十一章　学习摘要

11.1　趋势稳定过程与单位根过程

来自[?]

定义 11.1 (TSP)　趋势稳定指对某些数据，可以通过退化趋势而获得稳定，即对

$$y_t = \alpha + \beta t + u_t$$

其中 $u_t \sim I(0)$。

定义 11.2 (DSP)　差分稳定指

$$\Delta y_t = \alpha + u_t$$

其中 $u_t \sim I(0)$。

从图形上看，TSP 数据显然是围绕时间趋势 βt 波动且随着时间或增或减。相比较而言，DSP 数据图形不如 TSP 显著和有规律。

为检验数据是 TSP 还是 DSP，设定

$$y_t = \beta_0 + \beta_1 t + u_t, u_t = \rho u_{t-1} + e_t, e_t \sim iid \tag{11-1}$$

当 $\rho < 1$ 时，数据生成过程是 TSP。另一方面：

$$y_t - y_{t-1} = \beta_1 + u_t - u_{t-1}$$

有

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + u_t \\ &= \beta_0 + \beta_1 t + \rho u_{t-1} + e_t \\ &= \beta_0 + \beta_1 t + \rho[y_{t-1} - \beta_0 - \beta_1(t-1)] + e_t \end{aligned}$$

可以将原式约简为

$$y_t = \alpha + \delta t + \rho y_{t-1} + e_t \tag{11-2}$$

其中：

$$\begin{aligned} \alpha &= \beta_0 - \rho(\beta_0 - \beta_1) \\ \delta &= \beta_1(1-\rho) \end{aligned}$$

若 $\rho = 1$，则 $\delta = 0$。这样，可将原假设写为:

$$H_0 \qquad \rho = 1, \delta = 0 \tag{11-3}$$

接受原假设意味着数据为 DSP：拒绝原假设则意味着数据为 TSP。这个方法要求原式的误差结构为 $AR(1)$ 过程。但许多经济变量的时间序列数据并不满足这一条件。

11.2　结构突变的趋势稳定

对 y_t 是趋势稳定还是结构变化的趋势稳定可表述为：

$$\begin{aligned} & y_t = \alpha_t + \delta t + e_t \\ & \alpha_t = \alpha_{t-1} + \varepsilon_t \\ & \varepsilon_t \sim iid \end{aligned}$$

$$H_0 \ \sigma_\varepsilon^2 = 0$$

由于 $\alpha_t = \alpha$ (常数)与 σ_ε^2 等价，接受原假设意味着 y_t 是趋势稳定，拒绝原假设意味着数据生成过程是结构变化的趋势稳定。

$Kwiatkowski, Phillips, Schmidt, Shin$(1992)提出对原假设为结构突变的 H_0的检验统计量 LM。

$$\begin{aligned} LM &= \sum_{t=1}^{T} S_t^2/\hat{\sigma}_\varepsilon^2 \\ S_t &= \sum_{i=1}^{t} e_i \qquad t = 1, 2, \cdots, T \end{aligned} \tag{11-4}$$

$KPSS$ 检验的渐近临界值表可见[?]的第 135页。$Leybourne, McCabe$ 1994 发现该检验势偏低。

11.3 外生性结构突变的单位根及其检验

定义 11.3 (**外生性结构突变点**) 先验设定外生结构突变点即结构突变点已知时称其为外生性结构突变点。

假定发生结构突变的时点已知为t_B:

$$D_t = \begin{cases} 1 & t > t_B \\ 0 & else \end{cases}$$

若单位根过程具有截距突变:

$$y_t = \mu_0 + y_{t-1} + \mu_1 D_t + e_t \qquad e_t \sim I(0) \tag{11-5}$$

则称 y_t 为**具有结构突变的单位根过程**。

记确定性趋势为 $DT_t = \mu_0 + \delta t$，则截距突变表示为 $DT_t = \mu_0 + \delta t + \mu_1 D_t$，对应的模型为

$$y_t = \mu_0 + \delta t + \mu_1 D_t + e_t \tag{11-6}$$

当 $e(t) \sim I(1)$ 时，称 y_t 由结构变化的单位根过程所生成。这一模型亦称为**崩溃模型**$CrashModel$。这是因为结构变化之后，y_t 的均值不再返回结构变化之前的均值轨迹。

