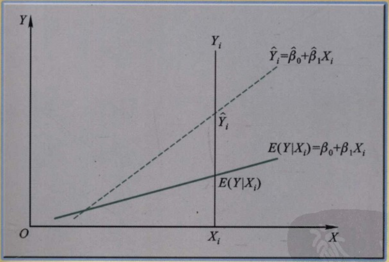
普通高等教育“十一五”国家级规划教材
面向21世纪课程教材

高等学校经济学类核心课程教材

计量经济学

（第三版）

李子奈　潘文卿　编著



高等教育出版社
HIGHER EDUCATION PRESS

★国家精品课程教材★

计量经济学、统计学主要课程教材

计量经济学(第三版)(十一五国家级规划　国家精品课程教材)	李子奈　潘文卿
计量经济学学习指南与练习（国家精品课程教材）	潘文卿　李子奈
计量经济学导论(十一五国家级规划)	王少平
计量经济学高级教程及Stata应用	陈　强
应用计量经济学：时间序列分析（第2版）	Walter Enders著　杜江 等译
计量经济学导论（十一五国家级规划　英文改编）	Jeffrey M.Wooldridge著
统计学（第三版）(十一五国家级规划)	袁　卫 等
统计学习题与案例	袁　卫 等
国民经济统计学(十一五国家级规划　国家精品课程教材)	邱　东 等
统计学	纪　宏 丁立宏
应用时间序列分析(十一五国家级规划)	史代敏
统计学基础	吴启富
统计学原理	范秀荣
统计学案例与分析	苏继伟

计量经济学课程教与学资源包


计量经济学
（第三版）

＋


计量经济学
学习指南与练习

＋


教学课件

＋


课程网站

ISBN 978-7-04-028961-9

9 787040 289619 >

定价 32.00元

普通高等教育“十一五”国家级规划教材

面向21世纪课程教材

高等学校经济学类核心课程教材

计量经济学

Jiliang Jingji Xue

（第三版）

李子奈　潘文卿　编著

高等教育出版社·北京
HIGHER EDUCATION PRESS　BEIJING

PDG

图书在版编目（CIP）数据

计量经济学 / 李子奈，潘文卿编著. —3版. —北京：
高等教育出版社，2010.3
ISBN 978-7-04-028961-9

Ⅰ.①计… Ⅱ.①李… ②潘… Ⅲ.①计量经济
学-高等学校-教材 Ⅳ.①F224.0

中国版本图书馆CIP数据核字（2010）第006799号

策划编辑 于　明　**责任编辑** 边晓娜　**封面设计** 杨立新　**责任绘图** 黄建英
版式设计 王　莹　**责任校对** 刘　莉　**责任印制** 尤　静

出版发行 高等教育出版社
社　　址 北京市西城区德外大街4号
邮政编码 100120
总　　机 010-58581000

经　　销 蓝色畅想图书发行有限公司
印　　刷 化学工业出版社印刷厂

开　　本 787×960　1/16
印　　张 25
字　　数 460 000

购书热线 010-58581118
咨询电话 400-810-0598
网　　址 http:// www.hep.edu.cn
http:// www.hep.com.cn
网上订购 http:// www.landraco.com
http:// www.landraco.com.cn
畅想教育 http:// www.widedu.com

版　　次 2000年7月第1版
2010年3月第3版
印　　次 2010年3月第1次印刷
定　　价 32.00元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。
版权所有　侵权必究

物料号　28961-00

内容简介

本书融计量经济学理论、方法与应用为一体；以中级水平内容为主，适当吸收初级和高级水平的内容；以经典线性模型为主，适当介绍一些适用的非经典模型。全书形成了一个独具特色的内容体系。

全书详细论述了经典的单方程计量经济学模型的理论方法，适当介绍了联立方程计量经济学模型和时间序列计量经济学模型的理论方法，并引入了几类扩展的单方程计量经济学模型。在计量经济学应用模型中，本书着重讨论了模型类型选择、模型变量选择、模型函数关系设定和模型变量性质设定的原则和方法。在详细介绍线性回归模型的数学过程的基础上，各章的重点不是理论方法的数学推导与证明，而是对实际应用中出现的实际问题的处理，并尽可能与中国的模型实例相结合。

本书既包含了由教育部经济学学科教学指导委员会制定的高等学校经济学科本科计量经济学课程教学基本要求的全部内容，又为学有余力者提供了进一步学习的指南。该书适合于作为各类高等院校经济、管理学科本科生的教材或教学参考书，也可供具有一定数学、经济学和经济统计学基础的经济管理人员和研究人员阅读和参考。

BRIEF INTRODUCTION

This book combines theories, methodologies with applications of econometrics. Based on middle level, some contents belong to preliminary and advanced textbooks of econometrics are included in it .It mainly introduces classical econometric models, besides, some extensive models are also introduced. So, this book has a special contents system.

In the theories and methodologies chapters of the book, the theories and methodologies about classical single-equation econometric models are discussed more detailed. The theories and methodologies about simultaneous-equations econometric models and time series econometric models are also discussed properly. Some kinds of extensive single-equation models are only introduced very briefly. In the applications chapter of the book, the principle and methods about models type specification, population regression models specification, models relationship specification and variables property specification are discussed. The mathematical process about linear regression as the basis of econometric methodologies is described very clearly. But for the mathematical process of other estimation methodologies, more attentions are paid to how to think them, how to solve the practice problems in their applications and how to combine with China's cases.

It is proper to select this book as the econometrics textbook for undergraduate students of every kinds of universities and colleges, because it covers all basic teaching contents required by the guideline and provides outstanding students with a good advanced learning materials .

总前言

高等学校经济学类核心课程和工商管理类核心课程是在高等教育面向21世纪教学内容和课程体系改革计划“经济学类专业课程结构、共同核心课程及主要教学内容改革研究与实践”和“工商管理类专业课程结构及主要教学内容改革研究与实践”两个项目调研基础上提出、经经济学学科教学指导委员会和工商管理类学科教学指导委员会讨论通过、教育部批准的必修课程。其中，经济学类各专业的核心课程共8门：政治经济学、西方经济学、计量经济学、国际经济学、货币银行学、财政学、会计学、统计学；工商管理类各专业的核心课程共9门：微观经济学、宏观经济学、管理学、管理信息系统、会计学、统计学、财务管理、市场营销学、经济法。这些课程确定后，教育部高教司组织有关专家制定了各门课程的教学基本要求，并组编了相应的各门教材。各门课程的教学基本要求及教材由高等教育出版社于2000年秋季出齐,供各高等学校选用。

教育部高等教育司
2000年3月

PDG

第三版序言

(一)

《计量经济学》(第二版)自2005年4月出版以来，受到广大读者，特别是高等学校教师和学生的广泛欢迎。大家在使用的过程中，通过各种方式对本书提出了许多宝贵的意见和建议。这些意见和建议中有些在第二版重印时已经被采纳并在书中作了相应修改，有些被第三版所吸收。值此第三版出版之际，我们对这些热心的读者表示最诚挚的谢意!

《计量经济学》教材的编写，一直得到教育部高等教育司、教育部高等学校经济学学科教学指导委员会和高等教育出版社的直接指导和大力支持。本书第一版被作为“面向21世纪课程教材”和“高等学校经济学类核心课程教材”;第二版被作为高等教育出版社“高等教育百门精品课程教材建设计划”的立项项目，并列入新闻出版总署“十五”国家重点图书出版计划;第三版又被列入教育部“普通高等教育‘十一五’国家级教材规划项目”。这些都是对我们极大的信任，他们的指导和支持使我们倍受鼓舞。也借第三版出版之际，对他们表示衷心的感谢!

《计量经济学》(第三版)吸收了我们近几年关于计量经济学模型方法论基础的部分研究成果。该项研究得到国家社会科学基金的大力支持，并于2008年作为国家社会科学基金重点项目(08AJY001，计量经济学模型方法论基础研究)立项资助。部分研究成果已经发表于《经济研究》、《经济学动态》、《统计研究》等刊物，以更广泛地听取读者的意见。也借第三版出版之际，对国家社会科学基金和有关刊物的支持表示衷心的感谢!

(二)

从20世纪70年代末、80年代初以来，我国计量经济学教学与研究的发展已经经历了引进、推广与普及的第一阶段，以及教学提高与应用扩张的第二阶段，现在已经进入提高与创新的第三阶段。从《计量经济学》(第二版)出版的2005年到现在的近5年时间内，最显著的变化就是计量经济学理论方法研究的加强和应用研究的普及。

在此期间，计量经济学理论方法研究更受重视，其研究水平也得到了进一步提高。研究生计量经济学高级课程普遍开设，一些在理论计量经济学领域有

成就的海外学者和一批在海外受到良好训练的年青学者回国，以及国内外学术交流的广泛开展，为计量经济学理论方法研究创造了基础和环境。计量经济学理论方法研究，既是学科发展的基础，又是学科水平的体现，只有加强理论方法研究，产生一批原创性成果，我们才可能融入世界计量经济学主流。这应该成为本阶段的任务之一。

在此期间，计量经济学模型方法在我国经济理论研究和经济问题分析中被普遍采用，并迅速向管理、劳动、教育、卫生、人口乃至社会等领域扩展。在我国经济类学术刊物上，以计量经济学模型方法作为主要分析方法的论文占全部论文的比例，已经迅速提高到50%以上。而且研究对象遍及经济的各个领域，所应用的模型方法遍及计量经济学的各个分支。不论以何种方式作出评价，计量经济学模型方法都已经成为我国经济研究的一种主流的实证研究方法。更为重要的是，计量经济学模型已经成为综合经济管理部门和经济类研究机构分析经济形势、研究实际经济问题和制定经济政策的常用工具，提高了经济预测和决策的水平。但是，严重问题仍然存在。没有实际意义的“自娱自乐”式的研究和存在问题甚至错误的“自欺欺人”式的研究并存，这并不是个别现象。所以，研究重要问题，采用正确的模型方法，并争取有所发现，应该成为本阶段的任务之一。

这些是《计量经济学》(第三版)修订出版的大背景。

(三)

与《计量经济学》(第二版)相比，第三版最重要的修订在于“计量经济学应用模型”。

在本书第一、二版中，经典计量经济学应用模型一直是重点内容之一。它主要通过对生产函数模型、需求函数模型、消费函数模型以及宏观计量经济模型的介绍，一方面使学生熟悉常用的计量经济学应用模型的理论模型和估计方法；另一方面，也是更重要的，使学生了解这些模型是如何提出与发展的，为学生在未来的实践中自己提出与发展新的模型打下基础。教学实践表明，这些目的是能够达到的。但是，如此设计应用模型教学内容，也存在一定的缺陷：对于更具一般意义的计量经济学模型方法论，特别是如何在应用研究中设定一个正确的模型，缺少系统的讨论。

一件小事给了我很大的教育。三年前的一天，一位同学到我的办公室，向我报告他的一个发现：需求法则在我国并不适用。因为他用我国的食品消费量作为被解释变量，以食品价格指数作为解释变量，建立了食品需求模型，经过模型估计发现，食品价格指数的参数为正，即食品消费量随着食品价格的上升而上升，与商品需求量随着价格上升而下降的需求法则相违背。我思考很久，出现这种问题的责任在哪里？不在学生，也不在教师，而是我们的教科书，至

少与教科书有关。国内外所有的计量经济学教科书，都是以模型的估计和检验为核心内容，尽管也介绍应用模型，也有大量来自于实际应用的例题，但是往往是摆出一个模型，告诉学生应该怎样估计，怎样检验，偏偏没有告诉学生这个模型是按照什么思路建立起来的。甚至许多计量经济学教科书认为，如何设定或者提出一个实际的应用模型，是理论经济学和统计学的任务，而不是计量经济学的任务。

我们讲授计量经济学的目的是什么？学生学习计量经济学的目的又是什么？为什么要将计量经济学作为一门本科学生必修的核心课程？诚然，为学生进一步学习高级课程和从事计量经济学理论方法研究打下坚实的基础，是重要的目的之一。但是，更重要的目的是为了使学生能够掌握这一主流的实证经济研究方法，正确地从事应用研究。那么为达到这一目的就应该在教学过程中告诉学生如何正确地设定或者提出一个实际的应用模型。这就是第三版将“计量经济学应用模型”作为最重要的修订内容的理由。

(四)

《计量经济学》(第三版)取消了第二版中的第七章“计量经济学应用模型”，重新编写了第九章“计量经济学应用模型”，名称虽相同，内容却有了根本的变化。在第九章中，按照计量经济学应用模型研究的步骤，设计了4节内容。将第二版第七章中生产函数、需求函数和消费函数模型中的部分内容，作为一般意义的计量经济学应用模型方法的案例纳入其中。

第1节是关于计量经济学应用模型的模型类型设定，讨论如何针对研究对象选择计量经济学模型类型，即确定所应该建立的是参数模型还是非参数模型，是单方程模型还是联立方程模型，是截面数据模型还是时间序列数据模型或者平行数据模型，是经典截面数据模型还是非经典的选择性样本模型、计数数据模型、离散选择模型或者持续时间数据模型，等等。本节还着重讨论了模型类型对数据类型的依赖性。这显然是应用模型设定的第一步。

第2节是关于计量经济学应用模型总体回归模型设定中的变量选择问题，讨论在模型类型确定之后，应该按照什么原则选择进入模型的变量。本节对“研究目的导向”、“先验理论导向”和“数据关系导向”进行了分析和批评，提出了应该按照“一般性”、“现实性”、“统计检验必要性”和“经济主体动力学关系导向”的原则选择变量。这是应用模型设定的第二步。

第3节是关于计量经济学应用模型函数关系设定，讨论如何在经济学理论和统计分析的指导下，设定模型中解释变量和被解释变量之间的关系，即模型的函数形式。这是应用模型设定的第三步。至此完成了一个应用研究的总体回归模型设定工作。

第4节是关于计量经济学应用模型变量性质设定，讨论如何确定被选择进

入模型的变量的性质，包括：它们对被解释变量具有直接影响还是间接影响？它们是内生变量还是外生变量？它们是随机变量还是确定性变量？另外，本节重点讨论了变量性质设定的相对性。这是进行模型估计之前必须进行的工作，是应用模型设定的第四步。

(五)

我从事计量经济学教学与研究已近 30 年，一直在思考两个问题并且仍然没有最终答案。一是计量经济学课程是否是经济学课程，以及如何才能使之成为真正的经济学课程？二是计量经济学模型方法是否是科学，以及如何才能使之成为真正的科学？

如上所述的计量经济学应用模型的章节设计，是目前国内外教科书中所没有见过的。从这个角度看，《计量经济学》(第三版)的内容体系设计，具有创新性。而其具体内容属于我们正在研究的计量经济学模型方法论基础的范畴，也是目前国内外教科书中所没有的，因而会有不成熟甚至存在问题之处。从这个意义讲，《计量经济学》(第三版)的具体内容，也具有创新性。那么，为什么要将本来已经被公认为比较成熟的教材进行如此修订，增加了不成熟的内容呢？其目的就是试图为上述两个问题寻求答案。

创新性的工作并不一定都是成功的。是否成功，需要实践的检验。我们真诚地希望，使用本教材的老师和学生，以及其他读者，能够就本书的体系设计和具体内容提出批评和建议。我们有决心和信心通过不断地修改完善，使之最终成为独具特色的中级计量经济学精品教材。

(六)

《计量经济学》(第三版)其他章节的主要修订包括：第二章将“一元线性回归模型的基本假设”专门列为一节，将所有基本假设按照对模型设定的假设、对解释变量的假设和对随机干扰项的假设进行分类，使之更加系统化。将第二章第 6 节标题修改为“实例及时间序列问题”，分别列举了截面数据模型和时间序列数据模型的实例，然后提出时间序列数据问题，与第八章相呼应。这一点十分重要，经典的线性回归模型理论是基于随机抽样的截面数据的，但是在实际应用中，包括在经典线性回归模型的教学中必然会大量采用时间序列数据，所以加以专门的说明是必要的。第五章“专门问题”中取消了原来的建模理论一节，相关内容已融入第九章“计量经济学应用模型”中。在第七章“扩展的单方程计量经济学模型”中增加了选择性样本模型，该类模型作为微观计量经济学模型体系中的最重要组成部分，应用十分广泛，同时该模型的介绍还可以帮助读者进一步深入理解已经学习的经典单方程计量经济学模型的理论基础。另外，本章将原来“扩展的单方程计量经济学模型”中的变参数模型和简单非线性模型进行适当简化后，融入经典单方程模型的内容之中。

《计量经济学》(第三版)对例题进行了大量的精心修改。除了进行必要的数据更新外，更重要的是尽量使例题与理论方法相一致，并使经典模型的例题与现代的模型理论方法之间不发生矛盾。

(七)

第三版仍然按照4学分70学时的课程设计教学内容，在总的内容和篇幅上与第二版相当。不同的学校可以根据学生的基础水平和学时限制，在教学安排中选择其中的部分或者全部内容。

大体上可以将教材内容分为两个层次。第一层次包括第一章至第六章和第九章，即绪论、经典单方程模型和联立方程计量经济学模型，以及计量经济学应用模型，不包括带“*”内容，这一层次属于计量经济学课程的一般教学要求。第二层次包括全部内容，但是压缩第二章至第五章的教学学时。这是本科计量经济学课程的较高教学要求。

李子奈

2009年8月于清华大学

第二版序言

(一)

计量经济学作为一门课程，在我国高等院校的经济学科、管理学科相关专业中开设，已经有20余年的历史，它的重要性也逐渐为人们所认识。1998年7月，教育部高等学校经济学学科教学指导委员会成立，在第一次会议上，讨论并确定了高等学校经济学类各专业的8门共同核心课程，其中包括“计量经济学”。2000年，我受教育部高等教育司和经济学学科教学指导委员会的委托，编著了高等学校经济学类核心课程教材《计量经济学》(第一版)，由高等教育出版社出版。

在第一版序言中，关于课程的教学目的和教材的设计原则，作了如下描述：

“试图通过课程教学，使学生达到：(1)了解现代经济学的特征，了解经济数量分析课程在经济学课程体系中的地位，了解经济数量分析在经济学科的发展和实际经济工作中的作用；(2)掌握基本的经典计量经济学理论与方法，并对计量经济学理论与方法的扩展和新发展有概念性了解；(3)能够建立并应用简单的计量经济学模型，对现实经济现象中的数量关系进行实际分析；(4)具有进一步学习与应用计量经济学理论、方法与模型的基础和能力。”

“本教材内容体系的设计原则是：(1)定位于初级与中级之间的水平。计量经济学按照内容深度一般分为初级、中级和高级三个层次。考虑到在我国高等院校本科阶段，一般只设置一个层次的计量经济学课程，而且学生具备数理统计学基础，所以将课程定位于初级与中级之间的水平。(2)理论与应用并重。计量经济学按照研究对象可以分为理论计量经济学和应用计量经济学。理论计量经济学以计量经济学的理论与方法为主要内容，强调方法的数学基础，侧重于模型方法的数学证明与推导；应用计量经济学则以计量经济学的理论与方法的应用为主要内容，强调应用模型的经济学和经济统计学基础，侧重于建立与应用模型过程中实际问题的处理。本课程将在初级与中级之间的水平上理论与应用并重。(3)在理论方法部分，重在基本原理和方法思路，尽量精简复杂的数学推导与证明。(4)必需的数学基础知识，包括矩阵运算和数理统计中的回归分析、假设检验等，属于经济类专业本科生数学课程的基本要求，不出现在课程内容中，由学生自己学习与复习。(5)属于中、高级的，但是十分重要的内容和非经

典的理论方法，在课程中作概念性介绍，为学生进一步学习建立一个基础。(6)加强综合练习。通过综合练习，给学生以理论、方法与应用的综合能力，并学会使用计量经济学软件包。综合练习不占课内学时。(7)具有较宽的适用面。不同的学校、不同的专业、不同的先修课程基础，以及不同的学时，对课程教学的要求是不同的。在保证基本教学要求的情况下，整章、整节的舍弃，不影响教学内容体系的完整和前后衔接。”

这些，仍然是我们编写本书的指导思想和原则。

(二)

《计量经济学》(第一版)作为“面向21世纪课程教材”和“高等学校经济学类核心课程教材”，于2000年7月出版以来，被广泛采用。在使用过程中，众多高等院校教师就教材的内容体系和具体章节中存在的问题，提出了很多宝贵的意见。在这四年中，计量经济学的理论方法和应用研究也有了新的发展，最具有代表性的是计量经济学家两度获得诺贝尔经济学奖。2000年的诺贝尔经济学奖授予在微观计量经济学领域作出突出贡献的赫克曼(J.Heckman)和麦克法登(D.McFadden)，恩格勒(R. F. Engle)和格兰杰(C.W. J. Granger)由于在时间序列计量经济学领域的贡献而于2003年获奖。这极大地推动了计量经济学课程教学的发展，并在相当大的程度上改变了计量经济学的课程教学。更为重要的是，在这四年中，计量经济学课程在我国众多高等院校中已经普遍开设，教师的水平有了显著提高，学生的知识基础，尤其是数学和理论经济学基础得到了加强；而应用研究也已经普遍开展，翻开国内主要的经济类学术期刊，可以看到，建立计量经济学模型研究分析中国现实经济问题已经成为论文的主体。所有这些，都对修订《计量经济学》(第一版)提出了迫切的需求。

与第一版相比较，本书有以下几方面变化：

第一，加强了基础内容。经典的单方程计量经济学模型是最基本和应用最普遍的计量经济学模型，其理论方法也是联立方程计量经济学模型和后来发展的各种现代计量经济学模型的基础，毫无疑问应该成为课程教学的重点。在第一版中，单方程计量经济学模型理论方法部分较为简洁，适于具有较好的应用数理统计学基础的学生采用。但是，我国大部分高等院校的经济院系，并没有专门开设应用数理统计学课程。为此，在本书中，将原来的第二章“单方程计量经济学模型理论与方法”扩充为第二、三、四、五章，增加了单方程计量经济学模型理论方法的数理统计学基础和模型设定与检验的有关专题内容，使得这部分内容更加充实与系统。

第二，引入了学科前沿内容。与第一版比较，本书引入了属于微观计量经济学的离散选择模型和平行数据模型，尽管只是最简单的部分，但是为学生了解这些模型打下了基础。将原来只有两节的时间序列计量经济学内容扩充为完

整的一章，比较系统地介绍了发展迅速且应用领域广阔的这一现代计量经济学的分支。另外，对诸如广义矩估计等新近发展的理论方法，也作为经典理论方法的延伸而作了概念性介绍。

第三，增加了实际例题。在理论与应用的结合上，除了保留第一版中专门设计的应用模型一章外，将重点放在精心编写的实际例题上。专门的应用模型章节，目的是训练学生分析经济行为，建立理论模型的能力。而紧随各部分理论方法的应用实例，对于学生正确地理解和应用这些理论方法，是十分必要的。书中的例题都是中国的实际经济问题的分析，有些例题贯穿全章，甚至几章，随着理论方法的深入而反复采用，对于教师的“教”和学生的“学”都是十分有益的。

第四，改变了应用软件。编写第一版时，在征求部分高等院校教师意见的基础上，选择了当时大家普遍拥有的TSP6.5作为教学软件，模型方法和例题都结合TSP6.5讲授。现在，EViews作为目前世界上最流行的计量经济学软件之一，已经普遍应用，而且它继承了TSP的优点，功能齐全，操作简单、灵活。所以，在本书中选择EViews作为教学软件。

第五，压缩了联立方程计量经济学模型。联立方程计量经济学模型是经典计量经济学内容体系的重要组成部分，它的应用领域主要是宏观经济模型。考虑到教材篇幅和本科生需要掌握的知识重点，在本书中只保留联立方程计量经济学模型理论方法中的几种单方程估计方法，同时将原来的宏观计量经济学模型由一章压缩成一节。

(三)

本书按照4学分70学时的课程设计教学内容，在总的内容和篇幅上多于第一版，这样，可以使教材更具有适用性。不同的学校可以根据学生的基础水平和学时限制，在教学安排中选择其中的部分或者全部内容。

本书内容大体上可以分为三个层次。

第一层次包括第一章至第六章，即绪论、经典单方程计量经济学模型和联立方程计量经济学模型，不含带“*”的部分。这是本科生必须掌握的计量经济学中最基础和最成熟的内容，适合于3学分课程且学生的数学和理论经济学背景较弱的情况。这是计量经济学课程的最低教学要求。

第二层次包括第一章至第七章，以及第八章中的一部分，不含带“*”的部分，相当于第一版教材的基本要求，适合于4学分课程且学生的数学和理论经济学背景一般的情况。这是计量经济学课程的一般教学要求。

第三个层次包括全部内容，但要压缩第二章至第五章的教学学时，适合于4学分课程且学生的数学和理论经济学背景较强的情况。这是计量经济学课程的较高教学要求。

（四）

本书共分九章。

第一章，绪论，是本书的纲。通过教学，要求学生达到：了解计量经济学的基本概念；了解计量经济学的内容体系，以及本课程涉及的内容；理解计量经济学是一门经济学科，以及它在经济学科中的地位；了解计量经济学的主要应用；了解建立与应用经典计量经济学模型的工作步骤，以及在每一步骤中应注意的关键。对于未接触过计量经济学的学生来讲，并不能全部理解，也不要求学生全部理解，只需要建立一个最基本的概念，对于学习整个课程是大有益处的。

第二章和第三章，分别为经典单方程计量经济学模型的一元线性回归模型和多元线性回归模型，是本书最基础的内容。通过教学，要求学生达到：理解经典线性单方程计量经济学模型的数理统计学基础，包括回归分析、假设检验和区间估计；熟练掌握经典线性单方程计量经济学模型的理论与方法，包括基本假设、模型估计和统计检验；理解最小二乘原理和最大似然原理，以及在模型估计中的应用；能够运用矩阵描述、推导和证明与普通最小二乘法有关的估计过程和结论；能够应用计量经济学软件完成模型的估计和统计检验。在这两章结束时要求学生独立完成一个综合练习，自己选择研究对象，自己建立理论模型，自己收集样本数据，进行模型的估计和统计检验。

第四章，放宽基本假定的经典单方程计量经济学模型，即经典单方程计量经济学模型的计量经济学检验，也是课程的基础内容。通过教学，要求学生达到：了解实际经济分析中计量经济学模型违背各个基本假设的经济背景；从经济学和数学两个方面理解违背基本假设的后果；理解并熟练掌握常用的检验方法；熟悉各种基本假设违背情况下模型最有效和最常用的估计方法，如加权最小二乘法、可行的广义最小二乘法、差分法与广义差分法、工具变量法等，以及它们在应用软件中的实现。在本章结束时要求学生对前面完成的综合练习进行计量经济学检验，重新估计模型，对结果进行分析，并提交一篇报告。

第五章，经典单方程计量经济学模型的几个专门问题，作为前面三章经典单方程计量经济学模型理论方法的补充，在理论和应用上都是不可缺少的。通过教学，要求学生达到：理解在模型中引入虚拟变量和滞后变量的问题背景、引入原则和方法；熟悉分布滞后模型和自回归模型及其参数估计方法；熟练应用格兰杰检验于模型变量选择和变量关系分析；理解模型的变量选择和关系设定可能带来模型的确定性偏误；掌握常用的检验方法。

第六章，联立方程计量经济学模型理论与方法，是课程的重点内容之一。通过教学，要求学生达到：理解线性联立方程计量经济学模型的基本概念和有关模型识别、检验的理论与方法；熟练掌握几种主要的单方程估计方法，能够

运用矩阵描述、推导和证明与这些方法有关的过程和结论；能够独立完成由3~5个方程组成的简单联立方程计量经济学模型的建模全过程工作；能够应用计量经济学软件。在本章结束前要求学生独立完成一个综合练习，建立一个3~5个方程的中国宏观经济模型，自己建立理论模型，收集样本数据，用几种方法进行模型的估计，对结果进行分析，最后提交一篇报告。

第七章，经典计量经济学应用模型，是课程的重点内容之一。通过本章前三节的教学，一方面使学生熟悉常用的计量经济学应用模型的理论模型和估计方法；另一方面，也是更重要的方面，使学生了解这些模型是如何提出与发展的，为学生在未来的实践中自己提出与发展新的模型打下方法论基础。所以在本章的每一节都应有不同的建模方法论重点。例如，在生产函数模型中，着重介绍各种生产函数模型是如何沿着要素之间替代性质的描述和技术要素的描述这两条线索逐渐发展的；在需求函数模型中，着重介绍各种需求函数模型是如何依赖于效用函数而发展的；在消费函数模型中，着重介绍各种消费函数模型是如何依赖于各种消费理论假设而提出的，等等。在教学内容安排上，视学生的宏观经济学和微观经济学水平及专业方向而有所取舍。本章中的宏观计量经济学模型是课程的选学内容，可以视学时安排和教学要求选择全部或部分内容，或者不选。通过教学，使学生达到：了解计量经济学模型的一个重要研究与应用领域——宏观经济；掌握宏观计量经济学模型的设定理论；了解中国宏观计量经济学模型的主要特征、总体结构和主要模块与方程的设计；能够看懂和应用已有的宏观计量经济学模型。

第八章，扩展的单方程计量经济学模型，是课程的选学内容，可以视学生的基础水平和教学要求选择全部或部分内容，或者不选。通过本章教学，一方面扩展学生的知识面，为学生今后进一步学习和应用计量经济学理论与方法打下基础，使学生理解：单方程计量经济学模型是一个内容广泛的体系，经典的线性模型是其中最基本和最重要的一部分，以及几类扩展模型的研究对象、基本理论和方法思路。另一方面，使学生掌握一些重要的知识点。例如，确定性变参数模型的经济含义和估计方法；非线性普通最小二乘法的原理及其在应用软件中的实现；二元离散选择模型的实际应用价值，从原始模型到效用模型的原理，二元Probit模型和Logit模型的参数估计方法及其在应用软件中的实现；平行数据(panel data)模型的设定检验，固定影响变截距模型的最小二乘虚拟变量估计方法和固定影响变系数模型的可行广义最小二乘估计方法。

第九章，时间序列计量经济学模型。虽然是课程的选学内容，但它是现代计量经济学的重要组成部分，已经形成了独立的分支和课程，在学生的基础水平和学时允许的情况下应尽可能选学。通过教学，要求学生达到：了解时间序列平稳性的概念、重要性和检验方法，尤其是单位根检验；掌握三类常用的随

机时间序列模型的识别、估计和检验方法；了解协整的概念、重要性和检验方法；了解误差修正模型的经济意义和建立误差修正模型的全过程，并能够建立实际的误差修正模型；熟悉应用软件中时间序列分析的基本功能，并能够应用软件完成时间序列平稳性检验、单位根检验和协整检验。

（五）

本书作为高等教育出版社“高等教育百门精品课程教材建设计划”立项项目，已列入新闻出版总署“十五”国家重点图书出版规划。同时，本书还是国家精品课程配套教材。本书尚有专门的习题集与之配套，所以在教材的每章只附有少量的习题。这些习题只是为了帮助学生把握课程内容的重点和难点，并不足以帮助学生深入地理解和正确地应用计量经济学的理论方法，即使是习题集也是这样。而要做到这一点，综合练习是不可缺少的。在课程学习的同时，选择适当的现实经济问题，建立计量经济学模型，完成建模的全过程，是对课程内容最好的复习，是最好的“习题”。

本书配有相关的电子课件，包括电子教案、数据集和相关试卷等，读者可通过书后所附回执免费索取。在电子课件的制作过程中，我们注重的是内容，而在技术上着力不够。教材的修订往往要间隔一段时间，但是电子课件更新是很方便的，我们努力争取在本书每一次重印时，都会更新电子课件的内容。同时，读者还可通过书后配套的学习卡登录高等教育出版社的网站(http://4a.hep.edu.cn，http://la.hep.com.cn)，浏览相关的网络课程并进行教学答疑。

（六）

在本书编著过程中，参考了国内外许多计量经济学教科书，包括国外最新的教材，在本书的参考文献中列出了书名。

“计量经济学”作为清华大学首批重点建设的精品课程之一，得到了学校教学管理部门的指导和支持。本书作为课程建设成果之一，通过了学校组织的专家评审，并已经在清华大学的课堂上使用两届，专家和学生们提出了许多宝贵的意见。高等教育出版社将本书列入“高等教育百门精品课程教材建设计划”，并给予大力指导和支持。在此，对于高等教育出版社、清华大学教学管理部门、有关参考书的作者、专家和学生们一并表示最衷心的感谢。

由于我们水平有限，即使在计量经济学领域学识也很肤浅，书中定有不妥甚至错误之处，恳请读者批评指正。

李子奈

2004 年 9 月于清华大学

第一版序言

计量经济学作为一门课程，在我国一部分高等院校的经济学科、管理学科相关专业中开设，已经有近20年的历史，它的重要性也逐渐为人们所认识。1996年7月，我作为召集人承担了教育部(原国家教委)“高等教育面向21世纪教学内容和课程体系改革计划”的重点项目“经济类专业数量分析系列课程设置和教学内容研究”的研究工作，在广泛调查研究的基础上，提出了课程设置的初步方案；1997年7月，利用中国数量经济学会年会的机会，在近百所高校教师中进行了充分讨论，正式提出了“经济类专业数量分析系列课程设置研究报告”，建议将计量经济学列入经济类专业核心课程，所有专业都要开设。随后我即开始准备编写一本教材，作为项目的一个研究成果。1998年7月，教育部高等学校经济学学科教学指导委员会成立，在第一次会议上，讨论并确定了高等学校经济学门类各专业的8门共同核心课程，其中包括计量经济学。将计量经济学首次列入经济类专业核心课程，是我国经济学学科教学走向现代化和科学化的重要标志，必将对我国经济学人才培养质量产生重要影响，也使我受到很大的鼓舞，加快了编写该教材的步伐。

我自1986年起，一直从事计量经济学的教学工作。1992年由清华大学出版社出版的由我编著的《计量经济学——方法与应用》一书，属于中级水平的计量经济学教材，为许多学校所采用，并获得1995年国家教委优秀教材一等奖。1994年至1995年间，全国高等院校数量经济学会在原国家教委高教司的支持下，组织有关高校编写了计量经济学教学大纲，并于1995年5月在《数量经济技术经济研究》上发表。这两项成果为编写本教材提供了重要基础。同时，近年来在全国高校悄然兴起的关于教育思想的讨论，尤其是关于如何培养学生不断地学习新知识，从事新工作的能力的讨论，也为本教材的编写提供了指导原则。

本教材按照50～70课内学时，课内外学时比为1：2设计其内容体系。以微积分、线性代数、概率论与数理统计、微观经济学、宏观经济学和经济统计学为先修课程。试图通过课程教学，使学生达到：(1)了解现代经济学的特征，了解经济数量分析课程在经济学课程体系中的地位，了解经济数量分析在经济学科的发展和实际经济工作中的作用；(2)掌握基本的经典计量经济学理论与方

法，并对计量经济学理论与方法的扩展和新发展有概念性了解；(3)能够建立并应用简单的计量经济学模型，对现实经济现象中的数量关系进行实际分析；(4)具有进一步学习与应用计量经济学理论、方法与模型的基础和能力。

本教材内容体系的设计原则是：(1)定位于初级与中级之间的水平上。计量经济学按照内容深度一般分为初级、中级和高级三个层次。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计量经济学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计量经济学模型理论与方法，经典的线性联立方程计量经济学模型理论与方法，以及传统的应用模型为主要内容；高级以扩展的单方程计量经济学模型理论与方法，非线性模型理论与方法，以及动态计量经济学理论与方法为主要内容。考虑到在我国高等院校本科阶段，一般只设置一个层次的计量经济学课程，而且学生具备数理统计学基础，所以将课程定位于初级与中级之间的水平上。(2)理论与应用并重。计量经济学按照研究对象可以分为理论计量经济学和应用计量经济学。理论计量经济学以计量经济学的理论与方法为主要内容，强调方法的数学基础，侧重于模型方法的数学证明与推导；应用计量经济学则以计量经济学的理论与方法的应用为主要内容，强调应用模型的经济学和经济统计学基础，侧重于建立与应用模型过程中实际问题的处理。本课程将在初级与中级之间的水平上理论与应用并重。(3)在理论方法部分，重在基本原理和方法思路，尽量精简复杂的数学推导与证明。(4)必须的数学基础知识，包括矩阵运算和数理统计中的回归分析、假设检验等，属于经济类专业本科生数学课程的基本要求，不出现在课程内容中，由学生自己学习与复习。(5)属于中、高级的，但是十分重要的内容和非经典的理论方法，在课程中作概念性介绍，为学生进一步学习建立一个基础。(6)加强综合练习。通过综合练习，给学生以理论、方法与应用的综合能力，并学会使用计量经济学软件包。综合练习不占课内学时。(7)具有较宽的适用面。不同的学校、不同的专业、不同的先修课程基础，以及不同的学时，对课程教学的要求是不同的。在保证基本教学要求的情况下，整章(例如第三章、第六章)、整节(用“*”标出的)的舍弃，不影响教学内容体系的完整和前后衔接。

全书共分六章和附录。

第一章，绪论，是课程的纲。通过教学，要求学生达到：了解计量经济学的基本概念；了解计量经济学的内容体系以及本课程涉及的内容；理解计量经济学是一门经济学科以及在经济学科中的地位；了解计量经济学的主要应用；了解建立与应用计量经济学模型的工作步骤，以及在每一步骤应注意的关键。对于未接触过计量经济学的学生来讲，并不能全部理解，也不要求学生全部理解，只需要建立一个最基本的概念，对于学习整个课程是大有益处的。

第二章，单方程计量经济学模型理论与方法，是课程的重点和主要内容，

应占总课内学时的1/3以上。通过教学，要求学生达到：熟练掌握线性单方程计量经济学模型的理论与方法；能够运用矩阵描述、推导和证明与普通最小二乘法有关的过程和结论；能够独立完成建立线性单方程计量经济学模型的全过程工作；能够应用计量经济学软件。在教学中注意课堂讲授与课外练习的结合。在本章结束前要求学生独立完成一个综合练习，自己选择研究对象，自己建立理论模型，自己收集样本数据，进行模型的估计和检验，最后提交一篇报告。这对于课程内容的理解和能力的培养都是十分必要的。

第三章，扩展的单方程计量经济学模型理论与方法，是课程的选学内容，可以视学生的基础水平和教学要求选择全部、部分内容，或者不选。通过本章教学，一方面扩展学生的知识面，更重要的是为学生今后进一步学习和应用计量经济学理论与方法打下基础。使学生理解：单方程计量经济学模型是一个内容广泛的体系，经典的线性模型是其中最基本和最重要的一部分，以及几类扩展模型的研究对象、基本理论和方法思路。

第四章，联立方程计量经济学模型理论与方法，是课程的重点内容之一。通过教学，要求学生达到：理解线性联立方程计量经济学模型的基本概念和有关模型识别、检验的理论与方法；熟练掌握几种主要的单方程估计方法，能够运用矩阵描述、推导和证明与这些方法有关的过程和结论；能够独立完成由3～5个方程组成的简单联立方程计量经济学模型的建模全过程工作；能够应用计量经济学软件。在本章结束前要求学生独立完成一个综合练习，建立一个3～5个方程的中国宏观经济模型，自己建立理论模型，自己收集样本数据，用几种方法进行模型的估计，对结果进行分析，最后提交一篇报告。

第五章，单方程计量经济学应用模型，是课程的重点内容之一。通过本章教学，一方面使学生熟悉常用的计量经济学应用模型的理论模型和估计方法；另一方面，也是更重要的方面，使学生了解这些模型是如何提出与发展的，为学生在未来的实践中自己提出与发展新的模型打下方法论基础。所以在本章的每一节都有不同的建模方法论重点。例如，在生产函数模型中，着重介绍各种生产函数模型是如何沿着要素之间替代性质的描述和技术要素的描述这两条线索逐渐发展的；在需求函数模型中，着重介绍各种需求函数模型是如何依赖于效用函数而发展的；在消费函数模型中，着重介绍各种消费函数模型是如何依赖于各种消费理论假设而提出的；等等。在教学内容安排上，视学生的宏观、微观经济学水平和专业方向而有所取舍。

第六章，宏观计量经济学模型，是课程的选学内容，可以视学时安排和教学要求选择全部、部分内容，或者不选。通过教学，使学生达到：了解计量经济学模型的一个重要研究与应用领域——宏观经济；掌握宏观计量经济学模型的设定理论；了解不同体制、不同发展阶段下宏观计量经济学模型的异同；了

解中国宏观计量经济学模型的主要特征、总体结构和主要模块与方程的设计；能够看懂和应用已有的宏观计量经济学模型。

在附录中，除了几种必用的统计分布表外，还专门介绍了TSP6.5软件的应用。学习计量经济学课程，必须学会使用至少一种应用软件，这是一项基本教学要求。但是，学习使用软件不是依靠课堂，而是靠练习。计量经济学应用软件包种类很多，没有必要规定必须使用哪种，所以没有将最常用的TSP6.5软件的应用介绍放在本书的正文中。

在本书编著过程中，除了主要参考我本人编著的《计量经济学——方法与应用》外，还参考了《经济计量学》(张保法著，河南人民出版社，1992年)，《计量经济学》(张寿，于清文编著，上海交通大学出版社，1984年)，《计量经济学——理论、方法和模型》(唐国兴编著，复旦大学出版社，1988年)，《经济计量学》(G.C.Chow著，郑宗成等译，中国友谊出版公司，1988年)，《经济预测与决策技术》(冯文权编著，武汉大学出版社，1989年)，《经济计量学教科书》(L.Klein著，谢嘉译，商务印书馆，1983年)，《计量经济学》(陈正澄著，台湾三民书局，1980年)，《动态经济计量学》(D.Hendry，秦朵著，上海人民出版社，1998年)，《经济计量学理论与实践引论》(G.G.Judge等著，周逸江等译，中国统计出版社，1993年)，《宏观经济模型论述》(汪同三著，经济管理出版社，1992年)，《应用经济计量学教程》(吴承业，龚德恩编著，中国铁道出版社，1996年)，*Introductory Econometrics*: *Theory and Applications* (R.L.Thomas, Longman Inc.,1985)，*Econometric Models*, *Techniques*, *and Applications* (M.D.Intriligator, R.G.Bodkin,Cheng Hsiao, Prentice-Hall International Inc.,1996)，*Introduction to Econometrics* (G.S.Maddala, Prentice-Hall International Inc.,1992)，*Econometric Analysis* (W.H.Greene, Prentice-Hall International Inc.,1997)等教科书和专著，以及我曾经指导过的学生们的学位论文和综合练习。在此向有关作者表示感谢。

由于本人水平有限，即使在计量经济学领域学识也很肤浅，书中定有不妥甚至错误之处，恳请读者批评指正。

李子奈
1998年12月

目 录

第四章　经典单方程计量经济学模型：放宽基本假定的模型 …… 107

PDG

第七章　扩展的单方程计量经济学模型……229

CONTENTS

第一章　绪　论 1

本章是全书的纲，将对计量经济学进行总体上的介绍，并对建立与应用计量经济学模型的步骤和要点进行简要的说明。尽管第一次学习计量经济学的读者可能不能完全理解本章的内容，但是建立起一个概念对于学习全书是十分重要的。

§1.1　计量经济学

一、计量经济学

计量经济学是经济学的一个分支学科，是以揭示经济活动中客观存在的数量关系为内容的分支学科。

英文“econometrics”最早是由挪威经济学家弗里希(R.Frish)于1926年模仿“biometrics”(生物计量学)提出的，它的提出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它创办的学术刊物Econometrica于1933年正式出版，才标志着计量经济学作为一门独立学科正式诞生了。计量经济学从诞生之日起，就显示了极强的生命力，经过20世纪40年代和50年代的大发展及60年代的大扩张，已经在经济学科中占据极其重要的地位。正如著名计量经济学家、诺贝尔经济学奖获得者克莱因(R.Klein)在 *A Textbook of Econometrics* 的序言中所评价的：**“计量经济学已经在经济学科中居于最重要的地位”，“在大多数大学和学院中，计量经济学的讲授已经成为经济学课程表中最有权威的一部分”**。著名经济学家、诺贝尔经济学奖获得者萨缪尔森(P.Samuelson)甚至说：**“第二次世界大战后的经济学是计量经济学的时代”**。

弗里希将计量经济学定义为经济理论、统计学和数学三者的结合。1933年在 *Econometrica* 的创刊号社论中，弗里希写下了一段话：**“用数学方法探讨经济学可以从好几个方面着手，但任何一个方面都不能和计量经济学混为一谈。计量经济学与经济统计学绝非一码事；它也不同于我们所说的一般经济理论，尽管经济理论大部分具有一定的数量特征；计量经济学也不应视为数学应用于经济学的同义语。”“经验表明，统计学、经济理论和数学这三者对于真正了解**

现代经济生活的数量关系来说，都是必要的，但本身并非是充分条件。三者结合起来，就是力量，这种结合便构成了计量经济学。”

自20世纪80年代以来，计量经济学在我国得到迅速传播与发展。在有关的出版物和课程表中出现了“计量经济学”与“经济计量学”两种名称。“经济计量学”是由英文“econometrics”直译得到的，而且强调该学科的主要内容是经济计量的方法，是估计经济模型和检验经济模型；“计量经济学”则试图通过名称强调它是一门经济学科，强调它的经济学内涵与外延，本书以“计量经济学”为名，也在于此。但实际上，翻开两类不同名称的出版物，就会发现其内容并无区别。

二、计量经济学模型

模型，是对现实的描述和模拟。用各种不同的方法对现实进行描述和模拟，就构成了各种不同的模型，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。语义模型用语言来描述现实，例如，对处于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。物理模型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。几何模型用图形来描述现实，例如，一个零部件的加工图。数学模型用数学语言描述现实，也是一种重要的模型，由于它能够揭示现实活动中的数量关系，所以具有其特殊重要性。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。

经济数学模型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。在这里，我们着重区分数理经济模型和计量经济学模型。

数理经济模型揭示经济活动中各种因素之间的理论关系，用确定性的数学方程加以描述。例如，上述用语言描述的生产活动，可以用生产函数描述如下：

$$Q = f(T, K, L)$$

或者更具体地用某一种生产函数描述为

$$Q = Ae^{\gamma t} K^{\alpha} L^{\beta}$$

公式中用 Q 表示产出量，T 表示技术，K 表示资本，L 表示劳动。公式描述了技术、资本、劳动与产出量之间的理论关系，认为这种关系是准确实现的。利用数理经济模型，可以分析经济活动中各种因素之间的相互影响，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，因为在上面的公式中，参数α，β，γ是未知的。

计量经济学模型揭示经济活动中各种因素之间的定量关系，用随机性的数学方程加以描述。例如，上述生产活动中各因素之间的关系，用随机数学方程描述为

$$Q = A\mathrm{e}^{\gamma t} K^{\alpha} L^{\beta} \mu$$

其中μ为随机干扰项。这就是计量经济学模型的理论形式。例如，以中国全民所有制工业生产活动为研究对象，以 1964 年至 1984 年中国全民所有制工业生产活动的数据为样本，就可以应用计量经济学方法得到如下关系：

$$Q = 0.647\,9\mathrm{e}^{0.012\,8t} K^{0.360\,8} L^{0.675\,6}$$

上式揭示了这个特定问题中技术、资本、劳动与产出量之间的定量关系。利用这个关系，可以对研究对象进行进一步深入研究，如结构分析和生产预测等。这就是计量经济学模型得到高度重视和广泛应用的原因所在。

从上面的例子中也可以看到经济理论、数理经济学和计量经济学在经济研究中各自的位置和作用。

三、计量经济学的内容体系

计量经济学作为经济学的一个分支学科，在经济学科中居于最重要的位置，其理论方法已经形成了庞大的内容体系。在国内外的大学中，也绝非一门课程就可以涵盖其全部内容，一般分为多个层次的多门课程。于是，出现了关于计量经济学内容体系的各种分类和各种带有计量经济学名称的教科书。下面仅对此作最简单的介绍并借以说明本教材的定位。

1. 广义计量经济学和狭义计量经济学

广义计量经济学是利用经济理论、统计学和数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。在西方许多以“econometrics”为名称的书中，往往包含如此广泛的内容。这些方法，尽管都是经济理论、统计学和数学的结合，但是它们之间的区别是显而易见的。

狭义计量经济学，也就是我们通常所说的计量经济学，以揭示经济现象中的因果关系为目的，在数学上主要应用回归分析方法。本书中的计量经济学模型就是这个意义上的经济数学模型。

2. 初、中、高级计量经济学

计量经济学按照内容深度一般分为初级、中级和高级三个层次。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计量经济学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计量经济学模型理论与方法、经典的线性联立方程计量经济学模型理论与方法，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。

考虑到在我国高等院校本科阶段，一般只设置一个层次的计量经济学课程，而且学生具备数理统计学基础，所以本书定位于中级水平。大部分内容属于中级水平，少部分内容属于高级水平，但是它是站在整个学科内容体系的角度来安排的。这就可以使读者在学习与掌握中级内容的同时，对整个学科内容体系形成一个完整的认识，并对它的最新发展有所了解，为进一步学习打好基础。

3．理论计量经济学和应用计量经济学

计量经济学根据研究对象和内容侧重点不同，可以分为理论计量经济学和应用计量经济学。理论计量经济学以介绍、研究计量经济学的理论与方法为主要内容，侧重于理论与方法的数学证明与推导，与数理统计联系极为密切。理论计量经济学除了介绍计量经济学模型的数学理论基础和普遍应用的计量经济学模型的参数估计方法与检验方法外，还研究特殊模型的估计方法与检验方法，应用了广泛的数学知识。应用计量经济学则以建立与应用计量经济学模型为主要内容，强调应用模型的经济学和经济统计学基础，侧重于对建立与应用模型过程中实际问题的处理。本书是二者的结合。

4．经典计量经济学和非经典计量经济学

计量经济学于 20 世纪 20 年代末 30 年代初创立。经过 40 年代和 50 年代的发展及 60 年代的扩展，应该说，到 60 年代末，计量经济学作为一门学科已经成熟。自 20 世纪 70 年代以来，由于经济活动复杂性增强和计量经济学应用领域的扩展，计量经济学理论方法得到了很大的发展，并形成了微观计量经济学、非参数计量经济学、时间序列计量经济学和平行数据(Panel Data)计量经济学等新的分支。

经典计量经济学(classical econometrics)一般指 20 世纪 70 年代以前发展并广泛应用的计量经济学，它们具有显著的共同特征。其理论方法方面的特征是：(1)模型类型：采用随机模型；(2)模型导向：以经济理论为导向建立模型；(3)模型结构：变量之间的关系表现为线性或者可以化为线性，属于因果分析模型，解释变量具有同等地位，模型具有明确的形式和参数；(4)数据类型：以时间序列数据或者截面数据为样本，被解释变量为服从正态分布的连续随机变量；(5)估计方法：仅利用样本信息，采用最小二乘方法或者最大似然方法估计模型。其应用方面的特征是：(1)应用模型的方法论基础：实证分析，经验分析，归纳；(2)应用模型的功能：结构分析，政策评价，经济预测，理论检验与发展；(3)应用模型的领域：传统的应用领域，如生产、需求、消费、投资、货币需求，以及宏观经济等。

非经典计量经济学一般指 20 世纪 70 年代以后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，主要包括前面所提及的微观计量经济学、非参数计量经济学、时间序列计量经济学和平行数据计量经济学等。也可以按照界定经典模型理论方法的五个方面，即模型类型、模型导向、模型结构、数据类型和估计方法，将所有“非经典”的计量经济学问题分类，使之形成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五大类，构成非经典计量经济学的内容体系。实践表明，这样的划分与界定，至少对于“教”与“学”来讲，是十分有益的。

本书以经典计量经济学为主，适当引入一些简单的、应用较多的现代计量经济学理论方法。一方面，从理论方法角度，经典计量经济学理论方法是非经典计量经济学理论方法的基础；另一方面，从应用的角度，经典计量经济学模型仍然是目前应用最为普遍的计量经济学模型。所以，作为本科生教材，以经典计量经济学为主无疑是适当的。为了进一步学习和应用的需要，对某些应用较多的现代计量经济学理论方法作简单介绍，也是完全必要的。

5. 微观计量经济学和宏观计量经济学

类似于经济学中的微观经济学和宏观经济学，也可以按研究对象将计量经济学分为微观计量经济学(Microeconometrics)和宏观计量经济学(Macroeconometrics)。

2000 年诺贝尔经济学奖授予对微观计量经济学作出原创性贡献的经济学家赫克曼(J.Heckman)和麦克法登(D.McFaddan)。然而，在 2000 年诺贝尔经济学奖公布之前，学术界和文献中还没有正式提出“微观计量经济学”这一概念，正是在 2000 年诺贝尔经济学奖公报中才正式提出。在以《微观计量经济学和微观数据》为题的公报中，将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”，而“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数据。正因为近些年来关于个人和家庭的微观数据的显著增加，才使得微观计量经济学得到了很大的发展。2000 年以来，关于微观计量经济学的研究形成了新的高潮，以 *Microeconometrics*，*Advanced Microeconometrics*，*Applied Microeconometrics*，*Topics in Microeconometrics*，*Methods in Microeconometrics* 等为名的教科书纷纷出版，相关课程纷纷设立。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和选择性样本模型的理论方法。它们都属于非经典的现代计量经济学，本书仅仅介绍其中最简单的部分内容。

宏观计量经济学的名称由来已久，但是它的主要内容和研究方向发生了变化。利用计量经济学理论方法，建立宏观经济模型，并对宏观经济进行分析、评价和预测，一直是计量经济学的主要研究领域，因此，长期以来，经典的宏

观计量经济学模型理论、方法和应用构成宏观计量经济学的主要内容。但是近20多年来，单位根检验、协整理论以及动态计量经济学则成为宏观计量经济学的主要研究方向，以至在2001年*Journal of Econometrics*发行100期的纪念专辑上，特别邀请著名计量经济学家、单位根和协整理论的重要创始人、2003年诺贝尔经济学奖获得者格兰杰(C. W. J. Granger)和在动态时间序列分析领域作出突出贡献的著名计量经济学家斯托克(J. H. Stock)分别以“Macroeconometrics—Past and Future”和“Macroeconometrics”为题发表两篇综述性论文，他们都将单位根和协整理论作为宏观计量经济学的重要内容。本书将介绍经典的宏观计量经济学模型，同时对于宏观计量经济学的新发展，以及在时间序列分析部分的单位根检验和协整理论也作简单的介绍。

四、计量经济学是一门经济学科

经常遇到一些学过或者看过计量经济学教科书的人提出这样的问题：计量经济学属于经济学还是应用数学？或者说，学了计量经济学，方法知道了不少，就是不会用，也不知道用在哪里。这是一个重要而又实际的问题。

在本书开篇第一句，我们就指出：计量经济学是经济学的一门分支学科，即它是一门经济学科。为什么？

第一，从计量经济学的定义看。前面已经介绍，弗里希将计量经济学定义为经济理论、统计学和数学三者的结合，而且他明确提出：“计量经济学与经济统计学绝非一码事；它也不同于我们所说的一般经济理论，尽管经济理论大部分具有一定的数量特征；计量经济学也不应视为数学应用于经济学的同义语。”“经验表明，统计学、经济理论和数学这三者对于真正了解现代经济生活的数量关系来说，都是必要的，但本身并非是充分条件。三者结合起来，就是力量，这种结合便构成了计量经济学。”我们不妨把这种结合称之为定量化的经济学或者经济学的定量化。

第二，考察一下计量经济学在西方国家经济学科中的地位。如前所说，在西方国家，“计量经济学已经在经济学科中居于最重要的地位”，“在大多数大学和学院中，计量经济学的讲授已经成为经济学课程表中最有权威的一部分”，甚至说，“第二次世界大战后的经济学是计量经济学的时代”。在这里，可以用诺贝尔经济学奖获得者作为例证。从1969年诺贝尔经济学奖设立时起，至2008年，共有62位经济学家获奖，覆盖了经济学的各分支学科。直接因为对计量经济学的创立和发展作出贡献而获奖者达10人，居经济学各分支学科之首。1969年第一届获奖者，并不是萨缪尔森、希克斯这样的经济学大家，而是创立计量经济学的弗里希和推广应用计量经济学，建立了第一个用于研究经济周期理论的计量经济学模型的丁伯根(J. Tinbergen)。1973年，列昂惕夫(W. Leotief)作为

投入产出分析的创始人而获奖，投入产出分析也属于广义的计量经济学。1980年获奖者克莱因是经典计量经济学理论与应用的集大成者。1984年获奖者斯通(R. Stone)是一位统计学家，他的贡献之一是“极大地改善了计量经济分析的数据基础”。1989年，哈维尔莫(T. Haavelmo)因为他于1943年发表的论文奠定了计量经济学的概率论基础而获奖。2000年诺贝尔经济学奖授予两位对微观计量经济学作出原创性贡献的经济学家赫克曼和麦克法登，前者的贡献是选择性样本计量经济学模型，后者的贡献是离散选择模型。2003年，恩格尔(R. F. Engle)和格兰杰因为在时间序列计量经济学领域的贡献而获奖。除此之外，绝大多数诺贝尔经济学奖获得者，即使主要贡献不在计量经济学领域，但在他们的研究中都普遍应用了计量经济学方法。索罗(R.M.Solow)因他的经济增长理论而获得1987年诺贝尔经济学奖，而他的理论贡献得益于用计量经济学方法建立的总量生产函数及导出的增长方程；莫迪利尼亚(F.Modigliani)由于在家庭储蓄和金融市场作用方面的首创性研究而获得1985年诺贝尔经济学学奖，他曾是数学教师，担任过计量经济学会会长，并在研究中广泛应用了计量经济学实证分析方法；1993年诺贝尔经济学奖得主福格尔(R.W.Fogel)和诺斯(R.C.North)，属于新制度经济学派，主要研究经济史，但其获奖原因却是“在经济史研究中的定量研究领域所作出的贡献”。这些足以说明计量经济学属于经济学。

第三，计量经济学与数理统计学是有严格区别的。数理统计学作为一门数学学科，它可以应用于经济领域，也可以应用于其他领域，如社会学和自然科学等。但它与经济理论、经济统计学结合而形成的计量经济学，则只限于经济领域。

第四，也是最重要的，从建立与应用计量经济学模型的全过程可以看出，理论模型的设定和样本数据的收集，必须以对经济理论和所研究的经济现象的透彻认识为基础。即使是涉及数学方法较多的模型参数估计、模型检验等，单靠数学知识也是难以完成的。

诚然，“计量经济学的根本任务是估计经济模型和检验经济模型”，计量经济学方法，“从狭义上看，模型参数估计方法是它的核心内容”，这些写在一些教科书前言中的话都是对的。但是，离开方法提出的经济背景、方法本身的经济学解释和方法应用的经济对象，计量经济学方法将是一堆无用的数学符号。

综上所述，结论是十分清楚的：计量经济学是一门经济学科，而不是应用数学或其他。

五、计量经济学在经济学科中的地位

一般认为，1969年诺贝尔经济学奖的设立，标志着经济学已成为一门科学。在经济学不断科学化的过程中，计量经济学起到了特殊的作用。

这里需要考察一下现代经济学，主要是现代西方经济学的特征。现代西方经济学有许多特征，可以从不同的角度去归纳。从方法论的角度讲，主要有以下三个方面。一是越来越多地从方法论的角度去阐述和定义经济学，认为“经济学是一种思考社会问题的方法”，“经济学的主要贡献是它的分析框架”，“经济学是一套用以观察无限丰富和多变的世界的工具”；认为经济学是其他社会科学的基础，类似于物理学在自然科学中的地位。二是愈来愈重视研究方法的科学性，重实证分析，轻规范分析，认为“规范的方法显然是不科学的”，“经济学，对于规范的问题只能保持沉默”，“科学知识的占有尚不具备解决规范问题的能力”，“如果将价值判断引入经济理论，这种理论就不可能成为客观的科学”。这些认识虽然过于偏激，甚至存在谬误，在我们看来，经济学不能完全排斥规范分析，不能完全否定价值判断，但这些反映了西方经济学把自己定义为一门实证的社会科学的事实。三是数学的广泛应用已成为一个普遍趋势。经济学作为一门科学，如果从亚当·斯密(Adam Smith)1776年的《国富论》算起，也只有200多年的时间。经济学研究的数学化和定量化是经济学迅速科学化的重要标志。当然，数学仅仅是一种工具，而不是经济学理论本身。但正是这种工具，推动了经济学理论的发展，微分学与边际理论、优化方法与最优配置理论以及数理统计学与经济学的实证化就是例证。翻开任何一本经济学教科书或任何一份经济学刊物，无不用数学语言阐述经济理论，用定量的方法描述和讨论人们关心的经济现实问题。许多世界一流大学的经济学院系在其教学计划的培养目标中，都对学生应用数学工具的能力提出明确要求。例如，“现代经济学理论的一个显著特点是数学的广泛应用，学生必须学会用数学工具描述和发展经济学理论”，“教学计划的目标之一是教会学生将数学作为经济分析的一个基本工具，去思考和描述经济问题和政策”。于是，计量经济学成为学生必须学习的核心课程，而且从初级、中级到高级一直都是。以上这些特征，决定了计量经济学在西方经济学中的重要地位。

经济学科是否与许多自然科学学科一样，存在“世界先进水平”？是，又不是。讲不是，是指经济学理论与经济政策。各国国情不同，经济制度与体制不同，所处的发展阶段不同，指导发展的经济理论和实施的经济政策当然不同。在此方面，不会也不应有“世界先进水平”。讲是，是指经济学研究方法和经济分析方法存在“世界先进水平”，而在这个方面，我们落后了，而且落后了许多。

毫无疑问，我国的经济学需要科学化和现代化，要真正成为一门科学，成为一门能够指导中国社会主义市场经济体制的建立和经济发展的科学，那么，重要内容之一就是学习现代西方经济学先进的研究分析方法。所以，学习、跟踪、研究、发展计量经济学，是一个重要任务。

§1.2 建立经典单方程计量经济学模型的步骤和要点

本节以经典单方程计量经济学模型为对象，介绍建立计量经济学模型的过程。这里的计量经济学模型，按照上节的界定，是指揭示经济现象中客观存在的因果关系，并主要采用回归分析方法的经济数学模型。在学习全书之前，首先对建模过程有一个整体的了解，会使学生在学习具体内容时更具有目的性和针对性。凡是后续内容中要详细介绍的部分，在本节中只作为一个步骤列出，以示它在整体中的位置，具体内容在后面会详细介绍；凡是后续内容中不再介绍的部分，在本节中进行较为详细的讲解。

作为本书的特色和重点内容之一，第九章将专门讨论建立计量经济学模型中的若干方法论问题，包括模型类型的选择、总体模型的设定、模型变量的选择等。为了避免重复，本节只介绍建立模型的步骤，以及每个步骤中比较浅显的要点。至于每个步骤的逻辑学、经济学和统计学内涵以及比较深入的要点，只有在学习全书之后才能真正理解它们，因此也将在第九章对此进行较为详细的讨论。

一、理论模型的设计

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，也称总体回归模型。例如，上节中的生产函数

$$Q = A\mathrm{e}^{\gamma t} K^{\alpha} L^{\beta}$$

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。**

理论模型的设计必须遵循“从一般到简单”的原则，即作为建模起点的总体模型必须能够包容所有经过约化得到的“简洁”的模型。具体讲，它应该包含所有对被解释变量产生影响的变量，尽管其中的某些变量会因为显著性不高或者不满足正交性条件等原因在后来的约化过程中被排除。关于这个原则，将在第九章进行讨论。

1．确定模型所包含的变量

在单方程计量经济学模型中，变量分为两类。作为研究对象的变量，也就是因果关系中的“果”，如生产函数中的产出量，是模型中的**被解释变量**；而作为“原因”的变量，如生产函数中的资本、劳动、技术，是模型中的**解释变量**。

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。

严格地说，上述生产函数中的产出量、资本、劳动、技术等，只能称为“因素”，这些因素间存在着因果关系。为了建立起计量经济学模型，必须选择适当的变量来表征这些因素，因而这些变量必须具有数据可得性。于是，可以用总产值来表征产出量，用固定资产原值来表征资本，用职工人数来表征劳动，用时间作为一个变量来表征技术，这样，最后建立的模型是关于总产值、固定资产原值、职工人数和时间变量之间关系的数学表达式。下面，为了叙述方便，我们将“因素”与“变量”间的区别暂时略去，都以“变量”来表示。

现在问题的关键在于，在确定了被解释变量之后，怎样才能正确地选择解释变量。

第一，需要正确理解和把握所研究的经济现象中暗含的经济学理论和经济行为规律。这是正确选择解释变量的基础。例如，在上述生产问题中，已经明确指出属于供给不足的情况，那么影响产出量的因素就应该在投入要素方面，而在当前，一般的投入要素主要是技术、资本与劳动。如果属于需求不足的情况，那么影响产出量的因素就应该在需求方面，而不在投入要素方面。这时，如果研究的对象是消费品生产，应该选择居民收入等变量作为解释变量；如果研究的对象是生产资料生产，应该选择固定资产投资总额等变量作为解释变量。由此可见，同样是建立生产模型，所处的经济环境不同，研究的行业不同，变量选择是不同的。

第二，选择变量要考虑数据的可得性。这就要求对经济统计学有透彻的了解。计量经济学模型要在样本数据，即变量的样本观测值的支持下，采用一定的数学方法估计参数，以揭示变量之间的定量关系。所以所选择的变量必须在统计指标体系中存在，并有可靠的数据来源。如果必须引入个别对被解释变量有重要影响的政策变量、条件变量，则应采用虚变量的样本观测值的选取方法。

第三，选择变量时要考虑所有入选变量之间的关系，使得每个解释变量都是独立的。这是计量经济学模型技术所要求的。当然，在开始时要做到这一点是困难的，如果在所有入选变量中出现相关的变量，可以在建模过程中检验并予以剔除。

从这里可以看出，**建立模型的第一步就已经体现了计量经济学是经济理论、统计学和数学三者结合的思想**。

在选择变量时容易发生错误。下面的例子都是从已有的计量经济学应用研究成果中发现的，代表了几类容易发生的错误。例如，

$$CZSR=4\,219.1+4.729\,0\cdot GPRZ$$

其中，CZSR 代表财政收入，GPRZ 代表股票融资额。这里遗漏了重要的变量。显然，影响财政收入的因素较多，最重要的因素是各项税收。股票融资额肯定不是重要的因素，更不是唯一因素。又如，

$$NFCK = -107.66 + 0.13 \cdot SSLS + 0.22 \cdot NFSG$$

其中，NFCK 代表农副产品出口额，SSLS 代表社会商品零售总额，NFSG 代表农副产品收购额。这里选择了无关的变量。因为社会商品零售总额与农副产品出口额无直接关系，更不是影响农副产品出口额的原因。再如，

$$SZJK = 0.73 \cdot QGTZ + 0.21 \cdot CK + 0.18 \cdot SCXF + 67.60 \cdot D$$

其中，SZJK 代表生产资料进口额，QGTZ 代表轻工业投资，CK 代表出口额，SCXF 代表生产消费，D 代表进出口政策。这里选择了不重要的变量。因为轻工业投资对生产资料进口额虽有影响，但不是重要的，或者说是不完全的，重要的是全社会固定资产投资额，应该选择这个变量。再看一个例子：

$$NYCZ = 0.78 + 0.24 \cdot LSCL + 0.05 \cdot NJDL - 0.21 \cdot SZMJ$$

其中，NYCZ 代表农业总产值，LSCL 代表粮食产量，NJDL 代表农机动力，SZMJ 代表受灾面积。这里选择了不独立的变量。因为粮食产量是受农机动力和受灾面积影响的，它们之间存在相关性。

值得注意的是上述几个模型都能很好地拟合样本数据，所以绝对不能把对样本数据的拟合程度作为判断模型变量选择是否正确的主要标准。

变量的选择不是一次完成的，往往需要经过多次反复进行。

2. 确定模型的数学形式

选择了适当的变量，接下来就要选择适当的数学形式描述这些变量之间的关系，即建立理论模型。

选择模型的数学形式的主要依据是经济行为理论。在数理经济学中，已经对常用的生产函数、需求函数、消费函数、投资函数等模型的数学形式进行了广泛的研究，可以借鉴这些研究成果。需要指出的是，现代经济学尤其注重实证研究，任何建立在一定经济学理论假设基础上的理论模型，如果不能很好地解释过去，尤其是历史统计数据，那么它就不能为人们所接受。这就要求理论模型的建立要在参数估计和模型检验的全过程中反复修改，以得到一种既能较好地解释经济行为又能较好地反映历史上已经发生的诸变量之间关系的数学模型。忽视任何一方面都是不对的。

也可以根据变量的样本数据作出解释变量与被解释变量之间关系的散点图，并将由散点图显示的变量之间的函数关系作为理论模型的数学形式。这也是人们在建模时经常采用的方法。

在某些情况下，如果无法事先确定模型的数学形式，那么就采用各种可能的形式进行试模拟，然后选择模拟结果较好的一种。

3．拟定理论模型中待估参数的理论期望值

理论模型中的待估参数一般都具有特定的经济含义，它们的数值要待模型估计和检验后，即经济数学模型完成后才能确定。但对于它们的数值范围，即理论期望值，可以根据它们的经济含义在开始时拟定。这一理论期望值可以用来检验模型的估计结果。

拟定理论模型中待估参数的理论期望值，关键在于理解待估参数的经济含义。例如，上述生产函数理论模型中有 4 个待估参数α，β，γ和A，其中，α是资本的产出弹性，β是劳动的产出弹性，γ近似为技术进步速度，A是效率系数。根据这些经济含义，它们的数值范围应该是

$$0<\alpha<1,\quad 0<\beta<1,\quad \alpha+\beta\approx 1$$

$$0<\gamma<1(\text{接近 }0),\quad A>0$$

二、样本数据的收集

样本数据的收集与整理，是建立计量经济学模型过程中最费时费力的工作，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的，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。

1．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。但是，一个重要的问题是：什么样的时间序列数据才能适合于经典计量经济学模型？它们必须是平稳的时间序列。如果是非平稳时间序列，它们之间必须存在经济上的均衡关系和统计上的协整关系。这个问题将在本书第八章专门讨论。利用时间序列数据作样本时，还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，建立我国纺织行业生产模型时，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 20 世纪 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进行估计的一个主要原因。四是模型随机干扰项的序列相关问题。用时间序列数据作样本，容易引起模型随机干扰项产生序列相关。这个问题后面还要专门讨论。

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。这些数据主要由统计部门提供。研究者也可以根据研究的需要，设计调查方案，进行实际调查，以获得截面数据。**经典计量经济学模型理论是基于随机抽样的截面数据而建立的，随机抽样是经典模型对截面数据的最重要和最基本的要求**。对于不满足要求的截面数据，必须发展专门的模型。本书第七章将专门介绍几类基于截面数据的非经典计量经济学模型。用截面数据作为计量经济学模型的样本数据，还应注意以下几个问题：一是样本与总体的一致性问题。计量经济学模型的参数估计，从数学上讲，是用从总体中随机抽取的个体样本估计总体的参数，这就要求总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。这样，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。二是模型随机干扰项的异方差问题。用截面数据作样本，容易使模型随机干扰项产生异方差。这个问题后面还要专门讨论。

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 1980 年以后的年份，该虚变量的样本观测值为 1，对于 1980 年以前的年份，该虚变量的样本观测值为 0。关于虚变量的设定，后面将专门讨论。

2. 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，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，样本点之间的联系紧密，那么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并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并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。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。为什么呢？其原因就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，产出量为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，有人直接用于模型的估计，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，即总体与样本的一致性。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，用人均收入与消费的数据作为总量消费函数模型的样本数据，用某些省份的数据作为全国总量模型的样本数据，等等。

三、模型参数的估计

模型参数的估计方法，是计量经济学的核心内容。在建立了理论模型并收集整理了符合模型要求的样本数据之后，就可以选择适当的方法估计模型，得到模型参数的估计量。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(对联立方程计量经济学模型而言)、估计方法的选择、软件的应用等内容。

在后面的章节中将详细讨论估计问题，在此不重复叙述。

四、模型的检验

在得到模型的参数估计量之后，可以说一个计量经济学模型就已经初步建立起来了。但是，它能否客观揭示所研究的经济现象中诸因素之间的关系，能否付诸应用，还要通过检验才能决定。一般讲，计量经济学模型必须通过四级检验，即经济意义检验、统计检验、计量经济学检验和模型预测检验。

1. 经济意义检验

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。其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。

首先检验参数估计量的符号。例如，有下列煤炭行业生产模型：

$$MCL = -108.5427 + 0.00067 \cdot GZZ + 0.01527 \cdot ZGS - 0.00681 \cdot DHL + 0.00256 \cdot MHL$$

其中，MCL 代表煤炭产量，GZZ 代表固定资产原值，ZGS 代表职工人数，DHL 代表电力消耗量，MHL 代表木材消耗量。在该模型中，电力消耗量前的参数估计量为负，意味着电力消耗越多，煤炭产量越低，从经济行为上无法解释该现象，所以此模型不能通过检验，应该找出原因重新建立模型。

如果所有参数估计量的符号都正确，则要进一步检验参数估计量的大小。例如，有下列煤炭企业生产函数模型：

$$\ln MCL = 2.69 + 1.85 \ln GZZ + 0.51 \ln ZGS$$

因为该模型是一个对数线性模型，所以在该模型中，固定资产原值前的参数的经济意义是明确的，即固定资产原值的产出弹性，表示当固定资产原值增加 1% 时煤炭产量增加的百分数。根据产出弹性的概念，该参数估计量应该是 0 与 1 之间的一个数，模型中的参数估计量虽然符号正确，但是数值范围与理论期望值不符，此模型不能通过检验，应该找出原因重新建立模型。

即使模型参数估计量的符号正确，数值范围适当，仍然不能说已经通过经济意义检验，还要对参数之间的关系进行检验。例如，有下列职工家庭日用品需求模型：

$$\ln GMZC = -3.69 + 1.20 \ln SR - 6.40 \ln JG$$

其中，被解释变量 GMZC 为人均购买日用品支出额，解释变量 SR 和 JG 分别表示人均收入和日用品类价格。该模型也是一个对数线性模型，所以在该模型中，人均收入和日用品类价格前的参数的经济意义是明确的，即它们各自的需求弹性。这两个参数估计量的符号是正确的，数值范围大体适当。但是根据经济意义，这两个参数估计量之和应该在 1 左右，因为当收入增长 1%，价格增长 1%时，人均购买日用品支出额也应该增长 1%左右。显然该模型的参数估计

量不能通过检验，应该找出原因重新建立模型。

只有当模型中的参数估计量通过所有经济意义的检验，方可进行下一步检验。模型参数估计量的经济意义检验是一项最基本的检验，经济意义不合理，不管其他方面的质量有多高，模型也是没有实际价值的。

2．统计检验

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。应用最广泛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。

3．计量经济学检验

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。最主要的检验准则有随机干扰项的序列相关性检验和异方差性检验，解释变量的多重共线性检验等。

4．模型预测检验

模型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，即所谓的模型的超样本特性。具体检验方法为：(1) 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；(2) 将所建立的模型用于样本以外某一时期的实际预测，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。

经历并通过了上述步骤的检验后，可以说已经建立了所需要的计量经济学模型，并可以将它应用于预定的目的。

五、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和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他经济学分支学科的主要特征；数据，即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，更广义讲就是信息，是计量经济学研究的原料**。这三方面缺一不可。

一般情况下，在计量经济学研究中，方法的研究是人们关注的重点，方法的水平往往成为衡量一项研究成果水平的主要依据，这是正常的。计量经济学理论方法的研究是计量经济学研究工作者义不容辞的义务。但是，不能因此而忽视对经济学理论的探讨，一个不懂得经济学理论，不了解经济行为的人，是无法从事计量经济学研究工作的，是不可能建立起一个哪怕极其简单的计量经济学模型的。所以，计量经济学家首先应该是一个经济学家。相比之下，人们

对数据，尤其是数据质量问题的重视更显不足。在申请一项研究项目或评审一项研究成果时，对数据的可得性、可用性、可靠性缺乏认真的推敲；在研究过程中出现问题时，较少从数据质量方面去找原因。而目前的实际情况是，数据已经成为制约计量经济学发展的重要问题。

六、计量经济学应用软件介绍

随着计量经济学理论与方法的发展，其数学过程也越来越复杂，从而推动了计算机应用软件的发展。反过来，也正是有了方便的应用软件，才使计量经济学有今天的繁荣。常用的计量经济学软件很多，虽然它们的侧重面不同，但都具有基本的计量经济学分析功能。

1. EViews

EViews (Econometric Views) 是目前世界上最流行的计量经济学软件之一。EViews 具有数据处理、作图、统计分析、建模分析、预测和模拟等功能，在建模分析方面，包括单方程的线性模型和非线性模型，联立方程计量经济学模型，时间序列分析模型，分布滞后模型，向量自回归模型，误差修正模型，离散选择模型等有多种估计方法。EViews 的操作简单、灵活，使用的命令接近自然语言，具有丰富的多层次的菜单提示，使用者不需要编写程序，只要根据需要逐层选择菜单中所列的项目即能完成分析工作。

2. SPSS/PC

SPSS/PC 的原意是统计分析软件包，是 20 世纪 70～80 年代国际上广泛流行的统计分析软件包之一。它提供了经典计量经济学分析的大部分功能，但它并不局限于计量经济学分析，而是面向一般的社会科学，如社会学、人口学、气象学等，即凡是有关的统计分析问题，均可以使用该软件包进行各种分析。它还特别适用于对截面资料或调查资料的数理统计分析。SPSS/PC 曾经于 20 世纪 80 年代初在我国的一些计算机上安装过，但在我国的应用并未得到推广。

3. SAS

SAS 的原意是统计分析系统，于 1976 年商品化以来，以其超凡的功能和可靠的技术支持著称于世。经过多年的完善与发展，它在国际上已经被誉为数据分析的标准软件，在各个领域得到广泛的应用。SAS 是集数据管理、数据分析和信息处理为一体的应用软件系统。它是一种集成软件，用户可以将各种模块适当组合以满足各自不同的需要。将其用于计量经济学分析，不仅能完成经典计量经济学模型的估计和检验，而且还可进行模型诊断，例如，检查数据中的异常点，指出模型中需要增加的变量等。SAS 已经在我国一些单位得到应用。

4. GAUSS

GAUSS 的原意是一种程序语言，是一种为矩阵运算而设计的计算机语言。

通常也把用这种语言编写的应用软件称为 GAUSS，这些软件都具有极强的矩阵运算功能。计量经济学分析应用广泛的矩阵运算，所以 GAUSS 为计量经济学分析与应用提供了强有力的技术支持。LSQ/GAUSS，即集中于基本计量经济学分析的 GAUSS 软件，它在使用方便和计算快捷方面较其他软件具有明显的优越性。对于非线性计量经济学模型的估计，GAUSS 更具有其他软件不可比拟的优点。GAUSS 已经在我国一些单位的教学与研究中得到应用。

5. PC-GIVE

PC-GIVE 是一个独具特色的计量经济学分析软件。它是根据 Hendry 学派的理论方法研制而成的，于 1984 年推出，主要用于动态计量经济学分析，包括经济数据的分析，计量经济学模型的评估，动态计量经济学模型的建立等主要功能。它所提供的多种综合统计检验量可以帮助用户选择模型最合适的动态形式。PC-GIVE 也是一个人机交互系统，在菜单提示下可进行方便的操作。PC-GIVE 也已经在我国一些单位的教学与研究中得到应用。

6. Stata

Stata 是一个用于分析和管理数据的功能强大且小巧玲珑的实用统计分析软件，由美国计算机资源中心研制。从 1985 到现在，已连续推出它的多个版本，通过不断更新和扩充，其内容日趋完善。它同时具有数据管理软件、统计分析软件、绘图软件、矩阵计算软件和程序语言的特点。在统计分析中，几乎具有所有计量经济学模型估计和检验的功能，特别在平行数据分析方面具有优势。Stata 也采用命令方式进行操作，使用上远比 SAS 简单。目前，Stata 在一般计量经济学应用模型研究中被广泛使用。

实际应用的计量经济学软件还有很多，以上列举的只是我们曾经使用过的几种，当然也是最为流行的几种。

建立与应用计量经济学模型必须掌握至少一种计量经济学软件。学习应用软件的最好方法是实践，也就是自己实际采用一种软件去建立模型。因为读者所使用的软件各不相同，所以在本书中并不对软件的应用作更多的介绍，尽管它是十分重要的。本书第一版中的所有例题都是采用 TSP 6.5 完成的，而在第二、三版中，所有例题都是采用 EViews 完成的。

§1.3 计量经济学模型的应用

经济系统中各部分之间，经济过程中各环节之间和经济活动中各因素之间，除了存在经济行为理论上的相互联系外，还存在数量上的相互依存关系。研究客观存在的这些数量关系，是经济研究的一项重要任务，是经济决策的一项基础性工作，是发展经济理论的一种重要手段。所以说，计量经济学是经济

数量分析的最重要的分支学科。

计量经济学模型的应用大体可以被概括为四个方面：结构分析，经济预测，政策评价，检验与发展经济理论。在本书后续章节中将结合具体计量经济学模型来解释每个方面的应用，这里，仅作一些概念性介绍，以期对后续课程的学习起到某些指导作用。

一、结构分析

经济学中的结构分析是对经济现象中变量之间相互关系的研究。它不同于人们通常所说的结构分析，如产业结构、产品结构、消费结构和投资结构中的结构分析。它研究的是当一个变量或几个变量发生变化时会对其他变量乃至经济系统产生什么样的影响。从这个意义上讲，我们所进行的经济系统的定量研究工作，说到底，就是结构分析。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。在经济研究中，除了需要研究经济系统中变量绝对量之间的关系，还要掌握变量的相对变化所带来的相互影响，以掌握经济活动的数量规律和有效地控制经济系统。计量经济学模型结构式揭示了变量之间的直接因果关系，从模型出发进一步揭示变量相对变化量之间的关系是十分方便的。

乘数，也是经济学中的一个重要概念，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。它直接度量经济系统中变量之间的相互影响，经常被用来研究外生变量的变化对内生变量的影响，对于实现经济系统的调控有重要作用。乘数可以从计量经济学模型的简化式很方便地求得。关于计量经济学模型的结构式和简化式的概念，将在第六章专门介绍，简单地说，结构式的解释变量中可以出现内生变量，而简化式的解释变量中全部为外生或滞后内生变量。

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。显然，弹性分析和乘数分析都是比较静力分析的形式。计量经济学模型为比较静力分析提供了一个基础，如果没有定量描述变量之间关系的计量经济学模型，比较静力分析将无从着手。

结构分析过去是，现在是，将来也仍然是计量经济学模型应用的一个主要方面。

二、经济预测

计量经济学模型作为一类经济数学模型，是从经济预测，特别是短期预测

发展起来的。在 20 世纪 50 年代与 60 年代，计量经济学模型在西方国家经济预测中不乏成功的实例，成为经济预测的一种主要模型方法。但是，进入 20 世纪 70 年代后，人们对计量经济学模型的预测功能提出了质疑，起因并不是它未能对发生于 1973 年和 1979 年的两次“石油危机”提出预报，而是几乎所有的模型都无法预测“石油危机”对经济造成的影响。对计量经济学模型的预测功能的批评是有道理的，或者说计量经济学模型的预测功能曾经被夸大了。应该看到，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的。于是，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，计量经济学模型显得无能为力。同时，还应该看到，20 世纪 40～60 年代甚至后来建立的计量经济学模型都是以凯恩斯理论为经济理论基础的，而经济理论本身已经有了很大的发展，滞后于经济现实与经济理论的模型在应用中当然要遇到障碍。

为了适应经济预测的需要，计量经济学模型技术也在不断发展。将计量经济学模型与其他经济数学模型相结合，是一个重要的发展方向。

三、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。从宏观经济领域到微观经济领域，每时每刻都存在政策评价的问题。经济政策具有不可试验性。当然，有时在采取某项政策前，可在局部范围内先进行试验，然后推行，但即使如此，在局部可行的在全局上并不一定可行，这就使得政策评价显得尤其重要。

经济数学模型可以起到“经济政策实验室”的作用。尤其是计量经济学模型，揭示了经济系统中变量之间的相互联系，将经济目标作为被解释变量，经济政策作为解释变量，可以很方便地评价各种不同政策对目标的影响。将计量经济学模型和计算机技术结合起来，可以建成名副其实的“经济政策实验室”。

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。

四、检验与发展经济理论

实践的观点是唯物辩证法的首要和基本的观点，实践是检验真理的唯一标准。任何经济学理论，只有当它成功地解释了过去，才能为人们所接受。计量经济学模型提供了一种检验经济理论很好的方法。从建立计量经济学模型的步

骤中不难发现，一个成功的模型，必须很好地拟合样本数据，而样本数据则是已经发生的经济活动的客观再现，所以在模型中表现出来的经济活动的数量关系，则是经济活动所遵循的经济规律，即理论的客观再现。于是，就提出了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，如果拟合得很好，这种经济理论就得到了检验，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，就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。

本章练习题

1．什么是计量经济学？计量经济学方法与一般经济数学方法有什么区别？

2．计量经济学的研究对象和内容是什么？计量经济学模型研究的经济关系有哪两个基本特征？

3．为什么说计量经济学在当代经济学科中占据重要地位？当代计量经济学发展的基本特征与动向是什么？

4．建立与应用计量经济学模型的主要步骤有哪些？

5．计量经济学模型主要有哪些应用领域？各自的原理是什么？

6．模型的检验包括几个方面？其具体含义是什么？

7．下列假想模型是否属于揭示因果关系的计量经济学模型？为什么？

(1) $S_t = 112.0 + 0.12R_t$，其中S_t为第t年农村居民储蓄增加额(单位：亿元)，R_t为第t年城镇居民可支配收入总额(单位：亿元)。

(2) $S_{t-1} = 4\ 432.0 + 0.30R_t$，其中$S_{t-1}$为第$t-1$年底农村居民储蓄余额(单位：亿元)，$R_t$为第$t$年农村居民纯收入总额(单位：亿元)。

8．指出下列假想模型中的错误，并说明理由：

$$\mathrm{RS}_t = 8\ 300.0 - 0.24 \cdot \mathrm{RI}_t + 1.12 \cdot \mathrm{IV}_t$$

其中，RS_t为第t年社会消费品零售总额(单位：亿元)，RI_t为第t年居民收入总额(单位：亿元)(城镇居民可支配收入总额与农村居民纯收入总额之和)，IV_t为第t年全社会固定资产投资总额(单位：亿元)。

PDG

第二章　经典单方程计量经济学模型：一元线性回归模型

2

单方程计量经济学模型是相对于联立方程计量经济学模型而言的，它以单一经济现象为研究对象，模型中只包括一个方程，是应用最为普遍的计量经济学模型。经典单方程计量经济学模型的理论与方法，不仅是计量经济学内容体系中最重要的组成部分，也是联立方程计量经济学模型理论与方法的基础。本章首先从简单的一元线性回归模型入手，介绍经典单方程计量经济学模型的设定与估计问题，为以后各章的学习打下基础。

§2.1　回归分析概述

一、回归分析基本概念

1. 变量间的相互关系

无论是自然现象之间还是社会经济现象之间，大都存在着不同程度的联系。计量经济学的主要问题之一就是要探寻各种经济变量之间的相互联系程度、联系方式及其运动规律。各种经济变量间的关系可分为两类：一类是确定的函数关系，另一类是不确定的统计相关关系。

确定性现象间的关系常常表现为函数关系。例如，圆面积 S 与圆半径 r 间的关系，只要给定半径值 r，与之对应的圆面积 S 也就随之确定：$S=\pi r^2$。

非确定性现象间的关系常常表现为统计相关关系。例如，农作物产量 Y 与施肥量 X 间的关系，其特点是：农作物产量 Y 随着施肥量 X 的变化呈现某种规律性的变化，在适当的范围内，随着 X 的增加，Y 也增加。但与前述函数关系不同的是，给定施肥量 X，与之对应的农作物产量 Y 并不能确定。主要原因在于，除了施肥量，还有诸如阳光、气温、降雨等其他许多因素都在影响着农作物的产量。这时，我们无法确定农作物产量与施肥量间确定的函数关系，但却能通过统计计量等方法研究它们间的统计相关关系。农作物产量 Y 作为非确定性变量，也被称为随机变量。

当然，变量间的函数关系与相关关系并不是绝对的，在一定条件下两者可相互转化。例如，在对确定性现象的观测中，往往存在测量误差，这时函数关系常会通过相关关系表现出来；反之，如果对非确定性现象的影响因素能够一

一辨认出来，并全部纳入到变量间的依存关系式中，则变量间的相关关系就会向函数关系转化。相关分析与回归分析主要研究非确定性现象间的统计相关关系。

2. 相关分析与回归分析

变量间的统计相关关系可以通过相关分析与回归分析来研究。相关分析(correlation analysis)主要研究随机变量间的相关形式及相关程度。

从变量间相关的表现形式看，有线性相关与非线性相关之分，前者往往表现为变量的散点图接近于一条直线。变量间线性相关程度的大小可通过相关系数来测量，两个变量 X 和 Y 的总体相关系数为

$$\rho_{XY}=\frac{\operatorname{Cov}(X,\ Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}} \tag{2.1.1}$$

其中，$\operatorname{Cov}(X, Y)$是变量 X 和 Y 的协方差，$\operatorname{Var}(X)$和 $\operatorname{Var}(Y)$分别是变量 X 和 Y 的方差。

如果给出 X 与 Y 的一组样本$(X_i,\ Y_i)$，$i=1,\ 2,\ \cdots,\ n$，则样本相关系数为

$$r_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\cdot\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \tag{2.1.2}$$

其中，$\bar{X}$ 与 $\bar{Y}$ 分别是变量 X 与 Y 的样本均值。

多个变量间的线性相关程度，可用复相关系数与偏相关系数来度量。

具有相关关系的变量间有时存在着因果关系，这时，我们可以通过回归分析(regression analysis)来研究它们间的具体依存关系。例如，根据经济学理论，消费支出与可支配收入之间不但密切相关，而且有着因果关系，即可支配收入的变化往往是消费支出变化的原因。这时，不仅可以通过相关分析研究两者间的相关程度，而且可以通过回归分析研究两者间的具体依存关系，即考察可支配收入每 1 元的变化所引起的消费支出的平均变化。

回归分析是研究一个变量关于另一个(些)变量的依赖关系的计算方法和理论。其目的在于通过后者的已知或设定值，去估计和(或)预测前者的(总体)均值。前一个变量称为被解释变量(explained variable)或因变量(dependent variable)，后一个变量称为解释变量(explanatory variable)或自变量(independent variable)。

相关分析与回归分析既有联系又有区别。首先，两者都是研究非确定性变量间的统计依赖关系，并能度量线性依赖程度的大小。其次，两者间又有明显的区别。相关分析仅仅是从统计数据上测度变量间的相关程度，而无须考察两者间是否有因果关系，因此，变量的地位在相关分析中是对称的，而且都是随

机变量；回归分析则更关注具有统计相关关系的变量间的因果关系分析，变量的地位是不对称的，有解释变量与被解释变量之分，而且解释变量也往往被假设为非随机变量。再次，相关分析只关注变量间的联系程度，不关注具体的依赖关系；而回归分析则更加关注变量间的具体依赖关系，因此可以进一步通过解释变量的变化来估计或预测被解释变量的变化，达到深入分析变量间依存关系，掌握其运动规律的目的。

回归分析构成计量经济学的方法论基础，其主要内容包括：

(1) 根据样本观察值对计量经济学模型参数进行估计，求得回归方程；

(2) 对回归方程、参数估计值进行显著性检验；

(3) 利用回归方程进行分析、评价及预测。

二、总体回归函数

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

例 2.1.1

一个假想的社区是由 99 户家庭组成的总体，研究该社区每月家庭消费支出 Y 与每月家庭可支配收入 X 的关系，即根据家庭的每月可支配收入，考察该社区家庭每月消费支出的平均水平。为研究方便，将该 99 户家庭组成的总体按可支配收入水平划分为 10 组，并分别分析每一组的家庭消费支出(表 2.1.1)。

表 2.1.1　某社区家庭每月可支配收入与消费支出统计表　　单位：元

每月家庭可支配收入 X	800	1 100	1 400	1 700	2 000	2 300	2 600	2 900	3 200	3 500
每月家庭消费支出 Y	561	638	869	1 023	1 254	1 408	1 650	1 969	2 090	2 299
	594	748	913	1 100	1 309	1 452	1 738	1 991	2 134	2 321
	627	814	924	1 144	1 364	1 551	1 749	2 046	2 178	2 530
	638	847	979	1 155	1 397	1 595	1 804	2 068	2 266	2 629
		935	1 012	1 210	1 408	1 650	1 848	2 101	2 354	2 860
		968	1 045	1 243	1 474	1 672	1 881	2 189	2 486	2 871
			1 078	1 254	1 496	1 683	1 925	2 233	2 552	
			1 122	1 298	1 496	1 716	1 969	2 244	2 585	
			1 155	1 331	1 562	1 749	2 013	2 299	2 640	

续表

每月家庭消费支出 Y			1 188	1 364	1 573	1 771	2 035	2 310		
			1 210	1 408	1 606	1 804	2 101			
				1 430	1 650	1 870	2 112			
				1 485	1 716	1 947	2 200			
						2 002				
共计	2 420	4 950	11 495	16 445	19 305	23 870	25 025	21 450	21 285	15 510

由于不确定因素的影响，对同一可支配收入水平 X，不同家庭的消费支出不完全相同，但由于调查的完备性，给定可支配收入水平 X 的消费支出 Y 的分布是确定的，即以 X 的给定值为条件的 Y 的**条件分布**(conditional distribution)是已知的，如 $P(Y=561|X=800)=1/4$。因此，给定收入 X 的值，可得消费支出 Y 的**条件均值**(conditional mean)或**条件期望**(conditional expectation)，如 $E(Y|X=800)=605$。表 2.1.2 给出了 10 组可支配收入水平下相应家庭消费支出的条件概率，以及各可支配收入水平组家庭消费支出的条件均值。

表 2.1.2 各可支配收入水平组相应家庭消费支出的条件概率与条件均值

单位：元

收入水平	800	1 100	1 400	1 700	2 000	2 300	2 600	2 900	3 200	3 500
条件概率	1/4	1/6	1/11	1/13	1/13	1/14	1/13	1/10	1/9	1/6
条件均值	605	825	1 045	1 265	1 485	1 705	1 925	2 145	2 365	2 585

以表 2.1.1 中的数据绘出可支配收入 X 与家庭消费支出 Y 的散点图(图 2.1.1)。从该散点图可以看出，虽然不同的家庭消费支出存在差异，但平均来说，随着可支配收入的增加，家庭消费支出也在增加。进一步，这个例子中 Y 的条件均值恰好落在一根正斜率的直线上，这条直线称为**总体回归线**。

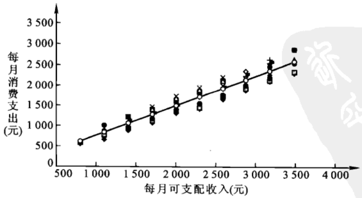


图 2.1.1 不同可支配收入水平组家庭消费支出的条件分布图

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

$$\mathrm{E}(Y \mid X)=f(X) \tag{2.1.3}$$

称为(双变量)**总体回归函数(Population Regression Function, PRF)**。

总体回归函数表明被解释变量 Y 的平均状态(总体条件期望)随解释变量 X 变化的规律。至于具体的函数形式，是由所考察总体固有的特征来决定的。由于实践中总体往往无法全部考察到，因此总体回归函数形式的选择就是一个经验方面的问题，这时经济学等相关学科的理论就显得很重要。例如，生产函数常以 Cobb–Douglas 幂函数的形式出现，U 形边际成本函数以二次多项式的形式出现，等等。将居民消费支出看成是其可支配收入的线性函数时，(2.1.3)式可进一步写成

$$\mathrm{E}(Y \mid X)=\beta_0+\beta_1 X \tag{2.1.4}$$

其中，β_0，β_1 是未知参数，称为**回归系数**(regression coefficients)。(2.1.4)式也称为**线性总体回归函数**。线性函数形式最为简单，其中参数的估计与检验也相对容易，而且多数非线性函数可转换为线性形式，因此，为了研究的方便，计量经济学中总体回归函数常设定成线性形式。

需注意的是，经典计量经济方法中所涉及的线性函数，指回归系数是线性的，即回归系数只以它的一次方出现，对解释变量则可以不是线性的。

三、随机干扰项

在上述家庭可支配收入–消费支出的例子中，总体回归函数描述了所考察总体的家庭消费支出平均说来随可支配收入变化的规律，但对某一个个别家庭，其消费支出 Y 不一定恰好就是给定可支配收入水平 X 下的消费的平均值 $\mathrm{E}(Y \mid X)$。图 2.1.1 显示，个别家庭消费支出 Y 聚集在给定可支配收入水平 X 下所有家庭平均消费支出 $\mathrm{E}(Y \mid X)$ 的周围。

对每个个别家庭，记

$$\mu=Y-\mathrm{E}(Y \mid X) \tag{2.1.5}$$

称 μ 为观察值 Y 围绕它的期望值 $\mathrm{E}(Y \mid X)$ 的**离差**(deviation)，它是一个不可观测的随机变量，称为**随机误差项**(stochastic error)，通常又不加区别地称为随机干扰项（stochastic disturbance）。

由(2.1.5)式，个别家庭的消费支出为

$$Y=\mathrm{E}(Y \mid X)+\mu \tag{2.1.6}$$

或者在线性假设下

$$Y=\beta_0+\beta_1 X+\mu \tag{2.1.7}$$

即给定可支配收入水平 X，个别家庭的消费支出可表示为两部分之和：(1)该收入水平下所有家庭的平均消费支出 $\mathrm{E}(Y \mid X)$，称为**系统性**(systematic)**部分**或**确定性**(deterministic)**部分**；(2)其他**随机部分**或**非系统性**(nonsystematic)**部分** μ。

(2.1.6)式或(2.1.7)式称为**总体回归函数的随机设定形式**，它表明被解释变量 Y 除了受解释变量 X 的系统性影响外，还受其他未包括在模型中的诸多因素的随机性影响，μ 即为这些影响因素的综合代表。由于方程中引入了随机干扰项，成为计量经济学模型，因此也称为**总体回归模型**(population regression model)。

在总体回归函数中引入随机干扰项，主要有以下几方面的原因。

(1) 代表未知的影响因素。由于对所考察总体认识上的非完备性，许多未知的影响因素还无法引入模型，因此，只能用随机干扰项代表这些未知的影响因素。

(2) 代表残缺数据。即使所有的影响变量都能被包括在模型中，也会有某些变量的数据无法取得。例如，经济理论指出，居民消费支出除受可支配收入的影响外，还受财富拥有量的影响，但后者在实践中往往是无法收集到的。这时，模型中不得不省略这一变量，而将其归入随机干扰项。

(3) 代表众多细小影响因素。有一些影响因素已经被认识，而且其数据也可以收集到，但它们对被解释变量的影响却是细小的。考虑到模型的简洁性，以及取得诸多变量数据可能带来的较大成本，建模时往往省掉这些细小变量，而将它们的影响综合到随机干扰项中。

(4) 代表数据观测误差。由于某些主客观的原因，在取得观测数据时，往往存在测量误差，这些观测误差也被归入随机干扰项。

(5) 代表模型设定误差。由于经济现象的复杂性，模型的真实函数形式往往是未知的，因此，实际设定的模型可能与真实的模型有偏差。随机干扰项包含了这种模型设定误差。

(6) 变量的内在随机性。即使模型没有设定误差，也不存在数据观测误差，由于某些变量所固有的内在随机性，也会对被解释变量产生随机性影响。这种影响只能被归入到随机干扰项中。

总之，随机干扰项具有非常丰富的内容，在计量经济学模型的建立中起着重要的作用。如果进一步分析，可以发现，当随机干扰项仅包含上述(3)和(6)时，称之为“原生”的随机干扰，是模型所固有的；当随机干扰项包含上述(1)、(2)、(4)、(5)时，称之为“衍生”的随机误差，是在模型设定过程中产生的，是可以避免的。尽管本书对此不加区别，但认识这一点是重要的。

四、样本回归函数

尽管总体回归函数揭示了所考察总体被解释变量与解释变量间的平均变化规律，但总体的信息往往无法全部获得，因此，总体回归函数实际上是未知

的。现实的情况往往是，通过抽样，得到总体的样本，再通过样本的信息来估计总体回归函数。

仍以例 2.1.1 中社区家庭可支配收入与消费支出的关系为例，假设从该总体中按每组可支配收入水平各取一个家庭进行观测，得到表 2.1.3 所示的一个样本。问题归结为：能否从该样本中预测整个总体对应于选定 X 的平均每月消费支出，即能否从该样本估计总体回归函数？

表 2.1.3　家庭消费支出与可支配收入的一个随机样本　　单位：元

X	800	1 100	1 400	1 700	2 000	2 300	2 600	2 900	3 200	3 500
Y	638	935	1 155	1 254	1 408	1 650	1 925	2 068	2 266	2 530

该样本的散点图如图 2.1.2 所示，可以看出，该样本散点图近似于一条直线。画一条直线尽可能地拟合该散点图。由于样本取自总体，可用该直线近似地代表总体回归线。该直线称为**样本回归线**(sample regression line)，其函数形式记为

$$\hat{Y} = f(X) = \hat{\beta}_0 + \hat{\beta}_1 X \tag{2.1.8}$$

称之为**样本回归函数**(Sample Regression Function，SRF)。

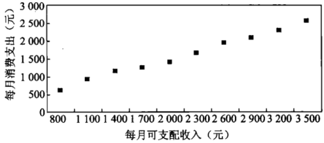


图 2.1.2　家庭可支配收入与消费支出的样本散点图

将(2.1.8)式看成(2.1.7)式的近似替代，则 $\hat{Y}$ 就为 $\mathrm{E}(Y \mid X)$ 的估计量，$\hat{\beta}_0$ 为 β_0 的估计量，$\hat{\beta}_1$ 为 β_1 的估计量。

同样地，样本回归函数也有如下的随机形式：

$$Y = \hat{Y} + \hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 X + e \tag{2.1.9}$$

其中，e 称为**(样本)残差(或剩余)项**(residual)，代表了其他影响 Y 的随机因素的集合，可看成是 μ 的估计量 $\hat{\mu}$。由于方程中引入了随机项，成为计量经济学模型，因此也称之为**样本回归模型**(sample regression model)。

回归分析的主要目的，就是根据样本回归函数，估计总体回归函数。也就

是根据

$$Y = \hat{Y} + e = \hat{\beta}_0 + \hat{\beta}_1 X + e$$

估计

$$Y = \mathrm{E}(Y \mid X) + \mu = \beta_0 + \beta_1 X + \mu$$

即设计一种“方法”构造 SRF，以使 SRF 尽可能“接近”PRF，或者说使 $\hat{\beta}_j (j = 0,1)$ 尽可能接近 $\beta_j (j = 0,1)$。图 2.1.3 绘出了总体回归线与样本回归线的基本关系。

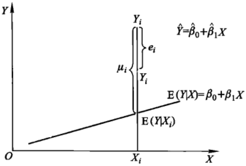


图 2.1.3　总体回归线与样本回归线的基本关系

§2.2　一元线性回归模型的基本假设

单方程计量经济学模型分为线性模型和非线性模型两大类。在线性模型中，变量之间的关系呈线性关系；在非线性模型中，变量之间的关系呈非线性关系。线性回归模型是线性模型的一种，它的数学基础是回归分析，即用回归分析方法建立的线性模型，用以揭示经济现象中的因果关系。

一元线性回归模型是最简单的计量经济学模型，在模型中只有一个解释变量，其一般形式是

$$Y = \beta_0 + \beta_1 X + \mu \tag{2.2.1}$$

其中，Y 为被解释变量，X 为解释变量，β_0 与 β_1 **为待估参数**，μ **为随机干扰项**。在有 n 个样本观测点 $\{(X_i, Y_i): i = 1,2,\cdots,n\}$ 的情况下，(2.2.1)式也可写成如下形式：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i,\ \ i = 1,2,\cdots,n \tag{2.2.2}$$

回归分析的主要目的是要通过样本回归函数(模型)尽可能准确地估计总体回归函数(模型)。**为保证参数估计量具有良好的性质，通常对模型提出若干基本假设。**

对模型(2.2.1)或(2.2.2)，基本假设包括对模型设定的假设、对解释变量 X 的假设、以及对随机干扰项 μ 的假设。

一、对模型设定的假设

假设 1：回归模型是正确设定的。

计量经济模型是对所关注经济现象或经济理论进行经验研究的基本工具，因此刻画经济现象或描述经济理论的计量模型的正确设定最为重要。模型的正确设定主要包括两方面的内容：(1)模型选择了正确的变量；(2)模型选择了正确的函数形式。

模型选择了正确的变量指在设定总体回归函数时，既没有遗漏重要的相关变量，也没有多选无关的变量。模型选择了正确的函数形式是指当被解释量与解释变量间呈现某种函数形式时，我们所设定的总体回归方程恰为该函数形式。例如在生产函数的设定中，如果产出量与资本投入及劳动投入的关系呈现幂函数的形式，我们在总体回归模型的设定中就设定了该幂函数的形式。

当假设 1 满足时，称为模型没有**设定偏误**(specification error)，否则就会出现模型的设定偏误。第五章将详细讨论模型的设定偏误问题。

二、对解释变量的假设

假设 2：解释变量 X 是确定性变量，不是随机变量，在重复抽样中取固定值。

在实验或可控条件下获得数据的情况中，X 的非随机性能够得到满足。但在有些情形中，X 往往也是随机的。这里将解释变量认定为非随机的，可以简化对后续参数估计性质的讨论。

假设 3：解释变量 X 在所抽取的样本中具有变异性，而且随着样本容量的无限增加，解释变量 X 的样本方差趋于一个非零的有限常数，即

$$\sum_{i=1}^{n}(X_i-\bar{X})^2/n \to Q,\ n\to\infty \tag{2.2.3}$$

在以因果关系为基础的回归分析中，往往就是通过解释变量 X 的变化来解释被解释变量 Y 的变化的，因此，解释变量 X 要有足够的变异性。而对其样本方差的极限为非零的有限常数的假设，则旨在排除时间序列数据出现持续上升或下降的变量作为解释变量，因为这类数据不仅使大样本统计推断变得无效，而且往往产生所谓的**伪回归问题**(spurious regression problem)。关于伪回归的确切含义将在第九章中讨论。

三、对随机干扰项的假设

假设 4：随机误差项μ具有给定X条件下的零均值、同方差以及不序列相关性，即

$$E(\mu_i \mid X_i)=0 \tag{2.2.4}$$

$$\text{Var}(\mu_i \mid X_i)=\sigma^2 \tag{2.2.5}$$

$$\text{Cov}(\mu_i,\mu_j \mid X_i,X_j)=0,\ i\neq j \tag{2.2.6}$$

随机误差项μ的条件零均值假设意味着μ的期望不依赖于X的变化而变化，且总为常数零。该假设表明μ与X不存在任何形式的相关性，因此该假设成立时也往往称X为**外生解释变量**(exogenous explanatory variable)，否则称X为**内生解释变量**(endogenous explanatory variable)。该假设最为重要，只有该假设成立时，总体回归函数的随机形式(2.1.7)式才能等价于非随机形式(2.1.4)式。

随机误差项μ的条件同方差假设意味着μ的方差不依赖于X的变化而变化，且总为常数σ^2。在μ的条件零均值与条件同方差假设下，总体回归函数可显示为图 2.2.1。

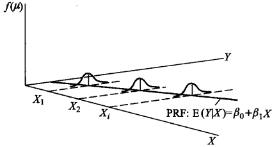


图 2.2.1　μ的条件零均值与条件同方差假设下的总体回归函数

需要注意的是，当随机误差项μ的条件零均值假设成立时，根据**期望迭代法则**(law of iterated expectation)一定有如下非条件零均值性质：

$$E(\mu_i)=0 \tag{2.2.7}$$

同样地，随机误差项μ的条件同方差假设成立时，根据期望迭代法则一定有如下非条件同方差性质：

$$\text{Var}(\mu_i)=\sigma^2 \tag{2.2.8}$$

另外，在随机误差项零均值的假设下，同方差还可写成如下的表达式：

$$\begin{aligned}\text{Var}(\mu_i \mid X_i)&=E(\mu_i^2 \mid X_i)-[E(\mu_i \mid X_i)]^2\\&=E(\mu_i^2 \mid X_i)=\sigma^2\end{aligned} \tag{2.2.9}$$

或
$$\mathrm{Var}(\mu_i)=\mathrm{E}(\mu_i^2)-[\mathrm{E}(\mu_i)]^2=\mathrm{E}(\mu_i^2)=\sigma^2 \tag{2.2.10}$$

随机误差项 μ 的条件不序列相关性表明在给定解释变量任意两个不同的值时，对应的随机误差项不相关。同样地，(2.2.6)式可等价地表示为
$$\mathrm{Cov}(\mu_i,\mu_j\mid X_i,X_j)=\mathrm{E}[(\mu_i\mid X_i)(\mu_j\mid X_j)]=0 \tag{2.2.11}$$

假设 5：随机误差项与解释变量之间不相关，即
$$\mathrm{Cov}(X_i,\mu_i)=0 \tag{2.2.12}$$

当随机误差项 μ 的条件零均值假设成立时，该假设一定成立，因为从 μ 的条件零均值假设(2.2.4)式可得到 μ 的非条件零均值特性(2.2.7)式。从而有
$$\mathrm{Cov}(X_i,\mu_i)=\mathrm{E}(X_i\mu_i)-\mathrm{E}(X_i)\mathrm{E}(\mu_i)=\mathrm{E}(X_i\mu_i)=0$$
其中最后一个等式仍可通过期望迭代法则推出。因此，假设 5 并非是必需的。但由于这一特征在回归分析中十分重要，尤其是在模型参数的估计中扮演着重要的角色，因此这里仍将其独立列出。

假设 6：随机误差项服从零均值、同方差的正态分布，即
$$\mu_i\mid X_i\sim N(0,\sigma^2) \tag{2.2.13}$$

假设 6 是为通过样本回归函数推断总体回归函数的需要而提出的，尤其是在小样本下，该假设显得十分重要。在大样本的情况下，正态性假设可以放松，因为根据中心极限定理，当样本容量趋于无穷大时，在大多数情况下，随机误差项的分布会越来越接近正态分布。

以上假设也称为线性回归模型的**经典假设**(calssical assumption)，满足该假设的线性回归模型，也称为**经典线性回归模型**(Classical Linear Regression Model, CLRM)。而前 4 个假设也专门称为**高斯-马尔可夫假设**(Gauss-Markov assumption)，这些假设能够保证下节介绍的估计方法具有良好的效果。

最后需要指出，在上述经典假设下，线性回归模型(2.2.1)中被解释变量 Y 具有如下条件分布特征：
$$Y\mid X\sim N(\beta_0+\beta_1X,\ \sigma^2) \tag{2.2.14}$$

图 2.2.2 描绘出了总体回归线与 Y 的条件分布状况。

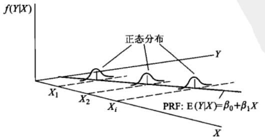


图 2.2.2 总体回归函数与 Y 的条件分布图

在实际建立模型的过程中，除了随机误差项的正态性假设外，对模型是否满足其他假设都要进行检验。这就是“建立计量经济学模型步骤”中“计量经济学检验”的任务。对于随机误差项的正态性假设，根据中心极限定理，如果仅包括源生性的随机干扰，当样本容量趋于无穷大时，都是满足的。如果包括衍生的随机误差，即使样本容量趋于无穷大，正态性假设也经常是不满足的。但是在初、中级教材中，一般将它忽略。

§2.3 一元线性回归模型的参数估计

一元线性回归模型的参数估计，是在一组样本观测值$\{(X_i,Y_i): i=1,2,\cdots,n\}$下，通过一定的参数估计方法，估计出样本回归线。常见的估计方法有三种：普通最小二乘法(OLS)、最大似然法(ML)与矩估计法(MM)。

一、参数估计的普通最小二乘法(OLS)

已知一组样本观测值$\{(X_i,Y_i): i=1,2,\cdots,n\}$，普通最小二乘法(Ordinary Least Squares, OLS)要求样本回归函数尽可能好地拟合这组值，即样本回归线上的点$\hat{Y}_i$与真实观测点Y_i的“总体误差”尽可能地小。普通最小二乘法给出的判断标准是：被解释变量的估计值与实际观测值之差的平方和

$$Q=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-(\hat{\beta}_0+\hat{\beta}_1X_i))^2 \tag{2.3.1}$$

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

为什么用平方和？因为样本回归线上的点$\hat{Y}_i$与真实观测点Y_i之差可正可负，简单求和可能将很大的误差抵消掉，只有平方和才能反映二者在总体上的接近程度，这就是**最小二乘原理**。