当突变发生在斜率（时间趋势系数）而截距不变，即 $DT_t = \mu_0 + \delta_0 t + \delta_1 t^*$，当 $t > t_B$ 时，$t^* = t - t_B$；当 $t < t_B$ 时，$t^* = 0$。其对应的模型为

$$y_t = \mu_0 + \delta_0 t + \delta_1 t^* + e_t \tag{11-7}$$

由于斜率反映增长率，因此也称模型为**变化的增长率模型**。

当截距和斜率同时具有结构变化，即 $DT_t = \mu_0 + \mu D_t + \delta_0 t + \delta_1 D_t t^* + e_t$，对应的模型为:

$$y_t = \mu_0 + \mu_1 D_t + \delta_0 t + \delta_1 D_t t^* + e_t \tag{11-8}$$

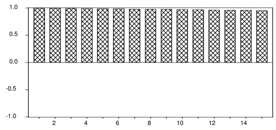
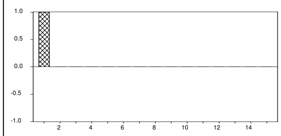
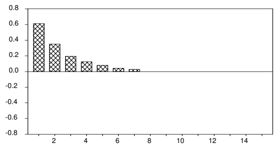
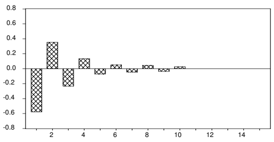
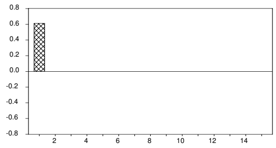
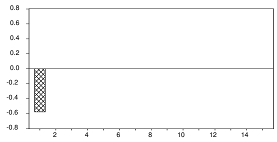
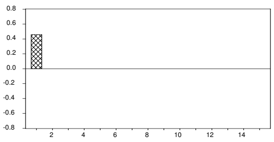
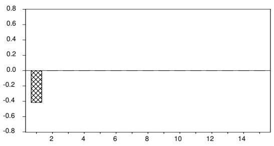
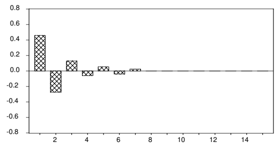
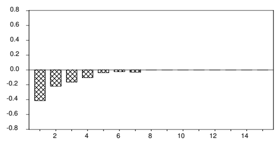
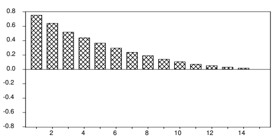
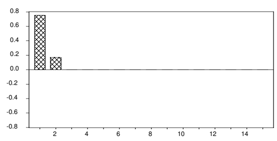
对于以上三个模型，当 $r(t) \sim I(1)$ 时，y_t 具有结构变化的单位根；而 $e(t) \sim I(0)$ 时，y_t 是为结构变化的趋势稳定。对这三个模型其原假设为 $H_0 : e_t \sim I(1)$，备择假设为 $H_1 : e_t \sim I(0)$。这三个模型也称**外层追加模型**$additiveoutlier$，因为结构变化之后的数据与变化之前相比较，有一个垂直横轴的位移，故这些数据就像是原始数据的外层，从而将模型称为外层追加模型。

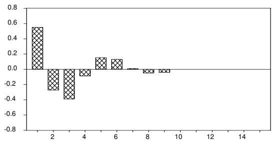
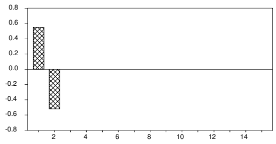
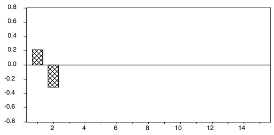
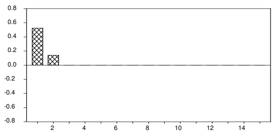
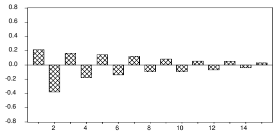
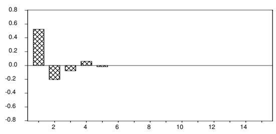
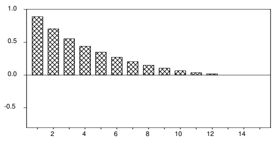
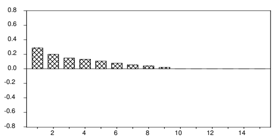
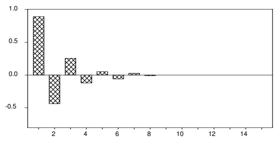
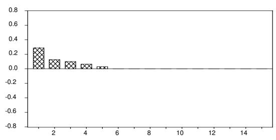
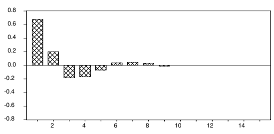
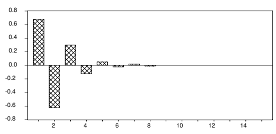
附录 A　图表