根据微积分学的运算，当Q对$\hat{\beta}_0$，$\hat{\beta}_1$的一阶偏导数为0时，Q达到最小，即

$$\begin{cases}\dfrac{\partial Q}{\partial \hat{\beta}_0}=0\\[2ex]\dfrac{\partial Q}{\partial \hat{\beta}_1}=0\end{cases}$$

可推得用于估计$\hat{\beta}_0$，$\hat{\beta}_1$的下列方程组：

$$\begin{cases}\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)=0\\\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)X_i=0\end{cases} \tag{2.3.2}$$

或

$$\begin{cases}\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2\end{cases} \tag{2.3.3}$$

解得

$$\begin{cases}\hat{\beta}_0 = \dfrac{\sum X_i^2 \sum Y_i - \sum X_i \sum Y_i X_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} \\ \hat{\beta}_1 = \dfrac{n\sum Y_i X_i - \sum Y_i \sum X_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}\end{cases} \tag{2.3.4}$$

方程组(2.3.2)或(2.3.3)称为**正规方程组**(normal equations)。记

$$\begin{aligned}\sum x_i^2 &= \sum (X_i - \bar{X})^2 \\ &= \sum X_i^2 - \frac{1}{n}\left(\sum X_i\right)^2 \\ \sum x_i y_i &= \sum (X_i - \bar{X})(Y_i - \bar{Y}) \\ &= \sum X_i Y_i - \frac{1}{n}\sum X_i \sum Y_i\end{aligned}$$

方程组(2.3.4)的参数估计量可以写成

$$\begin{cases}\hat{\beta}_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}\end{cases} \tag{2.3.5}$$

称为普通最小二乘法估计量的**离差形式**(deviation form)。在计量经济学中，往往以小写字母表示对均值的离差。由于 $\hat{\beta}_0$，$\hat{\beta}_1$ 的估计结果是从最小二乘原理得到的，故称为**普通最小二乘估计量**(ordinary least squares estimator)。

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

$$\begin{aligned}\hat{y}_i &= (\hat{\beta}_0 + \hat{\beta}_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{X} + \bar{e}) \\ &= \hat{\beta}_1 (X_i - \bar{X}) - \frac{1}{n}\sum e_i\end{aligned}$$

可得

$$\hat{y}_i = \hat{\beta}_1 x_i \tag{2.3.6}$$

其中，用到了正规方程组的第一个方程

$$\sum e_i = \sum [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)] = 0$$

(2.3.6)式也称为样本回归函数的**离差形式**。

最后，需要交代一个重要的概念，即“估计量”(estimator)和“估计值”

PDG

(estimate)的区别。由(2.3.4)式或(2.3.5)式给出的参数估计结果是由一个具体样本资料计算出来的，它是一个“估计值”，或者“点估计”，是参数估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的一个具体数值；但从另一个角度，仅仅把(2.3.4)式或(2.3.5)式看成 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的一个表达式，那么，它就成为 Y_i 的函数，而 Y_i 是随机变量，所以 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 也是随机变量，因而从这个角度考虑，就称之为“估计量”。在本章后续内容中，有时把 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 作为随机变量，有时又把 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 作为确定的数值，道理就在于此。

二、参数估计的最大似然法(ML)

最大似然法(Maximum Likelihood, ML)，也称**最大或然法**，是不同于普通最小二乘法的另一种参数估计方法，是从最大似然原理出发发展起来的其他估计方法的基础。虽然其应用没有普通最小二乘法普遍，但在计量经济学理论上占据很重要的地位，因为最大似然原理比最小二乘原理更本质地揭示了通过样本估计总体参数的内在机理。计量经济学理论的发展，更多的是以最大似然原理为基础的，对于一些特殊的计量经济学模型，只有最大似然方法才是很成功的估计方法。

对于普通最小二乘法，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；而对于最大似然法，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。显然，这是从不同原理出发的两种参数估计方法。

从总体中经过 n 次随机抽取得到样本容量为 n 的样本观测值，在任一次随机抽取中，样本观测值都以一定的概率出现。如果已经知道总体的参数，当然由变量的频率函数可以计算其概率。如果只知道总体服从某种分布，但不知道其分布参数，通过随机样本可以求出总体的参数估计量。以正态分布的总体为例，每个总体都有自己的分布参数的期望和方差，如果已经得到 n 组样本观测值，在这些可供选择的总体中，哪个总体最可能产生已经得到的 n 组样本观测值呢？显然，要对每个可能的正态总体估计取得 n 组样本观测值的联合概率，然后选择其参数能使观测值的联合概率为最大的那个总体。将样本观测值联合概率函数称为变量的**似然函数**。在已经取得样本观测值的情况下，使似然函数取极大值的总体分布参数所代表的总体具有最大的概率取得这些样本观测值，该总体参数即是所要求的参数。通过似然函数最大化以求得总体参数估计量的方法称为最大似然法。

在满足基本假设条件下，对一元线性回归模型

$$Y = \beta_0 + \beta_1 X + \mu$$

随机抽取 n 组样本观测值 $\{(X_i,Y_i): i=1,\ 2,\ \cdots,\ n\}$，由于 Y_i 服从如下的正态分布：

$$Y_i \sim N(\beta_0+\beta_1 X_i,\sigma^2)$$

于是，Y_i 的概率函数为

$$P(Y_i)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2\sigma^2}(Y_i-\beta_0-\beta_1X_i)^2}\ , \qquad i=1,2,\cdots,n$$

因为 Y_i 是相互独立的，所以 Y 的所有样本观测值的联合概率，也即**似然函数**为

$$\begin{aligned}L(\beta_0,\beta_1,\sigma^2)&=P(Y_1,Y_2,\cdots,Y_n)\\&=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum(Y_i-\beta_0-\beta_1X_i)^2}\end{aligned} \tag{2.3.7}$$

将该似然函数最大化，即可求得模型参数的**最大似然估计量**。

由于似然函数的最大化与似然函数的对数的最大化是等价的，所以取对数似然函数如下：

$$L^*=\ln L=-n\ln(\sqrt{2\pi}\sigma)-\frac{1}{2\sigma^2}\sum(Y_i-\beta_0-\beta_1X_i)^2 \tag{2.3.8}$$

对 L^* 求最大值，等价于对 $\sum(Y_i-\beta_0-\beta_1X_i)^2$ 求最小值。设 $\hat{\beta}_0,\hat{\beta}_1$ 满足该最值条件，即

$$\begin{cases}\dfrac{\partial}{\partial\hat{\beta}_0}\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2=0\\[2ex]\dfrac{\partial}{\partial\hat{\beta}_1}\sum(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i)^2=0\end{cases}$$

解得模型的参数估计量为

$$\begin{cases}\hat{\beta}_0=\dfrac{\sum X_i^2\sum Y_i-\sum X_i\sum Y_iX_i}{n\sum X_i^2-(\sum X_i)^2}\\[3ex]\hat{\beta}_1=\dfrac{n\sum Y_iX_i-\sum Y_i\sum X_i}{n\sum X_i^2-(\sum X_i)^2}\end{cases}$$

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

三、参数估计的矩法(MM)

普通最小二乘法是通过得到一个关于参数估计值的**正规方程组**并对它进

PDG

行求解而完成的。正规方程组(2.3.2)或(2.3.3)可以通过矩估计（Method of Moment, MM）的思想来导出。矩估计的基本原理是用相应的样本矩来估计总体矩。

在本章 § 2.2 对一元回归模型的假设中，已给出了如下两个基本的总体矩条件：

$$\mathrm{E}(\mu_i)=0$$

$$\mathrm{Cov}(X_i,\mu_i)=\mathrm{E}(X_i\mu_i)=0$$

于是，相应的样本矩条件可写成

$$\frac{1}{n}\sum\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)=0 \tag{2.3.9}$$

$$\frac{1}{n}\sum\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)X_i=0 \tag{2.3.10}$$

以上述方程组成的方程组，各自去掉$\frac{1}{n}$后不改变该方程组的解，而去掉$\frac{1}{n}$后该方程组恰为普通最小二乘法中的正规方程组(2.3.2)式。因此，解与普通最小二乘法以及极大似然法的结果相同。这种估计样本回归函数的方法称为矩估计法。

例 2.3.1

在上述家庭可支配收入-消费支出例子中，对于所抽出的一组样本数，参数估计的计算可通过表 2.3.1 进行。

表 2.3.1　参数估计的计算表

	X_i	Y_i	x_i	y_i	x_iy_i	x_i^2	y_i^2	X_i^2	Y_i^2
1	800	638	−1 350	−945	1 275 615	1 822 500	892 836	640 000	407 044
2	1 100	935	−1 050	−648	680 295	1 102 500	419 774	1210 000	874 225
3	1 400	1 155	−750	−428	320 925	562 500	183 098	1960 000	1 334 025
4	1 700	1 254	−450	−329	148 005	202 500	108 175	2 890 000	1 572 516
5	2 000	1 408	−150	−175	26 235	22 500	30 590	4 000 000	1 982 464
6	2 300	1 650	150	67	10 065	22 500	4 502	5 290 000	2 722 500
7	2 600	1 925	450	342	153 945	202 500	117 032	6 760 000	3 705 625
8	2 900	2 068	750	485	363 825	562 500	235 322	8 410 000	4 276 624
9	3 200	2 266	1 050	683	717 255	1 102 500	466 626	10 240 000	5 134 756
10	3 500	2 530	1 350	947	1 278 585	1 822 500	896 998	12 250 000	6 400 900
求和	21 500	15 829			4 974 750	7 425 000	3 354 955	53 650 000	28 410 679
平均	2 150	1 583							

由(2.3.5)式计算得

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{4\ 974\ 750}{7\ 425\ 000} = 0.670$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 1\ 583 - 0.670 \times 2\ 150 = 142.4$$

因此，由该样本估计的回归方程为

$$\hat{Y}_i = 142.4 + 0.670 X_i$$

四、最小二乘估计量的统计性质

当估计出模型参数后，需考虑参数估计值的精度，即是否能代表总体参数的真值。一般地，由于抽样波动的存在，以及所选估计方法的不同，都会使估计的参数与总体参数的真值有差距，因此考察参数估计量的统计性质就成了衡量该估计量“好坏”的主要准则。

一个用于考察总体的估计量，可从如下几个方面考察其优劣性：(1)线性性，即它是否是另一个随机变量的线性函数；(2)无偏性，即它的均值或期望是否等于总体的真实值；(3)有效性，即它是否在所有线性无偏估计量中具有最小方差；(4)渐近无偏性，即样本容量趋于无穷大时，它的均值序列是否趋于总体真值；(5)一致性，即样本容量趋于无穷大时，它是否依概率收敛于总体的真值；(6)渐近有效性，即样本容量趋于无穷大时，它在所有的一致估计量中是否具有最小的渐近方差。

这里，前三个准则也称作估计量的**有限样本性质或小样本性质**(small-sample properties)，因为一旦某估计量具有该类性质，它是不以样本的大小而改变的。拥有这类性质的估计量称为**最佳线性无偏估计量**(Best Linear Unbiased Estimator，BLUE)。当然，在有限样本情形下，有时很难找到最佳线性无偏估计量，这时就需要考察样本容量无限增大时估计量的渐近性质。后三个准则称为估计量的无限样本性质或**大样本渐近性质**(large-sample asymptotic properties)。如果有限样本情况下不能满足估计的准则，则应扩大样本容量，考察参数估计量的大样本性质。

需要说明的是，从估计量统计性质的角度看，无偏性与有效性是小样本性质中最为重要的两个性质，线性性并不是必须的；而在大样本性质中，由于问题较为复杂，人们更多地关注一致性。

可以证明，在经典线性回归的假定下，最小二乘估计量是具有最小方差的线性无偏估计量。

1．线性性

线性性，即估计量 $\hat{\beta}_0$，$\hat{\beta}_1$ 是 Y_i 的线性组合。由(2.3.5)式知

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} \\ &= \frac{\sum x_i Y_i}{\sum x_i^2} - \frac{\bar{Y} \sum x_i}{\sum x_i^2} = \sum k_i Y_i\end{aligned}$$

其中，$k_i = \dfrac{x_i}{\sum x_i^2}$。同样可得

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

其中，$w_i = \dfrac{1}{n} - \bar{X} k_i$。

2．无偏性

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

$$\begin{aligned}\hat{\beta}_1 &= \sum k_i Y_i = \sum k_i (\beta_0 + \beta_1 X_i + \mu_i) \\ &= \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i \mu_i\end{aligned}$$

易知

$$\sum k_i = \frac{\sum x_i}{\sum x_i^2} = 0\,, \qquad \sum k_i X_i = 1$$

故

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

$$\begin{aligned}\mathrm{E}(\hat{\beta}_1) &= \mathrm{E}(\beta_1 + \sum k_i \mu_i) \\ &= \beta_1 + \sum k_i \mathrm{E}(\mu_i) = \beta_1\end{aligned}$$

同样地，容易得出

$$\begin{aligned}\mathrm{E}(\hat{\beta}_0) &= \mathrm{E}(\beta_0 + \sum w_i \mu_i) \\ &= \mathrm{E}(\beta_0) + \sum w_i \mathrm{E}(\mu_i) = \beta_0\end{aligned}$$

3．有效性(最小方差性)

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

首先，由 $\hat{\beta}_1$，$\hat{\beta}_0$ 是关于 Y_i 的线性函数，可求得它们的方差为

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= \mathrm{Var}(\sum k_i Y_i) = \sum k_i^2 \mathrm{Var}(\beta_0 + \beta_1 X_i + \mu_i) \\ &= \sum k_i^2 \mathrm{Var}(\mu_i) = \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 \sigma^2 = \frac{\sigma^2}{\sum x_i^2}\end{aligned} \tag{2.3.11}$$

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\sum w_i Y_i) = \sum w_i^2 \mathrm{Var}(\beta_0 + \beta_1 X_i + \mu_i) \\ &= \sum (\frac{1}{n} - \bar{X}k_i)^2 \sigma^2 = \sum \left[\left(\frac{1}{n} \right)^2 - 2\frac{1}{n}\bar{X}k_i + \bar{X}^2 k_i^2 \right] \sigma^2 \\ &= \left[\frac{1}{n} - \frac{2}{n}\bar{X}\sum k_i + \bar{X}^2 \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 \right] \sigma^2 \\ &= \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right) \sigma^2 = \frac{\sum x_i^2 + n\bar{X}^2}{n\sum x_i^2} \sigma^2 = \frac{\sum X_i^2}{n\sum x_i^2} \sigma^2\end{aligned} \tag{2.3.12}$$

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

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

其中，$c_i = k_i + d_i$，d_i 为不全为零的常数，则容易证明(参见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)

$$\mathrm{Var}(\hat{\beta}_1^*) \geqslant \mathrm{Var}(\hat{\beta}_1)$$

同理，设 $\hat{\beta}_0^*$ 是其他估计方法得到的关于 β_0 的线性无偏估计量，则有

$$\mathrm{Var}(\hat{\beta}_0^*) \geqslant \mathrm{Var}(\hat{\beta}_0)$$

由以上分析可以看出，普通最小二乘估计量具有线性性、无偏性和有效性等优良性质，是**最佳线性无偏估计量**，这就是著名的**高斯-马尔可夫定理**(Gauss-Markov theorem)。显然这些优良的性质依赖于对模型的基本假设。

对于线性回归模型的普通最小二乘估计量，除了拥有一个“好”的估计量所应具备的小样本性质外，它也拥有“好”的大样本性质。例如，对 $\hat{\beta}_1$ 的一致性来说，易知

$$\begin{aligned}P\lim(\hat{\beta}_1) &= P\lim(\beta_1 + \sum k_i \mu_i) \\ &= P\lim(\beta_1) + P\lim\left(\frac{\sum x_i \mu_i}{\sum x_i^2} \right) \\ &= \beta_1 + \frac{P\lim\left(\frac{\sum x_i \mu_i}{n} \right)}{P\lim\left(\frac{\sum x_i^2}{n} \right)}\end{aligned}$$

等式右边第二项分子是 X 与 μ 的样本协方差的概率极限，它等于总体协方差 $\mathrm{Cov}(X,\mu)$，根据基本假设，其值为 0；而分母是 X 的样本方差的概率极限，由基本假设为一有限常数 Q，因此有

$$P\lim(\hat{\beta}_1) = \beta_1 + \frac{0}{Q} = \beta_1$$

五、参数估计量的概率分布及随机干扰项方差的估计

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

为了达到对所估计参数精度测定的目的，还需进一步确定参数估计量的概率分布。由于普通最小二乘估计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 分别是 Y_i 的线性组合，因此 $\hat{\beta}_0$，$\hat{\beta}_1$ 的概率分布取决于 Y_i。在 μ_i 是正态分布的假设下，Y_i 是正态分布，则 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 也服从正态分布，其分布特征由其均值和方差唯一决定。由此可得

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

于是，$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的标准差分别为

$$\sigma_{\hat{\beta}_0} = \sqrt{\frac{\sigma^2 \sum X_i^2}{n\sum x_i^2}} \tag{2.3.13}$$

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sigma^2}{\sum x_i^2}} \tag{2.3.14}$$

标准差可用来衡量估计量接近其真实值的程度，进而判断估计量的可靠性(图 2.3.1)。

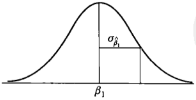


图 2.3.1

2. 随机干扰项 μ_i 的方差 σ^2 的估计

在估计的参数 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差表达式中，都含有随机干扰项的方差 σ^2。由

于σ^2实际上是未知的，因此$\hat{\beta}_0$和$\hat{\beta}_1$的方差实际上无法计算，这就需要对其进行估计。由于随机项μ_i不可观测，只能从μ_i的估计——残差e_i出发，对总体方差σ^2进行估计。可以证明σ^2的最小二乘估计量为(参见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \tag{2.3.15}$$

它是关于σ^2的无偏估计量。在最大似然估计法中，可通过对对数似然函数

$$L^* = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum(Y_i - \beta_0 - \beta_1 X_i)^2$$

关于σ^2求偏导，求得σ^2的如下最大似然估计量：

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{\sum e_i^2}{n} \tag{2.3.16}$$

在矩估计法中，由于有总体矩条件

$$\mathrm{Var}(\mu_i) = \mathrm{E}(\mu_i^2) = \sigma^2$$

其对应的样本矩条件即为

$$\hat{\sigma}^2 = \frac{1}{n}\sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \frac{\sum e_i^2}{n} \tag{2.3.17}$$

对照(2.3.15)式知，σ^2的最大似然估计量与矩估计量都不具无偏性，但却具有一致性。

在随机干扰项μ_i的方差σ^2估计出后，参数$\hat{\beta}_0$和$\hat{\beta}_1$的方差和标准差的估计量分别是：

$\hat{\beta}_1$的样本方差

$$S_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum x_i^2} \tag{2.3.18}$$

$\hat{\beta}_1$的样本标准差

$$S_{\hat{\beta}_1} = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}} \tag{2.3.19}$$

$\hat{\beta}_0$的样本方差

$$S_{\hat{\beta}_0}^2 = \frac{\hat{\sigma}^2 \sum X_i^2}{n \sum x_i^2} \tag{2.3.20}$$

$\hat{\beta}_0$ 的样本标准差

$$S_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \tag{2.3.21}$$

§ 2.4　一元线性回归模型的统计检验

回归分析是要通过样本所估计的参数来代替总体的真实参数，或者说是用样本回归线代替总体回归线。尽管从统计性质上已知，如果有足够多的重复抽样，参数的估计值的期望(均值)就等于其总体的参数真值，但在一次抽样中，估计值不一定就等于该真值。那么，在一次抽样中，参数的估计值与真值的差异有多大，是否显著，这就需要进一步进行统计检验，主要包括拟合优度检验、变量的显著性检验及参数检验的置信区间估计。

一、拟合优度检验

拟合优度检验，顾名思义，是检验模型对样本观测值的拟合程度。检验的方法是构造一个可以表征拟合程度的指标，在这里称为统计量，它是样本的函数。从检验对象中计算出该统计量的数值，然后与某一标准进行比较，得出检验结论。有人也许会问，采用普通最小二乘法进行估计，已经保证了模型最好地拟合了样本观测值，为什么还要检验拟合程度呢？问题在于，在一个特定的条件下做得最好的并不一定就是高质量的。普通最小二乘法所保证的最好的拟合，是同一个问题内部的比较，拟合优度检验结果所表示的优劣是不同问题之间的比较。例如，图 2.4.1 中的直线方程都是由散点表示的样本观测值用最小二乘法估计的结果，对于每个问题它们都满足残差的平方和最小，但是二者对样本观测值的拟合程度显然是不同的。

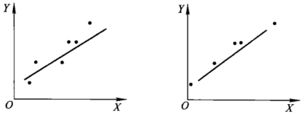
1. 总离差平方和的分解

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

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

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

$$y_i = Y_i - \overline{Y} = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y}) = e_i + \hat{y}_i \tag{2.4.1}$$



(a)　　(b)

图 2.4.1　OLS 法样本回归直线

图 2.4.2 表示了这种分解，其中，$\hat{y}_i = \hat{Y}_i - \overline{Y}$ 是样本回归线理论值(回归拟合值)与观测值 Y_i 的平均值之差，可认为是由回归线解释的部分；$e_i = Y_i - \hat{Y}_i$ 是实际观测值与回归拟合值之差，是回归线不能解释的部分。显然，如果 Y_i 落在样本回归线上，则 Y 的第 i 个观测值与样本均值的离差，全部来自样本回归拟合值与样本均值的离差，即完全可由样本回归线解释，表明在该点处实现完全拟合。

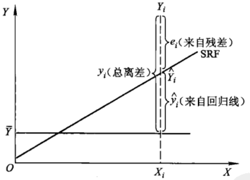


图 2.4.2　离差分解示意图

对于所有样本点，则需考虑这些点与样本均值离差的平方和。由于

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

可以证明 $\sum \hat{y}_i e_i = 0$，所以有

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 \tag{2.4.2}$$

记

$$\sum y_i^2 = \sum\ (Y_i - \bar{Y})^2 = \mathrm{TSS}$$

称为**总离差平方和**(total sum of squares)，反映样本观测值总体离差的大小；

$$\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \mathrm{ESS}$$

称为**回归平方和**(explained sum of squares)，反映由模型中解释变量所解释的那部分离差的大小；

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \mathrm{RSS}$$

称为**残差平方和**(residual sum of squares)，反映样本观测值与估计值偏离的大小，也是模型中解释变量未解释的那部分离差的大小。

(2.4.2)式表明Y **的观测值围绕其均值的总离差平方和可分解为两部分：一部分来自回归线，另一部分则来自随机势力。因此，可用来自回归线的回归平方和占** Y **的总离差平方和的比例来判断样本回归线与样本观测值的拟合优度。**

读者也许会问，既然 RSS 反映样本观测值与估计值偏离的大小，可否直接用它作为拟合优度检验的统计量呢？这里提出了一个普遍的问题，即**作为检验统计量的一般应该是相对量，而不能用绝对量**。因为用绝对量作为检验统计量，无法设置标准。在这里，残差平方和与样本容量关系很大，当 n 比较小时，它的值也较小，但不能因此而判断模型的拟合优度就好。

2. 可决系数 R^2 统计量

根据上述关系，可以用

$$R^2 = \frac{\mathrm{ESS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}} \tag{2.4.3}$$

检验模型的拟合优度，称 R^2 为**可决系数**(coefficient of determination)。显然，在总离差平方和中，回归平方和所占的比重越大，残差平方和所占的比重越小，回归直线与样本点拟合得越好。如果模型与样本观测值完全拟合，则有 $R^2 = 1$。当然，模型与样本观测值完全拟合的情况很少发生，R^2 等于 1 的情况较少。但毫无疑问的是该统计量越接近于 1，模型的拟合优度越高。

实际计算可决系数时，在 $\hat{\beta}_1$ 已经估计出后，一个较为简单的计算公式为

$$R^2 = \hat{\beta}_1^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right) \tag{2.4.4}$$

这里用到了样本回归函数的离差形式来计算回归平方和：

$$\mathrm{ESS} = \sum \hat{y}_i^2 = \sum (\hat{\beta}_1 x_i)^2 = \hat{\beta}_1^2 \sum x_i^2$$

在例 2.1.1 的可支配收入–消费支出例子中，

$$R^2 = \hat{\beta}_1^2 \frac{\sum x_i^2}{\sum y_i^2} = \frac{(0.670)^2 \times 7\ 425\ 000}{3\ 354\ 955} = 0.993\ 5$$

说明在线性回归模型中，家庭消费支出总离差中，由家庭可支配收入的离差解释的部分占99.35%，模型的拟合优度较高。

由(2.4.3)式知，可决系数的取值范围为$0 \leqslant R^2 \leqslant 1$，它是一个非负的统计量，随着抽样的不同而不同，即是随抽样而变动的统计量。为此，对可决系数的统计可靠性也应进行检验，这将在第三章中讨论。

二、变量的显著性检验

变量的显著性检验，旨在对模型中被解释变量与解释变量之间的线性关系是否显著成立作出推断，或者说考察所选择的解释变量是否对被解释变量有显著的线性影响。

从上面的拟合优度检验中可以看出，拟合优度高，则解释变量对被解释变量的解释程度就大，线性影响就强，可以推测模型线性关系成立；反之，就不成立。但这只是一个模糊的推测，不能给出一个统计上的严格的结论。因此，还必须进行变量的显著性检验。**变量的显著性检验所应用的方法是数理统计学中的假设检验。**

1．假设检验

假设检验是统计推断的一个主要内容，它的基本任务是根据样本所提供的信息，对未知总体分布的某些方面的假设作出合理的判断。

假设检验的程序是：先根据实际问题的要求提出一个论断，称为统计假设，记为H_0；然后根据样本的有关信息，对H_0的真伪进行判断，作出拒绝H_0或接受H_0的决策。

假设检验的基本思想是概率性质的反证法。为了检验原假设H_0是否正确，先假定这个假设是正确的，看由此能推出什么结果。如果导致一个不合理的结果，则表明“假设H_0为正确”是错误的，即原假设H_0不正确，因此要拒绝原假设H_0；如果没有导致一个不合理现象的出现，则不能认为原假设H_0不正确，因此不能拒绝原假设H_0。

概率性质的反证法的根据是小概率事件原理，该原理认为“小概率事件在一次试验中几乎是不可能发生的”。在原假设H_0下构造一个事件，这个事件在“假设H_0是正确”的条件下是一个小概率事件。随机抽取一组容量为n的样本观测值进行该事件的试验，如果该事件发生了，说明“假设H_0正确”是错误的，因为不应该出现的小概率事件出现了。因而应该拒绝原假设H_0；反之，如果该小概率事件没有出现，就没有理由拒绝原假设H_0，应该接受原假设H_0。

2．变量的显著性检验

用以进行变量显著性检验的方法主要有三种：F 检验，t 检验，z 检验。它们的区别在于构造的统计量不同。应用最为普遍的是 t 检验。几乎所有的计量经济学软件包中，都有关于 t 统计量的计算结果。我们在此只介绍 t 检验。

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

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

进一步根据数理统计学中的定义，如果真实的 σ^2 未知，而用它的无偏估计量 $\hat{\sigma}^2 = \dfrac{\sum e_i^2}{n-2}$ 替代时，可构造如下统计量：

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum x_i^2}}} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \tag{2.4.5}$$

则该统计量服从自由度为 $n-2$ 的 t 分布。因此，可用该统计量作为 β_1 显著性检验的 t 统计量。

如果变量 X 是显著的，那么参数 β_1 应该显著地不为 0。于是，在变量显著性检验中设计的原假设与备择假设分别为

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

给定一个显著性水平 α，比如 0.05，查 t 分布表(见附录)，得到一个临界值 $t_{\frac{\alpha}{2}}(n-2)$，则 $|t| > t_{\frac{\alpha}{2}}(n-2)$ (这里的 t 已不同于(2.4.5)式，其中 $\beta_1 = 0$)为原假设 H_0 下的一个小概率事件。

在参数估计完成后，可以很容易计算 t 的数值。如果发生了 $|t| > t_{\frac{\alpha}{2}}(n-2)$，则在 $1-\alpha$ 的置信度下拒绝原假设 H_0，即变量 X 是显著的，通过变量显著性检验；如果未发生 $|t| > t_{\frac{\alpha}{2}}(n-2)$，则在 $1-\alpha$ 置信度下不拒绝原假设 H_0，表明变量 X 是不显著的，未通过变量显著性检验。

对于一元线性回归方程中的 β_0，可构造如下 t 统计量进行显著性检验：

$$t = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\dfrac{\hat{\sigma}^2 \sum X_i^2}{n \sum x_i^2}}} = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \tag{2.4.6}$$

同样地，该统计量服从自由度为 $n-2$ 的 t 分布，检验的原假设一般仍为 $\beta_0 = 0$。

在例 2.1.1 及例 2.2.1 的可支配收入–消费支出例中，首先计算 σ^2 的估计值：

$$\hat{\sigma}^2=\frac{\sum e_i^2}{n-2}=\frac{\sum y_i^2-\hat{\beta}_1^2\sum x_i^2}{n-2}$$

$$=\frac{3\ 354\ 955-0.670^2\times 7\ 425\ 000}{10-2}=2\ 734$$

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

$$S_{\hat{\beta}_1}=\sqrt{\frac{\hat{\sigma}^2}{\sum x_i^2}}=\sqrt{\frac{2\ 734}{7\ 425\ 000}}=\sqrt{0.000\ 4}=0.019$$

$$S_{\hat{\beta}_0}=\sqrt{\frac{\hat{\sigma}^2\sum X_i^2}{n\sum x_i^2}}=\sqrt{\frac{2\ 734\times 53\ 650\ 000}{10\times 7\ 425\ 000}}=44.45$$

t 统计量的计算结果分别为

$$t_1=\frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}=\frac{0.670}{0.019}=34.92$$

$$t_0=\frac{\hat{\beta}_0}{S_{\hat{\beta}_0}}=\frac{-142.40}{44.45}=-3.20$$

给定一个显著性水平 $\alpha=0.05$，查 t 分布表中自由度为 8(在这个例中 $n-2=8$)，$\alpha=0.05$ 的临界值，得到 $t_{\frac{\alpha}{2}}(8)=2.306$。$|t_1|>t_{\frac{\alpha}{2}}(n-2)$，说明解释变量家庭可支配收入在 95%的置信度下显著，即通过了变量显著性检验；同样地，$|t_0|>t_{\frac{\alpha}{2}}(n-2)$,表明在 95%的置信度下，拒绝截距项为零的假设。

三、参数的置信区间估计

假设检验可以通过一次抽样的结果检验总体参数可能值的范围(最常用的假设为总体参数值为零)，但它并没有指出在一次抽样中样本参数值到底距总体参数的真值有多“近”。要判断样本参数的估计值在多大程度上可以“近似”地替代总体参数的真值，往往需要通过构造一个以样本参数的估计值为中心的“区间”，来考察它以多大的可能性(概率)包含着真实的参数值。这种方法就是参数检验的置信区间估计。

要判断估计的参数值 $\hat{\beta}_j$ 离真实的参数值 β_j 有多“近”（$j=0,1$），可预先选择一个概率 $\alpha(0<\alpha<1)$，并求一个正数 δ，使得**随机区间**(random interval)$(\hat{\beta}_j-\delta,\ \hat{\beta}_j+\delta)$ 包含参数 β_j 的真值的概率为 $1-\alpha$，即

$$\mathrm{P}(\hat{\beta}_j-\delta\leqslant\beta_j\leqslant\hat{\beta}_j+\delta)=1-\alpha$$

如果存在如上述这样的一个区间，称之为**置信区间**(confidence interval)；$1-\alpha$ 称

为**置信系数(置信度)**(confidence coefficient)，α 称为**显著性水平**(level of significance)；置信区间的端点称为**置信限**(confidence limit)或**临界值**(critical values)。

在变量的显著性检验中已经知道

$$t=\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}\sim t(n-2)\,,\qquad j=0,1$$

这就是说，如果给定置信度$1-\alpha$，从t分布表中查得自由度为$n-2$的临界值$t_{\frac{\alpha}{2}}$，那么t值处在$(-t_{\frac{\alpha}{2}},t_{\frac{\alpha}{2}})$的概率是$1-\alpha$，表示为

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

即

$$\mathrm{P}\left(-t_{\frac{\alpha}{2}}<\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}<t_{\frac{\alpha}{2}}\right)=1-\alpha$$

$$\mathrm{P}(\hat{\beta}_j-t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j}<\beta_j<\hat{\beta}_j+t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j})=1-\alpha$$

于是得到$1-\alpha$的置信度下β_j的置信区间是

$$(\hat{\beta}_j-t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j},\hat{\beta}_j+t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j}) \tag{2.4.7}$$

在例 2.1.1 与例 2.2.1 中，如果给定$\alpha=0.01$，查表得

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

从假设检验中已经得到

$$S_{\hat{\beta}_1}=0.019\,,\qquad S_{\hat{\beta}_0}=44.45$$

于是，根据(2.4.7)式计算得到β_1，β_0的置信区间分别为(0.605 6, 0.734 4)和(–6.719,291.52)。显然，参数β_1的置信区间小于β_0的置信区间。

由于置信区间在一定程度上给出了样本参数估计值与总体参数真值的“接近”程度，因此置信区间越小越好。**如何才能缩小置信区间呢？**从(2.4.7)式不难看出：(1)**增大样本容量**n。样本容量变大，可使样本参数估计量的标准差减小；同时，在同样的显著性水平下，n越大，t分布表中的临界值越小。(2)**提高模型的拟合优度**。因为样本参数估计量的标准差与残差平方和呈正比，模型的拟合优度越高，残差平方和应越小。

模型的参数一般具有特定的经济意义。例如，在例 2.1.1 中，参数β_1表示边际消费倾向。当经过模型估计得到$\hat{\beta}_1=0.670$后，我们能否说“边际消费倾向为 0.670”呢？不能。根据置信区间，我们只能说“边际消费倾向以 0.99 的

置信水平处于以 0.670 为中心的区间(0.605 6, 0.734 4)中”。

§2.5 一元线性回归分析的应用：预测问题

计量经济学模型的一个重要应用是经济预测。对于一元线性回归模型

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

如果给定样本以外的解释变量的观测值 X_0，可以得到被解释变量的预测值 $\hat{Y}_0$，可以此作为其条件均值 $\mathrm{E}(Y \mid X = X_0)$ 或个别值 Y 的一个近似估计。**严格地说，这只是被解释变量的预测值的估计值，而不是预测值。原因在于两方面：一是模型中的参数估计量是不确定的；二是随机干扰项的影响。所以，我们得到的仅是预测值的一个估计值，预测值仅以某一个置信度处于以该估计值为中心的一个区间中。预测在更大程度上说是一个区间估计问题。**

一、预测值是条件均值或个别值的一个无偏估计

在总体回归函数为 $\mathrm{E}(Y \mid X) = \beta_0 + \beta_1 X$ 的情况下，Y 在 $X = X_0$ 时的条件均值为

$$\mathrm{E}(Y \mid X = X_0) = \beta_0 + \beta_1 X_0$$

通过样本回归函数 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$，求得 $X = X_0$ 的拟合值为

$$\begin{aligned} \hat{Y}_0 &= \hat{\beta}_0 + \hat{\beta}_1 X_0 \\ \mathrm{E}(\hat{Y}_0) &= \mathrm{E}(\hat{\beta}_0 + \hat{\beta}_1 X_0) \\ &= \mathrm{E}(\hat{\beta}_0) + X_0 \mathrm{E}(\hat{\beta}_1) \\ &= \beta_0 + \beta_1 X_0 \end{aligned} \tag{2.5.1}$$

另一方面，在总体回归模型为 $Y = \beta_0 + \beta_1 X + \mu$ 的情况下，Y 在 $X = X_0$ 时的值为

$$\begin{aligned} Y_0 &= \beta_0 + \beta_1 X_0 + \mu \\ \mathrm{E}(Y_0) &= \mathrm{E}(\beta_0 + \beta_1 X_0 + \mu) \\ &= \beta_0 + \beta_1 X_0 + \mathrm{E}(\mu) \\ &= \beta_0 + \beta_1 X_0 \end{aligned} \tag{2.5.2}$$

(2.5.1)式与(2.5.2)式说明，在 $X = X_0$ 时，样本估计值 $\hat{Y}_0$ 是总体均值 $\mathrm{E}(Y \mid X = X_0)$ 和个别值 Y_0 的无偏估计，因此可用 $\hat{Y}_0$ 作为 $\mathrm{E}(Y \mid X = X_0)$ 与 Y_0 的预测值。

二、总体条件均值与个别值预测值的置信区间

1．总体条件均值预测值的置信区间

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

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

则
$$\mathrm{E}(\hat{Y}_0) = \mathrm{E}(\hat{\beta}_0) + X_0\mathrm{E}(\hat{\beta}_1) = \beta_0 + \beta_1 X_0$$
$$\mathrm{Var}(\hat{Y}_0) = \mathrm{Var}(\hat{\beta}_0) + 2X_0\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) + X_0^2\mathrm{Var}(\hat{\beta}_1)$$

可以证明(参见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2\bar{X}}{\sum x_i^2}$$

因此
$$\begin{aligned}\mathrm{Var}(\hat{Y}_0) &= \frac{\sigma^2\sum X_i^2}{n\sum x_i^2} - \frac{2X_0\bar{X}\sigma^2}{\sum x_i^2} + \frac{X_0^2\sigma^2}{\sum x_i^2}\\ &= \frac{\sigma^2}{\sum x_i^2}\left(\frac{\sum X_i^2 - n\bar{X}^2}{n} + \bar{X}^2 - 2X_0\bar{X} + X_0^2\right)\\ &= \frac{\sigma^2}{\sum x_i^2}\left[\frac{\sum x_i^2}{n} + (X_0 - \bar{X})^2\right]\\ &= \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]\end{aligned}$$

故
$$\hat{Y}_0 \sim N\left\{\beta_0 + \beta_1 X_0, \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]\right\} \tag{2.5.3}$$

将未知的σ^2代以它的无偏估计量$\hat{\sigma}^2$，则可构造t统计量：
$$t = \frac{\hat{Y}_0 - (\beta_0 + \beta_1 X_0)}{S_{\hat{Y}_0}} \sim t(n-2) \tag{2.5.4}$$

其中

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

于是，在$1-\alpha$的置信度下，总体均值$\mathrm{E}(Y \mid X_0)$的置信区间为

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times S_{\hat{Y}_0} < \mathrm{E}(Y \mid X_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times S_{\hat{Y}_0} \tag{2.5.5}$$

2．总体个别值预测值的置信区间

由$Y_0 = \beta_0 + \beta_1 X_0 + \mu$知
$$Y_0 \sim N(\beta_0 + \beta_1 X_0, \sigma^2)$$

PDG

于是
$$\hat{Y}_0 - Y_0 \sim N\left\{0, \sigma^2\left[1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right]\right\} \tag{2.5.6}$$

将未知的σ^2代以它的无偏估计量$\hat{\sigma}^2$，则可构造t统计量：
$$t=\frac{\hat{Y}_0-Y_0}{S_{\hat{Y}_0-Y_0}} \sim t(n-2) \tag{2.5.7}$$

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

从而在$1-\alpha$的置信度下，Y_0的置信区间为
$$\hat{Y}_0-t_{\frac{\alpha}{2}}\times S_{\hat{Y}_0-Y_0}<Y_0<\hat{Y}_0+t_{\frac{\alpha}{2}}\times S_{\hat{Y}_0-Y_0} \tag{2.5.8}$$

在例 2.1.1 及例 2.2.1 的可支配收入-消费支出例子中，得到的样本回归函数为
$$\hat{Y}_i=142.4+0.670X_i$$

则在X_0=1 000 处，
$$\hat{Y}_0=142.4+0.670\times 1\,000=812.4$$

它可作为总体均值$\mathrm{E}(Y \mid X=1\,000)$或 Y 的个别值在$X=1\,000$处的预测的估计值。而
$$\mathrm{Var}(\hat{Y}_0)=2\,734\left[\frac{1}{10}+\frac{(1\,000-2\,150)^2}{7\,425\,000}\right]=760.4$$
$$S(\hat{Y}_0)=27.6$$

因此，总体均值$\mathrm{E}(Y \mid X=1\,000)$的 95%的置信区间为
$$812.4-2.306\times 27.6<\mathrm{E}(Y \mid X=1\,000)<812.4+2.306\times 27.6$$

或为
$$(748.8, 875.9)$$

同样地，对于Y在$X=1\,000$的个别值Y_0，易知其 95%的置信区间为
$$812.4-2.306\times 59.1<Y\,|_{X=1\,000}<812.4+2.306\times 59.1$$

或为
$$(676.1, 948.7)$$

如图 2.5.1 所示，如果对每个X值求其总体均值$\mathrm{E}(Y \mid X)$的 95%的置信区间，将区间端点连接起来，可以得到关于总体回归函数的**置信带(域)(confidence band)**。同样地，对每个X值求Y的个别值的 95%的置信区间，将区间端点连接起来，可以得到关于个别值Y_0的**置信带(域)**。可以看出，Y的个别值Y_0的置信带比其总体均值的置信带宽。

对于Y的总体均值 $\mathrm{E}(Y_0)$与个别值Y_0的预测区间(置信区间)，有：(1)样本容量n越大，预测精度越高，反之预测精度越低；(2)样本容量一定时，置信带

的宽度在 X 的均值处最小，在其附近进行预测(插值预测)精度高；X 越远离其均值，置信带越宽，预测精度将下降。

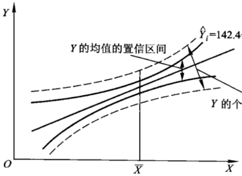


图 2.5.1 Y 均值与个别值的置信区间

§2.6 实例及时间序列问题

本节通过两个实例演示计量经济模型建立的一般过程。第一个例子是**截面数据**(cross-sectional data)的例子，第二个例子是**时间序列**(time series)的例子。在建立了时间序列的模型后，我们将初步讨论这类模型可能遇到的问题。

一、中国城镇居民人均消费支出模型：截面数据模型

例 2.6.1

为考察中国城镇居民 2006 年人均**可支配收入**与**消费支出**的关系，表 2.6.1 给出了中国内地 31 个省区以当年价测算的城镇居民家庭年人均收入(X)与年人均支出(Y)两组数据。由于表中是同一年份中不同地区居民家庭的人均可支配收入与人均消费支出，因此也称为是**截面数据**。

1. 建立模型

本例中我们假设拟建立如下一元回归模型：

$$Y = \beta_0 + \beta_1 X + \mu$$

表 2.6.2 给出了采用 Eviews 软件对表 2.6.1 中的数据进行回归分析的计算结果。

一般可写出如下回归分析结果：

$$\hat{Y}_i = 281.50 + 0.714\,6X_i$$

$$(1.05) \qquad (31.39)$$

$$R^2=0.971\,4 \qquad F=985.66 \qquad \text{D.W.}=1.46$$

表 2.6.1　中国内地各地区城镇居民家庭人均全年可支配收入与人均全年消费性支出(元)

地区	可支配收入 *X*	消费支出 *Y*	地区	可支配收入 *X*	消费支出 *Y*
北京	14 825.41	19 977.52	湖北	7 397.32	9 802.65
天津	10 548.05	14 283.09	湖南	8 169.30	10 504.67
河北	7 343.49	10 304.56	广东	12 432.22	16 015.58
山西	7 170.94	10 027.70	广西	6 791.95	9 898.75
内蒙古	7 666.61	10 357.99	海南	7 126.78	9 395.13
辽宁	7 987.49	10 369.61	重庆	9 398.69	11 569.74
吉林	7 352.64	9 775.07	四川	7 524.81	9 350.11
黑龙江	6 655.43	9 182.31	贵州	6 848.39	9 116.61
上海	14 761.75	20 667.91	云南	7 379.81	10 069.89
江苏	9 628.59	14 084.26	西藏	6 192.57	8 941.08
浙江	13 348.51	18 265.10	陕西	7 553.28	9 267.70
安徽	7 294.73	9 771.05	甘肃	6 974.21	8 920.59
福建	9 807.71	13 753.28	青海	6 530.11	9 000.35
江西	6 645.54	9 551.12	宁夏	7 205.57	9 177.26
山东	8 468.40	12 192.24	新疆	6 730.01	8 871.27
河南	6 685.18	9 810.26			

资料来源：《中国统计年鉴》(2007)。

其中括号内的数为相应参数的 t 检验值，R^2 是可决系数，F 与 D.W.是有关的两个检验统计量，其含义将在后面的章节中介绍。

表 2.6.2　中国内地城镇居民人均消费支出对人均可支配收入的回归

```
Dependent Variable:Y
Method:Least Squares
Date:11/02/08 Time:09:32
Sample:1 31
Included observations:31
------------------------------------------------------------------------
      Variable        Coefficient      Std.Error     t-Statistic     Prob
------------------------------------------------------------------------
          C            281.499 3       268.9497       1.046662      0.3039
          X            0.7145 54       0.022760       31.39525      0.0000
------------------------------------------------------------------------
R-squared               0.971419    Mean dependent var             8401.467
Adjusted R-squared      0.970433    S.D. dependent var             2388.455
S.E. of regression      410.6928    Akaike info criterion          14.93591
Sum squared resid        4891388    Schwarz criterion              15.02842
Log likelihood         –229.5066    F-statistic                    985.6616
Durbin-Watson stat      1.461502    Prob(F-statistic)              0.000000
------------------------------------------------------------------------
```

2．模型检验

从回归估计的结果看，模型拟合较好。可决系数 R^2 =0.971 4，表明城镇居民人均消费支出变化的 97.14%可由人均可支配收入的变化来解释。从斜率项的 t 检验值看，大于 5%显著性水平下自由度为 n–2=29 的临界值 $t_{0.025}$(29)=2.05，且该斜率值满足 0<0.714 6<1，符合经济理论中边际消费倾向在 0 与 1 之间的绝对收入假说，表明 2006 年，中国城镇居民家庭人均可支配收入每增加 1 元，人均消费支出增加 0.714 6 元。

3．预测

假设我们需要关注 2006 年人均可支配收入在 20 000 元这一档的中国城镇家庭的人均消费支出问题。由上述回归方程可得该类家庭人均消费支出的预测值：

$$\hat{Y}_0 = 281.50 + 0.714\ 6 \times 20\ 000 = 14\ 572.6\ (\text{元})$$

下面给出该类家庭人均消费支出 95%置信度的预测区间。

由于人均可支配收入 X 的样本均值与样本方差为

$$\mathrm{E}(X)=11\ 363.69 \qquad \mathrm{Var}(X)=10\ 853\ 528$$

于是，在 95%的置信度下，E(Y_0)的预测区间为

$$14\ 572.6 \pm 2.045 \times \sqrt{\frac{4\ 891\ 388}{31-2} \times \left(\frac{1}{31} + \frac{(20\ 000 - 11\ 363.69)^2}{(31-1) \times 10\ 853\ 528}\right)}$$

$$=14\ 572.6 \pm 429.3$$

或　　(14 143.3,15 001.9)

如果我们想知道某地区某城镇家庭人均可支配收入为 20 000 元时，该家庭人均消费支出的个值预测，则仍通过上述样本回归方程得到 14 572.6 元的消费支出预测值。

同样地，在 95%的置信度下，该家庭人均消费支出的预测区间为

$$14\ 572.6 \pm 2.045 \times \sqrt{\frac{4\ 891\ 388}{31-2} \times \left(1 + \frac{1}{31} + \frac{(20\ 000 - 11\ 363.69)^2}{(31-1) \times 10\ 853\ 528}\right)}$$

$$=14\ 572.6 \pm 943.2$$

或　　(13 629.3, 15 515.8)

PDG

二、中国居民总量消费函数：时间序列数据模型

例 2.6.2

建立总量消费函数(aggregate consumption function)是进行宏观经济管理的重要手段。为了从总体上考察中国居民收入与消费的关系，表 2.6.3 给出了中国名义支出法国内生产总值 GDP、名义居民总消费 CONS 以及表示宏观税赋的税收总额 TAX、表示价格变化的居民消费价格指数 CPI(1990=100)，并由这些数据整理出实际支出法国内生产总值 GDPC=GDP/CPI、居民实际消费总支出 *Y*=CONS/CPI，以及实际可支配收入 *X*=(GDP-TAX)/CPI。这些数据是 1978—2006 年的**时间序列数据**(time series data)，即观测值是连续不同年份中的数据。

表 2.6.3 中国居民总量消费支出与收入资料

单位：亿元

年份	GDP	CONS	CPI	TAX	GDPC	*X*	*Y*
1978	3 605.6	1 759.1	46.21	519.28	7802.5	6 678.8	3 806.7
1979	4 092.6	2 011.5	47.07	537.82	8694.2	7 551.6	4 273.2
1980	4 592.9	2 331.2	50.62	571.70	9073.7	7 944.2	4 605.5
1981	5 008.8	2 627.9	51.90	629.89	9651.8	8 438.0	5 063.9
1982	5 590.0	2 902.9	52.95	700.02	10 557.3	9 235.2	5 482.4
1983	6 216.2	3 231.1	54.00	775.59	11 510.8	10 074.6	5 983.2
1984	7 362.7	3 742.0	55.47	947.35	13 272.8	11 565.0	6 745.7
1985	9 076.7	4 687.4	60.65	2 040.79	14 966.8	11 601.7	7 729.2
1986	10 508.5	5 302.1	64.57	2 090.37	16 273.7	13 036.5	8 210.9
1987	12 277.4	6 126.1	69.30	2 140.36	17 716.3	14 627.7	8 840.0
1988	15 388.6	7 868.1	82.30	2 390.47	18 698.7	15 794.0	9 560.5
1989	17 311.3	8 812.6	97.00	2 727.40	17 847.4	15 035.5	9 085.5
1990	19 347.8	9 450.9	100.00	2 821.86	19 347.8	16 525.9	9 450.9
1991	22 577.4	10 730.6	103.42	2 990.17	21 830.9	18 939.6	10 375.8
1992	27 565.2	13 000.1	110.03	3 296.91	25 053.0	22 056.5	11 815.3
1993	36 938.1	16 412.1	126.20	4 255.30	29 269.1	25 897.3	13 004.7
1994	50 217.4	21 844.2	156.65	5 126.88	32 056.2	28 783.4	13 944.2
1995	63 216.9	28 369.7	183.41	6 038.04	34 467.5	31 175.4	15 467.9
1996	74 163.6	33 955.9	198.66	6 909.82	37 331.9	33 853.7	17 092.5
1997	81 658.5	36 921.5	204.21	8 234.04	39 988.5	35 956.2	18 080.6
1998	86 531.6	39 229.3	202.59	9 262.80	42 713.1	38 140.9	19 364.1
1999	91 125.0	41 920.4	199.72	10 682.58	45 625.8	40 277.0	20 989.3
2000	98 749.0	45 854.6	200.55	12 581.51	49 238.0	42 964.6	22 863.9
2001	108 972.4	49 213.2	201.94	15 301.38	53 962.5	46 385.4	24 370.1
2002	120 350.3	52 571.3	200.32	17 636.45	60 078.0	51 274.0	26 243.2
2003	136 398.8	56 834.4	202.73	20 017.31	67 282.2	57 408.1	28 035.0
2004	16 0280.4	63 833.5	210.63	24 165.68	76 096.3	64 623.1	30 306.2
2005	188 692.1	71 217.5	214.42	28 778.54	88 002.1	74 580.4	33 214.4
2006	221 170.5	80 120.5	217.65	34 809.72	10 1616.3	85 623.1	36 811.2

资料来源：根据《中国统计年鉴》(2001，2007)整理。

1. 建立模型

对时间序列数据，也可建立类似于截面数据的计量经济模型，并进行回归分析。本例中我们假设拟建立如下一元回归模型：

$$Y = \beta_0 + \beta_1 X + \mu$$

表 2.6.4 给出了采用 Eviews 软件对表 2.6.3 中的数据进行回归分析的计算结果，表明可建立如下中国居民消费函数：

$$\hat{Y} = 2\ 091.29 + 0.437\ 5X$$

表 2.6.4　中国居民总量消费 *Y* 对可支配收入 *X* 的回归(1978—2006)

Dependent Variable:Y
Method:Least Squares
Sample:1978 2006
Included observations:29

	Coefficient	Std.Error	t-Statistic	Prob.
C	2091.295	334.9869	6.242914	0.0000
X	0.437527	0.009297	47.05950	0.0000
R-squared	0.987955	Mean dependent var		14855.72
Adjusted R-squared	0.987509	S.D. dependent var		9472.076
S.E. of regression	1058.633	Akaike info criterion		16.83382
Sum squared resid	30259014	Schwarz criterion		16.92811
Log likelihood	−242.0903	Hannan-Quinn criter		16.86335
F-statistic	2214.596	Durbin-Watson stat		0.277155
Prob(F-statistic)	0.000000			

2. 模型检验

$\sqrt{\frac{R^2}{n-2}}$　　$\sqrt{\frac{R^2}{n-2}(\frac{1}{n}+\frac{(x-\bar{x})^2}{(n-1)\sigma}}$

从回归估计的结果看，模型拟合较好：可决系数 R^2=0.988 0，截距项与斜率项的 t 检验值均大于 5%显著性水平下自由度为 n–2=27 的临界值 $t_{0.025}(27)$=2.05，且斜率项符合经济理论中边际消费倾向在 0 与 1 之间的绝对收入假说，斜率项 0.438 表明，在 1978—2006 年间，以 1990 年价计的中国居民可支配总收入每增加 1 亿元，居民总量消费支出平均增加 0.438 亿元。

3. 预测

2007 年，以当年价计的中国 GDP 为 263 242.5 亿元，税收入总额 45 621.9 亿元，居民消费价格指数为 409.1，由此可得到以 1990 年价计的可支配总收入 X 约 95 407.4 亿元，由上述回归方程可得 2007 年居民总量消费预测的点估计值：

$$Y_{2007}=2\ 091.3 + 0.437\ 5\times 95\ 407.4 = 438\ 34.6(\text{亿元})$$

2007 年，中国名义居民消费总量为 93 317.2 亿元，以 1990 年为基准的居民消费价格指数为 228.1，由此可推得当年中国实际居民消费总量为 40 910.7 亿元，可见相对误差为 7.14%。

下面给出 2007 年中国居民总量消费的预测区间。由于在样本期内

$$\mathrm{E}(X)=29\ 174.1 \qquad \mathrm{Var}(X)=463\ 039\ 370$$

于是，在 95%的置信度下，$\mathrm{E}(Y_{2007})$的预测区间为

$$43\ 834.6\pm 2.051\times\sqrt{\frac{30\ 259\ 014}{29-2}\times\left(\frac{1}{29}+\frac{(95\ 407.4-29\ 174.1)^2}{(29-1)\times 463\ 039\ 370}\right)}$$

$$=43\ 834.6\pm 1\ 326.3$$

或 $(42\ 508.3,45\ 160.9)$

同样地，在 95%的置信度下，Y_{2001} 的预测区间为：

$$43\ 834.6\pm 2.051\times\sqrt{\frac{30\ 259\ 014}{29-2}\times\left(1+\frac{1}{29}+\frac{(95\ 407.4-29\ 174.1)^2}{(29-1)\times 463\ 039\ 370}\right)}$$

$$=43\ 834.6\pm 2\ 545.1$$

或 $(41\ 289.5, 46\ 379.7)$

三、时间序列问题

例 2.6.2 表明，时间序列完全可以进行类似于截面数据的回归分析。然而，在时间序列回归分析中，有两个需注意的问题。

第一，关于抽样分布的理解问题。我们能把表 2.6.3 中的数据理解为是从某个总体中抽出的一个样本吗？如果我们将该组数据理解成来自于产生该组数据的经济系统，而该经济系统能够产生任何可能的“实际”数据，即该组数据只是该经济系统“所有”可能产生的数据中的一组数据。这时，我们就认为该组数据是众多可能产生的数据中的一个样本。

第二，关于“伪回归问题”(spurious regression problem)。注意到我们对可决系数 R^2 的定义与解释，它被定义为回归平方和占总离差平方和的比重，解释为被解释变量 Y 的变化中可由解释变量 X 的变化“解释”的部分。我们并未将这里的“解释”替换为“引起”，因为因果关系不能通过回归分析本身来判断。然而，由于回归分析往往就是要对因果关系进行评判，人们自然倾向于认为一个高的可决系数就意味着 X 对 Y 的“影响”能力强。

在现实经济问题中，对时间序列数据作回归，即使两个变量间没有任何的实际联系，也往往会得到较高的可决系数。问题在于许多经济变量的时间序列往往具有相同的变动趋势，如价格指数、产出水平、就业水平等往往表现出在一段时间内持续上升或持续下降。对于具有共同变化趋势的时间序列，即使它们之间没有任何实际的联系，也会产生较高的可决系数。这意味着许多通过较