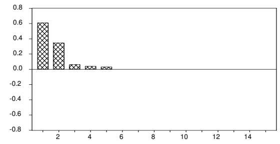
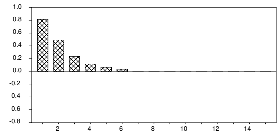
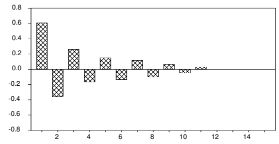
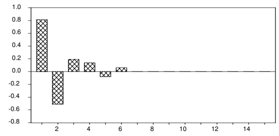
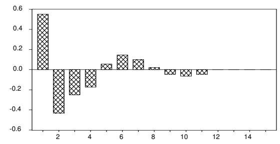
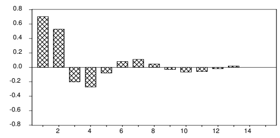
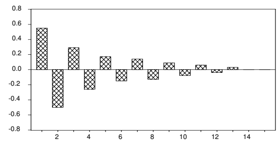
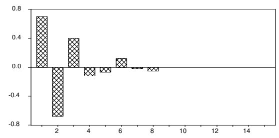
§A.1 自相关和偏自相关函数表

§A.2 ADF 检验表

§A.3 相关系数临界值检验表

模　　型	自相关函数特征	偏自相关函数特征
ARIMA(1,1,1) $\Delta x_t = \varphi_1 \Delta x_{t-1} + u_t + \theta_1 u_{t-1}$	缓慢地线性衰减 	
AR（1） $x_t = \varphi_1 x_{t-1} + u_t$	若$\varphi_1 > 0$，平滑地指数衰减  若$\varphi_1 < 0$，正负交替地指数衰减 	若$\varphi_{11} > 0$，k=1 时有正峰值然后截尾  若$\varphi_{11} < 0$，k=1 时有负峰值然后截尾 
MA（1） $x_t = u_t + \theta_1 u_{t-1}$	若$\theta_1 > 0$，k=1 时有正峰值然后截尾  若$\theta_1 < 0$，k=1 时有负峰值然后截尾 	若$\theta_1 > 0$，交替式指数衰减  若$\theta_1 < 0$，负的平滑式指数衰减 
AR（2） $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + u_t$	指数或正弦衰减  （两个特征根为实根）	k=1, 2 时有两个峰值然后截尾  （$\varphi_1 > 0$，$\varphi_2 > 0$）

	 （两个特征根为共轭复根）	 （$\varphi_1 > 0$，$\varphi_2 < 0$）
MA（2） $x_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$	k=1, 2 有两个峰值然后截尾  （$\theta_1 > 0$，$\theta_2 < 0$）  （$\theta_1 > 0$，$\theta_2 > 0$）	指数或正弦衰减  （$\theta_1 > 0$，$\theta_2 < 0$）  （$\theta_1 > 0$，$\theta_2 > 0$）
ARMA（1，1） $x_t = \varphi_1 x_{t-1} + u_t + \theta_1 u_{t-1}$	k=1 有峰值然后按指数衰减  （$\varphi_1 > 0$，$\theta_1 > 0$）  （$\varphi_1 > 0$，$\theta_1 < 0$）	k=1 有峰值然后按指数衰减  （$\varphi_1 > 0$，$\theta_1 > 0$）  （$\varphi_1 > 0$，$\theta_1 < 0$）
ARMA（2，1） $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + u_t + \theta_1 u_{t-1}$	k=1 有峰值然后按指数或正弦衰减  （$\varphi_1 > 0$，$\varphi_2 < 0$，$\theta_1 > 0$）	k=1, 2 有两个峰值然后按指数衰减  （$\varphi_1 > 0$，$\varphi_2 < 0$，$\theta_1 > 0$）
ARMA（1，2） $x_t = \varphi_1 x_{t-1} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$	k=1, 2 有两个峰值然后按指数衰减	k=1 有峰值然后按指数或正弦衰减

	 （$\varphi_1>0$，$\theta_1>0$，$\theta_2<0$）  （$\varphi_1>0$，$\theta_1>0$，$\theta_2>0$）	 （$\varphi_1>0$，$\theta_1>0$，$\theta_2<0$）  （$\varphi_1>0$，$\theta_1>0$，$\theta_2>0$）
ARMA（2，2） $x_t=\varphi_1x_{t-1}+\varphi_2x_{t-2}+u_t+\theta_1u_{t-1}+\theta_2u_{t-2}$	k=1, 2 有两个峰值然后按指数或正弦衰减  （$\varphi_1>0$，$\varphi_2<0$，$\theta_1>0$，$\theta_2<0$）  （$\varphi_1>0$，$\varphi_2<0$，$\theta_1>0$，$\theta_2>0$）	k=1, 2 有两个峰值然后按指数或正弦衰减  （$\varphi_1>0$，$\varphi_2<0$，$\theta_1>0$，$\theta_2<0$）  （$\varphi_1>0$，$\varphi_2<0$，$\theta_1>0$，$\theta_2>0$）

模型	统计量	样本容量	1%	2.5%	5%	10%
1	τ_ρ	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_ρ	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
3	τ_ρ	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
2	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_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
3	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

f	α				
	0.10	0.05	0.02	0.01	0.001
1	0.98769	0.99692	0.999507	0.999877	0.9999988
2	0.90000	0.95000	0.98000	0.99000	0.99900
3	0.8054	0.8783	0.93433	0.95873	0.99116
4	0.7293	0.8114	0.8822	0.91720	0.97406
5	0.6694	0.7545	0.8329	0.8745	0.95074
6	0.6215	0.7067	0.7887	0.8343	0.92493
7	0.5822	0.6664	0.7498	0.7977	0.8982
8	0.5494	0.6319	0.7155	0.7646	0.8721
9	0.5214	0.6021	0.6851	0.7348	0.8471
10	0.4933	0.5760	0.6581	0.7079	0.8233
11	0.4762	0.5529	0.6339	0.6835	0.8010
12	0.4575	0.5324	0.6120	0.6614	0.7800
13	0.4409	0.5139	0.5923	0.6411	0.7603
14	0.4259	0.4973	0.5742	0.6226	0.7420
15	0.4124	0.4821	0.5577	0.6055	0.7246
16	0.4000	0.4683	0.5425	0.5897	0.7084
17	0.3887	0.4555	0.5285	0.5751	0.6932
18	0.3783	0.4438	0.5155	0.5614	0.6787
19	0.3687	0.4329	0.5034	0.5487	0.6652
20	0.3598	0.4227	0.4921	0.5368	0.6524
25	0.3233	0.3809	0.4451	0.4869	0.5974
30	0.2960	0.3494	0.4093	0.4487	0.5541
35	0.2746	0.3246	0.3810	0.4182	0.5189
40	0.2573	0.3044	0.3578	0.3932	0.4896
45	0.2428	0.2875	0.3384	0.3721	0.4648
50	0.2306	0.2732	0.3218	0.3541	0.4433
60	0.2108	0.2500	0.2948	0.3248	0.4078
70	0.1954	0.2319	0.2737	0.3017	0.3799
80	0.1829	0.2172	0.2565	0.2830	0.3568
90	0.1726	0.2050	0.2422	0.2673	0.3375
100	0.1638	0.1946	0.2301	0.2540	0.3211

图 1-2 相关系数临界值检验表

附录 B　数学公式证明与推导

§B.1　自相关函数通解表达式的证明

对于 $AR(p)$ 过程

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + u_t \tag{2-1}$$

它的自相关函数满足下式

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p} \tag{2-2}$$

上式可改写为

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)\rho_k = 0 \tag{2-3}$$

则(2-2)有通解

$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k \tag{2-4}$$

证明：采用数学归纳法证明。

1。对于 $AR(1)$ 过程，它是显然成立的。

2。对于 $AR(2)$ 过程，有

$$(1 - G_1 L)(1 - G_2 L)\rho_k = 0 \tag{2-5}$$

令

$$(1 - G_2 L)\rho_k = y_k \tag{2-6}$$

根据(2-5)有

$$y_k = G_1 y_{k-1} = G_1^k y_0 \tag{2-7}$$

y_0 是系统决定的初始值。代入(2-6)并进行迭代

$$\begin{gathered}\rho_k \\ = G_2(G_2\rho_{k-2} + G_1^{k-1}y_0 + G_1^k y_0) \\ = G_2^k \rho_0 + y_0(G_1^k + G_2 G_1^{k-1} + \cdots + G_1 G_2^{k-1})\end{gathered} \tag{2-8}$$