高可决系数而“发现”的变量间的联系是虚假的(spurious)。例 2.6.2 中，可决系数 R^2 高达 0.988 0，然而，观察发现 X 与 Y 在该时间段中表现出持续的上升趋势(图 2.6.1)，这意味着我们需要谨慎对待所得到的较高的可决系数。这里得到的回归线对居民消费的“解释”可能存在部分的虚假性。

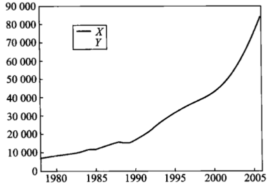


图 2.6.1　中国居民可支配总收入 X 与消费总支出 Y 变动图

为了避免时间序列作回归分析时，由于变量随时间的共同变化趋势而可能带来的误导，一种解决方案是在模型中引入代表时间变化的时间趋势项，这样可将变量共同变化的时间趋势分离出来，从而能够较准确地考察解释变量对被解释变量变动的解释能力，一定程度地消除可能存在的“伪回归”现象。在模型中引入更多解释变量就成为多元回归分析，也是后续章节的主要内容。当然，从理论上讲，只有“平稳”的时间序列才适用于经典的线性回归模型，而实际的时间序列经常是非平稳的。如何避免“伪回归”或“虚假回归”，如何利用非平稳时间序列建立经典线性回归模型，这些问题将在第八章中进一步讨论。

本章练习题

1．为什么计量经济学模型的理论方程中必须包含随机干扰项？

2．下列计量经济学方程哪些是正确的？哪些是错误的？为什么？

(1) $Y_t=\alpha+\beta X_t$，　$t=1,2,\cdots,n$；

(2) $Y_t=\alpha+\beta X_t+\mu_t$，　$t=1,2,\cdots,n$；

(3) $Y_t=\hat{\alpha}+\hat{\beta} X_t+\mu_t$，　$t=1,2,\cdots,n$；

(4) $\hat{Y}_t=\hat{\alpha}+\hat{\beta} X_t+\mu_t$，　$t=1,2,\cdots,n$；

(5) $Y_t=\hat{\alpha}+\hat{\beta} X_t$，　$t=1,2,\cdots,n$；

(6) $\hat{Y}_t=\hat{\alpha}+\hat{\beta} X_t$，　$t=1,2,\cdots,n$；

(7) $Y_t=\hat{\alpha}+\hat{\beta}X_t+\hat{\mu}_t$，　　$t=1,2,\cdots,n$；

(8) $\hat{Y}_t=\hat{\alpha}+\hat{\beta}X_t+\hat{\mu}_t$，　　$t=1,2,\cdots,n$。

其中带“ ^ ”者表示“估计值”。

3．一元线性回归模型的基本假设主要有哪些？违背基本假设的计量经济学模型是否就不可以估计？

4．线性回归模型

$$Y_i=\alpha+\beta X_i+\mu_i,\qquad i=1,2,\cdots,n$$

的零均值假设是否可以表示为$\frac{1}{n}\sum_{i=1}^{n}\mu_i=0$？为什么？

5．假设已经得到关系式$Y=\beta_0+\beta_1X$的最小二乘估计，试回答：

(1) 假设决定把变量X的单位扩大10倍，这样对原回归的斜率和截距会有什么样的影响？如果把变量Y的单位扩大10倍，又会怎样？

(2) 假定给X的每个观测值都增加2，对原回归的斜率和截距会有什么样的影响？如果给Y的每个观测值都增加2，又会怎样？

6．假使在回归模型$Y_i=\beta_0+\beta_1X_i+\mu_i$中，用不为零的常数$\delta$去乘每一个$X$值，这会不会改变$Y$的拟合值及残差？如果对每个$X$都加上一个非零常数$\delta$，又会怎样？

7．假设有人做了如下的回归：

$$y_i=\hat{\beta}_0+\hat{\beta}_1x_i+e_i$$

其中，y_i,x_i分别为Y_i,X_i关于各自均值的离差。问$\hat{\beta}_1$和$\hat{\beta}_0$将分别取何值？

8．令$\hat{\beta}_{YX}$和$\hat{\beta}_{XY}$分别为Y对X的回归和X对Y的回归中的斜率，证明：

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

其中r为X与Y之间的线性相关系数。

9．记样本回归模型为$Y_i=\hat{\beta}_0+\hat{\beta}_1X_i+e_i$，试证明：

(1) 估计的Y的均值等于实测的Y的均值：

$$\bar{\hat{Y}}=\bar{Y}$$

(2) 残差和为零，从而残差的均值为零：

$$\sum e_i=0,\qquad \bar{e}=0$$

(3) 残差项与X不相关：

$$\sum e_iX_i=0$$

(4) 残差项与估计的Y不相关：

$$\sum e_i\hat{Y}_i=0$$

10．证明：一元线性回归总离差平方和的分解式中：

$$\sum\hat{y}_ie_i=0$$

11．下面数据是依据10对X和Y的观察值得到的：

$$\sum Y_i=1\,110,\qquad \sum X_i=1\,680,\qquad \sum X_iY_i=204\,200$$

$$\sum X_i^2=315\,400,\qquad \sum Y_i^2=133\,300$$

假定满足所有的经典线性回归模型的假设。求：

(1) β_0，β_1的估计值及其标准差；

PDG

(2) 可决系数 R^2；

(3) 对 β_0，β_1 分别建立 95%的置信区间。利用置信区间法，你可以接受零假设：$\beta_1=0$ 吗？

12. 下表是中国内地 2007 年各地区税收 Y 和国内生产总值 GDP 的统计资料。

单位：亿元

地区	Y	GDP	地区	Y	GDP
北京	1 435.7	9 353.3	湖北	434.0	9 230.7
天津	438.4	5 050.4	湖南	410.7	9 200.0
河北	618.3	13 709.5	广东	2 415.5	31 084.4
山西	430.5	5 733.4	广西	282.7	5 955.7
内蒙古	347.9	6 091.1	海南	88.0	1 223.3
辽宁	815.7	11 023.5	重庆	294.5	4 122.5
吉林	237.4	5 284.7	四川	629.0	10 505.3
黑龙江	335.0	7 065.0	贵州	211.9	2 741.9
上海	1 975.5	12 188.9	云南	378.6	4 741.3
江苏	1 894.8	25 741.2	西藏	11.7	342.2
浙江	1 535.4	18 780.4	陕西	355.5	5 465.8
安徽	401.9	7 364.2	甘肃	142.1	2 702.4
福建	594.0	9 249.1	青海	43.3	783.6
江西	281.9	5 500.3	宁夏	58.8	889.2
山东	1 308.4	25 965.9	新疆	220.6	3 523.2
河南	625.0	15 012.5			

要求，以手工和运用 Eviews 软件(或其他软件)：

(1) 作出散点图，建立税收随国内生产总值 GDP 变化的一元线性回归方程，并解释斜率的经济意义；

(2) 对所建立的回归方程进行检验；

(3) 若 2008 年某地区国内生产总值为 8 500 亿元，求该地区税收收入的预测值及预测区间。

$$8500 \pm t\sqrt{\frac{R^2}{n-2}\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{(n-1)\sigma^2}\right)}$$

第三章　经典单方程计量经济学模型：多元线性回归模型 3

在实际经济问题中，一个变量往往受到多个变量的影响。例如，家庭消费支出，除了受家庭可支配收入的影响外，还受诸如家庭所拥有的财富、物价水平、金融机构存款利息，甚至广告、就业状况等多种因素的影响，表现在线性回归模型中的解释变量有多个。这样的模型被称为**多元线性回归模型**。多元线性回归模型参数估计的原理与一元线性回归模型相同，只是计算更为复杂。

§3.1　多元线性回归模型

一、多元线性回归模型

多元线性回归模型的一般形式为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu \tag{3.1.1}$$

其中k为解释变量的数目，$\beta_j\ (j = 1,2,\cdots,k)$称为**回归系数**(regression coefficient)。人们习惯上把常数项看作一个虚变量的参数，在参数估计过程中该虚变量的样本观测值始终取 1，这样，模型中解释变量的数目为$k+1$。

同一元回归分析一样，(3.1.1)式也被称为**总体回归函数**的随机表达形式。它的非随机表达式为

$$\mathrm{E}(Y \mid X_1, X_2, \cdots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{3.1.2}$$

可见，多元回归分析是以多个解释变量的给定值为条件的回归分析，(3.1.2)式表示各解释变量X值给定时Y的平均响应。β_j也被称为**偏回归系数**(partial regression coefficient)，表示在其他解释变量保持不变的情况下，X_j每变化一个单位时，Y的均值$\mathrm{E}(Y)$的变化，或者说β_j给出X_j的单位变化对Y均值的“直接”或“净”(不含其他变量)影响。

如果给出一组观测值$\{(X_{i1}, X_{i2}, \cdots, X_{ik}, Y_i): i = 1,2,\cdots,n\}$，则总体回归模型还可写成如下形式

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i \qquad i = 1,2\cdots,n \tag{3.1.3}$$

由(3.1.1)式表示的n个随机方程的矩阵表达式为

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\mu} \tag{3.1.4}$$

其中

$$\boldsymbol{Y}=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix}_{n\times1},\quad \boldsymbol{X}=\begin{pmatrix}1&X_{11}&X_{12}&\cdots&X_{1k}\\1&X_{21}&X_{22}&\cdots&X_{2k}\\\vdots&\vdots&\vdots&&\vdots\\1&X_{n1}&X_{n2}&\cdots&X_{nk}\end{pmatrix}_{n\times(k+1)}$$

$$\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\\beta_1\\\beta_2\\\vdots\\\beta_k\end{pmatrix}_{(k+1)\times1},\quad \boldsymbol{\mu}=\begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix}_{n\times1}$$

与一元回归分析相仿，在给出总体中的一个样本时，估计**样本回归函数**，并让它近似代表未知的总体回归函数。

样本回归函数可表示为

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2X_2+\cdots+\hat{\beta}_kX_k \tag{3.1.5}$$

其随机表达式为

$$Y=\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2X_2+\cdots+\hat{\beta}_kX_k+e \tag{3.1.6}$$

其中e称为**残差**或**剩余项**(residual)，可看成是总体回归函数中随机干扰项μ的近似替代。

在一个容量为n的样本下，样本回归函数(3.1.5)式与(3.1.6)式也可表示如下

$$\hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik}\quad i=1,2,\cdots,n \tag{3.1.7}$$

$$Y_i=\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik}+e_i\quad i=1,2,\cdots,n \tag{3.1.8}$$

同样地，(3.1.7)式与(3.1.8)式中样本回归函数的矩阵表达式分别为

$$\hat{\boldsymbol{Y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{3.1.9}$$

$$\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{e} \tag{3.1.10}$$

其中

$$\hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\beta}_0\\\hat{\beta}_1\\\vdots\\\hat{\beta}_k\end{pmatrix},\quad \boldsymbol{e}=\begin{pmatrix}e_1\\e_2\\\vdots\\e_n\end{pmatrix}$$

二、多元线性回归模型的基本假定

为了使参数估计量具有良好的统计性质，对多元线性回归模型可做出类似于一元线性回归分析那样的若干基本假设。

假设 1：回归模型是正确设定的。

假设 2：解释变量 $X_1, X_2, \cdots, X_k$ 是非随机的或固定的，且各 X_j 之间不存在严格线性相关性(无完全多重共线性)。

假设 3：各解释变量 X_j 在所抽取的样本中具有变异性，而且随着样本容量的无限增加，各解释变量的样本方差趋于一个非零的有限常数，即 $n \to +\infty$ 时，

$$\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2 \to Q_j$$

假设 4：随机误差项具有条件零均值、同方差及不序列相关性

$$\mathrm{E}(\mu_i \mid X_1, X_2, \cdots, X_k) = 0$$

$$\mathrm{Var}(\mu_i \mid X_1, X_2, \cdots, X_k) = \sigma^2$$

$$\mathrm{Cov}(\mu_i, \mu_j \mid X_1, X_2, \cdots, X_k) = 0 \qquad i \neq j$$

假设 5：解释变量与随机项不相关

$$\mathrm{Cov}(X_{ij}, \mu_i \mid X_1, X_2, \cdots, X_k) = 0 \qquad j = 1, 2, \cdots, k$$

假设 6：随机项满足正态分布

$$\mu_i \mid X_1, X_2, \cdots, X_k \sim N(0, \sigma^2)$$

与一元线性回归模型的假设相比，假设 2 是多元回归模型所特有的，该假设要求多个解释变量间不存在严格的线性相关性。假设 3 同样是为了使估计量有良好的大样本性质以及避免时间序列中可能出现的伪回归问题；假设 5 也并非是必需的，只要假设 4 中随机误差项的条件零均值假设成立，假设 5 一定成立。

为了书写方便，上述假设 2 至假设 6 还可用矩阵符号来表示：

假设 2：$n \times (k+1)$ 矩阵 $\boldsymbol{X}$ 的秩 $(\boldsymbol{X}) = k+1$，即 $\boldsymbol{X}$ 列满秩。

假设 3：$n \to +\infty$ 时，$\frac{1}{n}\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{Q}$,

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

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

假设 4：$\mathrm{E}(\boldsymbol{\mu} \mid \boldsymbol{X}) = \boldsymbol{0}$

$$\mathrm{Var}(\boldsymbol{\mu}\mid \boldsymbol{X})=\mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}'\mid \boldsymbol{X})=\mathrm{E}\left(\begin{pmatrix} \mu_1^2 & \cdots & \mu_1\mu_n \\ \vdots & & \vdots \\ \mu_n\mu_1 & \cdots & \mu_n^2 \end{pmatrix}\middle| \boldsymbol{X}\right)$$

$$=\begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}=\sigma^2\boldsymbol{I}$$

其中，$\boldsymbol{I}$ 为一 n 阶单位矩阵。

假设 5：$\mathrm{E}(\boldsymbol{X}'\boldsymbol{\mu}\mid \boldsymbol{X})=\boldsymbol{0}$，

假设 6：向量 $\boldsymbol{\mu}$ 服从一多维正态分布，即

$$\boldsymbol{\mu}\mid \boldsymbol{X}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)$$

类似于对一元回归模型假设的讨论，上述条件期望与条件方差的形式均可写成非条件的期望与方差的形式。

§ 3.2　多元线性回归模型的参数估计

同一元线性回归模型的参数估计一样，多元线性回归模型参数估计的任务仍有两项：一是求得反映变量之间数量关系的结构参数的估计量 $\hat{\beta}_j\,(j=0,1,\cdots,k)$；二是求得随机干扰项的方差估计 $\hat{\sigma}^2$。多元线性回归模型在满足 § 3.1 所列的基本假设的情况下，可以采用普通最小二乘法、最大似然法或者矩估计法估计参数。

一、普通最小二乘估计

1. 普通最小二乘估计及其矩阵表达

随机抽取 n 组样本观测值 $\{(X_{i1},X_{i2},\cdots X_{ik},Y_i):i=1,2,\cdots,n\}$，如果样本函数的参数估计值已经得到，则有

$$\hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik}\,,\qquad i=1,2,\cdots,n \tag{3.2.1}$$

根据最小二乘原理，参数估计值应使

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

$$=\sum_{i=1}^{n}[Y_i-(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})]^2 \tag{3.2.2}$$

达到最小。由微积分知识可知，只需求 Q 关于待估参数 $\hat{\beta}_j\,(j=0,1,\cdots,k)$ 的偏导数，并令其值为零，就可得到**待估参数估计值的正规方程组**：

$$\begin{cases}\sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})=\sum Y_i\\ \sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{i1}=\sum Y_iX_{i1}\\ \sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{i2}=\sum Y_iX_{i2}\\ \cdots\cdots\cdots\cdots\\ \sum(\hat{\beta}_0+\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})X_{ik}=\sum Y_iX_{ik}\end{cases} \tag{3.2.3}$$

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

(3.2.3)式的矩阵形式如下：

$$\begin{pmatrix} n & \sum X_{i1} & \cdots & \sum X_{ik}\\ \sum X_{i1} & \sum X_{i1}^2 & \cdots & \sum X_{i1}X_{ik}\\ \vdots & \vdots & & \vdots\\ \sum X_{ik} & \sum X_{ik}X_{i1} & \cdots & \sum X_{ik}^2\end{pmatrix}\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots\\ \hat{\beta}_k\end{pmatrix}=\begin{pmatrix}1 & 1 & \cdots & 1\\ X_{11} & X_{21} & \cdots & X_{n1}\\ \vdots & \vdots & & \vdots\\ X_{1k} & X_{2k} & \cdots & X_{nk}\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}$$

即

$$(\boldsymbol{X'X})\hat{\boldsymbol{\beta}}=\boldsymbol{X'Y} \tag{3.2.4}$$

由 $\boldsymbol{X}$ 的列满秩性可得 $\boldsymbol{X'X}$ 为满秩对称矩阵，故有

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \tag{3.2.5}$$

将上述过程用矩阵表示如下。

根据最小二乘原理，需寻找一组参数估计值 $\hat{\boldsymbol{\beta}}$，使得残差平方和

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

最小，即参数估计值应该是方程组

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

的解。求解过程如下：

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

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

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

即得到

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

于是，参数的最小二乘估计值为

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

例 3.2.1

在例 2.1.1 的家庭可支配收入-消费支出例中，

$$(\boldsymbol{X'X})=\begin{pmatrix}1 & 1 & \cdots & 1\\ X_1 & X_2 & \cdots & X_n\end{pmatrix}\begin{pmatrix}1 & X_1\\ 1 & X_2\\ \vdots & \vdots\\ 1 & X_n\end{pmatrix}$$

$$=\begin{pmatrix}n & \sum X_i\\ \sum X_i & \sum X_i^2\end{pmatrix}=\begin{pmatrix}10 & 21\,500\\ 21\,500 & 53\,650\,000\end{pmatrix}$$

$$\boldsymbol{X'Y}=\begin{pmatrix}1 & 1 & \cdots & 1\\ X_1 & X_2 & \cdots & X_n\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}$$

$$=\begin{pmatrix}\sum Y_i\\ \sum X_iY_i\end{pmatrix}=\begin{pmatrix}15\,829\\ 39\,007\,100\end{pmatrix}$$

可求得

$$(\boldsymbol{X'X})^{-1}=\begin{pmatrix}0.722\,6 & -0.000\,3\\ -0.000\,3 & 1.35\times10^{-7}\end{pmatrix}$$

于是

$$\hat{\boldsymbol{\beta}}=\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}0.722\,6 & -0.000\,3\\ -0.000\,3 & 1.35\times10^{-7}\end{pmatrix}\begin{pmatrix}15\,829\\ 39\,007\,100\end{pmatrix}=\begin{pmatrix}-142.4\\ 0.67\end{pmatrix}$$

---

2. 离差形式的普通最小二乘估计

对于正规方程组(3.2.3)式的矩阵形式

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

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

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

于是

$$\boldsymbol{X'e}=\boldsymbol{0} \tag{3.2.6}$$

或

$$\begin{cases}\sum e_i=0\\ \sum\limits_i X_{ij}e_i=0\,, \qquad j=1,2,\cdots,k\end{cases}$$

(3.2.6)式是多元线性回归模型正规方程组的另一种写法，由此容易得到多元回

归分析中的**样本回归模型的离差形式**：

$$y_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} + e_i , \qquad i=1,2, \cdots,n \tag{3.2.7}$$

其矩阵形式为

$$\boldsymbol{y} = \boldsymbol{x}\hat{\boldsymbol{\beta}} + \boldsymbol{e} \tag{3.2.8}$$

其中

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{i1} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

于是容易推出，离差形式下参数的最小二乘估计结果：

$$\begin{cases} \hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \cdots - \hat{\beta}_k \overline{X}_k \end{cases} \tag{3.2.9}$$

3. 随机干扰项μ的方差的普通最小二乘估计

可以证明(参见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)，在普通最小二乘法下，随机干扰项μ的方差的无偏估计量为

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1} \tag{3.2.10}$$

二、最大似然估计

对于多元线性回归模型(3.1.3)式，由于

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

所以

$$Y_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta},\sigma^2)$$

其中

$$\boldsymbol{X}_i = (1 \quad X_{i1} \quad X_{i2} \quad \cdots \quad X_{ik})$$

Y 的随机抽取的 *n* 组样本观测值的联合概率为

$$\begin{aligned} L(\boldsymbol{\beta},\sigma^2) &= \mathrm{P}(Y_1,Y_2,\cdots,Y_n) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum[Y_i-(\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik})]^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})} \end{aligned} \tag{3.2.11}$$

这就是变量$\boldsymbol{Y}$的似然函数。对数似然函数为

$$L^{*}=\ln L=-n\ln(\sqrt{2\pi}\sigma)-\frac{1}{2\sigma^{2}}(\boldsymbol{Y}-\boldsymbol{X\beta})'\ (\boldsymbol{Y}-\boldsymbol{X\beta}) \tag{3.2.12}$$

对似然函数求极大值，即对对数似然函数求极大值，也就是对

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

求极小值，就可以得到一组参数估计量$\hat{\boldsymbol{\beta}}$，即为参数的最大似然估计

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \tag{3.2.13}$$

显然，其结果与参数的普通最小二乘估计是相同的。

与一元回归相仿，容易得出多元线性回归下随机干扰项方差的估计如下：

$$\hat{\sigma}^{2}=\frac{(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{n}=\frac{\sum e_{i}^{2}}{n} \tag{3.2.14}$$

*三、矩估计

普通最小二乘估计是通过得到一个关于参数估计值的**正规方程组**并对它进行求解而完成的。正规方程组 (3.2.3)或(3.2.4)可以从矩估计的思路来导出。

矩估计的基本思想是寻找一组总体矩条件，并通过对应的样本矩条件来推导出未知参数的解。对多元线性总体回归模型(3.1.4)，存在如下一组矩条件：

$$\mathrm{E}(\boldsymbol{X}'\boldsymbol{\mu})=\boldsymbol{0} \tag{3.2.15}$$

于是，对应的样本矩条件可写为

$$\frac{1}{n}\boldsymbol{X}'(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})=\boldsymbol{0} \tag{3.2.16}$$

由此得到正规方程组(3.2.4)：

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

解此正规方程组即得样本估计参数$\hat{\boldsymbol{\beta}}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$。

可见矩估计的结果与普通最小二乘法以及最大似然估计法的结果一致。

值得一提的是，**矩估计法是工具变量方法**(Instrumental Variable, IV)和**广义矩估计法**(Generalized Moment Method, GMM)的基础。在矩估计法中关键是利用了基本假设

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

作为总体矩条件。**如果某个解释变量与随机项相关，只要能找到 1 个工具变量，仍然可以构成一组矩条件，这就是工具变量法。如果存在大于 k+1 个变量与随机项不相关，可以构成一组包含大于 k+1 个方程的矩条件，这就是广义矩估计法。**这些将在计量经济学高级课程中介绍。

四、参数估计量的统计性质

当多元线性回归模型满足基本假设时，其参数的普通最小二乘估计、最大似然估计及矩估计仍具有线性性、无偏性和有效性。同时，随着样本容量增加，即当 $n\to+\infty$ 时，参数估计量具有渐近无偏性、一致性及渐近有效性。

1. 线性性

由于

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

其中 $\boldsymbol{C}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 仅与固定的 $\boldsymbol{X}$ 有关。可见，参数估计量是被解释变量 $\boldsymbol{Y}$ 的线性组合。

2. 无偏性

参数估计量 $\hat{\boldsymbol{\beta}}$ 的无偏性证明如下：

$$\begin{aligned}\mathrm{E}(\hat{\boldsymbol{\beta}})&=\mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}]\\&=\mathrm{E}\,[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\mu})]\\&=\boldsymbol{\beta}+(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{\mu})\\&=\boldsymbol{\beta}\end{aligned}\tag{3.2.17}$$

这里利用了随机干扰项零均值的假设 $E(\boldsymbol{\mu})=\boldsymbol{0}$。

3. 有效性

首先给出参数估计量 $\hat{\boldsymbol{\beta}}$ 的**方差-协方差矩阵**：

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}})&=\mathrm{E}\{[\hat{\boldsymbol{\beta}}-\mathrm{E}(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}}-\mathrm{E}(\hat{\boldsymbol{\beta}})]'\}\\&=\mathrm{E}\,[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})']\\&=\mathrm{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]\\&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\&=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\\&=\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}\tag{3.2.18}$$

其中利用了

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

和

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

$\boldsymbol{I}$ 为单位矩阵。

根据高斯-马尔可夫定理，(3.2.18)式表示的方差在所有无偏估计量的方差中是最小的，所以该参数估计量具有有效性(证明见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)。

五、样本容量问题

模型参数估计是在样本观测值的支持下完成的。计量经济学模型，说到底是从表现已经发生的经济活动的样本数据中寻找经济活动中蕴涵的规律性，所以，它对样本数据具有很强的依赖性。收集与整理样本数据是一件非常困难的工作，于是，怎样选择合适的样本容量，使其既能满足建模的需要，又能减轻收集数据的困难，就成为一个重要的实际问题。

从建模需要来讲，样本容量越大越好，这是显而易见的。这里需要讨论的是最小样本容量和满足基本要求的样本容量。

1. 最小样本容量

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

从参数估计量

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

中可以看到，欲使$\hat{\boldsymbol{\beta}}$存在，必须使得$(\boldsymbol{X}'\boldsymbol{X})^{-1}$存在。为使得$(\boldsymbol{X}'\boldsymbol{X})^{-1}$存在，必须满足

$$|\boldsymbol{X}'\boldsymbol{X}| \neq 0$$

即矩阵$\boldsymbol{X}'\boldsymbol{X}$为$k+1$阶满秩矩阵。矩阵乘积的秩不超过各个因子矩阵的秩，即

$$R(\boldsymbol{AB}) \leqslant \min[R(\boldsymbol{A}), R(\boldsymbol{B})]$$

其中符号R表示矩阵的秩。所以，只有当

$$R(\boldsymbol{X}) \geqslant k+1$$

时，矩阵$\boldsymbol{X}'\boldsymbol{X}$才为$k+1$阶满秩矩阵。而$\boldsymbol{X}$为$n\times(k+1)$阶矩阵，其秩最大为$k+1$，此时必须有

$$n \geqslant k+1$$

即样本容量必须不少于模型中解释变量的数目(包括常数项)，这就是最小样本容量。

2. 满足基本要求的样本容量

虽然当$n \geqslant k+1$时可以得到参数估计量，但在n太小时，除了参数估计量质量不好外，一些建立模型所必需的后续工作也无法进行。例如，参数的统计检验要求样本容量必须足够大，z检验在$n<30$时不能应用；t检验为检验变量显著性的最常用方法，经验表明，当$n-k \geqslant 8$时t分布较为稳定，检验才较为有效。所以，**一般经验认为，当$n \geqslant 30$或者至少$n \geqslant 3(k+1)$时，才能说满足模型估计的基本要求。**

如果出现样本容量较小，甚至少于“最小样本容量”，那么只依靠样本信

息是无法完成模型估计的。这时需要引入非样本信息，如先验信息和后验信息，并采用其他估计方法，如贝叶斯(Bayes)估计方法，才能完成模型的参数估计。

六、多元线性回归模型的参数估计实例

例 3.2.2

在例 2.6.1 中，我们通过截面数据已建立了 2006 年中国内地城镇居民家庭全年人均消费支出的一元线性模型。这里我们再考虑建立多元线性模型。首先，城镇居民家庭全年人均可支配收入(X_1)仍是其消费支出(Y)的重要解释变量；另外，居民消费水平具有一定的惯性，也就是说，居民当年的消费支出在一定程度上受上一年已经实现了的消费支出的影响，因此，模型中可再引入前一年，即 2005 年内地城镇居民人均消费支出(X_2)作为另一解释变量。样本观测值见表 3.2.1。

表 3.2.1　中国内地各地区城镇居民家庭人均全年可支配收入与人均全年消费性支出(元)

地区	2006年消费支出Y	2006年可支配收入X_1	2005年消费支出X_2	地区	2006年消费支出Y	2006年可支配收入X_1	2005年消费支出X_2
北京	14 825.4	19 977.5	13 244.2	湖北	7 397.3	9 802.7	6 736.6
天津	10 548.1	14 283.1	9 653.3	湖南	8 169.3	10 504.7	7 505.0
河北	7 343.5	10 304.6	6 699.7	广东	12 432.2	16 015.6	11 809.9
山西	7 170.9	10 027.7	6 342.6	广西	6 792.0	9 898.8	7 032.8
内蒙古	7 666.6	10 358.0	6 928.6	海南	7 126.8	9 395.1	5 928.8
辽宁	7 987.5	10 369.6	7 369.3	重庆	9 398.7	11 569.7	8 623.3
吉林	7 352.6	9 775.1	6 794.7	四川	7 524.8	9 350.1	6 891.3
黑龙江	6 655.4	9 182.3	6 178.0	贵州	6 848.4	9 116.6	6 159.3
上海	14 761.8	20 667.9	13 773.4	云南	7 379.8	10 069.9	6 996.9
江苏	9 628.6	14 084.3	8 621.8	西藏	6 192.6	8 941.1	8 617.1
浙江	13 348.5	18 265.1	12 253.7	陕西	7 553.3	9 267.7	6 656.5
安徽	7 294.7	9 771.1	6 367.7	甘肃	6 974.2	8 920.6	6 529.2
福建	9 807.7	13 753.3	8 794.4	青海	6 530.1	9 000.4	6 245.3
江西	6 645.5	9 551.1	6 109.4	宁夏	7 205.6	9 177.3	6 404.3
山东	8 468.4	12 192.2	7 457.3	新疆	6 730.0	8 871.3	6 207.5
河南	6 685.2	9 810.3	6 038.0				

资料来源：根据《中国统计年鉴》(2006,2007)整理。

Eviews 软件估计结果如表 3.2.2 所示。两个解释变量前的参数估计值分别为 0.555 6 和 0.250 1，都为正数，且都处于 0 与 1 之间，常数项的估计值

PDG

也为正，这些参数估计值的经济含义是合理的。随机误差项的方差的估计值为

$$\hat{\sigma}^2=\frac{4\,170\,093}{31-3}=148\,931.9\text{。}$$

表 3.2.2 中国内地城镇居民人均消费支出二元回归估计

Dependent Variable:Y
Method:Least Squares
Included observations:31

Variable	Coefficient	Std.Error	t-Statistic	Prob
C	143.3265	260.4032	0.550402	0.5864
X1	0.555644	0.075308	7 378320	0 0000
X2	0.250085	0.113634	2.200791	0.0362
R-squared	0.975634	Mean dependent var	8401.468	
Adjusted R-squared	0.973893	S.D. dependent var	2388.459	
S.E. of regression	385.9169	Akaike info criterion	14.84089	
Sum squared resid	4170093	Schwarz criterion	14.97966	
Log likelihood	–227.0337	f-statistic	560.5650	
Durbin-Watson stat	1.843488	Prob(F-statistic)	0.000000	

§3.3 多元线性回归模型的统计检验

多元线性回归模型的参数估计出来后，即求出样本回归函数后，还需进一步对该样本回归函数进行统计检验，以判定估计的可靠程度，包括拟合优度检验、方程总体线性的显著性检验、变量的显著性检验，以及参数的置信区间估计等方面。

一、拟合优度检验

1. 可决系数与调整的可决系数

在一元线性回归模型中，使用可决系数 R^2 来衡量样本回归线对样本观测值的拟合程度。在多元线性回归模型中，也可用该统计量来衡量样本回归线对样本观测值的拟合程度。

记 $\mathrm{TSS}=\sum(Y_i-\bar{Y})^2$ 为总离差平方和，$\mathrm{ESS}=\sum(\hat{Y}_i-\bar{Y})^2$ 为回归平方和，$\mathrm{RSS}=\sum(Y_i-\hat{Y}_i)^2$ 为残差平方和，则

$$\begin{aligned}\mathrm{TSS}&=\sum(Y_i-\bar{Y})^2\\&=\sum[(Y_i-\hat{Y}_i)+(\hat{Y}_i-\bar{Y})]^2\\&=\sum(Y_i-\hat{Y}_i)^2+2\sum(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})+\sum(\hat{Y}_i-\bar{Y})^2\end{aligned}$$

由于

$$\begin{aligned}
&\sum(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})\\
&=\sum e_i(\hat{Y}_i-\bar{Y})\\
&=\hat{\beta}_0\sum e_i+\hat{\beta}_1\sum e_iX_{1i}+\cdots+\hat{\beta}_k\sum e_iX_{ki}+\bar{Y}\sum e_i\\
&=0
\end{aligned}$$

所以有

$$\mathrm{TSS}=\sum(Y_i-\hat{Y}_i)^2+\sum(\hat{Y}_i-\bar{Y})^2=\mathrm{RSS}+\mathrm{ESS} \tag{3.3.1}$$

即总离差平方和可分解为回归平方和与残差平方和两部分。回归平方和反映了总离差平方和中可由样本回归线解释的部分，它越大，残差平方和越小，表明样本回归线与样本观测值的拟合程度越高。因此，可用回归平方和占总离差平方和的比重来衡量样本回归线对样本观测值的拟合程度：

$$R^2=\frac{\mathrm{ESS}}{\mathrm{TSS}}=1-\frac{\mathrm{RSS}}{\mathrm{TSS}} \tag{3.3.2}$$

该统计量越接近于 1，模型的拟合优度越高。

在应用过程中发现，如果在模型中增加一个解释变量，R^2往往增大。这是因为残差平方和往往随着解释变量个数的增加而减少，至少不会增加。这就给人一个错觉：要使模型拟合得好，只要增加解释变量即可。但是，现实情况往往是，由增加解释变量个数引起的R^2的增大与拟合好坏无关，因此在多元回归模型之间比较拟合优度，R^2就不是一个合适的指标，必须加以调整。

在样本容量一定的情况下，增加解释变量必定使得自由度减少，所以调整的思路是将残差平方和与总离差平方和分别除以各自的自由度，以剔除变量个数对拟合优度的影响。记$\bar{R}^2$为**调整的可决系数**(adjusted coefficient of determination)，则有

$$\bar{R}^2=1-\frac{\mathrm{RSS}/(n-k-1)}{\mathrm{TSS}/(n-1)} \tag{3.3.3}$$

其中$n-k-1$为残差平方和的自由度，$n-1$为总离差平方和的自由度。显然，如果增加的解释变量没有解释能力，则对残差平方和 RSS 的减小没有多大帮助，但增加了待估参数的个数，从而使$\bar{R}^2$有较大幅度的下降。

调整的可决系数与未经调整的可决系数之间存在如下关系：

$$\bar{R}^2=1-(1-R^2)\frac{n-1}{n-k-1} \tag{3.3.4}$$

在实际应用中，$\bar{R}^2$达到多大才算模型通过了检验？没有绝对的标准，要看具体情况而定。模型的拟合优度并不是判断模型质量的唯一标准，有时甚至为

了追求模型的经济意义，可以牺牲一点拟合优度。在下一部分中，将推导出 $\bar{R}^2$ 与另一个统计量 F 的关系，那时会对 $\bar{R}^2$ 有新的认识。

在例 3.2.2 中，$\bar{R}^2=0.975\,6$，比例 2.6.1 中的 $R^2=0.971\,4$ 大，这应该说是很好的拟合结果了。

*2. 赤池信息准则和施瓦茨准则

为了比较所含解释变量个数不同的多元回归模型的拟合优度，常用的标准还有**赤池信息准则**(Akaike Information Criterion, AIC)和**施瓦茨准则**(Schwarz Criterion，SC)，其定义分别为

$$\mathrm{AIC}=\ln\frac{\boldsymbol{e}'\boldsymbol{e}}{n}+\frac{2(k+1)}{n} \tag{3.3.5}$$

$$\mathrm{SC}=\ln\frac{\boldsymbol{e}'\boldsymbol{e}}{n}+\frac{k}{n}\ln n \tag{3.3.6}$$

这两个准则均要求仅当所增加的解释变量能够减少 AIC 值或 SC 值时才在原模型中增加该解释变量。显然，与调整的可决系数相仿，如果增加的解释变量没有解释能力，则对残差平方和 $\boldsymbol{e}'\boldsymbol{e}$ 的减小没有多大帮助，但增加了待估参数的个数，这时可能导致 AIC 或 SC 的值增加。

在例 3.2.2 中，EViews 软件的估计结果显示 AIC 值与 SC 值分别为 14.84 与 14.98，分别小于例 2.6.1 中只包含人均国内生产总值一个解释变量时的相应值 14.94 与 15.03。从这点看，可以说前期人均居民消费可作为解释变量包括在模型中。

二、方程总体线性的显著性检验(F 检验)

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

从上面的拟合优度检验中可以看出，拟合优度高，解释变量对被解释变量的解释程度就高，可以推测模型总体线性关系成立；反之，就不成立。但这只是一个模糊的推测，不能给出一个在统计上严格的结论。这就要求进行方程的显著性检验。方程的显著性检验所应用的方法仍是数理统计学中的假设检验。

1. 方程显著性的 F 检验

方程显著性的 F 检验是要检验模型

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_kX_{ik}+\mu_i\ , \qquad i=1,2,\cdots,n$$

中参数 $\beta_1,\cdots,\beta_k$ 是否显著不为零。按照假设检验的原理与程序，原假设与备择假设分别为

$$H_0:\beta_1=0,\beta_2=0,\cdots,\beta_k=0$$
$$H_1:\beta_j\,(\,j=1,2,\cdots,k\,)\text{不全为零}$$

F 检验的思想来自于总离差平方和的分解式

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

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

$$\frac{\text{ESS}}{\text{RSS}}=\frac{\sum\hat{y}_i^2}{\sum e_i^2}$$

如果这个比值较大，则 X 的联合体对 Y 的解释程度高，可认为总体存在线性关系；反之总体上可能不存在线性关系。因此可通过该比值的大小对总体线性关系进行推断。

根据数理统计学中的知识，在原假设 H_0 成立的条件下，统计量

$$F=\frac{\text{ESS}/k}{\text{RSS}/(n-k-1)} \tag{3.3.7}$$

服从自由度为 $(k,n-k-1)$ 的 F 分布。因此，给定显著性水平 α，查表得到临界值 $F_\alpha(k,n-k-1)$，根据样本求出 F 统计量的数值后，可通过

$$F > F_\alpha(k,n-k-1) \quad (\text{或}\, F \leqslant F_\alpha(k,n-k-1))$$

来拒绝(或接受)原假设 H_0，以判定原方程总体上的线性关系是否显著成立。

对于例 3.2.2，计算得到 F=560.57，给定显著性水平 α=0.05，查 F 分布表，得到临界值 $F_{0.05}(2,28)$=3.34(例子中解释变量数目为 2，样本容量为 31)，显然有

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

表明模型的线性关系在 95%的置信水平下显著成立。

2. 关于拟合优度检验与方程总体线性的显著性检验关系的讨论

拟合优度检验和方程总体线性的显著性检验是从不同原理出发的两类检验，前者是从已经得到估计的模型出发，检验它对样本观测值的拟合程度，后者是从样本观测值出发检验模型总体线性关系的显著性。但是二者又是关联的，模型对样本观测值的拟合程度高，模型总体线性关系的显著性就强。那么，找出两个用作检验标准的统计量之间的数量关系，在实际应用中互为验证，是有实际意义的。

用(3.3.3)式和(3.3.7)式分别表示的两个统计量之间存在下列关系：

$$\overline{R}^2=1-\frac{n-1}{n-k-1+kF} \tag{3.3.8}$$

$$F=\frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{3.3.9}$$

由(3.3.9)式可知 F 与 R^2 同向变化：当 $R^2=0$ 时，$F=1$；R^2 越大，F 值也越

大；当 $R^2=1$ 时，F 为无穷大。因此，F 检验是所估计回归的总显著性的一个度量，也是 R^2 的一个显著性检验，亦即，检验原假设 $H_0:\beta_1=0,\beta_2=0,\cdots,\beta_k=0$，等价于检验 $R^2=0$ 这一虚拟假设。

对于例 3.2.2，给定显著性水平 $\alpha=0.05$ 时，查 F 分布表，得到临界值 $F_{0.05}(2,28)=3.34$，即是说，只要 F 统计量的值大于 3.34，模型的线性关系在 95% 的置信度下是显著成立的。将该数值代入(3.3.8)式，计算得到对应的 $\bar{R}^2$ 为 0.135 4。如果我们首先得到 $\bar{R}^2$ 为 0.135 4，肯定认为该模型质量不高，殊不知它的总体线性关系的显著性水平达到 95%。这样，**在应用中不必对 $\bar{R}^2$ 过分苛求，重要的是需考察模型的经济关系是否合理。**

三、变量的显著性检验(t 检验)

对于多元线性回归模型，方程的总体线性关系是显著的，并不能说明每个解释变量对被解释变量的影响都是显著的，因此，必须对每个解释变量进行显著性检验，以决定是否作为解释变量被保留在模型中。如果某个变量对被解释变量的影响并不显著，应该将它剔除，以建立更为简单的模型。变量显著性检验中应用最为普遍的是 t 检验，在目前使用的计量经济学软件包中，都有关于 t 统计量的计算结果。

1. t 统计量

在上一节中，已经导出了参数估计量的方差为

$$\mathrm{Var}(\hat{\boldsymbol{\beta}})=\sigma^2(\boldsymbol{X'X})^{-1}$$

以 c_{jj} 表示矩阵 $(\boldsymbol{X'X})^{-1}$ 主对角线上的第 j 个元素，于是参数估计量 $\hat{\beta}_j$ 的方差为

$$\mathrm{Var}(\hat{\beta}_j)=\sigma^2 c_{jj}\,,\qquad j=1,2,\cdots,k \tag{3.3.10}$$

其中 σ^2 为随机干扰项的方差，在实际计算时，用它的估计量 $\hat{\sigma}^2$ 代替。这样，当模型参数估计完成后，就可以计算每个参数估计量的方差值。

因为 $\hat{\beta}_j$ 服从如下正态分布：

$$\hat{\beta}_j \sim N(\beta_j,\sigma^2 c_{jj})$$

因此，可构造如下 t 统计量：

$$t=\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}=\frac{\hat{\beta}_j-\beta_j}{\sqrt{c_{jj}\dfrac{e'e}{n-k-1}}}\sim t(n-k-1) \tag{3.3.11}$$

该统计量即为用于变量显著性检验的 t 统计量。

2. t 检验

在变量显著性检验中，针对某变量 $X_j\,(j=1,2,\cdots,k)$设计的原假设与备择假设为

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

给定一个显著性水平α，得到临界值$t_{\frac{\alpha}{2}}(n-k-1)$，于是可根据

$$|t| > t_{\frac{\alpha}{2}}(n-k-1) \qquad (\text{或}|t| \leqslant t_{\frac{\alpha}{2}}(n-k-1))$$

来决定拒绝(或接受)原假设H_0，从而判定对应的解释变量是否应包含在模型中。

需注意的是，**在一元线性回归中，*t* 检验与 *F* 检验是一致的。**

一方面，t检验与F检验都是对相同的原假设$H_0 : \beta_1 = 0$进行检验；另一方面，两个统计量之间有如下关系：

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

$$= \frac{\hat{\beta}_1^2}{\sum e_i^2 / (n-2) \sum x_i^2} = \left[\frac{\hat{\beta}_1}{\sqrt{\sum e_i^2 / (n-2) \sum x_i^2}} \right]^2$$

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

在例 3.2.2 中，已经由应用软件计算出两个变量X_1，X_2的t值，分别为

$$|t_1| = 7.378, \qquad |t_2| = 2.201$$

给定显著性水平α=0.05，查t分布表中自由度为 28(在这个例子中$n-k-1$=28)的相应临界值，得到$t_{\frac{\alpha}{2}}(28)$=2.048。可见，两变量的t值都大于该临界值，所以拒绝原假设，即是说，模型中引入的 2 个解释变量都在 95%的水平下影响显著，都通过了变量的显著性检验。

经常遇到一些实际问题，各个变量的t值相差较大，有的在很高的显著性水平下影响显著，有的则在不太高的显著性水平下影响显著，是否都认为通过显著性检验？**没有绝对的显著性水平。关键仍然是考察变量在经济关系上是否对解释变量有影响，显著性检验起到验证的作用；同时还要看显著性水平不太高的变量在模型中及模型应用中的作用，不要简单地剔除变量。**

四、参数的置信区间

参数的假设检验用来判别所考察的解释变量是否对被解释变量有显著的线性影响，但并未回答在一次抽样中，所估计的参数值离参数的真实值有多“近”。这需要进一步通过对参数的置信区间的估计来考察。

在变量显著性检验中已经知道

$$t=\frac{\hat{\beta}_j-\beta_j}{S_{\hat{\beta}_j}}\sim t(n-k-1)$$

容易推出：在$1-\alpha$的置信度下β_j的置信区间是

$$(\hat{\beta}_j-t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j},\hat{\beta}_j+t_{\frac{\alpha}{2}}\times S_{\hat{\beta}_j}) \tag{3.3.12}$$

其中，$t_{\frac{\alpha}{2}}$为t分布表中显著性水平为α，自由度为$n-k-1$的临界值。

在例 3.2.2 中，如果给定α=0.05，查表得

$$t_{\frac{\alpha}{2}}(n-k-1)=t_{0.025}(28)=2.048$$

从回归计算中得到

$$\hat{\beta}_1=0.555\,6,\qquad S_{\hat{\beta}_1}=0.075\,3$$

$$\hat{\beta}_2=0.250\,1,\qquad S_{\hat{\beta}_2}=0.113\,6$$

根据(3.3.12)式计算得到β_1和β_2的置信区间分别为(0.401 4, 0.709 8)和(0.017 4, 0.482 8)。显然，参数β_1的置信区间比β_2要小，这意味着在同样的置信度下，β_1的估计结果精度更高一些。

同样地，在实际应用中，我们希望置信度越高越好，置信区间越小越好。**如何才能缩小置信区间呢？**从(3.3.12)式中可以看出：(1)**增大样本容量*n***，在同样的置信度下，n越大，临界值$t_{\frac{\alpha}{2}}$越小，同时，增大样本容量，在一般情况下可使$S_{\hat{\beta}_j}=\sqrt{c_{jj}\dfrac{\boldsymbol{e'e}}{n-k-1}}$减小，因为式中分母的增大是肯定的，分子并不一定增大；(2)更主要的是**提高模型的拟合优度**，以减小残差平方和$\boldsymbol{e'e}$，设想一种极端情况，如果模型完全拟合样本观测值，残差平方和为 0，则置信区间的长度也为 0；(3)**提高样本观测值的分散度**，在一般情况下，样本观测值越分散，c_{jj}越小。

值得注意的是，置信度的高低与置信区间的大小存在此消彼长的关系。置信度越高，在其他情况不变时，临界值$t_{\frac{\alpha}{2}}$越大，置信区间越大。如果要求缩小置信区间，在其他情况不变时，就必须降低对置信度的要求。

§ 3.4　多元线性回归模型的预测

计量经济学模型的一个重要应用是经济预测。对于模型

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

如果给定样本以外的解释变量的观测值$\boldsymbol{X}_0=(1,X_{01},X_{02},\cdots,X_{0k})$，可以得到被

解释变量的预测值

$$\hat{Y}_0 = \boldsymbol{X}_0\,\hat{\boldsymbol{\beta}}$$

同样地，**严格地说，这只是被解释变量预测值的估计值，而不是预测值。**原因在于模型中参数估计量的不确定性及随机干扰项的影响两个方面。因此，我们得到的仅是预测值的一个估计值。为了进行科学预测，还需求出预测值的置信区间，包括均值 $\mathrm{E}(Y_0)$ 和点预测值 Y_0 的置信区间。

一、E(Y₀)的置信区间

从参数估计量性质的讨论中易知

$$\mathrm{E}(\hat{Y}_0) = \mathrm{E}(\boldsymbol{X}_0\,\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_0\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_0\boldsymbol{\beta} = \mathrm{E}(Y_0)$$

$$\mathrm{Var}(\hat{Y}_0) = \mathrm{E}[(\boldsymbol{X}_0\,\hat{\boldsymbol{\beta}} - \boldsymbol{X}_0\boldsymbol{\beta})^2] = \mathrm{E}[\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$$

由于 $\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ 为标量，因此

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

容易证明

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

取随机干扰项的样本估计量 $\hat{\sigma}^2$，可构造如下 t 统计量：

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

于是，得到 $1-\alpha$ 的置信度下 $\mathrm{E}(\boldsymbol{Y}_0)$ 的置信区间：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} < E(Y_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} \qquad (3.4.1)$$

二、Y₀ 的置信区间

如果已经知道实际的预测值 Y_0，那么预测误差为

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

容易证明

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

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

e_0 服从正态分布，即

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

取随机干扰项的样本估计量 $\hat{\sigma}^2$，可得 e_0 的方差的估计量

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

构造 t 统计量

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

可得给定 $1-\alpha$ 的置信水平下 Y_0 的置信区间：

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1+\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1+\boldsymbol{X}_0(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0'} \qquad (3.4.2)$$

例 3.2.2 中，假设某城镇居民家庭 2006 年人均可支配收入为 20 000 元，其 2005 年人均消费支出为 14 000 元，则该家庭 2006 年人均居民消费支出的预测值为

$$\hat{Y}=143.3+0.555\,6\times 20\,000+0.250\,1\times 14\,000=14\,757(\text{元})$$

而就全国平均情况看，2006 年具有人均可支配收入 20 000 元，前一年人均消费支出 14 000 元的家庭，当年平均的人均消费支出预测值的置信区间可如下求出。

在 95%的置信度下，临界值 $t_{0.025}(28)$=2.048，随机扰动项方差的估计值为 $\hat{\sigma}^2$=148 931.9，由于

$$\boldsymbol{X}_0 = (1,\quad 20\,000,\quad 14\,000)$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 0.455\,307\,6 & -0.000\,004\,5 & -0.000\,047\,9 \\ -0.000\,004\,5 & 0.000\,000\,0 & -0.000\,000\,1 \\ -0.000\,047\,9 & -0.000\,000\,1 & 0.000\,000\,1 \end{pmatrix}$$

$$\boldsymbol{X}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_0 = 0.308\,8$$

于是 $\mathrm{E}(\hat{Y})$ 的 95%的置信区间为

$$14\,757 \pm 2.048 \times \sqrt{148\,931.9} \times \sqrt{0.308\,8}$$

或 (14 318.2，15 196.6)

同样地，就人均可支配收入 20 000 元，前一年人均消费支出 14 000 元的某单个家庭来说，也易得其 2006 年人均消费支出 $\hat{Y}$ 的 95%的置信区间：

$$14\,757 \pm 2.048 \times \sqrt{148\,931.9} \times \sqrt{1.308\,8}$$

或 (13 853.1, 15 661.7)

PDG

需要指出的是，经常听到这样的说法："如果给定解释变量值，根据模型就可以得到被解释变量的预测值为……"，这种说法是不科学的，也是计量经济学模型无法达到的。如果一定要给出一个具体的预测值，那么它的置信度则为0；如果一定要回答以100%的置信度处在什么区间中，那么这个区间是$(-\infty,+\infty)$。

§3.5　可化为线性的多元非线性回归模型

迄今为止，我们都假设未知的总体回归线是线性的，拟合优度检验及变量显著性检验也都是对函数形式的线性检验。然而，在实际经济活动中，经济变量的关系是复杂的，直接表现为线性关系的情况并不多见。例如，著名的恩格尔曲线(Engle Curve)表现为幂函数曲线形式，宏观经济学中的菲利普斯曲线(Pillips Cuve)表现为双曲线形式等。但是，它们中的大部分又可以通过一些简单的数学处理，使之化为数学上的线性关系，从而可以运用线性回归的方法建立线性计量经济学模型。下面通过一些常见的例子说明常用的数学处理方法。

一、模型的类型与变换

1. 倒数模型、多项式模型与变量的直接置换法

例如，商品的需求曲线是一种双曲线形式，商品需求量Q与商品价格P之间的关系表现为非线性关系：

$$\frac{1}{Q}=a+b\frac{1}{P}+\mu \tag{3.5.1}$$

显然，可以用$Y=\dfrac{1}{Q}$和$X=\dfrac{1}{P}$的置换，将方程变成

$$Y=a+bX+\mu \tag{3.5.2}$$

再如，著名的拉弗曲线(Laffer Curve)描述的税收s和税率r的关系是一种抛物线形式：

$$s=a+br+cr^2+\mu\,,\qquad c<0 \tag{3.5.3}$$

可以用$X_1=r,X_2=r^2$进行置换，将方程变成

$$s=a+bX_1+cX_2+\mu\,,\qquad c<0 \tag{3.5.4}$$

一般地，关于解释变量的非线性问题都可以通过变量置换变成线性问题。

2. 幂函数模型、指数函数模型与函数变换法

如果是关于参数的非线性问题，变量置换方法就无能为力了，函数变换是常用的方法。

例如，著名的 Cobb–Dauglas 生产函数将产出量 Q 与投入要素 (K,L) 之间的关系描述为幂函数的形式：

$$Q = AK^{\alpha}L^{\beta}\mathrm{e}^{\mu} \tag{3.5.5}$$

方程两边取对数后，即成为一个线性形式：

$$\ln Q = \ln A + \alpha\ln K + \beta\ln L + \mu \tag{3.5.6}$$

再如，生产中成本 C 与产量 Q 的关系呈现指数关系：

$$C = ab^{Q}\mathrm{e}^{\mu} \tag{3.5.7}$$

方程两边取对数后，即成为一个线性形式：

$$\ln C = \ln a + Q\ln b + \mu \tag{3.5.8}$$

3. 复杂函数模型与级数展开法

例如，著名的 CES 生产函数将产出量 Q 与投入要素 (K,L) 之间的关系描述为如下的复杂函数形式：

$$Q = A(\delta_1K^{-\rho} + \delta_2L^{-\rho})^{-\frac{1}{\rho}}\mathrm{e}^{\mu}, \qquad (\delta_1 + \delta_2 = 1) \tag{3.5.9}$$

方程两边取对数后，得到

$$\ln Q = \ln A - \frac{1}{\rho}\ln(\delta_1K^{-\rho} + \delta_2L^{-\rho}) + \mu \tag{3.5.10}$$

将式中 $\ln(\delta_1K^{-\rho} + \delta_2L^{-\rho})$ 在 $\rho = 0$ 处展开泰勒(Taylor)级数，取关于 ρ 的线性项，即得到一个线性近似式。如取 0 阶、1 阶、2 阶项，可得

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

当然，并非所有的非线性函数形式都可以线性化。无法线性化模型的一般形式为

$$Y = f(X_1, X_2, \cdots, X_k) + \mu \tag{3.5.11}$$

其中 $f(X_1, X_2, \cdots, X_k)$ 为非线性函数。形如

$$Q = AK^{\alpha}L^{\beta} + \mu \tag{3.5.12}$$

的生产函数模型就无法线性化，需要采用非线性方法估计其参数。

二、可化为线性的非线性回归实例

例 3.5.1

建立中国城镇居民食品消费需求函数模型。根据需求理论，居民对食品的消费需求函数大致为

$$Q = f(X, P_1, P_0) \tag{3.5.13}$$

其中，Q 为居民对食品的需求量，X 为消费者的消费支出总额，P_1 为食品

PDG

价格指数，P_0 为居民消费价格总指数。引入居民消费价格总指数 P_0 的原因主要在于研究居民其他消费对食品的替代性。需求理论同时指出，上述需求函数应具有零阶齐次性，即当所有商品价格和消费者货币支出总额按同一比例变动时，需求量保持不变，这就是所谓的消费者无货币幻觉。按照需求函数的这一特征，(3.5.13)式可写为

$$Q = f(\frac{X}{P_0}, \frac{P_1}{P_0}) \tag{3.5.14}$$

(3.5.14)式表明，居民对食品的消费需求，取决于居民的实际消费总支出 $\frac{X}{P_0}$ 以及食品的相对价格 $\frac{P_1}{P_0}$。显然，该式具有零阶齐次性。

为了进行比较，我们将同时估计(3.5.13)式与(3.5.14)式。首先确定具体的函数形式。根据恩格尔定律，随着居民消费支出的增加，居民对食品的消费支出也增加，但食品消费支出比例会逐渐下降。因此，居民对食品的消费支出与居民的总支出间呈幂函数的变化关系。同时，为了方便考察需求的价格弹性等相关问题，将(3.5.13)式具体写为

$$Q = AX^{\beta_1} P_1^{\beta_2} P_0^{\beta_3} \tag{3.5.15}$$

经对数变换，(3.5.15)式可用如下双对数线性回归模型进行估计：

$$\ln Q = \beta_0 + \beta_1 \ln X + \beta_2 \ln P_1 + \beta_3 \ln P_0 + \mu \tag{3.5.16}$$

式中，$\beta_0 = \ln A$。同样地，(3.5.14)式可用如下线性回归模型进行估计：

$$\ln Q = \beta_0 + \beta_1 \ln \frac{X}{P_0} + \beta_2 \ln \frac{P_1}{P_0} + \mu \tag{3.5.17}$$

采用双对数线性回归模型，能够方便地考察需求函数中零阶齐次性的特征。显然，对(3.5.16)式施加 $\beta_1 + \beta_2 + \beta_3 = 0$ 的约束，即可化为(3.5.17)式。因此，对(3.5.17)式进行回归，就意味着原需求函数满足零阶齐次性条件。

表 3.5.1 列出了用当年价测度的中国城镇居民人均消费支出(X)与人均食品消费支出(X_1)，表中 GP 表示中国城镇居民消费价格总指数。由于在 1995 年前没有城镇居民的食品消费价格指数，我们选取城镇食品零售价格指数(FP)作为城镇居民食品消费价格指数的近似替代。由这些数据容易推算出以 2000 年价测度的城镇居民人均食品消费支出(Q)，以及城镇居民消费价格缩减指数(P_0)与城镇居民食品消费价格缩减指数(P_1)。

表 3.5.1　中国城镇居民消费支出及价格指数　　单位：元

	X (当年价)	X_1 (当年价)	GP (上年=100)	FP (上年=100)	Q (2000年价)	P_0 (2000年=100)	P_1 (2000年=100)
1985	673.2	351.4	111.9	116.5	1 315.9	28.1	26.7
1986	799.0	418.9	107.0	107.2	1 463.3	30.1	28.6
1987	884.4	472.9	108.8	112.0	1 475.0	32.8	32.1
1988	1 104.0	567.0	120.7	125.2	1 412.5	39.5	40.1
1989	1 211.0	660.0	116.3	114.4	1 437.2	46.0	45.9
1990	1 278.9	693.8	101.3	98.8	1 529.2	46.6	45.4
1991	1 453.8	782.5	105.1	105.4	1 636.3	49.0	47.8
1992	1 671.7	884.8	108.6	110.7	1 671.4	53.2	52.9
1993	2 110.8	1 058.2	116.1	116.5	1 715.9	61.7	61.7
1994	2 851.3	1 422.5	125.0	134.2	1 718.7	77.2	82.8
1995	3 537.6	1 771.9	116.8	123.6	1 732.1	90.1	102.3
1996	3 919.5	1 904.7	108.8	107.9	1 725.6	98.1	110.4
1997	4 185.6	1 942.6	103.1	100.1	1 758.2	101.1	110.5
1998	4 331.6	1 926.9	99.4	96.9	1 799.8	100.5	107.1
1999	4 615.9	1 932.1	98.7	95.7	1 885.7	99.2	102.5
2000	4 998.0	1 971.3	100.8	97.6	1 971.3	100.0	100.0
2001	5 309.0	2 027.9	100.7	100.7	2 013.8	100.7	100.7
2002	6 029.9	2 271.8	99.0	99.9	2 258.3	99.7	100.6
2003	6 510.9	2 416.9	100.9	103.4	2 323.5	100.6	104.0
2004	7 182.1	2 709.6	103.3	109.9	2 370.2	103.9	114.3
2005	7 942.9	2 914.4	101.6	103.1	2 472.7	105.6	117.9
2006	8 696.6	3 111.9	101.5	102.6	2 573.4	107.2	120.9

资料来源：《中国统计年鉴》(1990—2007)

按(3.5.16)式回归，Eviews 软件的输出结果如表 3.5.2 所示。下面的(3.5.18)式给出了通常的报告式。

$$\ln\hat{Q} = 5.53 + 0.540\ln X - 0.258\ln P_1 - 0.288\ln P_0 \qquad (3.5.18)$$

$$(59.4) \quad (14.78) \quad (-1.45) \quad (-1.41)$$

$$R^2 = 0.977\,3 \qquad \bar{R}^2 = 0.973\,5 \qquad F = 258.84$$

表 3.5.2　中国城镇居民人均食品消费需求函数

```
Dependent Variable:LOG(Q)
Sample:1985 2006
Included observations:22

    Variable        Coefficient       Std.Error       t-Statistic     Prob.

       C             5.531950         0.093107        59.41489       0.0000
     LOG(X)          0.539917         0.036530        14.78015       0.0000
```

PDG

续表

LOG(P1)	−0.258012	0.178186	−1.447994	0.1648
LOG(P0)	−0.288561	0.205184	−1.406350	0.1766
R-squared	0.977345	Mean dependent var		7.493909
Adjusted R-squared	0.973569	S.D. dependent var		0.193147
S.E. of regression	0.031401	Akaike info criterion		−3.921001
Sum squared resid	0.017748	Schwarz criterion		−3.722630
Log likelihood	47.13101	F−statistic		258.8448
Durbin-Watson stat	0.696202	Prob(F−statistic)		0.000000

回归结果表明，在 1985—2006 年间，$\ln Q$ 变化的 97.7%可由其他三个变量的变化来解释。在 5%的显著性水平下，F 统计量的临界值为 $F_{0.05}(3,18)=3.16$，表明模型的线性关系显著成立。自由度 $n-k-1=18$ 的 t 统计量的临界值为 $t_{0.025}(18)=2.10$，因此 $\ln X$ 的参数显著地异于零，却不拒绝 $\ln P_1$ 与 $\ln P_0$ 前参数为零的假设，但这是否就意味着模型中应去掉价格因素呢？由于 $\ln P_1$ 与 $\ln P_0$ 的相关系数高达 0.997，表明两者间有较强的共线性，因此价格的影响或许存在，只不过该模型无法区分两者对被解释变量独立的影响。下节将进一步讨论两个价格因素是否可从模型中去掉，而解释变量间多重共线性的详细讨论将在第四章进行。

从 $\ln X$ 前的参数看，在 1985—2006 年间，中国城镇居民对食品的消费支出关于总消费支出的弹性为 0.54，表明中国城镇居民总的消费支出增加 1%，对食品消费支出平均增加 0.54%；虽然食品价格与物价总指数的变化对食品的消费需求的影响并不显著，但正如所期望的那样，在其他因素保持不变的情况下，食品价格的增加会减小城镇居民对食品的消费需求；同样地，$\ln P_0$ 前的参数为负，表明在以名义价格表示的居民消费总支出不变的情况下，居民消费价格总水平的上升会导致实际的居民消费总支出水平下降，所有的消费支出都会减少，其中包括对食品消费支出的减少。当然，其他物品与服务价格的上升，会一定程度地促使居民更多地消费食品。因此，$\ln P_0$ 前的符号主要得看这两种趋势对比的结果，这里显然是前者的力量超过了后者。

各变量的弹性和 $\hat{\beta}_1+\hat{\beta}_2+\hat{\beta}_3=-0.006$，很接近于零，但不为零。在下一节我们将进一步从统计学的意义上考察，看它是否为零，即估计的需求函数是否满足零阶齐次性特性。

按(3.5.17)式回归，Eviews 软件的输出结果如表 3.5.3 所示。

表 3.5.3 中国城镇居民人均食品消费需求函数

Dependent Variable:LOG(Q)
Sample:1985 2006
Included observations:22

Variable	Coefficient	Std.Error	t–Statistic	Prob.
C	5.524569	0.083108	66.47481	0.0000
LOG(X/P0)	0.534439	0.023198	23.03776	0.0000
LOG(P1/P0)	–0.275347	0.151143	–1.821763	0.0843

R–squared	0.977296	Mean dependent var	7.493909
Adjusted R–squared	0.974906	S.D. dependent var	0.193147
S.E. of regression	0.030596	Akaike info criterion	–4.009741
Sum squared resid	0.017787	Schwarz criterion	–3.860963
Log likelihood	47.10715	F–statistic	408.9291
Durbin–Watson stat	0.695256	Prob(F–statistic)	0.000000

$$\ln\hat{Q} = \underset{(66.47)}{5.52} + \underset{(23.03)}{0.534}\ln\frac{X}{P_0} - \underset{(-1.82)}{0.275}\ln\frac{P_1}{P_0} \qquad (3.5.19)$$

$$R^2=0.9773 \qquad \bar{R}^2=0.9749 \qquad F=408.9$$

模型拟合度较高，$\ln\dfrac{X}{P_0}$ 在 5%的显著性水平下显著，$\ln\dfrac{P_1}{P_0}$ 在 10%的显著性水平下显著。同样地，此期间中国城镇居民收入与消费支出总额的增加，会刺激对食品消费需求的增加，而食品相对价格的上升，对食品消费需求则起着抑制作用。

为了与(3.5.18)式作比较，将(3.5.19)式改写为

$$\begin{aligned}\ln\hat{Q} &= 5.52 + 0.534(\ln X - \ln P_0) - 0.275(\ln P_1 - \ln P_0) \\ &= 5.52 + 0.534\ln X - 0.275\ln P_1 - 0.259\ln P_0\end{aligned} \qquad (3.5.20)$$

可看出(3.5.20)式与(3.5.18)式比较接近，这意味着(3.5.18)式各变量的弹性和可能为零，即所建立的食品需求函数满足零阶齐次性特征。

*三、非线性普通最小二乘法

可化为线性的多元回归模型也可以直接采用非线性普通最小二乘法或者非线性最大似然法估计。下面简单介绍非线性普通最小二乘法的原理。

1. 普通最小二乘原理

将可化为线性的多元回归模型的一般形式表示为

$$Y = f(X_1, X_2, \cdots, X_k, \beta_0, \beta_1, \cdots, \beta_k) + \mu \qquad (3.5.21)$$

模型(3.5.21)中，如果随机误差项服从零均值、同方差的正态分布，且不序列相关，则可以从普通最小二乘原理出发，构造模型的估计方法。为了简单，

PDG

下面只对有一个参数的非线性模型进行讨论。

对于只有一个参数的非线性模型，在有 n 组观测值的情况下，(3.5.21)式可写成：

$$Y_i = f(X_i, \beta) + \mu_i \qquad i = 1, 2, \cdots, n \tag{3.5.22}$$

如果参数估计值已经得到，则应使得残差平方和最小，即

$$Q(\hat{\beta}) = \sum_{i=1}^{n} (Y_i - f(X_i, \hat{\beta}))^2 \tag{3.5.23}$$

最小。(3.5.23)式取极小值的一阶条件为

$$\frac{\mathrm{d}Q}{\mathrm{d}\hat{\beta}} = -2\sum_{i=1}^{n} (Y_i - f(X_i, \hat{\beta}) \left(\frac{-\mathrm{d}f(X_i, \hat{\beta})}{\mathrm{d}\hat{\beta}} \right) = 0$$

即

$$\sum_{i=1}^{n} (Y_i - f(X_i, \hat{\beta}) (\frac{\mathrm{d}f(X_i, \hat{\beta})}{\mathrm{d}\hat{\beta}}) = 0 \tag{3.5.24}$$

现在的问题在于如何求解非线性方程(3.5.24)。

对于多参数非线性模型，用矩阵形式表示(3.5.21)式为

$$\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{\mu} \tag{3.5.25}$$

其中各个符号的意义与线性模型相同。向量 $\boldsymbol{\beta}$ 的普通最小平方估计值 $\hat{\boldsymbol{\beta}}$ 应该使得残差平方和

$$Q(\hat{\boldsymbol{\beta}}) = (\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}))'(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}}))$$

达到最小值，即 $\hat{\boldsymbol{\beta}}$ 应该满足下列条件：

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(Q(\hat{\boldsymbol{\beta}})) = -2\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})) = \boldsymbol{0}$$

即

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')(\boldsymbol{Y} - f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})) = \boldsymbol{0} \tag{3.5.26}$$

其中 $\frac{\partial}{\partial \hat{\boldsymbol{\beta}}}(f(\boldsymbol{X}, \hat{\boldsymbol{\beta}})')$ 是一个 $k \times n$ 阶偏微分矩阵，其第 (j,i) 个元素为 $\frac{\partial}{\partial \hat{\beta}_j}(f(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}})')$。求解(3.5.26)式的原理和方法与求解(3.5.24)式相同，只是数学描述更为复杂。在下面关于求解方法的讨论中，我们只以(3.5.24)式为例，即以单参数非线性模型为例。

2．高斯-牛顿(Gauss-Newton)迭代法

对于非线性方程组(3.5.24)，直接解法已不适用，只能采用迭代解法，高斯-牛顿迭代法就是较为实用的一种。

(1)高斯-牛顿迭代法的原理

迭代是从(3.5.23)式出发的。

根据经验给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，将(3.5.23)式中的 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取一阶近似值，即有

$$f(X_i,\hat{\beta}) \approx f(X_i,\hat{\beta}_{(0)}) + \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\bigg|_{\hat{\beta}_{(0)}} (\hat{\beta}-\hat{\beta}_{(0)}) \tag{3.5.27}$$

令

$$Z_i(\hat{\beta}) = \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}$$

于是

$$Z_i(\hat{\beta}_{(0)}) = \frac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\bigg|_{\hat{\beta}_{(0)}}$$

代入(3.5.23)式，得

$$\begin{aligned} Q(\hat{\beta}) &= \sum_{i=1}^{n}(Y_i - f(X_i,\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})(\hat{\beta}-\hat{\beta}_{(0)}))^2 \\ &= \sum_{i=1}^{n}(Y_i - f(X_i,\hat{\beta}_{(0)}) + Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)} - Z_i(\hat{\beta}_{(0)})\hat{\beta})^2 \\ &= \sum_{i=1}^{n}(\tilde{Y}_i(\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})\hat{\beta})^2 \end{aligned} \tag{3.5.28}$$

其中，$\tilde{Y}_i(\hat{\beta}_{(0)}) = Y_i - f(X_i,\hat{\beta}_{(0)}) + Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)}$，可见，一旦给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，可以计算出(3.5.28)式中的 $\tilde{Y}_i(\hat{\beta}_{(0)})$ 和 $Z_i(\hat{\beta}_{(0)})$ 的确定的观测值。于是，将(3.5.23)式取极小值就变成了对(3.5.28)式取极小值。

如果有一个线性模型：

$$\tilde{Y}_i(\hat{\beta}_{(0)}) = Z_i(\hat{\beta}_{(0)})\beta + \varepsilon_i \tag{3.5.29}$$

很容易求得其参数 β 的普通最小二乘估计值 $\hat{\beta}_{(1)}$，该估计值使得残差平方和

$$Q(\hat{\beta}_{(1)}) = \sum_{i=1}^{n}(\tilde{Y}_i(\hat{\beta}_{(0)}) - Z_i(\hat{\beta}_{(0)})\hat{\beta}_{(1)})^2 \tag{3.5.30}$$

最小。比较(3.5.28)式与(3.5.30)式后发现，满足使(3.5.30)式达到最小的估计值 $\hat{\beta}_{(1)}$ 同时也是使(3.5.28)式达到最小的 $\hat{\beta}$。换句话说，线性模型(3.5.29)的普通最小二乘估计值就是模型(3.5.22)的一个近似估计值。因为它是在给定参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$ 的情况下得到的，将它记为参数估计值 $\hat{\beta}$ 的第一次迭代值 $\hat{\beta}_{(1)}$。它是通过对线性模型(3.5.29)进行普通最小二乘估计而得到的，而线性模型(3.5.29)实际上并不存在，故称之为**线性伪模型**。

将 $\hat{\beta}_{(1)}$ 作为 $\hat{\beta}$ 的新的给定值，将(3.5.23)式中的 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(1)}$ 处展开泰勒级数，取一阶近似值，又可以构造一个新的线性伪模型，对其进行普通最小二乘估计，得到 $\hat{\beta}$ 的第二次迭代值 $\hat{\beta}_{(2)}$……如此迭代下去，直到收敛(连续两次得到的参数估计值之差满足确定的标准)。至此完成了非线性模型(3.5.22)的普通最小二乘估计。

(2)高斯－牛顿迭代法的步骤

在对上述采用高斯－牛顿迭代法实现非线性模型参数最小二乘估计的原理了解之后，可以将高斯－牛顿迭代法的步骤简洁地归纳如下。

第一步：给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$，将 $f(X_i,\hat{\beta})$ 在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取一阶近似值；

第二步：计算 $z_i=\left.\dfrac{\mathrm{d}f(X_i,\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}(0)}$ 和 $\tilde{Y}_i=y_i-f(X_i,\hat{\beta}_{(0)})+Z_i\cdot\hat{\beta}_{(0)}$ 的样本观测值；

第三步：采用普通最小二乘法估计模型 $\tilde{Y}_i=Z_i\beta+\varepsilon_i$，得到 β 的估计值 $\hat{\beta}_{(1)}$；

第四步：用 $\hat{\beta}_{(1)}$ 代替第一步中的 $\hat{\beta}_{(0)}$，重复这一过程，直至收敛。

3．牛顿－拉弗森(Newton-Raphson)迭代法

牛顿－拉弗森迭代法作为高斯－牛顿迭代法的改进，当给出参数估计值 $\hat{\beta}$ 的初值 $\hat{\beta}_{(0)}$ 时，将(3.5.23)式在 $\hat{\beta}_{(0)}$ 处展开泰勒级数，取二阶近似值，即

$$Q(\hat{\beta})\approx Q(\hat{\beta}_{(0)})+\left.\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}(0)}(\hat{\beta}-\hat{\beta}_{(0)})+\frac{1}{2}\left.\frac{\mathrm{d}^2Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\right|_{\hat{\beta}(0)}(\hat{\beta}-\hat{\beta}_{(0)})^2 \qquad (3.5.31)$$

这里与高斯－牛顿迭代法有两点不同：一是直接对 $Q(\hat{\beta})$ 展开泰勒级数，而不是对其中的 $f(X_i,\hat{\beta})$ 展开；二是取二阶近似值，而不是取一阶近似值。

使(3.5.31)式达到极小的条件是

$$\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}=0$$

注意，这里的 $Q(\hat{\beta})$ 已经用(3.5.31)式的近似式代入，而不是(3.5.23)式。再对 $\dfrac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}$ 取一阶近似，则有

$$\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\approx\left.\frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\right|_{\hat{\beta}(0)}+\left.\frac{\mathrm{d}^2Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\right|_{\hat{\beta}(0)}\cdot(\hat{\beta}-\hat{\beta}_{(0)})=0$$

于是得到

$$\hat{\beta} = \hat{\beta}_{(0)} - \left(\frac{\mathrm{d}^2 Q(\hat{\beta})}{\mathrm{d}\hat{\beta}^2}\Big|_{\hat{\beta}_{(0)}}\right)^{-1} \cdot \frac{\mathrm{d}Q(\hat{\beta})}{\mathrm{d}\hat{\beta}}\Big|_{\hat{\beta}_{(0)}} \tag{3.5.32}$$

由(3.5.32)式得到的 $\hat{\beta}$ 并不是最后的参数估计值，将它作为第一次迭代值 $\hat{\beta}_{(1)}$，再进行上述过程，直至收敛。

无论是高斯-牛顿迭代法还是牛顿-拉弗森迭代法，都存在一个问题，即如何保证迭代所逼近的是总体极小值(即最小值)而不是局部极小值？这就需要选择不同的初值，进行多次迭代求解。

非线性普通最小二乘法早已出现在计量经济学应用软件中，即使是目前使用最为普遍、最为简单的 Eviews 中也有非线性普通最小二乘法估计方法。在选择了该估计方法、给定参数初始值后，只要将非线性方程的形式输入，就可以得到参数的估计量。

例 3.5.2

下面直接用非线性普通最小二乘法进行例 3.5.1 的估计。

首先，估计(3.5.16)式对应的非线性模型

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

这里需要将等式右边的 A 改写为 e^{β_0}。取各参数 β_0，β_1，β_2，β_3 的初值均为 1，Eviews 软件的估计结果如表 3.5.4 所示。

表 3.5.4 中国城镇居民人均食品消费支出的非线性估计

```
Dependent Variable：Q
Method:Least Squares
Sample:1985 2006
Included observations:22
Convergence achieved after 6 iterations
Q=EXP(C(1))*X^C(2)*P1^C(3)*P0^C(4)
      Variable          Coefficient        Std.Error        t-Statistic      Prob.
        C(1)             5.567708          0.083537          66.64931       0.0000
        C(2)             0.555715          0.029067          19.11874       0.0000
        C(3)            –0.190154          0.143823         –1.322146       0.2027
        C(4)            –0.394861          0.159291         –2.478866       0.0233
R–squared                0.983631      Mean dependent var              1830.000
Adjusted R–squared       0.980903      S.D. dependent var              365.1392
S.E. of regression       50.45954      Akaike info criterion           10.84319
Sum squared resid        45830.98      Schwarz criterion               11.04156
Log likelihood          –115.2751      Hannan–Quinn criter             10.88992
Durbin-Watson stat       0.672163
```

与原双对数线性模型的估计结果相比，常数项与 X 的参数的估计结果较为接近。虽然 P_1，P_0 对应的参数的估计结果有差异，但仍呈现 P_1 对应参数的估计结果小于 P_0 对应参数估计结果的特征。不同的是，这里 P_0 通过了5%显著性水平检验。

其次，估计(3.5.17)式对应的非线性模型

$$Q = A\left(\frac{X}{P_0}\right)^{\beta_1}\left(\frac{P_1}{P_0}\right)^{\beta_2} \tag{3.5.33}$$

这里仍需要将等式右边的 A 改写为 e^{β_0}。取各参数 β_0，β_1，β_2 的初值均为1，Eviews 软件的估计结果如表 3.5.5 所示。可以看出，这里 3 个参数的估计结果与对应的双对数线性模型的估计结果较为相似，尤其是常数项与 P_1/P_0 对应的参数，估计结果已非常接近了。

表 3.5.5　中国城镇居民人均食品消费支出的非线性估计

```
Dependent Variable：Q
Method:Least Squares
Sample:1985 2006
Included observations:22
Convergence achieved after 9 iterations
Q=EXP(C(1))*(X/P0)^C(2)*(P1/P0)^C(3)
      Variable          Coefficient        Std.Error         t–Statistic       Prob.
        C(1)              5.525965          0.072685           76.02666        0.0000
        C(2)              0.533824          0.019785           26.98163        0.0000
        C(3)             –0.242862          0.134014          –1.812218        0.0858
R–squared                 0.982669       Mean dependent var                  1830.000
Adjusted R–squared        0.980845       S.D. dependent var                  365.1392
S.E. of regression        50.53638       Akaike info criterion               10.80939
Sum squared resid         48524.59       Schwarz criterion                   10.95817
Log likelihood           –115.9033       Hannan–Quinn criter                 10.84444
Durbin–Watson stat        0.656740
```

§ 3.6　受约束回归

在建立回归模型时，有时根据经济理论需要对模型中变量的参数施加一定的约束条件。例如，上节建立中国城镇居民对食品的消费需求函数时，根据需求函数的一般理论，它应满足零阶齐次性条件，即双对数线性模型中各变量前的参数和为零。同样地，在估计以幂函数的形式表示的生产函数模型时，有时也施加产出关于资本与劳动的弹性和为 1 的约束。模型施加约束条件后进行回

归，称为**受约束回归**(restricted regression)，与此对应，不加任何约束的回归称为**无约束回归**(unrestricted regression)。

一、模型参数的线性约束

一般地，估计线性模型时可对模型参数施加若干个线性约束条件。例如，对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \tag{3.6.1}$$

可施加

$$\beta_1+\beta_2=1,\qquad \beta_{k-1}=\beta_k \tag{3.6.2}$$

于是，对(3.6.1)式的回归可转化为对施加上述条件后如下模型的回归：

$$Y=\beta_0+\beta_1X_1+(1-\beta_1)X_2+\cdots+\beta_{k-1}X_{k-1}+\beta_{k-1}X_k+\mu^* \tag{3.6.3}$$

或

$$Y^*=\beta_0+\beta_1X_1^*+\beta_3X_3+\cdots+\beta_{k-2}X_{k-2}+\beta_{k-1}X_{k-1}^*+\mu^* \tag{3.6.4}$$

其中

$$Y^*=Y-X_2,\qquad X_1^*=X_1-X_2,\qquad X_{k-1}^*=X_{k-1}+X_k$$

如果运用普通最小二乘法得到参数的估计结果 $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_3,\cdots,\hat{\beta}_{k-1}$，可由上述约束条件得到

$$\hat{\beta}_2=1-\hat{\beta}_1,\qquad \hat{\beta}_k=\hat{\beta}_{k-1}$$

然而，对所考察的具体问题能否施加约束条件，或者说能否直接对施加约束后的模型进行回归，还需进一步进行相应的检验。常用的检验有 F 检验、χ^2 检验与 t 检验，这里主要介绍 F 检验。

在同一数据样本下，记无约束样本回归模型的矩阵式为

$$\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{e} \tag{3.6.5}$$

记受约束样本回归模型的矩阵式为

$$\boldsymbol{Y}=\boldsymbol{X}\hat{\boldsymbol{\beta}}_*+\boldsymbol{e}_* \tag{3.6.6}$$

于是，受约束样本回归模型的残差项可写为

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

得到受约束样本回归模型的残差平方和 $\mathrm{RSS_R}$ 为

$$\boldsymbol{e}_*'\boldsymbol{e}_*=\boldsymbol{e}'\boldsymbol{e}+(\hat{\boldsymbol{\beta}}_*-\hat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}}_*-\hat{\boldsymbol{\beta}}) \tag{3.6.7}$$

式中第二项为一个非负标量，于是

$$\boldsymbol{e}_*'\boldsymbol{e}_*\geqslant\boldsymbol{e}'\boldsymbol{e} \tag{3.6.8}$$

其中，$\boldsymbol{e}'\boldsymbol{e}$ 为无约束样本回归模型的残差平方和 $\mathrm{RSS_U}$。

在(3.6.5)式与(3.6.6)式两个回归模型中，有着相同的被解释变量 Y 与相同的数据样本，于是 Y 的总离差平方和 TSS 也相同。(3.6.8)式表明受约束样本回归模型的残差平方和不小于无约束样本回归模型的残差平方和，于是，受约束样

本回归模型的回归平方和 $\mathrm{ESS_R}$ 不大于无约束样本回归模型的回归平方和 $\mathrm{ESS_U}$。这意味着，通常情况下，对模型施加约束条件会降低模型的解释能力。

但是，如果约束条件为真，则受约束回归模型与无约束回归模型具有相同的解释能力，从而使得 $\mathrm{RSS_U}$ 与 $\mathrm{RSS_R}$ 的差异变小。于是，可用 $\mathrm{RSS_R}-\mathrm{RSS_U}$ 的大小来检验约束条件的真实性。

根据数理统计学的知识，$\frac{\mathrm{RSS}}{\sigma^2}\sim\chi^2(n-k-1)$，其中 k 为回归模型中解释变量的个数，σ^2 为回归模型随机干扰项的方差。于是，

$$\frac{\mathrm{RSS_U}}{\sigma^2}\sim\chi^2(n-k_\mathrm{U}-1) \tag{3.6.9}$$

$$\frac{\mathrm{RSS_R}}{\sigma^2}\sim\chi^2(n-k_\mathrm{R}-1) \tag{3.6.10}$$

$$\frac{\mathrm{RSS_R}-\mathrm{RSS_U}}{\sigma^2}\sim\chi^2(k_\mathrm{U}-k_\mathrm{R}) \tag{3.6.11}$$

其中，k_U，k_R 分别为无约束与受约束回归模型的解释变量的个数(不包括常数项)。于是可通过计算(3.6.11)式的 χ^2 统计量来进行相应的 χ^2 检验。当然，由于随机干扰项的方差 σ^2 往往未知，检验时需用它的估计量 $\hat{\sigma}^2$ 替代。

当约束条件为真时，由(3.6.9)式与(3.6.11)式可进一步得到如下的 F 统计量：

$$F=\frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/(k_\mathrm{U}-k_\mathrm{R})}{\mathrm{RSS_U}/(n-k_\mathrm{U}-1)}\sim F(k_\mathrm{U}-k_\mathrm{R},n-k_\mathrm{U}-1) \tag{3.6.12}$$

F 统计量无需估计随机扰动项的方差 σ^2。根据该统计量，如果约束条件无效，则 $\mathrm{RSS_R}$ 与 $\mathrm{RSS_U}$ 的差异较大，计算的 F 值也较大。于是，可用计算的 F 统计量的值与所给定的显著性水平下的临界值作比较，来对约束条件的真实性进行检验。需注意的是，$k_\mathrm{U}-k_\mathrm{R}$ 恰为约束条件的个数。

例 3.6.1

在§3.5 中国城镇居民对食品的人均消费需求实例中，无约束回归模型(3.5.18)式的残差平方和 $\mathrm{RSS_U}$=0.017 748，受约束回归模型(3.5.19)式的残差平方和 $\mathrm{RSS_R}$=0.017 787，样本容量 n=22，无约束回归模型变量个数 k_U=3，约束条件个数 $k_\mathrm{U}-k_\mathrm{R}$=3–2=1。于是

$$F=\frac{(0.017\ 787-0.017\ 748)/1}{0.017\ 787/18}=0.039\ 5$$

在 5%的显著性水平下，自由度为(1，18)的 F 统计量的临界值为 $F_{0.05}$=4.41。计算的 F 值小于临界值，不能拒绝中国城镇居民对食品的人均消费需求函数具有零阶齐次特性这一假设。

需要指出的是，这里介绍的 F 检验适合所有关于参数线性约束的检验，§3.4 中对回归模型总体的线性检验，可以归结到这里的 F 检验上来。例如，对线性模型

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

的总体线性检验，就是要检验联合假设：

$$H_0: \beta_j = 0\,, \qquad j = 1, 2, \cdots, k$$

因此，受约束回归模型为

$$Y = \beta_0 + \mu.$$

由(3.6.12)式，检验的 F 统计量为

$$\begin{aligned} F &= \frac{(\mathrm{RSS_R} - \mathrm{RSS_U})/(k_\mathrm{U} - k_\mathrm{R})}{\mathrm{RSS_U}/(n - k_\mathrm{U} - 1)} \\ &= \frac{(\mathrm{TSS} - \mathrm{ESS_R} - \mathrm{RSS_U})/k}{\mathrm{RSS_U}/(n-k-1)} \\ &= \frac{(\mathrm{TSS} - \mathrm{RSS_U})/k}{\mathrm{RSS_U}/(n-k-1)} \\ &= \frac{\mathrm{ESS_U}/k}{\mathrm{RSS_U}/(n-k-1)} \end{aligned}$$

这里，运用了受约束回归模型的回归平方和 $\mathrm{ESS_R} = 0$。

二、对回归模型增加或减少解释变量

在建立回归模型时，一个重要的问题是如何判断增加重要的解释变量或去掉不必要的解释变量。t 检验可对单个变量的取舍进行判断，而上面介绍的 F 检验除能对单个变量的取舍进行判断外，还可对多个变量的同时取舍进行判断。

考虑如下两个回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \mu \tag{3.6.13}$$

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots + \beta_{k+q} X_{k+q} + \mu \tag{3.6.14}$$

(3.6.13)式可以看成是(3.6.14)式施加如下约束条件的受约束回归：

$$H_0:\ \ \beta_{k+1} = \beta_{k+2} = \cdots = \beta_{k+q} = 0 \tag{3.6.15}$$

相应的 F 统计量为

$$\begin{aligned} F &= \frac{(\mathrm{RSS_R} - \mathrm{RSS_U})/q}{\mathrm{RSS_U}/[n-(k+q+1)]} \\ &= \frac{(\mathrm{ESS_U} - \mathrm{ESS_R})/q}{\mathrm{RSS_U}/[n-(k+q+1)]} \sim F[q, n-(k+q+1)] \end{aligned} \tag{3.6.16}$$

如果约束条件为真，即额外的变量 $X_{k+1},\cdots,X_{k+q}$ 对 Y 没有解释能力，则 F 统计量较小；否则，约束条件为假，意味着额外的变量 $X_{k+1},\cdots,X_{k+q}$ 对 Y 有较强的解释能力，则 F 统计量较大。因此，可通过给定某一显著性水平下 F 分布的临界值与 F 统计量的计算值的比较，来判断额外变量 $X_{k+1},\cdots,X_{k+q}$ 是否应包括在模型中。

由(3.6.16)式可得到 F 统计量的另一个等价的式子：

$$F=\frac{(R_{\mathrm{U}}^2-R_{\mathrm{R}}^2)/q}{(1-R_{\mathrm{U}}^2)/[n-(k+q+1)]} \tag{3.6.17}$$

其中 R_{U}^2，R_{R}^2 分别为无约束回归与受约束回归方程的可决系数，表明通过变量增减前后回归方程的可决系数 R^2 是否有“足够大”的变化来判断变量的增减与否。

例 3.6.2

在§3.5 中国城镇居民对食品的人均消费需求实例中，t 检验发现食品价格与城镇居民的总消费价格的变动好像不对城镇居民人均食品消费产生影响，似乎可从模型中去掉。但另一方面，由于两价格指数具有很强的相关性，也可能使得模型无法准确地分辨出它们各自的影响。这里，我们可以通过 F 检验来判定是否可将两价格因素从模型中去掉，或者只去掉其中之一。

无约束模型仍是(3.5.18)式，其残差平方和为 $\mathrm{RSS_U}$ =0.017 748。去掉两价格因素的受约束模型回归结果如下：

$$\ln\hat{Q}=5.74+0.221\,3\ln X$$

$$(38.14)\quad(11.74)$$

$$R^2=0.873\,3\qquad \bar{R}^2=0.867\,0\qquad F=137.9\quad \mathrm{RSS}=0.099\,228$$

于是，可进行如下 F 检验

$$F=\frac{[0.099\,228-0.017\,748]/2}{0.017\,748/(22-4)}=41.318$$

该值远大于 5%显著性水平下相应的临界值 $F_{0.05}(2,18)$=3.55，因此，两价格变量不应同时去掉。类似地，还可检验是否可去掉一个价格因素。如去掉城镇居民消费价格指数 P_0 的受约束模型回归结果如下：

$$\ln\hat{Q}=5.45+0.524\,0\ln X-0.496\,3\ln P_1$$

$$(71.73)\quad(14.71)\quad(-8.76)$$

$$R^2=0.974\,8\qquad \bar{R}^2=0.972\,2\qquad F=368.3\quad \mathrm{RSS}=0.019\,698$$

相应的 F 检验为

$$F=\frac{[0.019\,698-0.017\,748]/1}{0.017\,748/(22-4)}=1.978$$

该值小于 5%显著性水平下相应的临界值 $F_{0.05}(1, 18)=4.41$，因此，可将 P_0 从原模型中去掉。读者可进一步检验，去掉 P_0 的模型是否仍满足需求方程的零阶齐次性条件。

*三、参数的稳定性

1. 邹氏参数稳定性检验

建立模型时往往希望模型的参数是稳定的，即所谓的结构不变，这将提高模型的预测与分析功能。然而，经济结构的变化往往导致计量模型结构也发生变化。如§3.5 中国城镇居民对食品的人均消费需求例子中，可检验 1998 年前后中国城镇居民的食品消费支出是否出现了结构变化。如图 3.6.1 所示，中国城镇居民人均食品消费虽然在 1985—2006 年间总体呈现较为一致的增长趋势，但在 1998 年前后确实存在着两条不同的趋势线(图 3.6.1 中的两条虚线)。显然，1998 年后中国城镇居民人均食品消费有着比前期更加快速的增长。下面给出一个结构变化的检验。

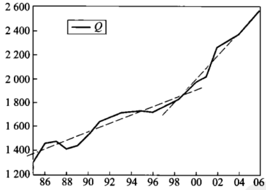


图 3.6.1 中国城镇居民人均食品消费（元）

假设需要建立的模型为

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \tag{3.6.1}$$

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

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_kX_k+\mu_1 \tag{3.6.18}$$

$$Y=\alpha_0+\alpha_1X_1+\cdots+\alpha_kX_k+\mu_2 \tag{3.6.19}$$

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

$$\begin{pmatrix}\boldsymbol{Y}_1\\\boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1 & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{X}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\\boldsymbol{\alpha}\end{pmatrix}+\begin{pmatrix}\boldsymbol{\mu}_1\\\boldsymbol{\mu}_2\end{pmatrix} \tag{3.6.20}$$

其中，$\boldsymbol{\beta}$，$\boldsymbol{\alpha}$ 分别是两时间序列对应模型中的参数列向量，$\boldsymbol{Y}_i(i=1,2)$ 是对应模型的被解释变量以其样本为元素的列向量，$\boldsymbol{X}_i(i=1,2)$ 是对应模型的解释变量矩阵。

如果$\boldsymbol{\beta}=\boldsymbol{\alpha}$，表示没有发生结构变化，因此可针对如下假设进行检验：

$$\boldsymbol{H}_0:\ \boldsymbol{\beta}=\boldsymbol{\alpha} \tag{3.6.21}$$

(3.6.20)式施加该约束条件后变换为受约束回归模型

$$\begin{pmatrix}\boldsymbol{Y}_1\\\boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1\\\boldsymbol{X}_2\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\boldsymbol{\mu}_1\\\boldsymbol{\mu}_2\end{pmatrix} \tag{3.6.22}$$

因此，仍可用如下 F 统计量进行检验：

$$F=\frac{(\mathrm{RSS}_\mathrm{R}-\mathrm{RSS}_\mathrm{U})/(k+1)}{\mathrm{RSS}_\mathrm{U}/[n_1+n_2-2(k+1)]}\sim F[k+1,n_1+n_2-2(k+1)] \tag{3.6.23}$$

其中，RSS_U 与 RSS_R 分别为对应于无约束模型(3.6.20)式与受约束模型(3.6.22)式的残差平方和。

记 RSS_1 与 RSS_2 为两时间序列对应的回归模型(3.6.18)式与(3.6.19)式在各自时间段上分别回归后所得的残差平方和，容易验证

$$\mathrm{RSS}_\mathrm{U}=\mathrm{RSS}_1+\mathrm{RSS}_2$$

于是，F 统计量可写为

$$F=\frac{[\mathrm{RSS}_\mathrm{R}-(\mathrm{RSS}_1+\mathrm{RSS}_2)]/(k+1)}{(\mathrm{RSS}_1+\mathrm{RSS}_2)/[n_1+n_2-2(k+1)]}\sim F[k+1,n_1+n_2-2(k+1)] \tag{3.6.24}$$

因此，对参数稳定性的原假设(3.6.21)式的检验步骤为：首先，分别以两个连续的时间序列作为两个样本运用(3.6.1)式进行回归，得到相应的残差平方和 RSS_1 与 RSS_2；然后，将两序列并为一个大样本后运用(3.6.1)式进行回归，得到大样本下的残差平方和 RSS_R；最后，通过(3.6.24)式的 F 统计量，在事先给定的显著性水平下进行假设检验。如果 F 大于相应的临界值，则拒绝原假设，认为发生了结构变化，参数是非稳定的。该检验方法也称为**邹氏参数稳定性检验**(Chow test for parameter stability)。

2. 邹氏预测检验

上述参数稳定性检验要求 $n_2>k$，即第二个时间段中样本数不能小于待估参数的个数。如果出现 $n_2<k$，则往往进行**邹氏预测检验**(Chow test for predictive failure)。

邹氏预测检验的基本思想是：先用前一时间段 n_1 个样本估计原模型；再用估计出的参数进行后一时间段 n_2 个样本的预测。如果预测误差较大，则说明参数发生了变化，否则说明参数是稳定的，即需用第一时间段估计的参数 $\hat{\beta}_j\,(j=1,2,\cdots,k)$，考察第二时间段预测误差 $Y_i-(\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+\cdots+\hat{\beta}_kX_{ik})$ 的大小。

分别以 $\boldsymbol{\beta}$，$\boldsymbol{\alpha}$ 表示第一时间段与第二时间段的参数，即假设允许预测期出现不同的参数向量，则完整的模型可写为

$$
\begin{aligned}
\boldsymbol{Y}_1 &= \boldsymbol{X}_1\boldsymbol{\beta}+\boldsymbol{\mu}_1\\
\boldsymbol{Y}_2 &= \boldsymbol{X}_2\,\boldsymbol{\alpha}+\boldsymbol{\mu}_2\\
&= \boldsymbol{X}_2\boldsymbol{\beta}+\boldsymbol{X}_2(\boldsymbol{\alpha}-\boldsymbol{\beta})+\boldsymbol{\mu}_2\\
&= \boldsymbol{X}_2\boldsymbol{\beta}+\boldsymbol{\gamma}+\boldsymbol{\mu}_2
\end{aligned}
$$

其中 $\boldsymbol{\gamma}=\boldsymbol{X}_2(\boldsymbol{\alpha}-\boldsymbol{\beta})$。如果 $\boldsymbol{\gamma}=\boldsymbol{0}$，则 $\boldsymbol{\alpha}=\boldsymbol{\beta}$，表明参数在估计期与预测期相同。

将上述模型写成矩阵式：

$$
\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1 & \boldsymbol{0}\\ \boldsymbol{X}_2 & \boldsymbol{I}_{n_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}+\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix} \tag{3.6.25}
$$

其中，$\boldsymbol{I}_{n_2}$ 为 n_2 阶单位矩阵。可见，用前 n_1 个样本估计可得前 k 个参数 $\boldsymbol{\beta}$ 的估计，而 $\boldsymbol{\gamma}$ 不外是用后 n_2 个样本测算的预测误差 $\boldsymbol{X}_2\boldsymbol{\alpha}-\boldsymbol{X}_2\boldsymbol{\beta}$。如果参数没有发生变化，则 $\boldsymbol{\gamma}=\boldsymbol{0}$，(3.6.25)式简化为

$$
\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix} \tag{3.6.26}
$$

(3.6.25)式与(3.6.26)式分别可看成无约束与受约束回归模型，因此，可用(3.6.12)式的 F 统计量进行约束的有效性检验：

$$
\begin{aligned}
F &= \frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/[(k+1+n_2)-(k+1)]}{\mathrm{RSS_U}/[n_1+n_2-(k+1+n_2)]}\\
&= \frac{(\mathrm{RSS_R}-\mathrm{RSS_1})/n_2}{\mathrm{RSS_1}/(n_1-k-1)}
\end{aligned} \tag{3.6.27}
$$

这里，由于检验的约束为 n_2 个元素的向量 $\boldsymbol{\gamma}=\boldsymbol{0}$，因此，约束共有 n_2 个。另外，在无约束模型(3.6.25)式中，已将后 n_2 个观测值的预测误差记入到了 $\boldsymbol{\gamma}$ 中，因此，对应的样本残差 $\boldsymbol{e}_2=\boldsymbol{0}$，这样无约束回归中的总的残差平方和 $\mathrm{RSS_U}$ 就等于用前 n_1 个样本估计的残差平方和 $\mathrm{RSS_1}$。

邹氏预测检验仍为三步：第一步，在两时间段的合成大样本下做 OLS 回归，得到受约束模型的残差平方和 $\mathrm{RSS_R}$；第二步，对前一时间段的 n_1 个子样做 OLS 回归，得到残差平方和 $\mathrm{RSS_1}$；第三步，计算(3.6.27)式的 F 统计量，在事先给

PDG

定的显著性水平下进行假设检验。如果 F 大于相应的临界值，则拒绝原假设，认为预测期发生了结构变化。

例 3.6.3

在§3.5 中国城镇居民对食品的人均消费需求实例中，检验是否 1998 年前后中国城镇居民食品消费支出发生了结构变化。为了避免城镇居民消费价格与城镇居民食品消费价格间的强相关性对估计带来的影响，我们的模型采用经济意义更为明确的如下形式：

$$\ln Q = \beta_0 + \beta_1 \ln \frac{X}{P_0} + \beta_2 \ln \frac{P_1}{P_0} + \mu$$

在 1985—1997 年的时间段里，有如下回归结果

$$\ln \hat{Q} = 4.633 + 0.799 \ln \frac{X}{P_0} - 1.046 \ln \frac{P_1}{P_0}$$

(12.9)　(7.59)　(−3.33)

$R^2=0.926\,2$　$\bar{R}^2=0.911\,5$　D.W.=1.07　F=62.77　$\text{RSS}_1=0.008\,3$

在 1998—2006 年的时间段里，有如下回归结果

$$\ln \hat{Q} = 5.223 + 0.606 \ln \frac{X}{P_0} - 0.224 \ln \frac{P_1}{P_0}$$

(54.8)　(24.9)　(−1.99)

$R^2=0.994\,0$　$\bar{R}^2=0.992\,0$　D.W.=1.87　F=497.6　$\text{RSS}_2=0.000\,8$

在例 3.5.1 中，已经得到 1985—2006 年时间段的回归结果，其中 $\text{RSS}_\text{U}=0.017\,8$，于是结构变化的 F 检验值为

$$F = \frac{[0.017\,8 - (0.008\,3 + 0.000\,8)]/3}{(0.008\,3 + 0.000\,8)/(22-6)} = 5.10$$

在 5%的显著性水平下，自由度为（3，16）的 F 分布的临界值为 $F_{0.05}$（3,16）=3.24，可见计算的 F 值远大于临界值，拒绝参数稳定的原假设，表明中国城镇居民对食品的人均消费需求行为在 1998 年前后发生了显著变化。

---

*四、非线性约束

估计线性模型时也可对模型参数施加非线性约束。例如，对模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu \tag{3.6.1}$$

施加非线性约束 $\beta_1\beta_2 = 1$。然而，这时再无法找到一个类似于(3.6.3)式或(3.6.4)式那样简单的受约束回归模型，只能得到如下受约束回归模型：

$$Y = \beta_0 + \beta_1 X_1 + \frac{1}{\beta_1} X_2 + \cdots + \beta_k X_k + \mu^* \tag{3.6.28}$$

该模型无法运用通常的普通最小二乘法进行估计，必须采用**非线性最小二乘法**(nonlinear least squares)进行估计。

非线性最小二乘法的主要问题在于其参数估计量并不必然具有期望的小样本性质。然而，从最小二乘估计原理与最大似然估计原理来看，受约束的残差平方和最小化问题，等价于受约束的最大似然函数最大化的问题，且最小二乘估计量就是最大似然估计量，因此，估计量具有一致性与渐近有效性等大样本性质。

有鉴于此，非线性约束检验是建立在最大似然原理基础上的。有三个著名的检验：**最大似然比检验**(Likelihood Ratio Test, LR)、**沃尔德检验**(Wald Test, WD)与**拉格朗日乘数检验**(Lagrange Multiplier Test, LM)。它们的共同特点是：在大样本下，以共同的 χ^2 检验为基础，而自由度就是约束条件的个数。

1. 最大似然比检验(LR)

最大似然比检验仍需估计无约束回归模型与受约束回归模型，但运用的估计方法是最大似然法，检验的是两个似然函数的值的差异是否“足够”大。

记 $L(\boldsymbol{\beta},\sigma^2)$ 为似然函数，对给定的样本数据，无约束回归就是求一组参数 $\hat{\boldsymbol{\beta}}$ 与 $\hat{\sigma}^2$，以使似然函数 $L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)$ 的值最大。如果给定约束条件 $\boldsymbol{g}(\boldsymbol{\beta})=\mathbf{0}$，有约束的回归则是在该约束条件下，求另一组参数 $\tilde{\boldsymbol{\beta}}$ 与 $\tilde{\sigma}^2$，以使似然函数 $L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$ 的值最大。根据拉格朗日乘数法，就是求如下函数的极值问题：

$$\Phi = L(\boldsymbol{\beta},\sigma^2) - \boldsymbol{\lambda}'\boldsymbol{g}(\boldsymbol{\beta}) \tag{3.6.29}$$

其中，$\boldsymbol{g}(\boldsymbol{\beta})$ 是以各约束条件为元素的列向量，$\boldsymbol{\lambda}'$ 则是以相应拉格朗日乘数为元素的行向量。

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

$$\frac{L(\tilde{\boldsymbol{\beta}},\ \tilde{\sigma}^2)}{L(\hat{\boldsymbol{\beta}},\ \hat{\sigma}^2)}$$

如果该比值很小，说明受约束似然函数值与无约束似然函数值差距较大，则应拒绝约束条件为真的假设；如果该比值接近于 1，则受约束似然函数值与无约束似然函数值很接近，应接受约束条件为真的假设。

在进行检验时，由于在大样本下

$$\mathrm{LR} = -2[\ln L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) - \ln L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)] \sim \chi^2(h) \tag{3.6.30}$$

其中 h 是约束条件的个数，则可以在给定的显著性水平下，通过最大似然比检验的计算值与相应的 χ^2 分布的临界值的比较，来判断是否拒绝给定的约束条件

的假设。

在例 3.5.1 中，检验需求函数的零阶齐次性条件，受约束回归模型的最大似然值的对数值为 $\ln L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)=47.11$，无约束回归模型的最大似然值的对数值为 $\ln L(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)=47.13$，于是

$$\mathrm{LR}=-2(47.11-47.13)=0.04$$

该值小于 5%显著性水平下自由度为 1 的 χ^2 分布的临界值 $\chi^2_{0.05}(1)=3.841$，不拒绝原约束的假设，表明中国城镇居民对食品的人均消费需求函数满足零阶齐次性条件。

2. 沃尔德检验(WD)

最大似然比检验不仅要估计无约束模型，还要估计受约束模型。而在沃尔德检验中，只需估计无约束模型。例如，对模型

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k+\mu \tag{3.6.1}$$

要检验 $\beta_1+\beta_2=1$ 的约束，只需对该模型进行回归，并判断 $\hat{\beta}_1+\hat{\beta}_2$ 与 1 的差距是否足够大。在所有基本假设都成立的条件下，容易证明

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

因此，在 $\beta_1+\beta_2=1$ 的约束条件下

$$z=\frac{\hat{\beta}_1+\hat{\beta}_2-1}{\sigma_{\hat{\beta}_1+\hat{\beta}_2}}\sim N(0,1) \tag{3.6.31}$$

由于 $\sigma^2_{\hat{\beta}_1+\hat{\beta}_2}$ 是 $\hat{\beta}_1+\hat{\beta}_2$ 的方差，其值与随机干扰项 μ 的方差有关：

$$\sigma^2_{\hat{\beta}_1+\hat{\beta}_2}=\mathrm{Var}(\hat{\beta}_1)+\mathrm{Var}(\hat{\beta}_2)+2\mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_2)=\sigma^2f(\boldsymbol{X}),$$

以 σ^2 的极大似然估计量 $\tilde{\sigma}^2=\dfrac{\boldsymbol{e}'\boldsymbol{e}}{n}$ 代入，并记 $\tilde{\sigma}^2_{\hat{\beta}_1+\hat{\beta}_2}=\tilde{\sigma}^2f(\boldsymbol{X})$，于是，可建立服从自由度为 1 的 χ^2 分布的沃尔德统计量

$$W=\frac{(\hat{\beta}_1+\hat{\beta}_2-1)^2}{\tilde{\sigma}^2_{\hat{\beta}_1+\hat{\beta}_2}}\sim\chi^2(1) \tag{3.6.32}$$

如果有 h 个约束条件，则可得到 h 个类似于(3.6.31)式的统计量 $z_1,z_2,\cdots,z_h$，当它们相互独立时，其平方和服从自由度为 h 的 χ^2 分布。然而，一般情况下，由 h 个约束条件得到的 h 个 z 统计量不相互独立，因此，无法得到精确的 χ^2 分布。但是在各约束条件为真的情况下，可建立大样本下的服从自由度为 h 的渐近 χ^2 分布统计量

$$W=\boldsymbol{Z}'\boldsymbol{C}^{-1}\boldsymbol{Z}\sim\chi^2(h) \tag{3.6.33}$$

其中，$\boldsymbol{Z}$ 为以 z_i 为元素的列向量，$\boldsymbol{C}$ 是 z_i 的方差-协方差矩阵。因此，统计量 W

从总体上测量了无约束回归不满足各约束条件的程度。

对非线性约束，也可类似地建立沃尔德统计量 W，但算法描述要复杂得多。例如，对(3.6.1)式，如果给出非线性约束 $\beta_1\beta_2=1$，也可以证明，在大样本下，

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

因此，如果 $\beta_1\beta_2=1$ 的约束为真，则在大样本下

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

这里 $\tilde{\sigma}^2_{\hat{\beta}_1\hat{\beta}_2}$ 的算法比较复杂，可用泰勒公式进行展开计算。

3. 拉格朗日乘数检验(LM)

沃尔德检验只需估计无约束模型，而拉格朗日乘数检验则只需估计受约束模型。

在受约束回归最大似然法的极值问题

$$\Phi = L(\boldsymbol{\beta}, \sigma^2) - \boldsymbol{\lambda}' g(\boldsymbol{\beta}) \tag{3.6.29}$$

中，$\boldsymbol{\lambda}'$ 是以各约束条件相应拉格朗日乘数为元素的行向量，各拉格朗日乘数 λ_i 的大小则衡量了各约束条件对最大似然函数值 $L(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)$ 的影响程度。如果某一约束为真，则该约束条件对最大似然函数值 $L(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)$ 的影响很小，于是，相应的拉格朗日乘数的值应接近于零。因此，拉格朗日乘数检验就是检验某些拉格朗日乘数的值是否“足够大”，如果“足够大”，则拒绝约束条件为真的假设。

拉格朗日统计量拉格朗日乘数检验本身是一个关于拉格朗日乘数的复杂的函数，在各约束条件为真的情况下，服从自由度恰为约束条件个数的渐近 χ^2 分布。同样地，如果为线性约束，拉格朗日乘数检验则服从于精确的 χ^2 分布：

$$\mathrm{LM} = nR^2 \tag{3.6.34}$$

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

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

这里，$\hat{e}_{\mathrm{R}}$ 为受约束回归模型的残差序列。

如果约束是非线性的，辅助回归方程的估计比较复杂，但仍可按(3.6.34)式计算拉格朗日乘数检验统计量的值。

最后，一般有 $\mathrm{LM} \leqslant \mathrm{LR} \leqslant W$，因此，在有限样本中，它们的数值结果会有所不同。

本章练习题

1. 多元线性回归模型的基本假设是什么？试说明在证明最小二乘估计量的无偏性和有

效性的过程中，哪些基本假设起了作用？

2. 在多元线性回归分析中，t 检验与 F 检验有何不同？在一元线性回归分析中二者是否有等价的作用？

3. 为什么说对模型参数施加约束条件后，其回归的残差平方和一定不比未施加约束的残差平方和小？在什么样的条件下，受约束回归与无约束回归的结果相同？

4. 在一项调查大学生一学期平均成绩(Y)与每周在学习(X_1)、睡觉(X_2)、娱乐(X_3)与其他各种活动(X_4)所用时间的关系的研究中，建立如下回归模型：

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\mu$$

如果这些活动所用时间的总和为一周的总小时数 168。问：保持其他变量不变，而改变其中一个变量的说法是否有意义？该模型是否有违背基本假设的情况？如何修改此模型以使其更加合理？

5. 考虑下列两个模型：

(a) $$Y_i=\alpha_0+\alpha_1X_{i1}+\alpha_2X_{i2}+u_i$$

(b) $$Y_i-X_{i1}=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+v_i$$

(1) 证明：

$$\hat{\beta}_1=\hat{\alpha}_1-1\,,\qquad \hat{\beta}_0=\hat{\alpha}_0\,,\qquad \hat{\beta}_2=\hat{\alpha}_2$$