当(2-5)有相同解时，显然(2-8)可以写成(2-2) 形式。

当(2-5)有相异解时，(2-8)可做如下变换

$$\begin{gathered}rho_k \\ = G_2^k \rho_0 + y_0 G_1 \tfrac{G_1^k - G_2^k}{G_1 - G_2} \\ = A_1 G_1^k + A_2 G_2^k\end{gathered} \tag{2-9}$$

3。对于 $AR(\rho - 1)$ 过程：

§B.2　密度函数估计的渐近无偏性与相合性证明

这是因为：

$$\begin{aligned}Ef_n(x) &= \frac{1}{h(n)} E\left\{K\left(\frac{x-X}{h(n)}\right)\right\} \\ &= \frac{1}{h(n)} \int_{-\infty}^{+\infty} K\left(\frac{x-t}{h(n)}\right) f(t)dt \\ &= \frac{1}{h(n)} \int_{-\infty}^{+\infty} K\left(\frac{y}{h(n)}\right) f(x-y)dy(x-t=y) \\ &= \int_{-\infty}^{+\infty} K\left(Z\right) f(x-h(n)Z)dZ \left(\frac{y}{h(n)} = Z\right)\end{aligned}$$

$$\int_{|Z|\geqslant T_0} K(Z)dZ \leqslant \frac{\varepsilon}{4}M$$

$$M = \sup_x f(x)$$

$$\lim_{h\to\infty} \int_{-T_0}^{T_0} K(Z)f(x-hZ)dZ = f(x)\int_{-T_0}^{T_0} K(Z)dZ$$

$$\begin{aligned}
&|Ef_n(x)-(x)| \\
&\leqslant \left|\int_{-T_0}^{T_0} K(Z)f(x-h(n)Z)dZ - \int_{-T_0}^{T_0} K(Z)f(x)dZ\right| + 2\int_{|Z|\geqslant T_0} K(Z)dZ\cdot M \\
&\leqslant \left|\int_{-T_0}^{T_0} K(Z)f(x-h(n)Z)dZ - f(x)\int_{-T_0}^{T_0} K(Z)dZ\right| + \frac{\varepsilon}{2} \\
&\to \frac{\varepsilon}{2}(n\to+\infty)
\end{aligned}$$

$$\lim_{n\to\infty} Ef_n(x) = f(x)$$

$$Var(f_n(x)) = \frac{1}{n}\cdot\frac{1}{h^2(n)}\left[EK^2\left(\frac{x-X}{h(n)}\right) - \left(EK\left(\frac{x-X}{h(n)}\right)\right)^2\right]$$

$$\lim_{n\to\infty}\frac{1}{h(n)}E\left[K^2\left(\frac{x-X}{h(n)}\right)\right] = f(x)\int_{-\infty}^{+\infty} K^2(x)dx$$

$$\lim_{n\to\infty} E(f_n(x)-f(x))^2 = \lim_{n\to\infty} Var(f_n(x)) + \lim_{n\to\infty}(Ef_n - f(x))^2 = 0$$

$$f_n(x) \xrightarrow{P} f(x)(n\to\infty)$$

附录 C　词汇表

§C.1　A

§C.2 B

dominant root	强根。特征方程中绝对值较大的根。

§C.3 C

compact	紧
curse of dimensionality	维数灾

§C.4 D

dynamic programming	动态规划
deterministic decision process	确定性决策过程

§C.5 E

exact differential equation 全微分方程，恰当微分方程

§C.6 I

information criterion	信息准则
intensive form	紧凑形式
imaginary number	虚数

§C.7 L

long range dependence 长程相依
leptokurtosis and fat-tail 高峰厚尾

§C.8 M

market clearing	市场出清
martigale	鞅
multistep decision process	多阶段决策过程

§C.9 O

over shooting	过度调整

§C.10 P

phase paths	相路径
phase trajectories	相轨道
principle of parsimony	节俭原则。模型参数尽可能地少并能适当地模仿出样本的性质。

§C.11 Q

quantile 分位数

§C.12 R

range	值域，范围，变程，区域；极差
ratchet effect	棘轮效应
recurrence	递归，循环
resource allocating Problem	资源分配问题