(2) 证明：两个模型的最小二乘残差相等，即对任何 i，有 $\hat{u}_i=\hat{v}_i$。

(3) 在什么条件下，模型(b)的 R^2 小于模型(a)的 R^2？

6. 考虑下列三个试验步骤：

(1) 对 $Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+u_i$ 进行回归；

(2) 对 $X_{i1}=\alpha_0+\alpha_1X_{i2}+v_i$ 进行回归，计算残差 $\hat{v}_i$；

(3) 对 $Y_i=\gamma_0+\gamma_1\hat{v}_i+\gamma_2X_{i2}+w_i$ 进行回归。

试证明 $\hat{\beta}_1=\hat{\gamma}_1$，并直观地解释该结果。

7. 考虑以下过原点回归：

$$Y_i=\hat{\beta}_1X_{i1}+\hat{\beta}_2X_{i2}+e_i$$

(1) 求参数的 OLS 估计量；

(2) 对该模型，是否仍有结论

$$\sum e_i=0\,,\qquad \sum e_iX_{i1}=0\,,\qquad \sum e_iX_{i2}=0$$

8. 对下列模型：

(a) $$Y_i=\alpha+\beta X_i+2Z_i+u_i$$

(b) $$Y_i=\alpha+\beta X_i-\beta Z_i+u_i$$

求出 β 的最小二乘估计值，并将结果与下面的三变量回归方程的最小二乘估计值作比较：

(c) $$Y_i=\alpha+\beta X_i+\gamma Z_i+u_i$$

你认为哪一个估计值更好？

9. 下表给出三变量模型 $Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+u_i$ 的回归结果：

方差来源	平方和(SS)	自由度(d.f.)	平方和的均值(MSS)
来自回归	65 965		
来自残差			
来自总离差	66 042	14	

(1) 求样本容量 n，残差平方和 RSS，回归平方和 ESS 及残差平方和 RSS 的自由度。

(2) 求拟合优度 R^2 及调整的拟合优度 $\overline{R}^2$ 。

(3) 检验假设：X_1 和 X_2 对 Y 无影响。应采用什么假设检验？为什么？

(4) 根据以上信息，你能否确定 X_1 和 X_2 各自对 Y 的影响？

10. 试由教材中的(3.6.16)式推导出(3.6.17)式。

11. 在一项对某社区家庭对某种消费品的消费需要调查中，得到下表所示的资料。

单位：元

序号	对某商品的消费支出 Y	商品单价 X_1	家庭月收入 X_2	序号	对某商品的消费支出 Y	商品单价 X_1	家庭月收入 X_2
1	591.9	23.56	7 620	6	644.4	34.14	12 920
2	654.5	24.44	9 120	7	680.0	35.30	14 340
3	623.6	32.07	10 670	8	724.0	38.70	15 960
4	647.0	32.46	11 160	9	757.1	39.63	18 000
5	674.0	31.15	11 900	10	706.8	46.68	19 300

请用手工与软件两种方式对该社区家庭对该商品的消费需求支出作二元线性回归分析，其中手工方式要求以矩阵表达式进行运算。

(1) 估计回归方程的参数及随机干扰项的方差 $\hat{\sigma}^2$，计算 R^2 及 $\overline{R}^2$ 。

(2) 对方程进行 F 检验，对参数进行 t 检验，并构造参数 95%的置信区间。

(3) 如果商品单价变为 35 元，则某一月收入为 20 000 元的家庭的消费支出估计是多少？构造该估计值的 95%的置信区间。

12. 检验教材例 3.6.2 去掉 P_0 的模型

$$\ln\hat{Q} = 5.45 + 0.524\,0\ln X - 0.496\,3\ln P_1$$

是否仍满足零阶齐次性条件？

13. 下表列出了中国某年按行业分的全部制造业国有企业及规模以上制造业非国有企业的工业总产值 Y，资产合计 K 及职工人数 L。

序号	工业总产值 Y/亿元	资产合计 K/亿元	职工人数 L/万人	序号	工业总产值 Y/亿元	资产合计 K/亿元	职工人数 L/万人
1	3 722.70	3 078.22	113	9	370.18	363.48	16
2	1 442.52	1 684.43	67	10	1 590.36	2 511.99	66
3	1 752.37	2 742.77	84	11	616.71	973.73	58
4	1 451.29	1 973.82	27	12	617.94	516.01	28
5	5 149.30	5 917.01	327	13	4 429.19	3 785.91	61
6	2 291.16	1 758.77	120	14	5 749.02	8 688.03	254
7	1 345.17	939.10	58	15	1 781.37	2 798.90	83
8	656.77	694.94	31	16	1 243.07	1 808.44	33

续表

序号	工业总产值 Y/亿元	资产合计 K/亿元	职工人数 L/万人	序号	工业总产值 Y/亿元	资产合计 K/亿元	职工人数 L/万人
17	812.70	1 118.81	43	25	5 364.83	8 129.68	244
18	1 899.70	2 052.16	61	26	4 834.68	5 260.20	145
19	3 692.85	6 113.11	240	27	7 549.58	7 518.79	138
20	4 732.90	9 228.25	222	28	867.91	984.52	46
21	2 180.23	2 866.65	80	29	4 611.39	18 626.94	218
22	2 539.76	2 545.63	96	30	170.30	610.91	19
23	3 046.95	4 787.90	222	31	325.53	1 523.19	45
24	2 192.63	3 255.29	163				

设定模型为

$$Y = AK^{\alpha}L^{\beta}\mathrm{e}^{\mu}$$

(1) 利用上述资料，进行回归分析。

(2) 回答：中国该年的制造业总体呈现规模报酬不变状态吗？

14．继续 13 题，如果将 Cobb-Dauglas 生产函数设定为 $Y = AK^{\alpha}L^{\beta} + \mu$，则模型是非线性的，而且无法线性化。试给出这一设定的非线性普通最小二乘估计结果，并与第 13 题的估计结果进行比较。

第四章　经典单方程计量经济学模型：放宽基本假定的模型

4

前述计量经济学模型的回归分析，是在对线性回归模型提出若干基本假定的条件下，应用普通最小二乘法得到了无偏且有效的参数估计量。但是，在实际的计量经济学问题中，完全满足这些基本假定的情况并不多见。不满足基本假定的情况，称为**基本假定违背**，主要包括：

(1) 随机干扰项序列存在异方差性；

(2) 随机干扰项序列存在序列相关性；

(3) 解释变量之间存在多重共线性；

(4) 解释变量是随机变量且与随机干扰项相关。

除此之外，还有模型设定有偏误和解释变量的方差随着样本容量的增加而不断增加这两类基本假定的违背。

在进行计量经济学模型的回归分析时，必须对所研究对象是否满足普通最小二乘法下的基本假定进行检验，即检验是否存在一种或多种违背基本假定的情况，这种检验称为**计量经济学检验**。经过计量经济学检验发现出现一种或多种基本假定违背时，则不能直接使用普通最小二乘法进行参数估计，而必须采取补救措施或发展新的估计方法。本章主要讨论基本假定违背的前四种情形，后两种将分别在第五章与第八章中探讨。

§4.1　异方差性

对于模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i, \qquad i = 1,2,\cdots,n \qquad (4.1.1)$$

同方差性假设为

$$\operatorname{Var}(\mu_i \mid X_{i1}, X_{i2}, \cdots, X_{ik}) = \sigma^2, \qquad i = 1,2,\cdots,n$$

如果出现

$$\operatorname{Var}(\mu_i \mid X_{i1}, X_{i2}, \cdots, X_{ik}) = \sigma_i^2, \qquad i = 1,2,\cdots,n$$

即对于不同的样本点，随机干扰项的方差不再是常数，而是互不相同，则认为

PDG

出现了异方差性(heteroscedasticity)。

一、异方差的类型

同方差性假定的意义是指，每个μ_i围绕其零平均值的方差并不随解释变量X的变化而变化，不论解释变量是大还是小，每个μ_i的方差保持相同，即

$$\sigma_i^2 = 常数 \neq f(X_i)$$

在异方差的情况下，σ_i^2已不是常数，它随X的变化而变化，即

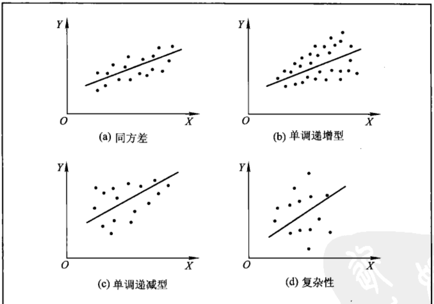
$$\sigma_i^2 = f(X_i)$$

异方差一般可归结为三种类型(图 4.1.1)：

(1) 单调递增型：σ_i^2随X的增大而增大；

(2) 单调递减型：σ_i^2随X的增大而减小；

(3) 复杂型：σ_i^2与X的变化呈复杂形式。



(a) 同方差　(b) 单调递增型

(c) 单调递减型　(d) 复杂性

图 4.1.1　异方差的类型

二、实际经济问题中的异方差性

在实际经济问题中，哪些情况容易出现异方差性？下面以三个例子加以说明。

PDG

例 4.1.1

以截面数据为样本研究居民家庭的储蓄行为

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

其中，Y_i 为第 i 个家庭的储蓄额，X_i 为第 i 个家庭的可支配收入。在该模型中，μ_i 项的方差为常数这一假定往往不符合实际情况。对高收入家庭来说，储蓄的差异较大；低收入家庭的储蓄则更有规律性(如为某一特定目的而储蓄)，差异较小。因此 μ_i 的方差往往随 X_i 的增加而增加，呈单调递增型变化。

例 4.1.2

以绝对收入假设为理论假设，以截面数据为样本建立居民消费函数(C)：

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

将居民按照收入 Y 等距离分成 n 组，取每组平均数为样本观测值。我们知道，一般情况下居民收入服从正态分布，所以处于每个收入组中的人数是不等的，处于中等收入组中的人数最多，处于两端收入组中的人数最少。人数多的组的平均数的误差小，人数少的组的平均数的误差大。所以样本观测值的观测误差随着解释变量观测值的不同而不同。如果样本观测值的观测误差构成随机干扰项的主要部分，那么对于不同的样本点，随机干扰项的方差互不相同，出现了异方差性。更进一步分析，在这个例子中，随机干扰项的方差随着解释变量 Y 的观测值的增大而呈 U 形变化，是复杂型的一种。

例 4.1.3

以某一行业的企业为样本建立企业生产函数模型

$$Y_i = \beta_0 A_i^{\beta_1} K_i^{\beta_2} L_i^{\beta_3} \mathrm{e}^{\mu_i}$$

产出量(Y)为被解释变量，选择资本(K)、劳动(L)、技术(A)等投入要素为解释变量，那么每个企业所处的外部环境对产出量的影响被包含在随机干扰项中。由于每个企业所处的外部环境对产出量的影响程度不同，造成了随机干扰项的异方差性。这时，随机干扰项的方差并不随某一个解释变量观测值的变化而呈规律性变化，为复杂型的一种。

一般经验告诉我们，对于采用截面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的差异较大，所以往往存在异方差性。

三、异方差性的后果

计量经济学模型一旦出现异方差性，如果仍采用普通最小二乘法估计模型参数，会产生一系列不良的后果。

1. 参数估计量非有效

根据§3.2 中关于参数估计量的无偏性和有效性的证明过程，可以看出，当计量经济学模型出现异方差性时，其普通最小二乘法参数估计量仍然具有线性性、无偏性，但不具有有效性。因为在有效性证明中利用了

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

而且，在大样本情况下，尽管参数估计量具有一致性，但仍然不具有渐近有效性。

2. 变量的显著性检验失去意义

在§3.3 关于变量的显著性检验中，构造了t统计量，它是建立在随机干扰项共同的方差σ^2不变而正确估计了参数方差$S_{\hat{\beta}_j}$的基础之上的。如果出现了异方差性，估计的$S_{\hat{\beta}_j}$出现偏误(偏大或偏小)，t检验失去意义。其他检验也是如此。

如对一元回归模型

$$Y_i=\beta_0+\beta_1X_i+\mu_i$$

的普通最小二乘估计有

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

可以证明，存在异方差的情况下正确的$\hat{\beta}_1$的方差应为

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{4.1.2}$$

而普通最小二乘法仍按下式给出$\hat{\beta}_1$的方差估计

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_i^2} \tag{4.1.3}$$

显然，只有同方差性满足时，(4.1.2)式与(4.1.3)式才会相同，否则普通最小二乘法给出的估计结果就会出现偏误，在有偏误的方差基础上构造的t检验也就失去了意义。

3. 模型的预测失效

一方面，由于上述后果，使得模型不具有良好的统计性质；另一方面，在预测值的置信区间中也包含有参数方差的估计量$S_{\hat{\beta}_j}$。所以，当模型出现异方差性时，仍然使用普通最小二乘估计量，将导致预测区间偏大或偏小，预测功能失效。

四、异方差性的检验

异方差性的检验方法是计量经济学中一个重要的课题。在一些计量经济学教科书和文献中，可以见到十多种检验方法，如图示检验法、等级相关系数法、戈里瑟(Gleiser)检验、巴特列特检验、G-Q 检验等，很难说哪种方法是最好的。这些方法尽管不同，但存在一个共同的思路。正如上面所指出的，**异方差性，即相对于不同的样本点，也就是相对于不同的解释变量观测值，随机干扰项具有不同的方差，那么检验异方差性，也就是检验随机干扰项的方差与解释变量观测值之间的相关性**。各种检验方法就是在这个思路下发展起来的。

问题在于用什么来表示随机干扰项的方差。一般的处理方法是首先采用普通最小二乘法估计模型，以求得随机干扰项的估计量(注意，该估计量是不严格的)，称之为“近似估计量”，用$\tilde{e}_i$表示。于是有

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

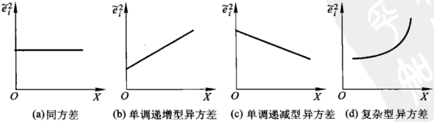
$$\tilde{e}_i = Y_i - (\hat{Y}_i)_{\mathrm{OLS}} \tag{4.1.4}$$

即用$\tilde{e}_i^2$来表示随机干扰项的方差。

下面有选择地介绍几种异方差性的检验方法。

1. 图示检验法

既可用 Y-X 的散点图进行判断，也可用某一个$\tilde{e}_i^2 - X$的散点图进行判断。对前者看是否存在明显的散点扩大、缩小或复杂型趋势(即不在一个固定的带型域中)，如图 4.1.1 所示；对后者看是否形成一条斜率为零的直线，如图 4.1.2 所示。



(a)同方差 (b) 单调递增型异方差 (c) 单调递减型异方差 (d) 复杂型异方差

图 4.1.2 不同异方差类型

图示检验法只能进行大概的判断，其他的统计检验方法则更为严格。

2. 帕克(Park)检验与戈里瑟检验

帕克检验与戈里瑟检验的基本思想是：以$\tilde{e}_i^2$或$|\tilde{e}_i|$为被解释变量，以原模型的某一解释变量X_j为解释变量，建立如下方程：

$$\tilde{e}_i^2 = f(X_{ij}) + \varepsilon_i$$

或

$$|\tilde{e}_i| = f(X_{ij}) + \varepsilon_i$$

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

$$f(X_{ij}) = \sigma^2 X_{ij}^{\alpha} \mathrm{e}^{\varepsilon_i}$$

或

$$\ln\ \tilde{e}_i^2 = \ln\ \sigma^2 + \alpha \ln\ X_{ij} + \varepsilon_i$$

进行检验，若α在统计上显著地异于零，表明存在异方差性。

当然，由于$f(X_j)$的具体形式未知，因此需要进行各种形式的试验。

3. G-Q(Goldfeld-Quandt)检验

帕克检验与戈里瑟检验的困难在于需要选择不同的解释变量，尝试各种不同的函数形式，进行多次反复试验，并且在进行试验的回归模型中，其随机干扰项本身就可能不满足普通最小二乘的经典假设。G-Q 检验则可同时克服这两大困难。

G-Q 检验以F检验为基础，适用于样本容量较大，异方差为单调递增或单调递减的情况。其基本思想是：先按某一解释变量对样本排序，再将排序后的样本一分为二，对两个子样分别进行普通最小二乘回归，然后利用两个子样的残差平方和之比构造F统计量进行异方差检验。G-Q 检验的步骤可描述如下。

(1) 将n组样本观测值按某一被认为有可能引起异方差的解释变量观测值的大小排队。

(2) 将序列中间的大约$c = \dfrac{n}{4}$个观测值除去，并将剩下的观测值划分为较小与较大的容量相同的两个子样本，每个子样样本容量均为$\dfrac{n-c}{2}$。

(3) 对每个子样分别进行普通最小二乘回归，并计算各自的残差平方和。分别用$\sum \tilde{e}_{1i}^2$与$\sum \tilde{e}_{2i}^2$表示较小与较大的残差平方和(自由度均为$\dfrac{n-c}{2} - k - 1$)。

(4) 在同方差性假定下，构造如下满足F分布的统计量：

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

(5) 给定显著性水平α，确定 F 分布表中相应的临界值$F_\alpha(v_1,v_2)$。若$F>F_\alpha(v_1,v_2)$，则拒绝同方差性假设，表明存在异方差性。当然，还可根据两个残差平方和对应的子样的顺序判断是单调递增异方差还是单调递减异方差。

4. 怀特(White)检验

G–Q 检验需要按某一被认为有可能引起异方差的解释变量观测值的大小排序，因此，可能需对各个解释变量进行轮流试验，而且，该方法只能检验单调递增或单调递减型异方差。怀特检验则不需要排序，且对任何形式的异方差都适用。

下面以两个解释变量的回归模型为例说明怀特检验的基本思想与步骤。

假设回归模型为

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\mu_i$$

可先对该模型作普通最小二乘回归，并得到$\tilde{e}_i^2$，然后作如下辅助回归：

$$\tilde{e}_i^2=\alpha_0+\alpha_1X_{1i}+\alpha_2X_{2i}+\alpha_3X_{1i}^2+\alpha_4X_{2i}^2+\alpha_5X_{1i}X_{2i}+\varepsilon_i$$

可以证明，在同方差性假设下，从该辅助回归得到的可决系数R^2与样本容量n的乘积，渐近地服从自由度为辅助回归方程中解释变量个数的χ^2分布：

$$nR^2\sim\chi^2$$

则可在大样本下，对统计量nR^2进行相应的χ^2检验。

需要注意的是，辅助回归仍是检验$\tilde{e}_i^2$与解释变量可能的组合的显著性，因此，辅助回归方程中还可引入解释变量的更高次方。如果存在异方差性，则表明$\tilde{e}_i^2$确与解释变量的某种组合有显著的相关性，这时往往显示出有较大的可决系数R^2，并且某一参数的t检验值较大。当然，在多元回归中，由于辅助回归方程中可能有太多解释变量，从而使自由度减少，有时可去掉交叉项。

五、异方差的修正

1. 加权最小二乘法(WLS)

如果模型被证明存在异方差性，则需要发展新的方法估计模型，最常用的方法是**加权最小二乘法**(Weighted Least Squares, WLS)。

加权最小二乘法是对原模型加权，使之变成一个新的不存在异方差性的模型，然后采用普通最小二乘法估计其参数。加权的基本思想是：在采用普通最

小二乘法时，对较小的残差平方e_i^2赋予较大的权数，对较大的e_i^2赋予较小的权数，以对残差提供的信息的重要程度作一番校正，提高参数估计的精度。

加权最小二乘法就是对加了权重的残差平方和实施普通最小二乘法：

$$\sum w_i e_i^2 = \sum w_i [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)]^2 \tag{4.1.5}$$

其中，w_i为权数。

例如，如果在检验过程中已经知道

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

即随机干扰项的方差与解释变量X_j之间存在相关性，那么可以用$\sqrt{f(X_j)}$去除原模型，使之变成如下形式的新模型：

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

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

在该模型中，存在

$$\mathrm{Var}\left[\frac{1}{\sqrt{f(X_{ji})}} \mu_i\right] = \left[\frac{1}{\sqrt{f(X_{ji})}}\right]^2 \mathrm{Var}(\mu_i)$$

$$= \frac{1}{f(X_{ji})} f(X_{ji})\sigma^2 = \sigma^2$$

即满足同方差性，于是可以用普通最小二乘法估计其参数，得到关于参数$\beta_0, \beta_1, \cdots, \beta_k$的无偏的、有效的估计量。这就是加权最小二乘法，在这里权数就是$\dfrac{1}{\sqrt{f(X_{ji})}}$。

加权最小二乘法具有比普通最小二乘法更普遍的意义，或者说普通最小二乘法只是加权最小二乘法中权恒取 1 时的一种特殊情况。从此意义看，加权最小二乘法也称为**广义最小二乘法**(generalized least squares, GLS)。

实施加权最小二乘法的关键是寻找适当的“权”，或者说是寻找模型中随机干扰项μ的方差与解释变量间的适当的函数形式。如果发现

$$\mathrm{Var}(\mu_i \mid X_{i1}, X_{i2}, \cdots, X_{ik}) = \sigma^2 f(X_{i1}, X_{i2}, \cdots, X_{ik})$$

则加权最小二乘法中的权即为$1/\sqrt{f(X_{i1}, X_{i2}, \cdots, X_{ik})}$。但如何寻找$\mu$的方差与各$X$间的函数关系呢？帕克检验指出可以进行各种形式的尝试，但下面给出一种相对灵活、有着广泛应用的方法。

假设μ的方差具有如下指数函数的形式：

$$\mathrm{Var}(\mu_i \mid X_{i1},\cdots,X_{ik})=\sigma^2\exp(\alpha_0+\alpha_1X_{i1}+\cdots+\alpha_kX_{ik}) \tag{4.1.6}$$

则可等价地写出

$$\mu_i^2=\sigma^2\exp(\alpha_0+\alpha_1X_{i1}+\alpha_2X_{i2}+\cdots+\alpha_kX_{ik})\varepsilon_i$$

其中，ε_i 可看成是条件均值为 1 的随机项。如果假设 ε_i 与各 X 独立，进一步有

$$\ln(\mu_i^2)=\delta_0+\alpha_1X_{i1}+\alpha_2X_{i2}+\cdots+\alpha_kX_{ik}+v_i \tag{4.1.7}$$

其中，v_i 为独立于各 X，且条件均值为 0 的随机项。由于(4.1.7)式满足普通最小二乘法的基本假设，当用可观测的值 $\tilde{e}_i$ 代替不可观测的 μ_i 时，用普通最小二乘法估计

$$\ln(\tilde{e}_i^2)=\delta_0+\alpha_1X_{i1}+\alpha_2X_{i2}+\cdots+\alpha_kX_{ik}+v_i \tag{4.1.8}$$

即可得到各 α_j 的无偏、一致且有效的估计 $\hat{\alpha}_j$（$j=1,2,\cdots,k$）。于是得到 μ 的方差估计：

$$\hat{\sigma}_i^2=\hat{\mu}_i^2=\hat{f}_i=\exp(\hat{\delta}_0+\hat{\alpha}_1X_{i1}+\hat{\alpha}_2X_{i2}+\cdots+\hat{\alpha}_kX_{ik}) \tag{4.1.9}$$

从而，估计的权为

$$\hat{w}_i=1/\hat{\sigma}_i=1/\sqrt{\hat{f}_i}=1/\sqrt{\exp(\hat{\delta}_0+\hat{\alpha}_1X_{i1}+\hat{\alpha}_2X_{i2}+\cdots+\hat{\alpha}_kX_{ik})} \tag{4.1.10}$$

最后需指出，(4.1.6)式的指数函数中只列出了各解释变量 X 的水平项，可根据估计的显著性，对各 X 进行取舍；此外，还可根据需要加入适当的 X 的高次方项。

由于加权最小二乘法中的权，或者说原模型中 μ 的方差与各 X 间适当的函数关系是估计出来的，因此这一广义最小二乘法也称为**可行的广义最小二乘法**(feasible GLS, FGLS)，由广义最小二乘法得到的原模型中的估计量称为**可行的广义最小二乘估计量**，广义最小二乘估计量具有 BLUE 的特征。

2．异方差稳健标准误法

加权最小二乘法的关键是寻找模型中随机扰动项 μ 的方差与解释变量间的适当的函数形式，而这并非一件易事。在有些情况下很难得到正确的 μ 的方差与解释变量的函数关系式，这时，可采用下面介绍的异方差稳健标准误法来消除异方差的存在带来的不良后果。

由于回归模型随机干扰项出现异方差时，普通最小二乘法只是影响到了参数估计量方差或标准差的正确估计，从而无法保证普通最小二乘估计量的有效性，但并不影响估计量的无偏性与一致性。因此，另一种针对异方差的修正的估计方法是：仍采用普通最小二乘估计量，但修正相应的方差。

如何修正普通最小二乘估计量相应的方差呢？怀特 1980 年提出的方法是，用普通最小二乘法估计的残差的平方 $\tilde{e}_i^2$ 作为相应 σ_i^2 的代表。如在一元线性回归中，估计的斜率 $\hat{\beta}_1$ 正确的方差应为

$$\mathrm{Var}(\hat{\beta}_1)=\frac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{4.1.2}$$

于是用普通最小二乘法估计的残差的平方 $\tilde{e}_i^2$ 作为相应 σ_i^2 的代表，即用下式作为 $\mathrm{Var}(\hat{\beta}_1)$ 的估计：

$$\frac{\sum x_i^2\tilde{e}_i^2}{\left(\sum x_i^2\right)^2} \tag{4.1.11}$$

怀特证明了大样本下，(4.1.11)式是(4.1.2)式的一致估计。(4.1.11)式的平方根称为 $\hat{\beta}_1$ 的**异方差稳健标准误**（heteroscedasticity-robust standard error），这种估计方法也被称为异方差稳健标准误法。

在存在异方差时，异方差稳健标准误法虽不能得到有效的估计量，但由于可以得到普通最小二乘估计量正确的方差估计，从而使得以估计量方差为基础的各种统计检验不再失效、建立的预测区间也更加可信，因此异方差稳健标准误法就成为在不能较好地实施加权最小二乘法时，消除异方差性不良后果的主要手段。

六、案例——中国农村居民人均消费函数

例 4.1.4

中国农村居民人均消费支出主要由人均纯收入来决定。农村人均纯收入除从事农业经营的收入外，还包括从事其他产业的经营性收入以及工资性收入、财产收入和转移支付收入等。为了考察从事农业经营的收入和其他收入对中国农村居民消费支出增长的影响，可使用如下双对数模型：

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

其中，Y 表示农村家庭人均消费支出，X_1 表示从事农业经营的纯收入，X_2 表示其他来源的纯收入。表 4.1.1 列出了中国内地 2006 年各地区农村居民家庭人均纯收入及消费支出的相关数据。

表 4.1.1　中国内地 2006 年各地区农村居民家庭人均纯收入与消费支出　单位：元

地区	人均消费支出 Y	从事农业经营的纯收入 X_1	其他来源的纯收入 X_2	地区	人均消费支出 Y	从事农业经营的纯收入 X_1	其他来源的纯收入 X_2
北　京	5 724.5	958.3	7 317.2	湖　北	2 732.5	1 934.6	1 484.8
天　津	3 341.1	1 738.9	4 489.0	湖　南	3 013.3	1 342.6	2 047.0
河　北	2 495.3	1 607.1	2 194.7	广　东	3 886.0	1 313.9	3 765.9
山　西	2 253.3	1 188.2	1 992.7	广　西	2 413.9	1 596.9	1 173.6

续表

地区	人均消费支出 Y	从事农业经营的纯收入 X_1	其他来源的纯收入 X_2
内蒙古	2 772.0	2 560.8	781.1
辽　宁	3 066.9	2 026.1	2 064.3
吉　林	2 700.7	2 623.2	1 017.9
黑龙江	2 618.2	2 622.9	929.5
上　海	8 006.0	532	8 606.7
江　苏	4 135.2	1 497.9	4 315.3
浙　江	6 057.2	1 403.1	5 931.7
安　徽	2 420.9	1 472.8	1 496.3
福　建	3 591.4	1 691.4	3 143.4
江　西	2 676.6	1 609.2	1 850.3
山　东	3 143.8	1 948.2	2 420.1
河　南	2 229.3	1 844.6	1 416.4
海　南	2 232.2	2 213.2	1 042.3
重　庆	2 205.2	1 234.1	1 639.7
四　川	2 395.0	1 405	1 597.4
贵　州	1 627.1	961.4	1 023.2
云　南	2 195.6	1 570.3	680.2
西　藏	2 002.2	1 399.1	1 035.9
陕　西	2 181.0	1 070.4	1 189.8
甘　肃	1 855.5	1 167.9	966.2
青　海	2 179.0	1 274.3	1 084.1
宁　夏	2 247.0	1 535.7	1 224.4
新　疆	2 032.4	2 267.4	469.9

注：从事农业经营的纯收入由从事第一产业的经营总收入与从事第一产业的经营支出之差计算，其他来源的纯收入由总纯收入减去从事农业经营的纯收入后得到。

资料来源：《中国农村住户调查年鉴(2007)》、《中国统计年鉴(2007)》。

普通最小二乘法的估计结果如下：

$$\ln\hat{Y} = 3.266 + 0.150\,2\ln X_1 + 0.477\,5\ln X_2$$

$$(3.14)\quad (1.38)\qquad (9.25)$$

$$R^2 = 0.779\,8 \qquad \text{D.W.}=1.78 \qquad F=49.60 \qquad \text{RSS}=0.835\,7$$

估计结果显示，即使在 10%的显著性水平下，都不拒绝从事农业经营的纯收入前参数为零的假设，因此可以认为，其他来源的纯收入而不是从事农业经营的纯收入的增长，对农户人均消费支出的增长更有刺激作用。下面对该模型进行异方差性检验。

可以认为不同地区农村人均消费支出的差别主要来源于非农经营收入及工资收入、财产收入等其他收入的差别上，因此，如果存在异方差性，则可能是 X_2 引起的。模型普通最小二乘回归得到的残差平方项 $\tilde{e}_i^2$ 与 $\ln X_2$ 的散点图表明(图 4.1.3)，存在递增型异方差性。

再进行进一步的统计检验。首先采用 G-Q 检验。

将原始数据按 X_2 排成升序，去掉中间的 7 个数据，得两个容量为 12 的子样本。对两个子样本分别作普通最小二乘回归，求各自的残差平方和 RSS_1 和 RSS_2：

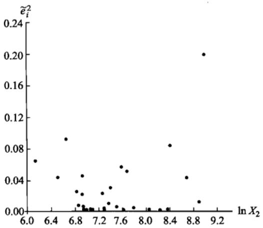


图 4.1.3 异方差性检验图

子样本 1：

$$\ln\hat{Y} = 3.14 + 0.398\ln X_1 + 0.235\ln X_2$$

(2.80) (5.05) (2.14)

$$R^2 = 0.739\,7, \qquad \mathrm{RSS}_1 = \sum e_i^2 = 0.070\,2$$

子样本 2：

$$\ln\hat{Y} = 3.99 - 0.114\ln X_1 + 0.620\ln X_2$$

(2.12) (−0.71) (5.55)

$$R^2 = 0.876\,9, \qquad \mathrm{RSS}_2 = \sum e_i^2 = 0.191\,2$$

计算 F 统计量：$F = \mathrm{RSS}_2 / \mathrm{RSS}_1 = 0.191\,2/0.070\,2 = 2.73$

在 5%与 10%的显著性水平下，自由度为(9,9)的 F 分布的临界值分别为 $F_{0.05}=3.18$ 与 $F_{0.10}=2.44$。因此 5%显著性水平下不拒绝两组子样方差相同的假设，但在 10%的显著性水平下拒绝。

再采用怀特检验。记 $\tilde{e}_i^2$ 为对原始模型进行普通最小二乘回归得到的残差平方项，将其与 X_1，X_2 及其平方项与交叉项作辅助回归，得

$$\tilde{e}^2 = 10.24 - 2.33\ln X_1 - 0.46\ln X_2 + 0.15(\ln X_1)^2 + 0.02(\ln X_2)^2 + 0.02\ln X_1 \ln X_2$$

(1.87) (−2.09) (−1.01) (2.56) (1.58) (0.47)

$$R^2 = 0.662\,9$$

怀特统计量 $nR^2 = 31 \times 0.662\,9 = 20.55$，该值大于 5%显著性水平下、自由度为 5 的 χ^2 分布的相应临界值 $\chi^2_{0.05} = 11.07$，因此，拒绝同方差的原假设。

去掉交叉项后的辅助回归结果为

$$\tilde{e}^2 = 7.763 - 1.851\ln X_1 - 0.258\ln X_2 + 0.126(\ln X_1)^2 + 0.017(\ln X_2)^2$$

$$(5.64)(-4.14)\qquad(-1.64)\qquad(4.10)\qquad(1.67)$$

$$R^2=0.659\,9$$

显然，其他收入X_2项与X_2的平方项的参数的t检验是显著的，且怀特统计量$nR^2=31\times0.659\,9=20.46$，因此，在5%的显著性水平下，仍是拒绝同方差这一原假设。

下面我们采用加权最小二乘法对原模型进行回归。

经试算，发现原模型普通最小二乘回归残差平方项的对数$\ln\tilde{e}_i^2$与$\ln X_2$及其平方项有显著的回归关系：

$$\ln\tilde{e}^2=93.20-25.981\ln X_2+1.701(\ln X_2)^2$$

$$(2.47)\qquad(-2.63)\qquad(2.65)$$

$$R^2=0.202\,2$$

于是，用$w_i=1/\sqrt{\hat{f}_i}=1/\sqrt{\exp(93.20-25.981\ln X_{i2}+1.701(\ln X_{i2})^2)}$作为适当的权，对原模型进行加权最小二乘估计（WLS）得到

$$\ln\hat{Y}=2.34+0.317\ln X_1+0.429\ln X_2$$

$$(3.23)\quad(3.82)\qquad(9.67)$$

$$R^2=0.782\,7\qquad \text{D.W.}=1.36\qquad F=50.40$$

可以看出，$\ln X_1$参数的t统计量的值有了显著的改进，这表明即使在1%显著性水平下，都不能拒绝从事农业生产带来的纯收入对农户人均消费支出有着显著影响的假设。当然，虽然$\ln X_1$的参数值较普通最小二乘估计有了较大程度的提高，但仍没有$\ln X_2$的参数估计值大，说明其他来源的纯收入确实比来自农业经营的纯收入对农户人均消费支出的影响更大一些。

下面我们检验是否经加权的回归模型已不存在异方差性。记经w_i加权的回归模型为

$$w\ln Y=\beta_0 w+\beta w\ln X_1+\beta_2 w\ln X_2+\mu$$

该模型的普通最小二乘回归结果为

$$\widehat{w\ln Y}=2.34w+0.317w\ln X_1+0.429w\ln X_2$$

记该回归模型的残差估计的平方为$\tilde{\tilde{e}}^2$，将其与w，$w\ln X_1$，$w\ln X_2$及其平方项作辅助回归，得

$$\tilde{\tilde{e}}^2=6.28-8.22w+0.61w\ln X_1+0.41w\ln X_2+0.26w^2$$
$$-0.003(w\ln X_1)^2-0.001(w\ln X_2)^2$$

$$R^2=0.265\,4$$

怀特统计量$nR^2=31\times0.265\,4=8.23$，该值在小于5%显著性水平下、自由度为6的$\chi^2$分布的相应临界值$\chi^2_{0.05}=12.59$，因此，不拒绝同方差的原假设。

最后，给出异方差稳健标准误法修正的结果：

$$\ln\hat{Y}=3.266+0.150\ 2\ln X_1+0.477\ 5\ln X_2$$

$$(2.59)\qquad(0.94)\qquad(8.32)$$

$$R^2=0.779\ 8\qquad \text{D.W.}=1.78\quad F=49.60\qquad \text{RSS}=0.835\ 7$$

可以看出，估计的参数与普通最小二乘法的结果相同，只是由于参数的标准差得到了修正，从而使得 t 检验值与普通最小二乘法的结果不同。当然，这里异方差稳健标准误法得到的结论并没改变农业经营所带来的纯收入不影响农户人均消费支出这一原结论。

§4.2 序列相关性

多元线性回归模型的基本假设之一是模型的随机干扰项相互独立或不相关。如果模型的随机干扰项违背了相互独立的基本假设，称为存在**序列相关性**(serial correlation)。

一、序列相关性

对于模型

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\cdots+\beta_kX_{ik}+\mu_i\,,\qquad i=1,2,\cdots,n \tag{4.2.1}$$

在其他假设仍成立的条件下，随机干扰项序列相关即意味着

$$\text{Cov}(\mu_i,\mu_j)=\text{E}(\mu_i\mu_j)\neq 0$$

或

$$\text{Var}(\boldsymbol{\mu})=\text{E}(\boldsymbol{\mu}\,\boldsymbol{\mu}')=\begin{pmatrix}\sigma^2 & \cdots & \text{E}(\mu_1\mu_n)\\ \vdots & & \vdots\\ \text{E}(\mu_n\mu_1) & \cdots & \sigma^2\end{pmatrix}=\begin{pmatrix}\sigma^2 & \cdots & \sigma_{1n}\\ \vdots & & \vdots\\ \sigma_{n1} & \cdots & \sigma^2\end{pmatrix}$$

$$=\sigma^2\boldsymbol{\Omega}\neq\sigma^2\boldsymbol{I} \tag{4.2.2}$$

如果仅存在

$$\text{E}(\mu_i\mu_{i+1})\neq 0\,,\qquad i=1,2,\cdots,n-1 \tag{4.2.3}$$

则称为**一阶序列相关**或**自相关**(autocorrelation)，这是最常见的一种序列相关问题。自相关往往可写成如下形式：

$$\mu_i=\rho\mu_{i-1}+\varepsilon_i\,,\qquad -1<\rho<1 \tag{4.2.4}$$

其中 ρ 称为**自协方差系数**(coefficient of autocovariance)或**一阶自相关系数**(first-order coefficient of autocorrelation)，ε_i 是满足以下标准普通最小二乘法假定的随机干扰项：

$$\text{E}(\varepsilon_i)=0\,,\qquad \text{Var}(\varepsilon_i)=\sigma^2\,,\qquad \text{Cov}(\varepsilon_i,\varepsilon_{i-s})=0\,(s\neq 0)$$

由于**序列相关性经常出现在以时间序列数据为样本的模型中**，因此，本节将代表不同样本点的下标 i 用 t 表示。

二、实际经济问题中的序列相关性

实际经济问题中，序列相关性产生的原因主要来自以下三个方面。

1. 经济变量固有的惯性

大多数经济时间数据都有一个明显的特点，就是它的惯性，表现在时间序列数据不同时间的前后关联上。例如，以绝对收入假设为理论假设，以时间序列数据为样本建立居民总消费函数模型：

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

我们知道，一般情况下居民总消费(C)除受总收入(Y)影响外，还受其他因素影响，如消费习惯等。但这些因素没有包括在解释变量中，它们对消费量的影响被包含在随机干扰项中。如果该项影响构成随机干扰项的主要部分，则可能出现序列相关性，即对于不同的年份，由于消费习惯等因素的惯性，它们对消费量的影响也是具有内在联系的。于是在不同的样本点之间，随机干扰项出现了相关，从而产生了序列相关性。更进一步分析，在这个例子中，随机干扰项之间表现为正相关。

又比如，在如下农产品供给模型中：

$$Q_t = \beta_0 + \beta_1 P_{t-1} + \mu_t$$

农产品供给(Q)对价格(P)的反映本身存在一个滞后期，这意味着，农户在第 t 年的过量生产(使该期价格下降)很可能导致在第 t+1 年削减产量；反之，第 t 年的减产又导致第 t+1 年的增产。这时，随机干扰项往往表现出负相关的特征。

2. 模型设定的偏误

所谓**模型设定偏误**(specification error)是指所设定的模型“不正确”，主要表现在模型中丢掉了重要的解释变量或模型函数形式有偏误。例如，本来应该估计的模型为

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \mu_t$$

但在模型设定中作了下述回归：

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + v_t$$

因此，该式中 $v_t = \beta_3 X_{t3} + \mu_t$。于是在 X_3 确实影响 Y 的情况下，这种模型设定的偏误往往是导致随机干扰项中一个重要的系统性影响因素，使其呈序列相关性。

又如，如果真实的边际成本回归模型应为

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t1}^2 + \mu_t$$

其中，Y 代表边际成本，X_1 代表产出量。但建模时设立了如下模型：

$$Y_t = \beta_0 + \beta_1 X_{t1} + v_t$$

因此，由于 $v_t = \beta_2 X_{t1}^2 + \mu_t$，包含了产出的平方对随机干扰项的系统性影响，随机干扰项也呈现序列相关性。

3. 数据的“编造”

书法的趣味地

在实际经济问题中，有时为了需要，有些数据是通过已知数据生成的。因此，新生成的数据与原数据间就有了内在的联系，表现出序列相关性。例如，季度数据来自月度数据的简单平均，这种平均的计算减弱了每月数据的波动而引进了数据中的匀滑性，这种匀滑性本身就能使随机干扰项中出现系统性的因素，从而出现序列相关性。另外，两个时间点之间的“内插”技术也会导致随机干扰项的序列相关性。

书法中奥

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。

三、序列相关性的后果

计量经济学模型一旦出现序列相关性，如果仍采用普通最小二乘法估计模型参数，会产生许多不良后果。

即时通信

1. 参数估计量非有效

根据普通最小二乘估计

随机应变

人的思维本来就是超文本 <发散.

根据普通最小二乘估计中关于参数估计量的无偏性和有效性的证明过程可以看出，当计量经济学模型出现序列相关性时，其普通最小二乘参数估计量仍然具有线性无偏性，但不具有有效性。因为在有效性证明中利用了

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

即同方差性和相互独立性条件。而且，在大样本情况下，参数估计量虽然具有一致性，但仍然不具有渐近有效性。

规

即兴创造

2. 变量的显著性检验失去意义

化成身体的一部分

在变量的显著性检验中，t 统计量是建立在参数方差正确估计基础之上的，这只有当随机干扰项具有同方差性和相互独立性时才能成立。如果存在序列相关性，估计的参数方差 $S_{\hat{\beta}_j}$ 出现偏误(偏大或偏小)，t 检验就失去意义。其他检验也是如此。

书法对理的突灵性.

如对一元回归模型

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

的普通最小二乘估计有

邯郸学步

PDG

$$\hat{\beta}_1 = \beta_1 + \sum k_t \mu_t = \beta_1 + \frac{\sum x_t \mu_t}{\sum x_t^2}$$

可以证明，存在(4.2.4)式所示的一阶序列相关的情况下正确的 $\hat{\beta}_1$ 的方差应为

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_t^2} + \frac{2\sigma^2}{\sum x_t^2}\left[\rho\frac{\sum\limits_{t=1}^{n-1} x_t x_{t+1}}{\sum x_t^2} + \rho^2\frac{\sum\limits_{t=1}^{n-2} x_t x_{t+2}}{\sum x_t^2} + \cdots + \rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right] \qquad (4.2.5)$$

而普通最小二乘法仍按下式给出 $\hat{\beta}_1$ 的方差估计

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_t^2} \qquad (4.2.6)$$

显然，只有序列无关性满足时，(4.2.5)式与(4.2.6)式才会相同，否则普通最小二乘法给出的估计结果就会出现偏误，在有偏误的方差基础上构造的 t 检验也就失去了意义。

3. 模型的预测失效

区间预测与参数估计量的方差有关，在方差估计有偏误的情况下，预测估计就不准确，预测精度降低。所以，当模型出现序列相关性时，它的预测功能失效。

四、序列相关性的检验

序列相关性的检验方法有多种，如冯诺曼比检验法、回归检验法、D.W.检验法等。**这些检验方法的共同思路是：首先采用普通最小二乘法估计模型，以求得随机干扰项的“近似估计量”，用 $\tilde{e}_t$ 表示：**

$$\tilde{e}_t = Y_t - (\hat{Y}_t)_{\mathrm{OLS}}$$

然后通过分析这些“近似估计量”之间的相关性以达到判断随机干扰项是否具有序列相关性的目的。下面介绍几种常用的检验方法。

1. 图示法

由于残差 $\tilde{e}_t$ 可以作为 μ_t 的估计，因此，如果 μ_t 存在序列相关性，必然会由残差项 $\tilde{e}_t$ 反映出来，因此可利用 $\tilde{e}_t$ 的变化图形来判断随机干扰项的序列相关性，如图 4.2.1 所示。

2. 回归检验法

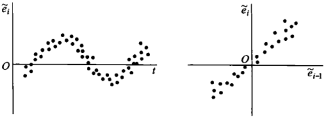
以 $\tilde{e}_t$ 为被解释变量，以各种可能的相关量，诸如 $\tilde{e}_{t-1}$，$\tilde{e}_{t-2}$，$\tilde{e}_t^2$ 等为解释变量，建立各种方程：

$$\tilde{e}_t = \rho\tilde{e}_{t-1} + \varepsilon_t, \qquad t=2,\cdots,n$$

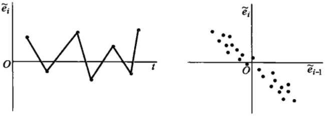
$$\tilde{e}_t = \rho_1\tilde{e}_{t-1} + \rho_2\tilde{e}_{t-2} + \varepsilon_t, \qquad t=3,\cdots,n$$

…………

PDG



(a) 正序列相关（正自相关）



(b) 负序列相关（负自相关）

图 4.2.1 残差项的序列相关性

对方程进行估计并进行显著性检验，如果存在某一种函数形式，使得方程显著成立，则说明原模型存在序列相关性。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。

3. D.W.检验法

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

(1) 解释变量 X 非随机；

(2) 随机干扰项 μ_t 为一阶自回归形式：

$$\mu_t = \rho\mu_{t-1} + \varepsilon_t$$

(3) 回归模型中不应含有滞后应变量作为解释变量，即不应出现下列形式：

$$Y_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \gamma Y_{t-1} + \mu_t$$

(4) 回归模型含有截距项。

杜宾和瓦森针对原假设 $H_0:\rho = 0$， 即 μ_t 不存在一阶自回归，构造如下统计量：

$$\text{D.W.} = \frac{\sum_{t=2}^{n}(\tilde{e}_t - \tilde{e}_{t-1})^2}{\sum_{t=1}^{n}\tilde{e}_t^2} \tag{4.2.7}$$

该统计量的分布与出现和给定样本中的 X 值有复杂的关系，因此其精确的分布很难得到。但他们成功地导出了临界值的上限 d_U 与下限 d_L，且这些上下限只与样本容量 n 和解释变量的个数 k 有关，而与解释变量的取值无关。因此，在检验时，只须计算该统计量的值，再根据样本容量 n 和解释变量数目 k 查 D.W.分布表，得到临界值 d_L 和 d_U，然后按照下列准则考察计算得到的 D.W.值，以判断模型的自相关状态：

若 0<D.W.<d_L，则存在正自相关；

若 d_L<D.W.<d_U，则不能确定；

若 d_U<D.W.<4−d_U，则无自相关；

若 4−d_U<D.W.<4−d_L 则不能确定；

若 4−d_L<D.W.<4，则存在负自相关。

也就是说，当 D.W.值在 2 附近时，模型不存在一阶自相关。其证明过程如下：

展开 D.W.统计量：

$$\text{D.W.}=\frac{\sum_{t=2}^{n}\tilde{e}_t^2+\sum_{t=2}^{n}\tilde{e}_{t-1}^2-2\sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1}}{\sum_{t=1}^{n}\tilde{e}_t^2} \tag{4.2.8}$$

当 n 较大时，$\sum_{t=2}^{n}\tilde{e}_t^2,\sum_{t=2}^{n}\tilde{e}_{t-1}^2,\sum_{t=1}^{n}\tilde{e}_t^2$ 大致相等，则(4.2.8)式可以化简为

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

其中，$\frac{\sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1}}{\sum_{t=1}^{n}\tilde{e}_t^2}\approx\frac{\sum_{t=2}^{n}\tilde{e}_t\tilde{e}_{t-1}}{\sum_{t=2}^{n}\tilde{e}_t^2}=\rho$ 为一阶自相关模型(4.2.4)式的参数估计。

如果存在完全一阶正相关，则 $\rho\approx 1$，　$\text{D.W.}\approx 0$；

如果存在完全一阶负相关，则 $\rho\approx -1$，　$\text{D.W.}\approx 4$；

如果完全不相关，则 $\rho=0$，　$\text{D.W.}=2$。

从判断准则中看到，存在一个不能确定的 D.W.值区域，这是这种检验方法的一大缺陷。而且 D.W.检验只能检验一阶自相关，并且对存在滞后被解释变量的模型无法检验。

4. 拉格朗日乘数（LM）检验

拉格朗日乘数检验克服了 D.W.检验的缺陷，适合于高阶序列相关及模型中

存在滞后被解释变量的情形。它是由布劳殊(Breusch)与戈弗雷(Godfrey)于 1978 年提出的，也称为 GB 检验。

对于模型(4.2.1)式，如果怀疑随机干扰项存在 p 阶序列相关：

$$\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + \cdots + \rho_p\mu_{t-p} + \varepsilon_t \tag{4.2.9}$$

拉格朗日乘数检验就可用来检验如下受约束回归方程：

$$Y_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \rho_1\mu_{t-1} + \cdots + \rho_p\mu_{t-p} + \varepsilon_t \tag{4.2.10}$$

约束条件为

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_p = 0 \tag{4.2.11}$$

如果约束条件 H_0 为真，则 LM 统计量服从大样本下自由度为 p 的渐近 χ^2 分布：

$$\mathrm{LM} = nR^2 \sim \chi^2(p) \tag{4.2.12}$$

其中，n, R^2 分别为如下辅助回归的样本容量与可决系数：

$$\tilde{e}_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \rho_1\tilde{e}_{t-1} + \cdots + \rho_p\tilde{e}_{t-p} + \varepsilon_t \tag{4.2.13}$$

$\tilde{e}_t$ 为原模型(4.2.1)式经普通最小二乘估计的残差项。给定显著性水平 α，查自由度为 p 的 χ^2 分布的相应临界值 $\chi^2_\alpha(p)$，如果计算的 LM 统计量的值超过该临界值，则拒绝约束条件为真的原假设，表明可能存在直到 p 阶的序列相关性。在实际检验中，可从 1 阶、2 阶……逐次向更高阶检验，并用辅助回归(4.2.13)式中各 $\tilde{e}_t$ 前参数的显著性来帮助判断序列相关的阶数。

五、序列相关的补救

如果模型被检验证明存在序列相关性，则需要发展新的方法估计模型。与模型出现异方差的情形相类似，有两种解决途径：一是变换原模型为不存在序列相关的新模型，再采用普通最小二乘法估计，这就是所谓的**广义最小二乘法**和**广义差分法**(generalized difference method)；另一条途径是仍采用普通最小二乘法估计原模型，之后再对参数估计量的方差或标准差进行修正，称为**序列相关稳健估计法**（serial correlation-robust method）。

1. 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。

一般情况下，对于模型

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\mu} \tag{4.2.14}$$

如果存在序列相关性，同时存在异方差性，即有

PDG

$$\mathrm{Cov}(\boldsymbol{\mu},\boldsymbol{\mu}')=\mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}')=\begin{pmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n}\\ \vdots & \vdots & & \vdots\\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2\end{pmatrix}=\sigma^2\boldsymbol{\Omega}$$

显然，$\boldsymbol{\Omega}$ 是对称正定矩阵，因此存在一可逆矩阵 $\boldsymbol{D}$，使得

$$\boldsymbol{\Omega}=\boldsymbol{D}\boldsymbol{D}'$$

用 $\boldsymbol{D}^{-1}$ 左乘(4.2.14)式两边，得到一个新的模型：

$$\boldsymbol{D}^{-1}\boldsymbol{Y}=\boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{D}^{-1}\boldsymbol{\mu} \tag{4.2.15}$$

即

$$\boldsymbol{Y}_*=\boldsymbol{X}_*\boldsymbol{\beta}+\boldsymbol{\mu}_*$$

该模型具有同方差性和随机干扰项相互独立性。因为

$$\begin{aligned}\mathrm{E}(\boldsymbol{\mu}_*\boldsymbol{\mu}_*')&=\mathrm{E}[\boldsymbol{D}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}'(\boldsymbol{D}^{-1})']=\boldsymbol{D}^{-1}\mathrm{E}(\boldsymbol{\mu}\boldsymbol{\mu}')(\boldsymbol{D}^{-1})'\\&=\boldsymbol{D}^{-1}\sigma^2\boldsymbol{\Omega}(\boldsymbol{D}^{-1})'=\boldsymbol{D}^{-1}\sigma^2\boldsymbol{D}\boldsymbol{D}'(\boldsymbol{D}')^{-1}=\sigma^2\boldsymbol{I}\end{aligned}$$

于是，可以用普通最小二乘法估计模型(4.2.15)式，记参数估计量为 $\hat{\boldsymbol{\beta}}_*$，则

$$\begin{aligned}\hat{\boldsymbol{\beta}}_*&=(\boldsymbol{X}_*'\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*'\boldsymbol{Y}_*\\&=[\boldsymbol{X}'(\boldsymbol{D}^{-1})'\boldsymbol{D}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{D}^{-1})'\boldsymbol{D}^{-1}\boldsymbol{Y}\\&=(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y}\end{aligned} \tag{4.2.16}$$

这就是原模型(4.2.14)式的广义最小二乘估计量，是无偏的、有效的估计量。

由上面的推导过程可知，只要知道随机干扰项的方差-协方差矩阵 $\sigma^2\boldsymbol{\Omega}$，就可采用广义最小二乘法得到参数的最佳线性无偏估计量。然而若只有 n 个样本点，要对包括各 β_j 在内的 $\frac{n(n-1)}{2}+k+2$ 个未知参数进行估计是困难的。这就需要对随机干扰项自相关的结构事先给出必要的假设。最常见的是假设随机干扰项具有一阶序列相关性：

$$\mu_t=\rho\mu_{t-1}+\varepsilon_t,\qquad -1<\rho<1 \tag{4.2.17}$$

这时，可以证明(参见《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)

$$\mathrm{Var}(\mu_t)=\frac{1}{1-\rho^2}\sigma_\varepsilon^2=\sigma^2$$

$$\mathrm{Cov}(\mu_t,\mu_{t-s})=\rho^s\frac{1}{1-\rho^2}\sigma_\varepsilon^2=\rho^s\sigma^2$$

于是

$$\operatorname{Var}(\boldsymbol{\mu}) = \frac{\sigma_\varepsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix} = \sigma^2\boldsymbol{\Omega} \tag{4.2.18}$$

易知

$$\boldsymbol{\Omega}^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}$$

从而

$$\boldsymbol{D}^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\rho & 1 \end{pmatrix} \tag{4.2.19}$$

2. 广义差分法

广义差分法是一类克服序列相关性的有效方法，被广泛地采用。它是将原模型变换为满足普通最小二乘法的差分模型，再进行普通最小二乘估计。

如果原模型存在

$$\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + \cdots + \rho_p\mu_{t-p} + \varepsilon_t \tag{4.2.20}$$

可以将原模型变换为

$$\begin{aligned} & Y_t - \rho_1 Y_{t-1} - \cdots - \rho_p Y_{t-p} \\ & = \beta_0(1-\rho_1-\cdots-\rho_p) + \beta_1(X_{t1} - \rho_1 X_{t-1,1} - \cdots - \rho_p X_{t-p,1}) + \cdots \\ & \qquad + \beta_k(X_{tk} - \rho_1 X_{t-1,k} - \cdots - \rho_p X_{t-p,k}) + \varepsilon_t \\ & \qquad\qquad t = 1+p, 2+p, \cdots, n \end{aligned} \tag{4.2.21}$$

模型(4.2.21)式为广义差分模型，该模型不存在序列相关性问题。采用普通最小二乘法估计该模型得到的参数估计量，即为原模型参数无偏且有效的估计量。

需要指出的是，广义差分法就是前面讲述的广义最小二乘法，但是却损失

了部分样本观测值。例如，在一阶序列相关的情况下，广义差分是对下面的差分模型进行普通最小二乘回归：

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{t1} - \rho X_{t-1,1}) + \cdots + \beta_k(X_{tk} - \rho X_{t-1,k}) + \varepsilon_t\,, \quad t = 2,3,\cdots,n$$

或

$$Y_t^* = \beta_0(1-\rho) + \beta_1 X_{t1}^* + \cdots + \beta_k X_{tk}^* + \varepsilon_t\,, \qquad t = 2,3,\cdots,n \qquad (4.2.22)$$

这一变换相当于(4.2.19)式的 $\boldsymbol{D}^{-1}$ 去掉第一行后左乘原模型(4.2.14)式，即运用了广义最小二乘法，但第一次观测值被排除了。

尽管在大样本中广义差分法与广义最小二乘法的估计结果相近，但在小样本中，观测值的损失可能会对估计结果有所影响。因此，在广义差分变换中，有时需弥补这一损失。例如，在一阶序列相关情况下，对损失的第一次观测值可进行如下的**普莱斯-温斯特变换**(Prais-Winsten transformation)：

$$Y_1^* = \sqrt{1-\rho^2}\,Y_1\,, \qquad X_{1j}^* = \sqrt{1-\rho^2}\,X_{1j} \quad (\,j = 1,2,\cdots,k\,)$$

这样，广义差分法的估计结果完全等同于广义最小二乘估计量。

3. 随机干扰项相关系数的估计

无论应用广义最小二乘法，还是应用广义差分法，都必须已知不同样本点之间随机干扰项的相关系数 $\rho_1,\rho_2,\cdots,\rho_p$。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，但基本思路大都是采用普通最小二乘法估计原模型，得到随机干扰项的“近似估计值”，然后利用该“近似估计值”求得随机干扰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。下面介绍常用的科克伦-奥科特(Cochrane-Orcutt)迭代法。

首先，采用普通最小二乘法估计原模型，得到随机干扰项的“近似估计值”，以之作为方程(4.2.20)的样本观测值，采用普通最小二乘法估计该方程，得到 $\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_p$，作为随机干扰项的相关系数 $\rho_1,\rho_2,\cdots,\rho_p$ 的第一次估计值。然后，将上述 $\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_p$ 代入(4.2.21)式，并对之进行普通最小二乘估计，得到 $\hat{\hat{\beta}}_0,\hat{\hat{\beta}}_1,\cdots,\hat{\hat{\beta}}_k$。将 $\hat{\hat{\beta}}_0,\hat{\hat{\beta}}_1,\cdots,\hat{\hat{\beta}}_k$ 代回原模型，求出原模型随机干扰项的新的“近似估计值”，并以之作为方程(4.2.20)的样本观测值，采用普通最小二乘法估计该方程，得到 $\hat{\hat{\rho}}_1,\hat{\hat{\rho}}_2,\cdots,\hat{\hat{\rho}}_p$，作为随机干扰项的相关系数 $\rho_1,\rho_2,\cdots,\rho_p$ 的第二次估计值。重复上述过程，可得到 $\rho_1,\rho_2,\cdots,\rho_p$ 的多次迭代值。

关于迭代的次数，可根据具体的问题来定。一般是事先给出一个精度，当相邻两次的 $\rho_1,\rho_2,\cdots,\rho_p$ 的估计值之差小于这一精度时，迭代终止。实践中，有时只要迭代两次，就可得到较满意的结果。两次迭代过程也被称为**科克伦-奥科特两步法**。

需要指出的是，如果各序列相关系数是被估计出来的，则模型参数的估计结果不再是广义最小二乘估计量，而是**可行的广义最小二乘估计量**，该估计方法也被称为**可行的广义最小二乘估计**。可行的广义最小二乘估计量不再是无偏的，但却是一致的，而且在科克伦-奥科特迭代法下，估计量也具有渐近有效性。

4. 广义差分法在计量经济学软件中的实现

在 EViews 计量经济学软件包中，可以采用很简单的方法实现广义差分法参数估计。(4.2.21)式可以改写为

$$\begin{aligned} Y_t = {} & \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \\ & \rho_1 (Y_{t-1} - \beta_0 - \beta_1 X_{t-1,1} - \cdots - \beta_k X_{t-1,k}) + \cdots + \\ & \rho_p (Y_{t-p} - \beta_0 - \beta_1 X_{t-p,1} - \cdots - \beta_k X_{t-p,k}) + \varepsilon_t \\ & t = 1+p, 2+p, \cdots, n \end{aligned}$$

即

$$Y_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \rho_1 \mu_{t-1} + \rho_2 \mu_{t-2} + \cdots + \rho_p \mu_{t-p} + \varepsilon_t$$
$$t = 1+p, 2+p, \cdots, n$$

当选择普通最小二乘法估计参数时，如果同时选择常数项和 $X_1, X_2, \cdots, X_k, \mathrm{AR}(1), \mathrm{AR}(2), \cdots, \mathrm{AR}(p)$ 作为解释变量，即可得到参数 $\beta_0, \beta_1, \cdots, \beta_k, \rho_1, \rho_2, \cdots, \rho_p$ 的估计值。其中 AR(p)表示随机干扰项的 p 阶自回归。在估计过程中自动完成了 $\rho_1, \rho_2, \cdots, \rho_p$ 的迭代，并显示总迭代次数。

至于选择几阶随机干扰项的自回归项作为解释变量，主要判断依据是 D.W. 统计量。所以，一般是先不引入自回归项，采用普通最小二乘法估计参数；根据显示的 D.W.统计量，逐次引入 AR(1), AR(2) ……直到满意为止。

5. 序列相关稳健标准误法

与回归模型随机误差项出现异方差的情况相类似，当模型随机误差项出现序列相关时，普通最小二乘法只是影响到了参数估计量方差或标准差的正确估计，从而无法保证普通最小二乘估计量的有效性，但并不影响估计量的无偏性与一致性。因此，与解决出现异方差时的情况相仿，另一种针对序列相关的修正的估计方法是：仍采用普通最小二乘估计量，但修正其相应的方差。

如何修正普通最小二乘估计量相应的方差呢？尼威(Newey)和韦斯特(West)于 1987 年提出了类似于怀特提出的解决模型出现异方差时的方法，即计算出参数估计量正确的标准差。换言之，在一元线性回归模型中，对斜率项的估计量 $\hat{\beta}_1$ 的方差，需按(4.2.5)式

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_t^2} + \frac{2\sigma^2}{\sum x_t^2} \left[\rho \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum x_t^2} + \rho^2 \frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum x_t^2} + \cdots + \rho^{n-1} \frac{x_1 x_n}{\sum x_t^2} \right]$$

进行估计，而不是按普通最小二乘法中的(4.2.6)式

$$\operatorname{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_i^2}$$

进行估计。当然，尼威和韦斯特给出的计算公式要复杂得多，本教材不再列出。

尼威和韦斯特提出的修正普通最小二乘参数估计量标准误的方法，不仅能在模型随机误差项只存在序列相关时得到参数估计量的正确标准误，而且当模型随机误差项同时存在异方差与序列相关时，也能得到参数估计量的正确标准误，因此该标准误也被称为异方差-序列相关一致标准误（heteroscedasticity-autocorrelation-consistent standard error），或简称为尼威-韦斯特标准误（Newey-West standard error)，该估计参数的方法也称为**序列相关稳健标准误法**（method of serial correlation-rodust standard error)。

可以证明，大样本下尼威-韦斯特标准误是普通最小二乘参数估计量标准误的一致估计。

与存在异方差时的情形相类似，序列相关稳健标准误法虽不能得到有效的估计量，但由于可以得到普通最小二乘估计量正确的方差估计，从而使得以估计量方差为基础的各种统计检验不再失效、建立的预测区间也更加可信，因此序列相关稳健标准误法就成为在不能较好地实施广义最小二乘法时，消除异方差性不良后果的主要手段。

六、虚假序列相关问题

由于随机干扰项的序列相关往往是在模型设定中遗漏了重要的解释变量或对模型的函数形式设定有误时出现的，这种情形可称为**虚假序列相关**，应在模型设定中排除。因此，这里有两个层次的问题需要判断。第一层次的问题是如果检验出模型存在序列相关现象，需判断在模型的设定中是否是由于遗漏了重要的解释变量或对模型的函数形式设定有误而引起的虚假序列相关，这称为**模型的设定偏误检验**，将在第五章进行讨论。如果经检验不存在由于模型设定偏误而导致的虚假序列相关，即模型存在的序列相关是真实的序列相关或纯序列相关，则通过相应的修正方法进行修正。第二层次的问题是如何在设定模型时就避免产生虚假序列相关问题，或者说如何避免出现模型设定偏误问题，一个基本的建模规则就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，这将在第九章中进一步阐述。

七、案例——中国居民总量消费函数

例 4.2.1

§2.6 中我们曾在例 2.6.2 中通过普通最小二乘法建立了如下中国居民总量消费函数：

$$\hat{Y} = 2\,091.3 + 0.437\,5X \qquad (4.2.23)$$

$$(6.24) \qquad (47.10)$$

$$R^2 = 0.988\,0 \qquad \bar{R}^2 = 0.987\,5 \qquad F = 2\,214.6 \qquad \text{D.W.} = 0.277$$

1. 进行序列相关性检验

从残差项 $\tilde{e}_t$ 与时间 t 以及 $\tilde{e}_t$ 与 $\tilde{e}_{t-1}$ 的关系图(图 4.2.2)看，随机项呈现正序列相关性。

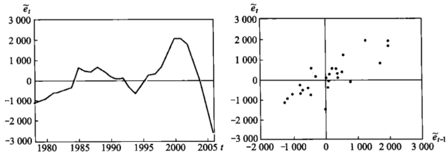


图 4.2.2 残差相关图

D.W.检验结果表明，在 5%显著性水平下，n=29，k=2(包含常数项)，查表得 d_L=1.34，d_U=1.48，由于 D.W.=0.277<d_L，故存在正自相关。

那么这里首先要问的一个问题是，序列相关是纯序列相关，还是由于模型设定有偏误而导致的虚假序列相关？

正如§2.6 讨论的那样，由于时间序列容易出现伪回归现象，因此做回归分析时须格外谨慎。本例中，Y 与 X 都是时间序列，而且它们确实表现出共同的变动趋势，因此有理由怀疑较高的 R^2 部分地是由这一共同的变化趋势带来的。为了排除时间序列模型中这种随时间变动而具有的共同变化趋势的影响，一种解决方案是在模型中引入时间趋势项，将这种影响分离出来。本例中，由于代表可支配收入的 X 与代表消费支出的 Y 均呈现非线性变化态势(图 2.6.1)，我们引入的时间变量 $T(T = 1,2,\cdots,\ 29)$以平方的形式出现，回归函数变化为

$$\hat{Y} = 3\,328.1 + 0.176\,2X + 21.656T^2 \qquad (4.2.24)$$

$$(17.06) \qquad (6.78) \qquad (10.19)$$

$$R^2 = 0.997\,6 \qquad \bar{R}^2 = 0.997\,4 \qquad F = 5\,380.4 \qquad \text{D.W.} = 0.442$$

这里，D.W.值仍然较低，没有通过 5%显著性水平下的 D.W.检验，因此判断(4.2.24)式仍存在正自相关性。

我们将在第五章进一步检验，未引入时间趋势项的模型(4.2.23)式存在着设定偏误，而引入时间趋势项的模型(4.2.23)式不再存在设定偏误问题。因此(4.2.23)式中较低的 D.W.值部分地是由模型设定偏误而引起的，即存在着序假序列相关的成分；而(4.2.23)式中较低的 D.W.值表明的是纯序列相关。

下面再对(4.2.24)式进行序列相关性的拉格朗日乘数检验。含 1 阶滞后残差项的辅助回归为

$$\tilde{e}_t = -47.09 + 0.019X - 1.62T^2 + 0.761\tilde{e}_{t-1}$$

$$(-0.39) \quad (1.18) \quad (-1.22) \quad (6.22)$$

$$R^2 = 0.619\,0$$

于是，LM=28×0.619 0=17.33，该值大于显著性水平为 5%、自由度为 1 的 χ^2 分布的临界值 $\chi^2_{0.05}(1)=3.84$，由此判断原模型存在 1 阶序列相关性。

含 2 阶滞后残差项的辅助回归为

$$\tilde{e}_t = -61.3 + 0.017X - 1.421T^2 + 1.056\tilde{e}_{t-1} - 0.363\tilde{e}_{t-2}$$

$$(-0.51) \quad (1.10) \quad (-1.10) \quad (5.28) \quad (-1.90)$$

$$R^2 = 0.656\,8$$

于是，LM=27×0.656 8=17.73，该值大于显著性水平为 5%、自由度为 2 的 χ^2 分布的临界值 $\chi^2_{0.05}(2)=5.99$，仍说明原模型存在序列相关性，但 $\tilde{e}_{t-2}$ 的参数未通过 5%的显著性检验，表明并不存在 2 阶序列相关性。结合 1 阶滞后残差项的辅助回归情况，可判断(4.2.24)式存在显著的 1 阶序列相关性。

2. 运用广义差分法进行自相关的处理

在 Eviews 软件包下，2 阶广义差分的估计结果为

$$\hat{Y}_t = 3\,505.7 + 0.199\,6X_t + 19.24T^2 + 0.748\,0AR(1) \qquad (4.2.25)$$

$$(8.69) \quad (6.59) \quad (6.57) \quad (5.93)$$

$$R^2 = 0.999\,1 \qquad \overline{R}^2 = 0.999\,0 \qquad \text{D.W.} = 1.39$$

式中，AR(1)前的参数值即为随机扰动项的 1 阶序列相关系数。在 5% 的显著性水平下，1.18=d_L<D.W.<d_U=1.65(样本容量为 28)，无法判断经广义差分变换后的模型是否已不存在序列相关性。但拉格朗日检验值为 LM=$n \cdot R^2$=27×0.091 7=2.48，小于显著性水平为 5%、自由度为 1 的 χ^2 分布的临界值 $\chi^2_{0.05}(1)=3.84$，表明模型干扰项已不存在自相关性。这里 R^2=0.091 7 是如下辅助回归的可决系数：

$$\tilde{\tilde{e}}_t = \alpha_0 + \alpha_1(x_t - 0.748\,0x_{t-1}) + \alpha_2(T^2 - 0.748\,0(T-1)^2) + \tilde{\tilde{e}}_{t-1}$$

其中，$\tilde{\tilde{e}}_t$ 是回归式(4.2.25)中的残差序列。

3. 序列相关稳健估计法

当模型存在序列相关性时，也可采用尼威-韦斯特的序列相关一致方差估计，即进行所谓的序列相关稳健估计，以达到对普通最小二乘法中参数的不正确方差估计的修正。本例中，通过 Eviews 软件给出的序列相关稳健估计结果为

$$\hat{Y} = 3\,643.0 + 0.165\,0X + 24.059T^2 \qquad (4.2.26)$$

$$(16.16) \quad (6.92) \quad (9.81)$$

$$R^2 = 0.997\,4 \quad \bar{R}^2 = 0.997\,2 \quad F = 4\,962.0 \quad \text{D.W.} = 0.426$$

可以看出，估计的参数与普通最小二乘法的结果相同，只是由于参数的标准差得到了修正，从而使得 t 检验值与普通最小二乘法的结果不同，但差异并不大。

§4.3　多重共线性

在讨论了回归模型随机干扰项违背同方差性和相互独立性假设时的检验方法和修正方法后，在本节和下节中将讨论模型的解释变量违背基本假设的问题。

一、多重共线性

对于模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i\,, \qquad (4.3.1)$$

其基本假设之一是解释变量 $X_1, X_2, \cdots, X_k$ 是相互独立的。**如果某两个或多个解释变量之间出现了相关性，则称为存在多重共线性**(multicollinearity)。

如果存在

$$c_1 X_{i1} + c_2 X_{i2} + \cdots + c_k X_{ik} = 0\,, \qquad (4.3.2)$$

其中 c_i 不全为 0，即某一个解释变量可以用其他解释变量的线性组合表示，则称为解释变量间存在**完全共线性**(perfect multicollinearity)。如果存在

$$c_1 X_{i1} + c_2 X_{i2} + \cdots + c_k X_{ik} + v_i = 0\,, \qquad (4.3.3)$$

其中 c_i 不全为 0，v_i 为随机干扰项，则称为**近似共线性**(approximate multicollinearity)或**交互相关**(intercorrelated)。

在矩阵表示的线性回归模型

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

中，完全共线性指秩 $R(\boldsymbol{X}) < k+1$，即矩阵

$$X=\begin{pmatrix}1 & X_{11} & X_{12} & \cdots & X_{1k}\\ 1 & X_{21} & X_{22} & \cdots & X_{2k}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk}\end{pmatrix}$$

中，至少有一个列向量可由其他列向量(不包括第一列)线性表出。例如，$\boldsymbol{X}_2=\lambda\boldsymbol{X}_1$，这时 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 的相关系数为 1，解释变量 $\boldsymbol{X}_2$ 对被解释变量的作用完全可由 $\boldsymbol{X}_1$ 代替。

完全共线性的情况并不多见，一般出现的是在一定程度上的共线性，即近似共线性。

二、实际经济问题中的多重共线性

一般地，产生多重共线性的主要原因有以下三个方面：

1. 经济变量相关的共同趋势

时间序列样本中发生多重共线性的主要原因在于许多基本经济变量存在相关的共同趋势。例如，经济繁荣时期，各基本经济变量(收入、消费、投资、价格)都趋于增长；经济衰退时期，又同时趋于下降。这些变量的样本数据往往呈现某些近似的比例关系。

截面数据也有可能产生多重共线性。例如，以某一行业的企业为样本建立企业生产函数模型，以产出量为被解释变量，选择资本、劳动力、技术等投入要素为解释变量。这些投入要素的数量往往与产出量成正比，产出量高的企业，投入的各种要素都比较多，这就使得投入要素之间出现线性相关性。如果以简单线性关系作为模型的数学形式，那么多重共线性是难以避免的。

2. 滞后变量的引入

在计量经济学模型中，往往需要引入滞后经济变量来反映真实的经济关系。例如，以相对收入假设为理论假设，则居民消费 C_t 的变动不仅受当期收入 Y_t 的影响，还受前期消费 C_{t-1} 的影响，于是建立如下模型：

$$C_t=\beta_0+\beta_1Y_t+\beta_2C_{t-1}+\mu_t$$

显然，模型中引入的当期收入和前期消费之间有较强的线性相关性。

3. 样本资料的限制

由于完全符合理论模型所要求的样本数据较难收集，在现有数据条件下，特定样本可能存在某种程度的多重共线性。

一般经验告诉我们，对于采用时间序列数据作样本，以简单线性形式建立的计量经济学模型，往往存在多重共线性；以截面数据作样本时，问题不那么严重，但仍然是存在的。

三、多重共线性的后果

计量经济学模型一旦出现多重共线性，如果仍采用普通最小二乘法估计模型参数，会产生下列不良后果：

1. 完全共线性下参数估计量不存在

多元线性回归模型

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

的普通最小二乘参数估计量为

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

如果出现完全共线性，则$(\boldsymbol{X'X})^{-1}$不存在，无法得到参数的估计量。

例如，对二元线性回归模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu \tag{4.3.4}$$

如果两个解释变量完全相关，如$X_2 = \lambda X_1$，该二元线性回归模型退化为一元线性回归模型

$$Y = \beta_0 + (\beta_1 + \lambda\beta_2)X_1 + \mu$$

这时，只能确定综合参数$\beta_1 + \lambda\beta_2$的估计值

$$\widehat{\beta_1 + \lambda\beta_2} = \frac{\sum x_{i1} y_i}{\sum x_{i1}^2}$$

却无法确定β_1, β_2各自的估计值。

2. 近似共线性下普通最小二乘法参数估计量的方差变大

在近似共线性下，虽然可以得到普通最小二乘参数估计量，但是由参数估计量方差的表达式

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X'X})^{-1}$$

可见，由于此时$|\boldsymbol{X'X}| \approx 0$，引起$(\boldsymbol{X'X})^{-1}$主对角线元素较大，使得参数估计量的方差增大，从而不能对总体参数作出准确推断。

仍以二元线性回归模型(4.3.4)式为例。离差形式下容易推出$\hat{\beta}_1$的方差为

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= \frac{\sigma^2 \sum x_{i2}^2}{\sum x_{i1}^2 \sum x_{i2}^2 - (\sum x_{i1} x_{i2})^2} \\ &= \frac{\dfrac{\sigma^2}{\sum x_{i1}^2}}{1 - \dfrac{(\sum x_{i1} x_{i2})^2}{\sum x_{i1}^2 \sum x_{i2}^2}} = \frac{\sigma^2}{\sum x_{i1}^2} \cdot \frac{1}{1 - r^2}\end{aligned} \tag{4.3.5}$$

其中，$\dfrac{(\sum x_{i1}x_{i2})^2}{\sum x_{i1}^2 \sum x_{i2}^2}$ 恰为 X_1 与 X_2 的线性相关系数的平方 r^2，由于 $r^2 \leqslant 1$，故 $\dfrac{1}{1-r^2} \geqslant 1$。

当完全不共线时，

$$r^2=0, \qquad \mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_{i1}^2}$$

当近似共线时，

$$0<r^2<1, \qquad \mathrm{Var}(\hat{\beta}_1)=\frac{\sigma^2}{\sum x_{i1}^2}\cdot\frac{1}{1-r^2}>\frac{\sigma^2}{\sum x_{i1}^2}$$

即多重共线性使参数估计量的方差增大，**方差膨胀因子**(variance inflation factor, VIF)为

$$\mathrm{VIF}(\hat{\beta}_1)=\frac{1}{1-r^2} \tag{4.3.6}$$

其增大趋势如表 4.3.1 所示。

当完全共线时，

$$r^2=1, \qquad \mathrm{Var}(\hat{\beta}_1)=+\infty$$

表 4.3.1　方差膨胀因子表

相关系数平方	0	0.5	0.8	0.9	0.95	0.96	0.97	0.98	0.99	0.999
方差膨胀因子	1	2	5	10	20	25	33	50	100	1 000

3. 参数估计量经济含义不合理

如果模型中两个解释变量具有线性相关性，如 X_1 和 X_2，那么它们中的一个变量可以由另一个变量表征。这时，X_1 和 X_2 前的参数并不反映各自与被解释变量之间的结构关系，而是反映它们对被解释变量的共同影响，所以各自的参数已经失去了应有的经济含义，于是经常表现出似乎反常的现象，例如估计结果本来应该是正的，结果却是负的。**经验告诉我们，在多元线性回归模型的估计中，如果出现参数估计值的经济意义明显不合理的情况，应该首先怀疑是否存在多重共线性。**

4. 变量的显著性检验和模型的预测功能失去意义

存在多重共线性时，参数估计值的方差与标准差变大，从而容易使通过样本计算的 t 值小于临界值，误导作出参数为零的推断，可能将重要的解释变量排除在模型之外。

变大的方差容易使预测值区间预测的“区间”变大，使预测失去意义。

四、多重共线性的检验

由于多重共线性表现为解释变量之间具有相关关系，所以用于多重共线性的检验方法主要是统计方法，如判定系数检验法、逐步回归检验法等。多重共线性检验的任务是：(1) 检验多重共线性是否存在；(2) 判明存在多重共线性的范围。

1. 检验多重共线性是否存在

(1) 对两个解释变量的模型，采用简单相关系数法

求出 X_1 与 X_2 的简单相关系数 r，若 $|r|$ 接近 1，则说明两变量存在较强的多重共线性。

(2) 对多个解释变量的模型，采用综合统计检验法

若在普通最小二乘法下，模型的 R^2 与 F 值较大，但各参数估计值的 t 检验值较小，说明各解释变量对 Y 的联合线性作用显著，但各解释变量间存在共线性而使得它们对 Y 的独立作用不能分辨，故 t 检验不显著。

2. 判明存在多重共线性的范围

如果存在多重共线性，需进一步确定多重共线性究竟由哪些变量引起。

(1) 判定系数检验法

使模型中每个解释变量分别以其余解释变量为解释变量进行回归计算，并计算相应的拟合优度，也称为判定系数。如果在某一种形式中判定系数较大，则说明在该形式中作为被解释变量的 X_j 可以用其他解释变量的线性组合代替，即 X_j 与其他解释变量间存在共线性。

可进一步对上述出现较大判定系数的回归方程作 F 检验：

$$F_j = \frac{R_{j.}^2/(k-1)}{(1-R_{j.}^2)/(n-k)} \sim F(k-1, n-k) \tag{4.3.7}$$

其中 $R_{j.}^2$ 为第 j 个解释变量对其他解释变量的回归方程的决定系数。若存在较强的共线性，则 $R_{j.}^2$ 较大且接近于 1，这时 $1-R_{j.}^2$ 较小，从而 F_j 的值较大。因此，可以给定显著性水平 α，通过计算的 F 值与相应的临界值的比较来进行检验。此时，原假设为 X_j 与其他解释变量间不存在显著的线性关系。

另一等价的检验是：在模型中排除某个解释变量 X_j，估计模型，如果拟合优度与包含 X_j 时十分接近，则说明 X_j 与其他解释变量之间存在共线性。

(2) 逐步回归法

以 Y 为被解释变量，逐个引入解释变量，构成回归模型，进行模型估计。根据拟合优度的变化决定新引入的变量是否可以用其他变量的线性组合代替，

而不是作为独立的解释变量。如果拟合优度变化显著，则说明新引入的变量是一个独立解释变量；如果拟合优度变化很不显著，则说明新引入的变量不是一个独立解释变量，它可以用其他变量的线性组合代替，也就是说它与其他变量之间存在共线性的关系。

五、克服多重共线性的方法

如果模型被证明存在多重共线性，则需要发展新的方法估计模型，最常用的方法有三类。

1. 第一类方法：排除引起共线性的变量

找出引起多重共线性的解释变量，将它排除出去，是最为有效地克服多重共线性问题的方法，所以逐步回归法得到了最为广泛的应用。但是，需要特别注意的是，当排除了某个或某些变量后，保留在模型中的变量的系数的经济意义将发生变化，其估计值也将发生变化。例如，在对数线性生产函数模型中，当包含资本、劳动、技术等投入要素时，资本的系数表示资本的产出弹性；但是，当资本和劳动存在共线性因而排除劳动时，资本的系数所表示的经济意义就不是资本的产出弹性，其估计值也将大于资本的产出弹性。

2. 第二类方法：差分法

对于以时间序列数据为样本，以直接线性关系为模型关系形式的计量经济学模型，将原模型变换为差分模型

$$\Delta Y_i = \beta_1 \Delta X_{i1} + \beta_2 \Delta X_{i2} + \cdots + \beta_k \Delta X_{ik} + \mu_i - \mu_{i-1}, \qquad i = 2,\cdots,n$$

可以有效地消除存在于原模型中的多重共线性。这是由经济时间序列数据的内在性质决定的。一般讲，增量之间的线性关系远比总量之间的线性关系弱一些。下面以例 2.6.2 中 1978—2006 年中国居民总量消费 CONS 与 GDP 的数据加以说明。

容易验证，CONS 与 GDP 间的相关系数为 0.998 9；而它们增量间的相关系数，即 ΔCONS 与 ΔGDP 间的相关系数为 0.945 1。如果分别作 CONS 关于 GDP 以及 ΔCONS 关于 ΔGDP 的线性回归，可得两回归模型的可决系数分别为 0.997 8 与 0.893 2。显然变量增量间的相关性要弱一些。

*3. 第三类方法：减小参数估计量的方差

多重共线性的主要后果是参数估计量具有较大的方差。若采取适当方法减小参数估计量的方差，虽然没有消除模型中的多重共线性，却能消除多重共线性造成的后果。例如，增加样本容量，可使参数估计量的方差减小。

20 世纪 70 年代发展的岭回归法(ridge regression)，以引入偏误为代价减小参数估计量的方差。具体方法是：引入矩阵 $\boldsymbol{D}$，使参数估计量为

PDG

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X} + \boldsymbol{D})^{-1}\boldsymbol{X'Y} \tag{4.3.8}$$

矩阵$\boldsymbol{D}$一般选择为主对角矩阵，即

$$\boldsymbol{D} = l\boldsymbol{I} \tag{4.3.9}$$

其中l为大于0的常数。显然，与普通最小二乘估计量相比，(4.3.8)式的估计量有较小的方差。

如何选择l是一个复杂的问题，何瑞尔(Hoerl)和肯纳德(Kennard)于1975年提出一种估计方法。首先对原模型的解释变量与被解释变量的离差形式进行标准化处理：

$$x_{ik}^{*} = \frac{x_{ik}}{\sqrt{\sum x_{ik}^{2}}}, \qquad y_{ik}^{*} = \frac{y_{ik}}{\sqrt{\sum y_{ik}^{2}}}$$

得到下列模型：

$$y_i^{*} = \beta_1^{*}x_{i1}^{*} + \beta_2^{*}x_{i2}^{*} + \cdots + \beta_k^{*}x_{ik}^{*} + \mu_i^{*}, \qquad i = 1,2,\cdots,n$$

用普通最小二乘法估计该模型，得到参数与随机干扰项方差的估计值$\hat{\beta}_1^{*},\hat{\beta}_2^{*},\cdots,\hat{\beta}_k^{*}$和$\hat{\sigma}^2$。选择

$$\hat{l} = \frac{(k-1)\hat{\sigma}^2}{\sum_{j=1}^{k}(\hat{\beta}_j^{*})^2}$$

作为(4.3.9)式中l的估计值。

最后需要指出的是，多重共线性是一种样本现象。同一个模型在一个样本下可能表现出多重共线性，而在另一个样本下可能就不存在多重共线性，因此增加样本容量就有可能消除多重共线性。

另外，多重共线性的主要问题在于使参数估计量的方差变大，而从(4.3.5)式知，随机干扰项的方差、变量的变异程度与方差膨胀因子一起决定着参数估计量的方差。如果存在多重共线性，但随机干扰项的方差很小，或变量的变异程度很大，都可能得到较小的参数估计量的方差。这时，即使有较严重的多重共线性，也不会带来不良后果。因此，只要回归方程估计的参数标准差较小，t统计值较大，就没有必要过于关心是否存在多重共线性的问题。

六、案例——中国粮食生产函数

例 4.3.1

根据理论和经验分析，影响粮食生产(Y)的主要因素有：农业化肥施用量(X_1)、粮食播种面积(X_2)、成灾面积(X_3)、农业机械总动力(X_4)、农业劳动力(X_5)，其中，成灾面积的符号为负，其余均应是正。表4.3.2列出了中国粮食生产的相关数据，拟建立中国粮食生产函数。

PDG

表 4.3.2 中国粮食生产与相关投入资料

年份	粮食产量（万吨）	农业化肥施用量（万公斤）	粮食播种面积（千公顷）	成灾面积（公顷）	农业机械总动力（万千瓦）	农业劳动力（万人）
1983	38 728	1 660	114 047	16 209	18 022	31 151
1984	40 731	1 740	112 884	15 264	19 497	30 868
1985	37 911	1 776	108 845	22 705	20 913	31 130
1986	39 151	1 931	110 933	23 656	22 950	31 254
1987	40 208	1 999	111 268	20 393	24 836	31 663
1988	39 408	2 142	110 123	23 945	26 575	32 249
1989	40 755	2 357	112 205	24 449	28 067	33 225
1990	44 624	2 590	113 466	17 819	28 708	38 914
1991	43 529	2 806	112 314	27 814	29 389	39 098
1992	44 264	2 930	110 560	25 895	30 308	38 699
1993	45 649	3 152	110 509	23 133	31 817	37 680
1994	44 510	3 318	109 544	31 383	33 802	36 628
1995	46 662	3 594	110 060	22 267	36 118	35 530
1996	50 454	3 828	112 548	21 233	38 547	34 820
1997	49 417	3 981	112 912	30 309	42 016	34 840
1998	51 230	4 084	113 787	25 181	45 208	35 177
1999	50 839	4 124	113 161	26 731	48 996	35 768
2000	46 218	4 146	108 463	34 374	52 574	36 043
2001	45 264	4 254	106 080	31 793	55 172	36 513
2002	45 706	4 339	103 891	27 319	57 930	36 870
2003	43 070	4 412	99 410	32 516	60 387	36 546
2004	46 947	4 637	101 606	16 297	64 028	35 269
2005	48 402	4 766	104 278	19 966	68 398	33 970
2006	49 804	4 928	104 958	24 632	72 522	32 561
2007	50 160	5 108	105 638	25 064	76 590	31 444

注：这里由于没有从事粮食生产的农业劳动力数据，故用第一产业劳动力替代。

资料来源：《中国统计年鉴》(1995，2008)。

设粮食生产函数为

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_4 \ln X_4 + \beta_5 \ln X_5 + \mu$$

1. 用普通最小二乘法估计模型

$$\hat{Y} = -4.16 + 0.382 \ln X_1 + 1.222 \ln X_2 - 0.081 \ln X_3 - 0.048 \ln X_4 - 0.102 \ln X_5$$

$$(-2.16) \quad (7.59) \quad (9.03) \quad (-5.30) \quad (-1.06) \quad (-1.76)$$

$$R^2 = 0.981\,6 \quad \bar{R}^2 = 0.976\,8 \quad F = 202.77 \quad \text{D.W.} = 1.79$$

PDG

由于R^2较大且接近于1，而且F=202.77>$F_{0.05}$(5,19)=2.74，故认为粮食生产与上述解释变量间总体线性关系显著。但由于其中X_4，X_5前参数估计值未能通过t检验，而且符号的经济意义也不合理，故认为解释变量间存在多重共线性。

2. 检验简单相关系数

$\ln X_1$，$\ln X_2$，$\ln X_3$，$\ln X_4$，$\ln X_5$的相关系数如表4.3.3所示。

表4.3.3 相关系数表

	$\ln X_1$	$\ln X_2$	$\ln X_3$	$\ln X_4$	$\ln X_5$
$\ln X_1$	1.000 0	−0.568 7	0.451 7	0.964 4	0.440 3
$\ln X_2$	−0.568 7	1.000 0	−0.214 1	−0.697 6	−0.073 3
$\ln X_3$	0.451 7	−0.214 1	1.000 0	0.398 8	0.411 3
$\ln X_4$	0.964 4	−0.697 6	0.398 8	1.000 0	0.279 5
$\ln X_5$	0.440 3	−0.073 3	0.411 3	0.279 5	1.000 0

由表中数据发现$\ln X_1$与$\ln X_4$间存在高度相关性。

3. 找出最简单的回归形式

分别作$\ln Y$与$\ln X_1$，$\ln X_2$，$\ln X_4$，$\ln X_5$间的回归：

(1) $\ln\hat{Y} = 8.902 + 0.224\ln X_1$

$$(43.2) \quad (8.78)$$

$$R^2 = 0.770\,2 \qquad \text{D.W.} = 0.94$$

(2) $\ln\hat{Y} = 15.15 - 0.384\ln X_2$

$$(2.56) \quad (-0.75)$$

$$R^2 = 0.024\,0 \qquad \text{D.W.} = 0.34$$

(3) $\ln\hat{Y} = 8.949 + 0.167\ln X_4$

$$(30.0) \quad (5.91)$$

$$R^2 = 0.602\,6 \qquad \text{D.W.} = 0.63$$

(4) $\ln\hat{Y} = 5.601 + 0.489\ln X_5$

$$(2.28) \quad (2.08)$$

$$R^2 = 0.158\,7 \qquad \text{D.W.} = 0.33$$

可见，粮食生产受农业化肥施用量的影响最大，与经验相符合，因此选(1)为初始的回归模型。

4. 逐步回归

将其他解释变量分别导入上述初始回归模型，寻找最佳回归方程(表4.3.4)。

表 4.3.4 逐步回归

	C	$\ln X_1$	$\ln X_2$	$\ln X_3$	$\ln X_4$	$\ln X_5$	$\bar{R}^2$	D.W.
$Y=f(X_1)$	8.902	0.224					0.760 2	0.94
t 值	(43.2)	(8.78)						
$Y=f(X_1, X_2)$	–6.293	0.298	1.258				0.940 2	1.59
t 值	(–3.47)	(19.2)	(8.38)					
$Y=f(X_1, X_2, X_3)$	–5.996	0.323	1.290	–0.087			0.975 5	1.41
t 值	(–5.16)	(29.8)	(13.4)	(–5.72)				
$Y=f(X_1, X_2, X_3, X_4)$	–6.04	0.322	1.294	–0.086	0.001		0.974 3	1.41
t 值	(–3.59)	(8.22)	(9.55)	(–5.51)	(0.03)			
$Y=f(X_1, X_2, X_3, X_5)$	–5.80	0.330	1.322	–0.081		–0.063	0.976 6	1.63
t 值	(–5.07)	(28.6)	(13.7)	(–5.28)		(–1.40)		

讨论：

第一步，在初始模型中引入 X_2，模型拟合优度提高，且参数符号合理，变量也通过了 t 检验，D.W.检验也表明不存在 1 阶序列相关性；

第二步，引入 X_3，拟合优度再次提高，且参数符号合理，变量也通过了 t 检验；只是 D.W.值落入了无法判断的区域，但由 LM 检验知仍不存在 1 阶自相关性；

第三步，引入 X_4，修正的拟合优度反而略有下降，同时 X_4 的参数未能通过 t 检验；

第四步，去掉 X_4，引入 X_5，拟合优度虽有所提高，但 X_5 的参数未能通过 t 检验，且参数符号与经济意义不符。

第三步与第四步表明，X_4 与 X_5 是多余的。同样还可继续验证，如果用与 X_1 高度相关的 X_4 替代 X_1，则 X_4 与 X_2，X_3，X_5 间的任意线性组合，均达不高以 X_1，X_2，X_3 为解释变量的回归效果。因此，最终的粮食生产函数应以 $Y=f(X_1,X_2,X_3)$ 为最优，拟合结果如下：

$$\ln\hat{Y}=-5.996+0.323\ln X_1+1.290\ln X_2-0.087\ln X_3$$

§4.4　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，并且与随机干扰项不相关。违背这一基本假设的问题被称为随机解释变量问题。

一、随机解释变量问题

对于模型

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i \text{ ,} \tag{4.4.1}$$

其基本假设之一是解释变量 $X_1, X_2, \cdots, X_k$ 是确定性变量。如果存在一个或多个随机变量作为解释变量，则称原模型存在**随机解释变量问题**。为讨论方便，假设(4.4.1)式中 X_2 为随机解释变量。对于随机解释变量问题，又分三种不同情况：

1. 随机解释变量与随机干扰项独立

即

$$\mathrm{Cov}(X_2, \mu) = \mathrm{E}(X_2 \mu) = \mathrm{E}(X_2)\mathrm{E}(\mu) = 0 \tag{4.4.2}$$

2. 随机解释变量与随机干扰项同期无关但异期相关

即

$$\mathrm{Cov}(X_{i2}, \mu_i) = \mathrm{E}(X_{i2}\mu_i) = 0 \text{ ,} \tag{4.4.3}$$

$$\mathrm{Cov}(X_{i2}, \mu_{i-s}) = \mathrm{E}(X_{i2}\mu_{i-s}) \neq 0 \text{ ,} \qquad s \neq 0 \tag{4.4.4}$$

3. 随机解释变量与随机干扰项同期相关

即

$$\mathrm{Cov}(X_{i2},\ \mu_i) = \mathrm{E}(X_{i2}\mu_i) \neq 0 \text{ ,} \tag{4.4.5}$$

如果某解释变量是确定性变量而不是随机变量，则该解释变量一定与随机误差项独立；如果解释变量是随机变量，则 § 3.1 给出的基本假设 5

$$\mathrm{Cov}(X_{ij}, \mu_i \mid X_1, X_2, \cdots X_k) = 0 \qquad j = 1, 2, \cdots, k$$

显得异常重要。该假设要求随机解释变量与随机误差项同期无关，这时随机解释变量被称为是**同期外生的**(contemporaneously exogenous)。如果随机解释变量与随机误差项既不同期相关，也不异期相关，则称该随机解释变量是**严格外生的**(strictly exogenous)。

二、实际经济问题中的随机解释变量问题

在实际经济问题中，经济变量往往都具有随机性。但是在单方程计量经济学模型中，凡是外生变量都被认为是确定性的。于是随机解释变量问题主要表

现于用滞后被解释变量作为模型的解释变量的情况。由于经济活动具有连续性，使得这类模型在以时间序列数据作样本的模型中占据较大份额。例如，消费不仅受收入的影响，还受前期消费水平的影响；投资不仅受收入的影响，还受前期投资水平的影响，等等。但是，并不是所有包含滞后被解释变量的模型都带来“随机解释变量问题”，下面通过两个例子简单予以说明，详细建立模型的过程将在第五章中讨论。

著名的“耐用品存量调整模型”可表示为

$$Q_t = \beta_0 + \beta_1 Y_t + \beta_2 Q_{t-1} + \mu_t, \qquad t = 1,2,\cdots,T \tag{4.4.6}$$

该模型表示，耐用品的存量由前一个时期的存量和当期收入共同决定。这是一个滞后被解释变量作为解释变量的模型。但是，如果模型不存在随机干扰项的序列相关性，那么随机解释变量Q_{t-1}只与μ_{t-1}相关，与μ_t不相关，属于随机解释变量与随机干扰项同期无关，但异期相关的情况。

著名的“合理预期消费函数模型”首先认为消费C_t是由对收入的预期Y_t^e所决定的：

$$C_t = \beta_0 + \beta_1 Y_t^e + \mu_t$$

在预期收入Y_t^e与实际收入Y之间存在假设

$$Y_t^e = (1-\lambda)Y_t + \lambda Y_{t-1}^e$$

的情况下，容易推出合理预期消费函数模型：

$$\begin{aligned} C_t &= \beta_0 + \beta_1(1-\lambda)Y_t + \beta_1\lambda Y_{t-1}^e + \mu_t \\ &= \beta_0 + \beta_1(1-\lambda)Y_t + \lambda(C_{t-1} - \beta_0 - \mu_{t-1}) + \mu_t \\ &= \beta_0(1-\lambda) + \beta_1(1-\lambda)Y_t + \lambda C_{t-1} + \mu_t - \lambda\mu_{t-1} \end{aligned} \tag{4.4.7}$$

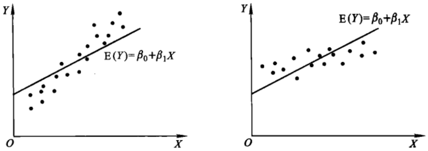
在该模型中，作为解释变量的C_{t-1}不仅是一个随机解释变量，而且与模型的随机干扰项$\mu_t - \lambda\mu_{t-1}$高度相关(因为C_{t-1}与μ_{t-1}高度相关)，属于随机解释变量与随机干扰项同期相关的情况。

三、随机解释变量的后果

计量经济学模型一旦出现随机解释变量，且与随机干扰项相关的话，如果仍采用普通最小二乘法估计模型参数，不同性质的随机解释变量会产生不同的后果。下面以一元线性回归模型为例进行说明。

从图形上看(图 4.4.1)，如果随机解释变量与随机干扰项正相关，则在抽取样本时，容易出现X值较小的点在总体回归线下方，而X值较大的点在总体回归线上方的情况，因此，拟合的样本回归线则可能低估(underestimate)截距项，而高估(overestimate)斜率项。反之，如果随机解释变量与随机干扰项负相关，

则往往导致拟合的样本回归线高估截距项而低估斜率项。



(a) 正相关　　(b) 负相关

图 4.4.1　随机解释变量与随机干扰项相关图

对一元线性回归模型

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

在第二章曾得到如下最小二乘估计量：

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \beta_1 + \frac{\sum x_i \mu_i}{\sum x_i^2} \tag{4.4.8}$$

随机解释变量 X 与随机干扰项 μ 的关系不同，参数普通最小二乘估计量的统计性质也会不同，同样分三种不同情况。

(1) **如果 X 与 μ 相互独立，得到的参数估计量仍然是无偏一致估计量。**这在第二章中已经得到证明。

(2) **如果 X 与 μ 同期不相关，而异期相关，得到的参数估计量有偏，但却是一致的。**由(4.4.8)式易知

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

尽管 X_i 与 μ_i 同期无关，但对任一 μ_i，k_i 的分母中一定包含不同期的 X；由异期相关性知 k_i 与 μ_i 相关，导致 $\mathrm{E}(\hat{\beta}_1) \neq \beta_1$，即参数估计量是有偏的。但是

$$\mathrm{P}\lim_{n\to\infty}\left(\beta_1 + \frac{\sum x_i \mu_i}{\sum x_i^2}\right) = \beta_1 + \frac{\mathrm{P}\lim\left(\frac{1}{n}\sum x_i \mu_i\right)}{\mathrm{P}\lim\left(\frac{1}{n}\sum x_i^2\right)}$$

$$= \beta_1 + \frac{\mathrm{Cov}(X_i, \mu_i)}{\mathrm{Var}(X_i)} = \beta_1$$

即 $\hat{\beta}_1$ 是 β_1 的一致估计。

(3) **如果 X 与 μ 同期相关，得到的参数估计量有偏且非一致**。这在上面的证明中已看得比较清楚。由于随机解释变量与随机干扰项同期相关时，会对普通最小二乘估计带来严重的不良后果，这时我们也称该随机解释变量具有内生性。

需要说明的是，**如果模型中带有滞后被解释变量作为解释变量，则当该滞后被解释变量与随机干扰项同期相关时，普通最小二乘估计量是有偏的且非一致的。即使同期无关，其普通最小二乘估计量也是有偏的，因为此时肯定出现异期相关。**

四、工具变量法

模型中出现随机解释变量并且与随机干扰项相关时，普通最小二乘估计量是有偏的。如果随机解释变量与随机干扰项异期相关，则可以通过增大样本容量的办法来得到一致的估计量；但如果是同期相关，即使增大样本容量也无济于事。这时，最常用的估计方法是**工具变量(instrument variable)法**。

1. 工具变量的选取

工具变量，顾名思义是在模型估计过程中被作为工具使用，以替代与随机干扰项相关的随机解释变量。如果选 Z 作为 X_j 的工具变量，Z 必须满足以下条件：

(1) 与所替代的随机解释变量高度相关：$\mathrm{Cov}(Z, X_j)=0$；

(2) 与随机干扰项不相关：$\mathrm{Cov}(Z, \mu)=0$；

(3) 与模型中其他解释变量不相关，以避免出现多重共线性。

2. 工具变量的应用

工具变量法是克服解释变量与随机干扰项相关影响的一种参数估计方法。下面仍以一元回归模型为例说明。

记一元线性回归模型如下：

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \tag{4.4.9}$$

矩估计是在两个重要的假设条件 $\mathrm{E}(\mu_i)=0$ 与 $\mathrm{E}(\mu_i X_i)=0$ 下，以之作为总体矩条件，并写出相应的样本矩条件

$$\frac{1}{n}\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0\,, \quad \frac{1}{n}\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)X_i = 0$$

后得到一个关于参数估计量的正规方程组：

$$\begin{cases} \sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \\ \sum X_i Y_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 \end{cases} \tag{4.4.10}$$

求解该正规方程组，得到

$$\hat{\beta}_1=\frac{\sum x_i y_i}{\sum x_i^2}, \qquad \hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}$$

然而，如果 X_i 与 μ_i 相关，则无法得到(4.4.10)式。

如果按照工具变量的选择条件选择 Z 为 X 的**工具变量**，则有总体矩条件

$$\mathrm{E}(\mu_i)=0, \qquad \mathrm{Cov}(X_i,\mu_i)=\mathrm{E}(\mu_i X_i)=0$$

于是，在一组容量为 n 的样本下，可写出相应的样本矩条件

$$\frac{1}{n}\sum\left(Y_i-\tilde{\beta}_0-\tilde{\beta}_1 X_i\right)=0, \qquad \frac{1}{n}\sum\left(Y_i-\tilde{\beta}_0-\tilde{\beta}_1 X_i\right)Z_i=0$$

并由此得到一个关于参数估计量的正规方程组：

$$\begin{aligned}\sum Y_i&=n\tilde{\beta}_0+\tilde{\beta}_1\sum X_i\\ \sum Z_i Y_i&=\tilde{\beta}_0\sum Z_i+\tilde{\beta}_1\sum Z_i X_i\end{aligned} \tag{4.4.11}$$

于是得到

$$\tilde{\beta}_1=\frac{\sum z_i y_i}{\sum z_i x_i}, \quad \tilde{\beta}_0=\bar{Y}-\tilde{\beta}_1\bar{X} \tag{4.4.12}$$

这种求模型参数估计量的方法称为**工具变量法**，$\tilde{\beta}_0,\tilde{\beta}_1$ 称为**工具变量法估计量**(instrumental variable estimator)。

对于多元线性回归模型，其矩阵形式为

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

采用工具变量法(假设 $\boldsymbol{X}_2$ 与随机干扰项相关，用工具变量 $\boldsymbol{Z}$ 替代)得到的正规方程组为

$$\boldsymbol{Z'Y}=\boldsymbol{Z'X\beta}$$

参数估计量为

$$\tilde{\boldsymbol{\beta}}=(\boldsymbol{Z'X})^{-1}\boldsymbol{Z'Y} \tag{4.4.13}$$

其中

$$\boldsymbol{Z}=\begin{pmatrix}1 & X_{11} & Z_1 & \cdots & X_{1k}\\ 1 & X_{21} & Z_2 & \cdots & X_{2k}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & X_{n1} & Z_n & \cdots & X_{nk}\end{pmatrix}$$

通常，对于没有选择另外的变量作为工具变量的解释变量，可以认为用自身作为工具变量。于是 $\boldsymbol{Z}$ 称为工具变量矩阵。

3. 工具变量法估计量是一致估计量

一元回归中，用工具变量法所求的参数估计量 $\tilde{\beta}_1$ 与总体参数真值 β_1 之间的关系为

$$\tilde{\beta}_1=\frac{\sum z_i y_i}{\sum z_i x_i}=\frac{\sum z_i Y_i}{\sum z_i x_i}=\frac{\sum z_i(\beta_0+\beta_1 X_i+\mu_i)}{\sum z_i x_i}$$

$$=\frac{\beta_1\sum z_i x_i}{\sum z_i x_i}+\frac{\sum z_i\mu_i}{\sum z_i x_i}=\beta_1+\frac{\sum z_i\mu_i}{\sum z_i x_i}$$

两边取概率极限得

$$\mathrm{P}\lim(\tilde{\beta}_1)=\beta_1+\frac{\mathrm{P}\lim\left(\frac{1}{n}\sum z_i\mu_i\right)}{\mathrm{P}\lim\left(\frac{1}{n}\sum z_i x_i\right)}$$

如果工具变量 Z 选取恰当，即有

$$\mathrm{P}\lim\left(\frac{1}{n}\sum z_i\mu_i\right)=\mathrm{Cov}(Z_i,\mu_i)=0$$

$$\mathrm{P}\lim\left(\frac{1}{n}\sum z_i x_i\right)=\mathrm{Cov}(Z_i,X_i)\neq 0$$

因此，

$$\mathrm{P}\lim(\tilde{\beta}_1)=\beta_1$$

尽管工具变量法估计量在大样本下具有一致性，但容易验证在小样本下，由于

$$\mathrm{E}\left(\frac{1}{\sum z_i x_i}\sum z_i\mu_i\right)\neq\mathrm{E}\left(\frac{1}{\sum z_i x_i}\right)\mathrm{E}\left(\sum z_i\mu_i\right)=0$$

工具变量法估计量仍是有偏的。

对工具变量法，有三点需要特别指出。

第一，经常产生一种误解，以为采用工具变量法是将原模型中的随机解释变量换成工具变量，即改变了原来的模型。实际上，从上面一元线性回归模型的例子中可以看出，**工具变量法并没有改变原模型，只是在原模型的参数估计过程中用工具变量“替代”随机解释变量**。或者说，上述工具变量法估计过程可等价地分解成下面两个阶段的普通最小二乘回归：

第一阶段，用普通最小二乘法进行 X 关于工具变量 Z 的回归：

$$\hat{X}_i=\hat{\alpha}_0+\hat{\alpha}_1 Z_i \tag{4.4.14}$$

第二阶段，以第一步得到的 $\hat{X}_i$ 为解释变量，进行如下普通最小二乘回归：

$$\hat{Y}_i=\tilde{\beta}_0+\tilde{\beta}_1\hat{X}_i \tag{4.4.15}$$

容易验证，(4.4.15)式中的参数 $\tilde{\beta}_1$ 与(4.4.12)式相同。(4.4.15)式表明，工具变量法仍是 Y 对 X 的回归，而不是对 Z 的回归。这里采用两个阶段的普通最小二乘法来估计模型参数，也被称为**两阶段最小二乘法**(two stage least squares, 2SLS)。

第二，如果一个随机解释变量可以找到多个相互独立的工具变量，人们希

望充分利用这些工具变量的信息，就形成了**广义矩方法**(Generalized Method of Moments, GMM)。在 GMM 中，矩条件大于待估参数的数量，于是如何求解成为它的核心问题。GMM 是近 20 年计量经济学理论方法发展的重要方向之一。工具变量法是 GMM 的一个特例；同样，普通最小二乘法也可看成是工具变量法的特例。

第三，要找到与随机干扰项不相关而又与随机解释变量相关的工具变量并不是一件很容易的事，但如果考虑到随机解释变量与随机干扰项相关的主要来源是由于同期测量误差引起的，就可以用滞后一期的随机解释变量作为原解释变量的工具变量。

五、解释变量的内生性检验

回归模型的基本假设要求随机解释变量与模型的随机干扰项至少不存在同期相关性，即随机解释变量至少是同期外生变量。那么如何判断所设定的模型中各解释变量是同期外生变量呢？经济学的相关知识能帮助我们作出一些基本的判断，如由于消费惯性的存在，可以认为前期的消费支出对当期的消费支出有着一定的影响，但不能反过来说当期的消费支出对前期的消费支出有影响。除此之外，豪斯曼(Hausman)从计量技术上给出了一个检验随机解释变量是否是同期外生变量的方法。

假设有如下设定的二元线性回归模型

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{i1} + \mu_i \tag{4.4.16}$$

其中，X 与 Z_1 是随机解释变量，而且明确知道 Z_1 是外生变量，但怀疑 X 是同期内生变量。如何检验 X 是否具有内生性呢？豪斯曼提出的检验的基本思想是：如果 X 是内生变量，则需寻找一外生变量 Z_2 作为工具变量并对(4.4.16)式进行工具变量法估计，将工具变量法的估计结果与对(4.4.16)式直接进行普通最小二乘法的估计结果对比，看差异是否显著。如果两者有显著的差异，则表明 X 是内生变量。由于工具变量法等价于两阶段最小二乘法，因此该检验法可具体如下进行。

第一步，将怀疑是内生变量的 X 关于外生变量 Z_1, Z_2 作普通最小二乘估计：

$$X_i = \alpha_0 + \alpha_1 Z_{i1} + \alpha_2 Z_{i2} + v_i \tag{4.4.17}$$

得到残差项 $\hat{v}$。这里假设随机干扰项 v 满足所有线性回归基本假设。该普通最小二乘回归的目的是为了得到残差项 $\hat{v}$，因此可认为是辅助回归。

第二步，将第一步得到的残差项 $\hat{v}$ 加入到原模型后，再进行普通最小二乘估计：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_{i1} + \delta \hat{v}_i + \varepsilon_i \tag{4.4.18}$$

仍假设随机干扰项 ε 满足所有线性回归基本假设，并与 v 不同期相关。如果 $\hat{v}$ 前

的参数δ显著为零，则表明(4.4.17)式的随机干扰项v与Y同期无关，进而与原模型(4.4.16)式的随机干扰项μ同期无关，而Z_1，Z_2是外生变量，它们肯定与μ同期无关，由(4.4.17)式知X与μ同期无关。因此，(4.4.18)式的普通最小二乘回归不拒绝$\delta=0$的假设，则可判断原模型(4.4.16)式中的解释变量X是内期外生变量，否则判断X是同期内生变量。

最后，有两点需要说明。

第一，由(4.4.17)式知，判断X与μ是否同期相关，等价于判断v与μ是否同期相关；而对(4.4.18)式的普通最小二乘回归，等价于对下式进行普通最小二乘回归：

$$\mu_i = \delta v_i + \varepsilon_i$$

第二，如果原回归模型有多个随机解释变量被怀疑与随机干扰项同期相关，则需寻找多个外生变量，并将每个所怀疑的解释变量与所有外生变量(包括原模型中已有的外生变量)作普通最小二乘回归，取得各自的残差项，并将它们全部引入到原模型中再进行普通最小二乘估计，通过t检验或多种情形的受约束F检验，可判断哪些解释变量确实是内生变量。

六、案例——中国城镇居民人均消费函数

例 4.4.1

在例 3.2.2 的中国城镇居民人均消费函数的估计中，采用普通最小二乘法估计了下面的模型：

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

其中，Y代表中国城镇居民人均消费支出，X_1代表城镇居民人均可支配收入，X_2代表前一年城镇居民人均消费支出。然而，如果考虑到在居民人均消费支出(Y)由人均可支配收入(X_1)决定的同时，人均消费支出(Y)又反过来影响着同期居民人均可支配收入，因此，那些未在模型中独立列出而纳入到随机干扰项的影响居民人均消费支出的因素，也影响着居民人均可支配收入，据此有理由怀疑居民人均可支配收入(X_1)与随机干扰项μ同期相关。对前一年城镇居民人均消费支出(X_2)而言，它影响着当年的消费支出，而当年的消费支出不会反过来影响前一年的消费支出，因此可认为X_2是同期外生变量。

下面再用豪斯曼检验来判定城镇居民人均可支配收入X_1是否确实是内生变量。

可选择前一年的城镇居民人均可支配收入Z作为X_1的工具变量，显然，前一年的居民人均可支配收入会与当年的居民人均可支配收入有较强的相

关性，但由于当年的消费支出不会影响前一年的可支配收入，因此 Z 与原模型的随机干扰项 μ 也不会存在同期相关性。表 4.4.1 列出了 2005 年中国内地 31 个省区城镇居民人均可支配收入(Z)的数据，将 X_1 关于 X_2，Z 进行普通最小二乘估计得

$$\hat{X}_1 = 132.74 - 0.470\,9X_2 + 1.460\,5Z \qquad (4.4.19)$$

$$(0.68) \quad (-4.08) \quad (16.98)$$

记录残差序列 $\hat{v}$，并将其加入原模型后进行普通最小二乘估计得：

$$\hat{Y} = 155.70 + 0.450\,2X_1 + 0.402\,6X_2 + 1.191\,0\hat{v} \qquad (4.4.20)$$

$$(1.11) \quad (10.60) \quad (6.31) \quad (8.35)$$

t 检验表明，$\hat{v}$ 前的参数显著不为 0，因此判断城镇居民人均可支配收入确实是内生变量。从而原模型的普通最小二乘估计量有偏并且是非一致的，必须采用工具变量法进行估计。