§C.13 S

stationary state	静止状态。相关变量保持不变，增长率为零
steady state	稳定状态。所有有关变量均按同一比率增长
sufficient and necessary condition	充要条件
supremum	上确界
stochastic decision process	随机性决策过程
shortest Path Problem	最短路线问题
singular	奇异的
sparce matrix	稀疏矩阵
splitting	分赃
stiffness matrix	刚度矩阵
stakeholders	利益相关者
subregion	子域

§C.14 T

tradeoff	折衷
triangular mesh	三角形网格

§C.15 U

uncertainty	不确定性 多种出现概率不知的结果发生的可能
undetermined	未定的
uniform refinement	均匀加密
uniform triangle net	均匀三角形网络
utility	效用 从消费一种物品或从事一项活动中所得到的满意程度
utility possibilities frontier	从两人效用水平的角度评价的所有有效率的资源配置的曲线

§C.16 V

variable cost	可变成本 随产出水平变动而变动的一项成本
variational theory	变分法
variance proportion	方差分量
volatility clustering	波动集群

§C.17 W

wave equation	波动方程
welfare economics	福利经济学 对市场和经济政策的规范评价
welfare effects	政府政策所导致的利益与损失

参 考 文 献

[1] 李子奈, 叶阿忠. 高等计量经济学. 清华大学出版社, 2002

[2] 李子奈编著. 计量经济学. 北京: 高等教育出版社, 2000

[3] Robert S. Pindyck, Daniel L. Rubinfeld. 计量经济模型与经济预测. 北京: 机械工业出版社, 1999

[4] 伍超标. 经济计量学导论. 北京: 中国统计出版社, 1998

[5] 胡定国等编著. 简明数学词典. 北京: 科学出版社, 2000

[6] 蒋中一. 数理经济学的基本方法. 北京: 商务印书馆, 2004

[7] Maurice Weir. 托马斯微积分（第十版）. 北京: 高等教育出版社, 2003

[8] 李庆杨, 王能超, 易大义编. 数值分析. 武汉: 华中科技大学出版社, 1986

[9] [美]蒋中一, 王永宏. 动态最优化基础. 北京: 商务印书馆, 1999

[10] [美] A. 帕普里斯, S. U. 佩莱. 概率、随机变量与随机过程(第四版). 西安: 西安交通大学出版社, 2004

[11] Knut Sydsaeter, Arne Strom, Peter Berck. 经济学家数学手册. 上海: 复旦大学出版社, 2001

[12] 安吉尔・德・拉・弗恩特. 经济数学方法与模型. 上海: 上海财经大学出版社, 2003

[13] [俄] A. D. 亚历山大洛夫. 数学它的内容,方法和意义(三卷本). 北京: 科学出版社, 2001

[14] [美] 约翰 L. 卡斯蒂. 20世纪数学的五大指导理论. 上海: 上海教育出版社, 2000

[15] [美]Walter Rudin(第三版). 数学分析原理. 北京: 机械工业出版社, 2004

[16] Yvonne M.M. Bishop, Stephen E. Fienberg, Paul W. Holland. 离散多元分析理论与实践. 上海: 中国统计出版社, 1998

[17] 张定胜. 计量经济学. 武汉: 武汉大学出版社, 2005

[18] 钱伯海主编. 统计学. 厦门: 厦门大学出版社, 2001

[19] 吴喜之. 现代贝叶斯统计学. 北京: 中国统计出版社, 2000

[20] 何晓群主编. 现代统计分析方法与应用. 北京: 中国人民大学出版社, 1998

[21] 孙尚拱. 隐变量分析法简介. 数理统计与管理. 1 2002, **21**(1)

[22] [美]M. R. 斯皮格尔, L. J. 斯蒂芬斯. 统计学(第三版). 北京: 科学出版社, 2002

[23] [美] 沃尔特. 恩特斯. 应用计量经济学(第二版）. 北京: 高等教育出版社, 2006

[24] [美] J. M. 伍德里奇. 计量经济学导论现代观点. 北京: 中国人民大学出版社, 2003

[25] 高惠璇等编译. SAS系统BASE SAS软件使用手册. 北京: 中国统计出版社, 1997

[26] 高惠璇等编译. SAS系统SAS/ETS软件使用手册. 北京: 中国统计出版社, 1998