表 4.4.1 2005 年中国内地 31 个省区城镇居民人均可支配收入 单位：元

地区	可支配收入 Z	地区	可支配收入 Z	地区	可支配收入 Z	地区	可支配收入 Z
北京	17 653.0	上海	18 645.0	湖北	8 785.9	云南	9 265.9
天津	12 638.6	江苏	12 318.6	湖南	9 524.0	西藏	9 431.2
河北	9 107.1	浙江	16 293.8	广东	14 769.9	陕西	8 272.0
山西	8 913.9	安徽	8 470.7	广西	9 286.7	甘肃	8 086.8
内蒙古	9 136.8	福建	12 321.3	海南	8 123.9	青海	8 057.9
辽宁	9 107.6	江西	8 619.7	重庆	10 243.5	宁夏	8 093.6
吉林	8 690.6	山东	10 744.8	四川	8 386.0	新疆	7 990.2
黑龙江	8 272.5	河南	8 668.0	贵州	8 151.1		

资源来源：《中国统计年鉴》2006 年。

以前一年(2005 年)中国城镇居民人均可支配收入 Z 作为当年(2006 年)城镇居民人均可支配收入 X_1 的工具变量，可得到如下工具变量估计结果：

$$\hat{Y} = 155.70 + 0.450\,2X_1 + 0.402\,6X_2 \qquad (4.4.21)$$

$$(0.58) \quad (5.52) \quad (3.28)$$

$$R^2 = 0.973\,9 \quad F=513.7 \quad \text{D.W.}=2.08 \quad \text{RSS}=4\,461\,857$$

为了与普通最小二乘估计进行比较，写出普通最小二乘估计结果：

$$\hat{Y} = 143.33 + 0.555\,6X_1 + 0.250\,1X_2$$

$$(0.55) \quad (7.37) \quad (2.20)$$

$$R^2 = 0.975\,6 \quad F=560.56 \quad \text{D.W.}=1.84 \quad \text{RSS}=4\,170\,093$$

由(4.4.20)式知，原模型的随机干扰项 μ 与辅助回归(4.4.19)式的随机干扰项 $\hat{v}$ 是正相关的，意味着同期内生变量 X_1 与 μ 具有正相关性。正如本节

讨论随机解释变量后果时所指出的，在随机解释变量与随机干扰项存在正相关的情形下，普通最小二乘估计量可能会低估截距项而高估斜率项。因此，正如所预期的那样，这里的工具变量法估计量，对普通最小二乘估计量对截距项的低估与斜率项 X_1 参数的高估作出了修正。

本章练习题

1. 对一元回归模型

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

(1) 假如其他基本假设全部满足，但 $\mathrm{Var}(\mu_i) = \sigma_i^2 \neq \sigma^2$，试证明估计的斜率项仍是无偏的，但方差变为

$$\mathrm{Var}(\tilde{\beta}_1) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$$

(2) 如果 $\mathrm{Var}(\mu_i) = \sigma^2 K_i$，试证明上述方差的表达式为

$$\mathrm{Var}(\tilde{\beta}_1) = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 K_i}{\sum x_i^2}$$

该表达式与同方差假定下的方差 $\mathrm{Var}(\hat{\beta}_1)$ 之间有何关系？分 K_i 大于 1 与小于 1 两种情况讨论。

2. 对习题 1 中的一元线性回归模型，如果已知 $\mathrm{Var}(\mu_i) = \sigma_i^2$，则可对原模型以权 $\frac{1}{\sigma_i}$ 相乘后变换成如下的二元模型：

$$\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{\mu_i}{\sigma_i}$$

对该模型进行普通最小二乘估计就是加权最小二乘法。试证明该模型的随机干扰项是同方差的，并求出 β_1 的上述加权最小二乘估计量。

3. 对一元线性回归模型

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

(1) 假如其他基本假设全部满足，但 $\mathrm{Cov}(\mu_i, \mu_j) \neq 0$，试证明，估计的斜率项仍是无偏的；

(2) 若自变量存在正相关，且随机干扰项存在如下一阶序列相关：

$$\mu_t = \rho \mu_{t-1} + \varepsilon_t$$

试证明估计的斜率项的方差为

$$\mathrm{Var}(\tilde{\beta}_1) = \frac{\sigma^2}{\sum x_t^2} + \frac{2\sigma^2}{\sum x_t^2} \left(\rho \frac{\sum_{t=1}^{n-1} x_t x_{t+1}}{\sum x_t^2} + \rho^2 \frac{\sum_{t=1}^{n-2} x_t x_{t+2}}{\sum x_t^2} + \cdots + \rho^{n-1} \frac{x_1 x_n}{\sum x_t^2} \right)$$

并就 $\rho>0$ 与 $\rho<0$，X_t 存在正序列相关或负序列相关时与模型满足所有基本假定下的普通最小二乘估计 $\mathrm{Var}(\hat{\beta}_1)$ 的大小进行比较。

4. 试证明：二元线性回归模型

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\mu_i$$

中变量 X_1 与 X_2 的参数的普通最小二乘估计可以写成

$$\hat{\beta}_1=\frac{(\sum y_ix_{i1})(\sum x_{i2}^2)-(\sum y_ix_{i2})(\sum x_{i1}x_{i2})}{\sum x_{i1}^2\sum x_{i2}^2(1-r^2)}$$

$$\hat{\beta}_2=\frac{(\sum y_ix_{i2})(\sum x_{i1}^2)-(\sum y_ix_{i1})(\sum x_{i1}x_{i2})}{\sum x_{i1}^2\sum x_{i2}^2(1-r^2)}$$

其中，r 为 X_1 与 X_2 的相关系数。讨论 r 等于或接近于 1 时，该模型的估计问题。

5. 对模型

$$Y_t=\beta_0+\beta_1X_{t1}+\beta_2X_{t2}+\beta_3Y_{t-1}+\mu_t$$

假设 Y_{t-1} 与 μ_t 相关。为了消除该相关性，采用工具变量法：先求 Y_t 关于 X_{t1} 与 X_{t2} 回归，得到 $\hat{Y}_t$，再作如下回归：

$$Y_t=\beta_0+\beta_1X_{t1}+\beta_2X_{t2}+\beta_3\hat{Y}_{t-1}+\mu_t$$

试问：这一方法能否消除原模型中 Y_{t-1} 与 μ_t 的相关性？为什么？

6. 对于一元回归模型

$$Y_t=\beta_0+\beta_1X_t^*+\mu_t$$

假设解释变量 X_t^* 的实测值 X_t 与之有偏误：$X_t=X_t^*+e_t$，其中 e_t 是具有零均值，不序列相关，且与 X_t^* 及 μ_t 不相关的随机变量。试问：

(1) 能否将 $X_t^*=X_t-e_t$ 代入原模型，使之变换成 $Y_t=\beta_0+\beta_1X_t+v_t$ 后进行估计？其中，v_t 为变换后模型的随机干扰项。

(2) 进一步假设 μ_t 与 e_t 之间，以及它们与 X_t^* 之间无异期相关，那么 $\mathrm{E}(X_{t-1}v_t)=0$ 成立吗？X_t 与 X_{t-1} 相关吗？

(3) 由(2)的结论，你能寻找什么样的工具变量对变换后的模型进行估计？

7. 验证教材(4.4.15)式中的参数估计与(4.4.12)式相同。

8. 下表列出了某年中国部分省市城镇居民每个家庭平均全年可支配收入 X 与消费性支出 Y 的统计数据。

单位：元

地区	可支配收入 X	消费性支出 Y	地区	可支配收入 X	消费性支出 Y
北　京	10 349.69	8 493.49	河　北	5 661.16	4 348.47
天　津	8 140.50	6 121.04	山　西	4 724.11	3 941.87
内蒙古	5 129.05	3 927.75	河　南	4 766.26	3 830.71
辽　宁	5 357.79	4 356.06	湖　北	5 524.54	4 644.50
吉　林	4 810.00	4 020.87	湖　南	6 218.73	5 218.79
黑龙江	4 912.88	3 824.44	广　东	9 761.57	8 016.91

续表

地区	可支配收入 X	消费性支出 Y	地区	可支配收入 X	消费性支出 Y
上海	11 718.01	8 868.19	陕西	5 124.24	4 276.67
江苏	6 800.23	5 323.18	甘肃	4 916.25	4 126.47
浙江	9 279.16	7 020.22	青海	5 169.96	4 185.73
山东	6 489.97	5 022.00	新疆	5 644.86	4 422.93

(1) 试用普通最小二乘法建立居民人均消费支出与可支配收入的线性模型；

(2) 检验模型是否存在异方差性；

(3) 如果存在异方差性，试采用适当的方法估计模型参数。

9. 中国 1980—2007 年全社会固定资产投资总额 X 与工业总产值 Y 的统计资料如下表所示。

单位：亿元

年份	全社会固定资产投资 X	工业增加值 Y	年份	全社会固定资产投资 X	工业增加值 Y
1980	910.9	1 996.5	1994	17 042.1	19 480.7
1981	961	2 048.4	1995	20 019.3	24 950.6
1982	1 230.4	2 162.3	1996	22 913.5	29 447.6
1983	1 430.1	2 375.6	1997	24 941.1	32 921.4
1984	1 832.9	2 789.0	1998	28 406.2	34 018.4
1985	2 543.2	3 448.7	1999	29 854.7	35 861.5
1986	3 120.6	3 967.0	2000	32 917.7	40 033.6
1987	3 791.7	4 585.8	2001	37 213.5	43 580.6
1988	4 753.8	5 777.2	2002	43 499.9	47 431.3
1989	4 410.4	6 484.0	2003	55 566.6	54 945.5
1990	4 517	6 858.0	2004	70 477.4	65 210.0
1991	5 594.5	8 087.1	2005	88 773.6	77 230.8
1992	8 080.1	10 284.5	2006	109 998.2	91 310.9
1993	13 072.3	14 188.0	2007	137 323.9	107 367.2

试问：

(1) 当设定模型为 $\ln Y_t = \beta_0 + \beta_1 \ln X_t + \mu_t$ 时，是否存在序列相关性？

(2) 若按一阶自相关假设 $\mu_t = \rho\mu_{t-1} + \varepsilon_t$，试用广义最小二乘法估计原模型。

(3) 采用差分形式 $X_t^* = X_t - X_{t-1}$ 与 $Y_t^* = Y_t - Y_{t-1}$ 作为新数据，估计模型 $Y_t^* = \alpha_0 + \alpha_1 X_t^* + v_t$，该模型是否存在序列相关？

10. 经济理论指出，家庭消费支出 Y 不仅取决于可支配收入 X_1，还决定于个人财富 X_2，即可设定如下回归模型：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \mu_i$$

试根据下表的资料进行回归分析，并说明估计的模型是否可靠，给出你的分析。

单位：元

编号	Y	X_1	X_2	编号	Y	X_1	X_2
1	700	800	8 100	6	1 150	1 800	18 760
2	650	1 000	10 090	7	1 200	2 000	20 520
3	900	1 200	12 730	8	1 400	2 200	22 010
4	950	1 400	14 250	9	1 550	2 400	24 350
5	1 100	1 600	16 930	10	1 500	2 600	26 860

PDG

第五章　经典单方程计量经济学模型：专门问题

5

在前面几章中，主要介绍了经典线性回归模型及其在若干基本假定下的估计问题，并分析了一个或多个假定不满足时所产生的后果及其可能的改进措施。然而上述方法还不能解决经济生活中遇到的全部问题。例如，如何考察某一突发事件对经济行为带来的影响，某变量的过去行为又是怎样影响变量当前变动路线的，等等。这需要建立专门的模型来进行研究。另一方面，直到目前，我们一直假设所设定的模型是正确的，即不存在模型设定偏误。然而，实际情况并非如此。经济理论并未告知变量间的具体关系应当是什么样的，比如应包括多少个解释变量，模型应选取线性形式还是双对数线性形式等。显然，如果模型应包括两个解释变量，而我们在建立模型时只包括了一个解释变量，这时就出现了所谓的**模型设定偏误**(model specification error)问题。对这一问题也需要进行专门的探讨。

本章将主要介绍经典单方程计量经济学模型中两类常见的专门问题：一类是模型中引入虚拟解释变量、滞后解释变量或(和)滞后被解释变量的问题；另一类是模型设定的偏误问题。

§5.1　虚拟变量模型

许多经济变量是可以定量度量的，如商品需求量、价格、收入、产量等，但也有一些影响经济变量的因素无法定量度量，如职业、性别对收入的影响，战争、自然灾害对 GDP 的影响，季节对某些产品(如冷饮)销售的影响等。为了能够在模型中反映这些因素的影响，并提高模型的精度，需要将它们“量化”，这种“量化”通常是通过引入“虚拟变量”来完成的。根据这些因素的属性类型，构造只取“0”或“1”的人工变量，通常称为**虚拟变量**(dummy variable)，记为 D。例如，反映文化程度的虚拟变量可取为

$$D=\begin{cases}1, & \text{本科学历}\\ 0, & \text{非本科学历}\end{cases}$$

一般地，在虚拟变量的设置中，基础类型和肯定类型取值为 1；比较类型和否定类型取值为 0。同时含有一般解释变量与虚拟变量的模型称为**虚拟变量**

PDG

模型。一个以性别为虚拟变量来考察职工薪金的模型如下：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \mu_i$$

其中，Y_i 为职工的薪金；X_i 为工龄；D_i=1 代表男性，D_i=0 代表女性。

一、虚拟变量的引入

虚拟变量作为解释变量引入模型有两种基本方式：加法方式和乘法方式。

1. 加法方式

上述职工薪金模型中性别虚拟变量的引入采取了加法方式，即模型中将虚拟变量以相加的形式引入模型。在该模型中，如果仍假定 $\mathrm{E}(\mu_i)$ =0，则女职工的平均薪金为

$$\mathrm{E}(Y_i \mid X_i, D_i = 0) = \beta_0 + \beta_1 X_i$$

男职工的平均薪金为

$$\mathrm{E}(Y_i \mid X_i, D_i = 1) = (\beta_0 + \beta_2) + \beta_1 X_i$$

从几何意义上看(图 5.1.1)，假定 $\beta_2 > 0$，则两个函数有相同的斜率，但有不同的截距。这意味着，男女职工平均薪金对工龄的变化率是一样的，但两者的平均薪金水平相差 β_2。可以通过传统的回归检验，对 β_2 的统计显著性进行检验，以判断男女职工的平均薪金水平是否有显著差异。

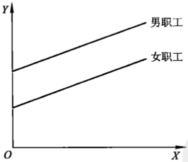


图 5.1.1 男女职工平均薪金示意图

又例如，在截面数据基础上，考虑个人保健支出对个人收入和教育水平的回归。教育水平考虑三个层次：高中以下，高中，大学及其以上。这时需要引入两个虚拟变量：

$$D_1 = \begin{cases} 1, & \text{高中} \\ 0, & \text{其他} \end{cases}$$

$$D_2 = \begin{cases} 1, & \text{大学及其以上} \\ 0, & \text{其他} \end{cases}$$

模型可设定如下：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \mu_i$$

在 $\mathrm{E}(\mu_i)=0$ 的初始假定下，容易得到高中以下、高中、大学及其以上教育水平个人保健支出的函数：

高中以下：　$\mathrm{E}(Y_i \mid X_i, D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 X_i$

高中：　$\mathrm{E}(Y_i \mid X_i, D_1 = 1, D_2 = 0) = (\beta_0 + \beta_2) + \beta_1 X_i$

大学及其以上：　$\mathrm{E}(Y_i \mid X_i, D_1 = 0, D_2 = 1) = (\beta_0 + \beta_3) + \beta_1 X_i$

假定 $\beta_3 > \beta_2$，则其几何意义如图 5.1.2 所示。

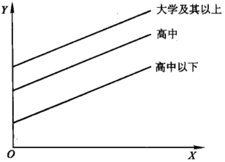


图 5.1.2　不同教育程度人员保健支出示意图

还可将多个虚拟变量引入模型中以考察多种“定性”因素的影响。例如，在职工薪金的例子中，再引入学历的虚拟变量 D_2：

$$D_2 = \begin{cases} 1, & \text{本科及以上学历} \\ 0, & \text{本科以下学历} \end{cases}$$

则职工薪金的回归模型可设计如下：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \mu_i$$

于是，不同性别、不同学历职工的平均薪金分别由下面各式给出：

女职工本科以下学历的平均薪金：

$$\mathrm{E}(Y_i \mid X_i, D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 X_i$$

男职工本科以下学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 1, D_2 = 0) = (\beta_0 + \beta_2) + \beta_1 X_i$$

女职工本科以上学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 0, D_2 = 1) = (\beta_0 + \beta_3) + \beta_1 X_i$$

男职工本科以上学历的平均薪金：

$$E(Y_i \mid X_i, D_1 = 1, D_2 = 1) = (\beta_0 + \beta_2 + \beta_3) + \beta_1 X_i$$

2．乘法方式

加法方式引入虚拟变量，可以考察截距的不同，而在许多情况下，往往是斜率有变化，或斜率、截距同时发生变化。斜率的变化可通过乘法的方式引入虚拟变量来测度。

例如，中国农村居民的边际消费倾向会与城镇居民的边际消费倾向不同吗？这种消费倾向的差异可通过在收入的系数中引入虚拟变量来考察。如，设

$$D_i = \begin{cases} 1, & \text{农村居民} \\ 0, & \text{城镇居民} \end{cases}$$

则全体居民的消费模型可建立如下：

$$C_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + \mu_i$$

这里，C，X 分别表示居民家庭人均年消费支出与年可支配收入，虚拟变量 D 以与 X 相乘的方式引入了模型中，从而可用来考察消费倾向的差异。在 $E(\mu_i) = 0$ 的假定下，上述模型所表示的函数可化为

农村居民：$E(C_i \mid X_i, D_i = 1) = \beta_0 + (\beta_1 + \beta_2) X_i$

城镇居民：$E(C_i \mid X_i, D_i = 0) = \beta_0 + \beta_1 X_i$

显然，如果 β_2 显著地异于 0，则可判定农村居民与城镇居民的边际消费倾向有差异。

如果采用的是时间序列，则可类似地引入某时点前后分别取值 1，0 的虚拟变量，来考察斜率项的变化。如以第 t^* 年为界，设置如下虚拟变量

$$D_t = \begin{cases} 0, & t^*\text{年前} \\ 1, & t^*\text{年后(含}t^*\text{年)} \end{cases}$$

可建立如下消费函数模型：

$$C_t = \beta_0 + \beta_1 X_t + \beta_2 D_t X_t + \mu_t$$

在 $E(\mu_t) = 0$ 的假定下，上述模型所表示的消费函数可化为

t^* 年前：$E(C_t \mid X_t, D_t = 1) = \beta_0 + \beta_1 X_t$

t^* 年后：$E(C_t \mid X_t, D_t = 0) = \beta_0 + (\beta_1 + \beta_2) X_t$

如果 β_2 显著地异于 0，则可判定斜率项在 t^* 年前后有显著差异。

例 5.1.1

表 5.1.1 中给出了中国内地 2007 年城镇居民家庭人均可支配收入与人均生活消费支出，以及农村居民家庭人均纯收入与人均生活消费支出的相关数据。可由这组数据来判断 2007 年中国内地农村居民与城镇居民边际消费倾向是否有差异。

表 5.1.1　2007 年中国内地居民人均可支配收入与生活消费支出数据(元)

	城镇居民		农村居民	
	人均消费	人均可支配收入	人均消费	人均纯收入
北京	15 330.4	21 988.7	6 399.3	9 439.6
天津	12 028.9	16 357.4	3 538.3	7 010.1
河北	8 235.0	11 690.5	2 786.8	4 293.4
山西	8 101.8	11 565.0	2 682.6	3 665.7
内蒙古	9 281.5	12 377.8	3 256.2	3 953.1
辽宁	9 429.7	12 300.4	3 368.2	4 773.4
吉林	8 560.3	11 285.5	3 065.4	4 191.3
黑龙江	7 519.3	10 245.3	3 117.4	4 132.3
上海	17 255.4	23 622.7	8 844.9	10 144.6
江苏	10 715.2	16 378.0	4 786.2	6 561.0
浙江	14 091.2	20 573.8	6 801.6	8 265.2
安徽	8 531.9	11 473.6	2 754.0	3 556.3
福建	11 055.1	15 506.1	4 053.5	5 467.1
江西	7 810.7	11 451.7	2 994.5	4 044.7
山东	9 666.6	14 264.7	3 621.6	4 985.3
河南	7 826.7	11 477.1	2 676.4	3 851.6
湖北	8 701.2	11 485.8	3 090.0	3 997.5
湖南	8 990.7	12 293.5	3 377.4	3 904.2
广东	14 336.9	17 699.3	4 202.3	5 624.0
广西	8 151.3	12 200.4	2 747.5	3 224.1
海南	8 292.9	10 996.9	2 556.6	3 791.4
重庆	9 890.3	12 590.8	2 526.7	3 509.3
四川	8 692.0	11 098.3	2 747.3	3 546.7
贵州	7 758.7	10 678.4	1 913.7	2 374.0
云南	7 921.8	11 496.1	2 637.2	2 634.1
西藏	7 532.1	11 130.9	2 217.6	2 788.2
陕西	8 427.1	10 763.3	2 559.6	2 644.7
甘肃	7 875.8	10 012.3	2 017.2	2 328.9
青海	7 512.4	10 276.1	2 446.5	2 683.8
宁夏	7 817.3	10 859.3	2 528.8	3 180.8
新疆	7 874.3	10 313.4	2 350.6	3 183.0

资料来源：《中国统计年鉴》(2008)。

以Y为人均消费，X为人均可支配收入，农村与城镇居民消费函数可写成：

农村居民：$Y_i = \alpha_1 + \alpha_2 X_i + \mu_{i1}$ $\qquad i = 1, 2, \cdots, n_1$

城镇居民：$Y_i = \beta_1 + \beta_2 X_i + \mu_{i2}$ $\qquad i = 1, 2, \cdots, n_2$

则有可能出现下述四种情况中的一种：

(1) $\alpha_1 = \beta_1$ 且 $\alpha_2 = \beta_2$，即两个回归相同，称为**重合回归**(coincident regression)；

(2) $\alpha_1 \neq \beta_1$ 但 $\alpha_2 = \beta_2$，即两个回归的差异仅在其截距，称为**平行回归**(parallel regression)；

(3) $\alpha_1 = \beta_1$ 但 $\alpha_2 \neq \beta_2$，即两个回归的差异仅在其斜率，称为**汇合回归**(concurrent regression)；

(4) $\alpha_1 \neq \beta_1$ 且 $\alpha_2 \neq \beta_2$，即两个回归完全不同，称为**相异回归**(dissimilar regression)。

可以通过第二章介绍的邹氏稳定性检验来考察是否有结构变化的问题。这一问题也可通过引入乘法形式的虚拟变量来解决。

将本例中的n_1与n_2次观察值合并，并用以估计以下回归模型：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_3 D_i + \beta_4 (D_i X_i) + \mu_i$$

其中D_i为引入的虚拟变量，农村居民取值 1，城镇居民取值 0，则

$$\mathrm{E}(Y_i \mid D_i = 0, X_i) = \beta_0 + \beta_1 X_i$$

$$\mathrm{E}(Y_i \mid D_i = 1, X_i) = (\beta_0 + \beta_3) + (\beta_1 + \beta_4) X_i$$

分别表示城镇居民消费函数与农村居民消费函数。在显著性检验中，如果β_4等于 0 的假设被拒绝，则说明农村居民与城镇居民边际消费倾向不同。

由表 5.1.1 中数据得到的具体回归结果为

$$\hat{Y}_i = 450.33 + 0.692\,0 X_i - 271.14 D_i + 0.027\,5 D_i X_i$$

$$(1.23) \qquad (26.61) \qquad (-0.62) \qquad (0.49)$$

$$\overline{R}^2 = 0.979\,9 \qquad F = 992.44 \qquad \text{D.W.} = 1.77$$

由β_3与β_4的t检验值可知，该两参数并非显著地不等于 0，显示 2007 年农村居民与城镇居民的边际消费倾向并无显著差异，他们有着共同的消费函数：

$$\hat{Y}_i = 253.39 + 0.705\,9 X_i$$

如果使用邹氏检验，可得F=0.19，该值小于 5%显著性水平下、自由度为(2，58)的 F 临界值$F_{0.05}(2,58)$=3.16，不拒绝中国农村居民与城镇居民消费行为无差异的假设。可见引入虚拟变量的检验与邹氏检验结果相同。但不同的是，引入虚拟变量的检验可以检验结构的变化来自截距项还是来自斜率项，而邹氏检验却不能。另外，引入虚拟变量的检验要比邹氏检验简单得多。

3. 临界指标的虚拟变量的引入

在经济发生转折时，可通过建立临界指标的虚拟变量模型来反映。例如，进口消费品数量Y主要取决于国民收入X的多少，中国在改革开放前后，Y对X的回归关系明显不同。这时，可以$t^*=1979$为转折期，以1979年的国民收入X_t^*为临界值，设如下虚拟变量：

$$D_t=\begin{cases}1, & t \geqslant t^* \\ 0, & t < t^*\end{cases}$$

则进口消费品的回归模型可建立如下：

$$Y_t=\beta_0+\beta_1 X_t+\beta_2(X_t-X_t^*)D_t+\mu_t$$

如果用普通最小二乘法得到该模型的回归方程为

$$\hat{Y}_t=\hat{\beta}_0+\hat{\beta}_1 X_t+\hat{\beta}_2(X_t-X_t^*)D_t$$

则两个时期进口消费品函数分别为

当$t<t^*=1979$时

$$\hat{Y}_t=\hat{\beta}_0+\hat{\beta}_1 X_t$$

当$t \geqslant t^*=1979$时

$$\hat{Y}_t=(\hat{\beta}_0-\hat{\beta}_2 X_t^*)+(\hat{\beta}_1+\hat{\beta}_2)X_t$$

其几何图形如图5.1.3所示。

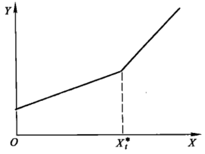


图 5.1.3 转折期回归示意图

二、虚拟变量的设置原则

虚拟变量的个数须按以下原则确定：每一定性变量所需的虚拟变量个数要比该定性变量的类别数少1，即如果有m个定性变量，只在模型中引入m–1个虚拟变量。

PDG

例如，已知冷饮的销售量Y除受k种定量变量X_k的影响外，还受春、夏、秋、冬四季变化的影响。要考察该四季的影响，只需引入三个虚拟变量即可：

$$D_{t1}=\begin{cases}1, & \text{春季}\\0, & \text{其他}\end{cases}$$

$$D_{t2}=\begin{cases}1, & \text{夏季}\\0, & \text{其他}\end{cases}$$

$$D_{t3}=\begin{cases}1, & \text{秋季}\\0, & \text{其他}\end{cases}$$

则冷饮销售量的模型为

$$Y_t=\beta_0+\beta_1X_{t1}+\cdots+\beta_kX_{tk}+\alpha_1D_{t1}+\alpha_2D_{t2}+\alpha_3D_{t3}+\mu_t$$

在上述模型中，若再引入第 4 个虚拟变量

$$D_{t4}=\begin{cases}1, & \text{冬季}\\0, & \text{其他}\end{cases}$$

则冷饮销售模型变量为

$$Y_t=\beta_0+\beta_1X_{t1}+\cdots+\beta_kX_{tk}+\alpha_1D_{t1}+\alpha_2D_{t2}+\alpha_3D_{t3}+\alpha_4D_{t4}+\mu_t$$

其矩阵形式为

$$\boldsymbol{Y}=(\boldsymbol{X}\quad\boldsymbol{D})\begin{pmatrix}\boldsymbol{\beta}\\\boldsymbol{\alpha}\end{pmatrix}+\boldsymbol{\mu}$$

如果只取 6 个观测值，其中春季与夏季取了两次，秋、冬各取到一次观测值，则其中

$$(\boldsymbol{X}\quad\boldsymbol{D})=\begin{pmatrix}1 & X_{11} & \cdots & X_{1k} & 1 & 0 & 0 & 0\\1 & X_{21} & \cdots & X_{2k} & 0 & 1 & 0 & 0\\1 & X_{31} & \cdots & X_{3k} & 0 & 0 & 1 & 0\\1 & X_{41} & \cdots & X_{4k} & 0 & 0 & 0 & 1\\1 & X_{51} & \cdots & X_{5k} & 0 & 1 & 0 & 0\\1 & X_{61} & \cdots & X_{6k} & 1 & 0 & 0 & 0\end{pmatrix}$$

$$\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\\beta_1\\\vdots\\\beta_k\end{pmatrix},\qquad\boldsymbol{\alpha}=\begin{pmatrix}\alpha_1\\\alpha_2\\\alpha_3\\\alpha_4\end{pmatrix}$$

显然，$(\boldsymbol{X}\quad\boldsymbol{D})$中的第 1 列可表示成后 4 列的线性组合，从而$(\boldsymbol{X}\quad\boldsymbol{D})$不满秩，参数无法唯一求出。这就是所谓的“虚拟变量陷阱”，应该避免这种情况发生。

§5.2 滞后变量模型

在经济活动中，广泛存在着时间滞后效应，即动态性。某些经济变量不仅受到同期各种因素的影响，而且也受到过去某些时期的各种因素甚至自身的过去值的影响。通常把这种过去时期的具有滞后作用的变量叫做**滞后变量**(lagged variable)，含有滞后变量的模型称为**滞后变量模型**。

滞后变量模型考虑了时间因素的作用，使静态分析的问题有可能成为动态分析。含有滞后被解释变量的模型，又称**动态模型**(dynamic model)。

一、滞后变量模型

1. 滞后效应与产生滞后效应的原因

一般说来，被解释变量与解释变量的因果关系不一定就在瞬时发生，可能存在时间的滞后，或者说解释变量的变化可能需要经过一段时间才能完全对被解释变量产生影响。同样地，被解释变量当前的变化也可能受其自身过去水平的影响，这种被解释变量受到自身或另一解释变量的前几期值影响的现象称为**滞后效应**，表示前几期值的变量称为**滞后变量**。例如，在研究消费函数时，通常认为，本期的消费除了受本期的收入水平影响之外，还受前一期收入以及前一期消费水平的影响：

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \beta_3 C_{t-1} + \mu_t$$

这就是含有滞后变量的模型，Y_{t-1}，C_{t-1}为滞后变量。

现实经济生活中，产生滞后效应的原因很多，主要有以下几个方面。

(1) 心理原因。由于人们固有的心理定式和行为习惯，其行为方式往往滞后于经济形势的变化，如中彩票的人不可能很快改变其生活方式。因此，以往的行为延续产生了滞后效应。

(2) 技术原因。在现实经济运行中，从生产到流通再到使用，每一个环节都需要一段时间，从而形成时滞。例如，工业生产中，当年的产出在某种程度上依赖于过去若干期内投资形成的固定资产。又如，当年农产品产量主要取决于过去一年价格的高低。

(3) 制度原因。契约、管理制度等因素也会造成经济行为的滞后。例如，定期存款到期才能提取，造成了它对社会购买力的影响具有滞后性；过去的订货合同影响着当前产品的产量等。

2. 滞后变量模型

以滞后变量作为解释变量，就得到**滞后变量模型**。它的一般形式为

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_q Y_{t-q} + \alpha_0 X_t + \alpha_1 X_{t-1} + \cdots + \alpha_s X_{t-s} + \mu_t \tag{5.2.1}$$

其中，q，s 为滞后时间间隔，Y_{t-q} 为被解释变量 Y 的第 q 期滞后，X_{t-s} 为解释变量 X 的第 s 期滞后。由于模型既含有 Y 对自身滞后变量的回归，还包括着解释变量 X 分布在不同时期的滞后变量，因此一般称为**自回归分布滞后模型**(Autoregressive Distributed-lag Model, ADL)。若滞后期长度有限，称模型为**有限自回归分布滞后模型**；若滞后期无限，称模型为**无限自回归分布滞后模型**。

(1) 分布滞后模型

如果滞后变量模型中没有滞后被解释变量，仅有解释变量 X 的当期值及其若干期的滞后值，称为**分布滞后模型**(distributed-lag model)。分布滞后模型的一般形式为

$$Y_t = \alpha + \sum_{i=0}^{s} \beta_i X_{t-i} + \mu_t \tag{5.2.2}$$

分布滞后模型的各系数体现了解释变量的当期值和各期滞后值对被解释变量的不同影响程度，因此也称为**乘数**(multiplier)。β_0 称为**短期**(short-run)或**即期乘数**(impact multiplier)，表示本期 X 变化一个单位对 Y 平均值的影响程度。$\beta_i\,(i=1,2,\cdots,s)$ 称为**动态乘数**或**延迟系数**，表示各滞后期 X 的变动对 Y 平均值影响的大小。$\sum_{i=0}^{s}\beta_i$ 则称为**长期**(long-run)或**均衡乘数**(total distributed-lag multiplier)，表示 X 变动一个单位，由于滞后效应而形成的对 Y 平均值总影响的大小。

由(5.2.2)式知，如果各期的 X 值保持不变，则 X 与 Y 间的长期或均衡关系即为

$$\mathrm{E}(Y) = \alpha + (\sum_{i=0}^{s} \beta_i) X \tag{5.2.3}$$

(2) 自回归模型

如果滞后变量模型中的解释变量仅包含 X 的当期值与被解释变量 Y 的一个或多个滞后值，则称为**自回归模型**(autoregressive model)。自回归模型的一般形式为

$$Y_t = \alpha_0 + \alpha_1 X_t + \sum_{i=1}^{q} \beta_i Y_{t-i} + \mu_t \tag{5.2.4}$$

其中，滞后期长度 q 也称为自回归模型的**阶数**(order)。而

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \mu_t \tag{5.2.5}$$

称为**一阶自回归模型**(first-order autoregressive model)。

二、分布滞后模型的参数估计

1．分布滞后模型估计的困难

对于无限期的分布滞后模型，由于样本观测值的有限性，使得无法直接对其进行估计。对于有限期的分布滞后模型，普通最小二乘回归也会遇到如下问题：

(1) 没有先验准则确定滞后期长度；

(2) 如果滞后期较长，将缺乏足够的自由度进行统计检验；

(3) 同名变量滞后值之间可能存在高度线性相关，即模型存在高度的多重共线性。

2．分布滞后模型的修正估计方法

针对上述困难，人们在大量研究的基础上提出了一系列的修正估计方法，但并不很完善。各种方法的基本思想大致相同，即都是通过对各滞后变量加权，组成线性合成变量而有目的地减少滞后变量的数目，以缓解多重共线性，保证自由度。

(1) 经验加权法

对于有限期分布滞后模型，往往根据实际问题的特点，以及人们的经验给各滞后变量指定权数，并按权数构成各滞后变量的线性组合，形成新的变量，再进行估计。权数的类型有以下三类。

第一类，递减型，即认为权数是递减的，X 的近期值对 Y 的影响较远期值大。例如，消费函数中，收入的近期值对消费的影响显然大于远期值的影响。一个滞后期为 3 的一组权数可取值如下：

$$\frac{1}{2},\ \frac{1}{4},\ \frac{1}{6},\ \frac{1}{8}$$

则新的线性组合变量为

$$W_{t1}=\frac{1}{2}X_t+\frac{1}{4}X_{t-1}+\frac{1}{6}X_{t-2}+\frac{1}{8}X_{t-3}$$

第二类，矩型，即认为权数是相等的，X 的逐期滞后值对 Y 的影响相同。例如，对滞后期为 3 的分布滞后模型，可指定相等权数为 $\frac{1}{4}$，则新的线性组合变量为

$$W_{t2}=\frac{1}{4}X_t+\frac{1}{4}X_{t-1}+\frac{1}{4}X_{t-2}+\frac{1}{4}X_{t-3}$$

第三类，倒 V 型，在这种形式中，假定权数先递增后递减呈倒“V”型。例如，在一个较长建设周期的投资中，历年投资 X 对产出 Y 的影响，往往是周期的期中投资额最大，因此对产出的贡献最大。设滞后期为 4，则一组权数可取为

$$\frac{1}{6},\ \frac{1}{4},\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{5}$$

于是新变量为

$$W_{t3}=\frac{1}{6}X_t+\frac{1}{4}X_{t-1}+\frac{1}{2}X_{t-2}+\frac{1}{3}X_{t-3}+\frac{1}{5}X_{t-4}$$

例 5.2.1

对一个分布滞后模型

$$Y_t=\alpha_0+\beta_0X_t+\beta_1X_{t-1}+\beta_2X_{t-2}+\beta_3X_{t-3}+\mu_t$$

给定递减权数

$$\frac{1}{2},\ \frac{1}{4},\ \frac{1}{6},\ \frac{1}{8}$$

构成新变量

$$W_{t1}=\frac{1}{2}X_t+\frac{1}{4}X_{t-1}+\frac{1}{6}X_{t-2}+\frac{1}{8}X_{t-3}$$

则原模型变为

$$Y_t=\alpha_0+\alpha_1W_{t1}+\mu_t$$

如果该模型满足普通最小二乘法的经典假设，就可进行普通最小二乘估计，估计出参数 $\hat{\alpha}_0$，$\hat{\alpha}_1$。

假设 $\hat{\alpha}_0$，$\hat{\alpha}_1$ 已估计出，分别为 0.5 与 0.8，则可写出原模型的估计结果：

$$\hat{Y}_t=0.5+\frac{0.8}{2}X_t+\frac{0.8}{4}X_{t-1}+\frac{0.8}{6}X_{t-2}+\frac{0.8}{8}X_{t-3}$$

或

$$\hat{Y}_t=0.5+0.4X_t+0.2X_{t-1}+0.133X_{t-2}+0.1X_{t-3}$$

经验加权法的优点是简单易行，缺点是设置权数的随意性较大。通常的做法是多选几组权数，分别估计出几个模型，然后根据各统计检验（R^2 检验，F 检验，t 检验，D.W. 检验)，从中选择最佳估计式。

(2) 阿尔蒙(Almon)多项式法

该方法的主要思想仍是针对有限滞后期模型，通过阿尔蒙变换，定义新变量，以减少解释变量个数，然后用普通最小二乘法估计参数。主要步骤如下。

第一步，阿尔蒙变换：对于分布滞后模型

$$Y_t=\alpha+\sum_{i=0}^{s}\beta_iX_{t-i}+\mu_t \tag{5.2.2}$$

假定其回归系数 β_i 可用一个关于滞后期 i 的适当阶数的多项式来表示，即

$$\beta_i=\sum_{k=0}^{m}\alpha_k(i)^k\ , \qquad i=0,1,\cdots,s \tag{5.2.6}$$

其中 $m<s$。阿尔蒙变换要求先验地确定适当阶数 m，如取 m=2，得

$$\beta_i=\sum_{k=0}^{2}\alpha_k(i)^k=\alpha_0+\alpha_1 i+\alpha_2(i)^2\ , \qquad i=0,1,\cdots,s \tag{5.2.7}$$

将(5.2.7)式代入(5.2.2)式得

$$\begin{aligned}Y_t&=\alpha+\sum_{i=0}^{s}\left[\sum_{k=0}^{2}\alpha_k(i)^k\right]X_{t-i}+\mu_t\\&=\alpha+\alpha_0\sum_{i=0}^{s}X_{t-i}+\alpha_1\sum_{i=0}^{s}(i)X_{t-i}+\alpha_2\sum_{i=0}^{s}(i)^2X_{t-2}+\mu_t\end{aligned}$$

定义新变量

$$W_{t0}=\sum_{i=0}^{s}X_{t-i}\ , \qquad W_{t1}=\sum_{i=0}^{s}(i)X_{t-i}\ , \qquad W_{t2}=\sum_{i=0}^{s}(i)^2X_{t-i}$$

将原模型转换为

$$Y_t=\alpha+\alpha_0W_{t0}+\alpha_1W_{t1}+\alpha_2W_{t2}+\mu_t \tag{5.2.8}$$

第二步，模型的普通最小二乘估计：对变换后的模型(5.2.8)式进行普通最小二乘估计。将得到的参数估计值 $\hat{\alpha}_0$，$\hat{\alpha}_1$，$\hat{\alpha}_2$ 代入(5.2.7)式，求出滞后分布模型参数的估计值 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_s$。

由于 $m<s$，可以认为原模型存在的自由度不足和多重共线性问题已得到改善。需注意的是，在实际估计中，阿尔蒙多项式的阶数 m 一般取 2 或 3，不超过 4，否则达不到减少变量个数的目的。

例 5.2.2

表 5.2.1 给出了中国电力行业基本建设投资 X 与发电量 Y 的相关资料，拟建立一个多项式分布滞后模型来考察两者的关系。

表 5.2.1　中国电力工业固定资产投资与发电量

年度	固定资产投资 (亿元)	发电量 (亿千瓦时)	年度	固定资产投资 (亿元)	发电量 (亿千瓦时)
1975	29	1 958	1979	48	2 820
1976	32	2 031	1980	41	3 006
1977	33	2 234	1981	34	3 093
1978	49	2 566	1982	42	3 277

续表

年度	固定资产投资 (亿元)	发电量 (亿千瓦时)	年度	固定资产投资 (亿元)	发电量 (亿千瓦时)
1983	56	3 514	1993	628	8 395
1984	72	3 770	1994	839	9 218
1985	97	4 107	1995	1 009	10 070
1986	133	4 495	1996	1 193	10 800
1987	176	4 973	1997	1 550	11 345
1988	215	5 452	1998	1 743	11 662
1989	222	5 848	1999	1 854	12 393
1990	300	6 212	2000	2 126	13 556
1991	354	6 775	2001	1 945	14 717
1992	445	7 539	2002	2 297	16 405

资料来源：电力行业固定资产投资来自《中国电力统计年鉴》，发电量来自《中国统计年鉴》。

为了测算电力行业固定资产投资增长与发电量增长间的变动关系，我们拟建立如下双对数线性模型：

$$\ln Y_t = \alpha + \sum_{i=0}^{s} \beta_i \ln X_{t-i} + \mu_t$$

由于无法预知电力行业基本建设投资对发电量影响的时滞期，需取不同的滞后期试算。经过试算发现，在2阶阿尔蒙多项式变换下，滞后期数取到第7期，估计结果的经济意义比较合理。2阶阿尔蒙多项式估计结果如下：

$$\ln \hat{Y}_t = 6.732 + 0.025W_{t0} - 0.023W_{t1} + 0.006W_{t2}$$

$$(203.5) \quad (2.29) \qquad (-7.48) \qquad (2.80)$$

$$\bar{R}^2 = 0.999\,64 \qquad F = 1\,764.9 \qquad \text{D.W.} = 0.80$$

通过(5.2.6)式求得的分布滞后模型参数估计值为

$$\hat{\beta}_0 = 0.150,\ \hat{\beta}_1 = 0.096,\ \hat{\beta}_2 = 0.054,\ \hat{\beta}_3 = 0.024,$$

$$\hat{\beta}_4 = 0.008,\ \hat{\beta}_5 = 0.003,\ \hat{\beta}_6 = 0.010,\ \hat{\beta}_7 = 0.029$$

最后得到分布滞后模型估计式为

$$\ln \hat{Y}_t = 6.732 + 0.150 \ln X_t + 0.096 \ln X_{t-1} + 0.054 \ln X_{t-2} + 0.024 \ln X_{t-3}$$

$$(203.5) \qquad (8.80) \qquad (15.77) \qquad (7.36) \qquad (2.29)$$

$$+0.008 \ln X_{t-4} + 0.003 \ln X_{t-5} + 0.010 \ln X_{t-6} + 0.029 \ln X_{t-7}$$

$$(0.71) \qquad (0.37) \qquad (1.41) \qquad (1.59)$$

需要说明的是，尽管从估计的α的标准差换算估计的β的标准差较为繁琐，但大多数应用软件都具有此功能，本例是Eviews软件的估计结果。另外，从估计的β及其t检验看，滞后期超过4年时，各参数不再显著，表明4年之前的固定资产投资的变动对当年电力增长的作用很有限。最后，为了比较，下面给出直接对滞后7期的模型进行普通最小二乘估计的结果：

$$\ln\hat{Y}_t = 6.74 + 0.124\ln X_t + 0.167\ln X_{t-1} - 0.010\ln X_{t-2} + 0.049\ln X_{t-3}$$

$$(119.4)\quad (1.36)\qquad (1.34)\qquad (-0.07)\qquad (0.37)$$

$$-0.002\ln X_{t-4} - 0.001\ln X_{t-5} + 0.047\ln X_{t-6} + 0.0003\ln X_{t-7}$$

$$(-0.02)\qquad (-0.02)\qquad (0.50)\qquad (0.01)$$

$$\overline{R}^2 = 0.997\,1 \qquad F = 480.2 \qquad \text{D.W.} = 0.84$$

可以看出，拟合优度有所下降，且所有变量均未通过 t 检验，而且负值的出现也与实际经济意义不相符。

(3) 科伊克(Koyck)方法

科伊克方法是将无限分布滞后模型转换为自回归模型，然后进行估计。对于无限分布滞后模型

$$Y_t = \alpha + \sum_{i=0}^{+\infty}\beta_i X_{t-i} + \mu_t \tag{5.2.9}$$

科伊克变换假设偏回归系数 β_i 随滞后期 i 按几何级数衰减：

$$\beta_i = \beta_0\lambda^i\,, \qquad i=0,1,2,\cdots \tag{5.2.10}$$

其中 $0<\lambda<1$，λ 称为分布滞后衰减率，$1-\lambda$ 称为调整速率(speed of adjustment)。

科伊克变换的具体做法如下。

将科伊克假定(5.2.10)式代入模型(5.2.9)式，得

$$Y_t = \alpha + \beta_0\sum_{i=0}^{+\infty}\lambda^i X_{t-i} + \mu_t \tag{5.2.11}$$

将(5.2.11)式滞后一期并乘以 λ，得

$$\lambda Y_{t-1} = \lambda\alpha + \beta_0\sum_{i=1}^{+\infty}\lambda^i X_{t-i} + \lambda\mu_{t-1} \tag{5.2.12}$$

将(5.2.11)式减去(5.2.12)式得科伊克变换模型

$$Y_t - \lambda Y_{t-1} = (1-\lambda)\alpha + \beta_0 X_t + \mu_t - \lambda\mu_{t-1} \tag{5.2.13}$$

整理得科伊克模型的一般形式

$$Y_t = (1-\lambda)\alpha + \beta_0 X_t + \lambda Y_{t-1} + v_t \tag{5.2.14}$$

其中，

$$v_t = \mu_t - \lambda\mu_{t-1}$$

科伊克模型有两个特点：一是以一个滞后被解释变量 Y_{t-1} 代替了大量的滞后解释变量 X_{t-i}，最大限度地节省了自由度，解决了滞后期长度 s 难以确定的问题；二是由于滞后一期的被解释变量 Y_{t-1} 与 X_t 的线性相关程度肯定可以小于 X 的各期滞后值之间的相关程度，从而缓解了多重共线性。

但科伊克变换同时也产生了两个新问题：一是模型存在随机干扰项 v_t 的一阶自相关性；二是滞后被解释变量 Y_{t-1} 与随机项 v_t 不独立，即 $\mathrm{Cov}(Y_{t-1}, v_t) \neq 0$。这些新问题需要进一步解决。

三、自回归模型的参数估计

1．自回归模型的构造

从上面的讨论中已看出，一个无限期分布滞后模型可以通过科伊克变换转化为自回归模型。事实上，许多滞后变量模型都可以转化为自回归模型，自回归模型是经济生活中更常见的模型。下面我们以自适应预期模型以及局部调整模型为例进行说明。

(1) 自适应预期(adaptive expectation)模型

在某些实际问题中，因变量Y_t并不取决于解释变量的当前实际值X_t，而取决于X_t的“预期水平”或“长期均衡水平” X_t^e。例如，家庭本期消费水平，取决于本期收入的预期值；市场上某种商品供求量，决定于本期该商品价格的均衡值。因此，自适应预期模型最初的表现形式是

$$Y_t = \beta_0 + \beta_1 X_t^e + \mu_t \tag{5.2.15}$$

由于预期变量是不可实际观测的，往往作如下自适应预期假定：

$$X_t^e - X_{t-1}^e = r(X_t - X_{t-1}^e) \tag{5.2.16}$$

其中r为**预期系数**(coefficient of expectation)，$0 \leqslant r \leqslant 1$。该式的经济含义为：“经济行为者将根据过去的经验修改他们的预期”，即本期预期值的形成是一个逐步调整的过程，本期预期值的增量是本期实际值与前一期预期值之差的一部分，其比例为r。这个假定还可写成

$$X_t^e = rX_t + (1-r)X_{t-1}^e \tag{5.2.17}$$

即本期预期值为本期真值与前期预期值的加权和。

将(5.2.17)式代入(5.2.15)式得

$$Y_t = \beta_0 + \beta_1[rX_t + (1-r)X_{t-1}^e] + \mu_t \tag{5.2.18}$$

将(5.2.15)式滞后一期并乘以$1-r$，得

$$(1-r)Y_{t-1} = \beta_0(1-r) + \beta_1(1-r)X_{t-1}^e + (1-r)\mu_{t-1} \tag{5.2.19}$$

以(5.2.18)式减去(5.2.19)式，整理得

$$Y_t = \beta_0 r + \beta_1 r X_t + (1-r)Y_{t-1} + v_t \tag{5.2.20}$$

其中，$v_t = \mu_t - (1-r)\mu_{t-1}$。可见自适应预期模型转化为了一个自回归模型。

(2) 局部调整(partial adjustment)模型

局部调整模型主要是用来研究物资储备问题的。例如，企业为了保证生产和销售，必须保持一定的原材料储备。对应于一定的产量或销售量X_t，存在着预期的最佳库存Y_t^e。局部调整模型的最初形式为

$$Y_t^e = \beta_0 + \beta_1 X_t + \mu_t \tag{5.2.21}$$

显然，Y_t^e不可观测。由于生产条件的波动，生产管理方面的原因，库存储备Y_t

PDG

的实际变化量只是预期变化的一部分。储备按预定水平逐步进行调整，故有如下局部调整假设：

$$Y_t - Y_{t-1} = \delta(Y_t^e - Y_{t-1}) \tag{5.2.22}$$

其中，δ 为调整系数，$0 \leqslant \delta \leqslant 1$。局部调整假设还可写成

$$Y_t = \delta Y_t^e + (1-\delta)Y_{t-1} \tag{5.2.23}$$

表明实际库存储备是本期最佳预期库存与上期实际库存的加权和。

将(5.2.21)式代入(5.2.23)式得

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + (1-\delta)Y_{t-1} + \delta\mu_t \tag{5.2.24}$$

可见，局部调整模型可转化为一个自回归模型。

2. 自回归模型的参数估计

对于自回归模型(5.2.4)式，估计时的主要问题在于，滞后被解释变量的存在可能导致它与随机干扰项相关，以及随机干扰项出现序列相关性。如科伊克模型(5.2.14)式与自适应预期模型(5.2.20)式，就存在着滞后被解释变量 Y_{t-1} 与随机干扰项的同期相关性，同时，随机干扰项还是自相关的。而局部调整模型(5.2.24)式则存在着滞后被解释变量 Y_{t-1} 与随机干扰项的异期相关性。因此，对自回归模型的估计主要需视滞后被解释变量与随机干扰项的不同关系进行估计。下面以一阶自回归模型为例说明。

(1) 工具变量法

对于一阶自回归模型

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + \mu_t \tag{5.2.5}$$

若滞后被解释变量 Y_{t-1} 与随机干扰项 μ_t 同期相关(如科伊克模型与自适应预期模型)，则普通最小二乘估计是有偏的，并且不是一致估计。因此，对上述模型，通常采用工具变量法，即寻找一个新的经济变量 Z_t 作为 Y_{t-1} 的工具变量进行估计。参数估计量具有一致性。

在实际估计中，一般用 $\hat{Y}_{t-1}$ 作为 Y_{t-1} 的工具变量，其中 $\hat{Y}_{t-1}$ 是 X 的若干滞后的线性组合：

$$\hat{Y}_{t-1} = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_s X_{t-s}$$

由于模型(5.2.5)式中已假设随机干扰项 μ_t 与解释变量 X 及其滞后项不存在相关性，因此(5.2.5)式中的 μ_t 与 $\hat{Y}_{t-1}$ 不再线性相关。一个更简单的情形是直接用 X_{t-1} 作为 Y_{t-1} 的工具变量。

(2) 普通最小二乘法

若滞后被解释变量 Y_{t-1} 与随机干扰项 μ_t 同期无关(如局部调整模型)，可直接使用普通最小二乘法进行估计，得到一致估计量。

需要指出的是，上述工具变量法只解决了解释变量与随机干扰项相关对参

数估计所造成的影响，但没有解决μ_t的自相关问题。事实上，对于自回归模型，随机干扰项的自相关问题始终是存在的，对于此问题，至今没有完全有效的解决方法。唯一可做的，就是尽可能地建立“正确”的模型，以使序列相关性的程度减轻。

例 5.2.3

建立中国长期货币流通量需求模型。考虑到适度的货币流通量是市场稳定的一个基本要素，而影响货币需求的因素，不仅在本期，而且在长期内发挥作用。中国改革开放以来，对货币需求量的影响因素，主要有资金运用中的贷款额以及反映价格变化的居民消费者价格指数。显然，贷款额的增加，将使贷款转化为现金投放的需求增加，而物价水平的上升，就需要有更多的货币来支付同等的商品购买量。表 5.2.2 列出了 1978 到 2007 年中国货币流通量、贷款额以及居民消费价格指数的相关数据。

表 5.2.2　中国货币流通量、贷款额和居民消费价格指数历史数据

年度	货币流通量Y（亿元）	居民消费价格指数P（1990年=100）	贷款额X（亿元）	年度	货币流通量Y（亿元）	居民消费价格指数P（1990年=100）	贷款额X（亿元）
1978	212.0	46.2	1 850.0	1993	5 864.7	126.2	32 943.1
1979	267.7	47.1	2 039.6	1994	7 288.6	156.7	39 976.0
1980	346.2	50.6	2 414.3	1995	7 885.3	183.4	50 544.1
1981	396.3	51.9	2 860.2	1996	8 802.0	198.7	61 156.6
1982	439.1	52.9	3 180.6	1997	10 177.6	204.2	74 914.1
1983	529.8	54.0	3 589.9	1998	11 204.2	202.6	86 524.1
1984	792.1	55.5	4 766.1	1999	13 455.5	199.7	93 734.3
1985	987.8	60.6	5 905.6	2000	14 652.7	200.6	99 371.1
1986	1 218.4	64.6	7 590.8	2001	15 688.8	201.9	112 314.7
1987	1 454.5	69.3	9 032.5	2002	17 278.0	200.3	131 293.9
1988	2 134.0	82.3	10 551.3	2003	19 746.0	202.7	158 996.2
1989	2 344.0	97.0	14 360.1	2004	21 468.3	210.6	178 197.8
1990	2 644.4	100.0	17 680.7	2005	24 031.7	214.4	194 690.4
1991	3 177.8	103.4	21 337.8	2006	27 072.6	217.7	225 347.2
1992	4 336.0	110.0	26 322.9	2007	30 375.2	228.1	261 690.9

资料来源：《中国统计年鉴》(2008)、《中国统计资料 50 年汇编》。

长期货币流通量模型可设定为

$$Y_t^e = \beta_0 + \beta_1 X_t + \beta_2 P_t + \mu_t \tag{5.2.25}$$

其中，Y_t^e为长期货币流通需求量。由于长期货币流通需求量不可观测，作局部调整

$$Y_t - Y_{t-1} = \delta(Y_t^e - Y_{t-1}) \qquad (5.2.26)$$

其中，Y_t 为实际货币流通量，表明每年货币流通量的调整只是预期调整的一部分。将(5.2.25)式代入(5.2.26)式得短期货币流通量需求模型

$$Y_t = \delta\beta_0 + \delta\beta_1 X_t + \delta\beta_2 P_t + (1-\delta)Y_{t-1} + \delta\mu_t \qquad (5.2.27)$$

(5.2.27)式为一局部调整模型，普通最小二乘估计结果如下：

$$Y_t = -202.5 + 0.035\,7X_t + 7.455\,7P_t + 0.723\,6Y_{t-1} \qquad (5.2.28)$$

$$(-0.91) \quad (2.84) \quad (2.43) \quad (5.44)$$

$$R^2=0.998\,5, \quad \bar{R}^2=0.998\,4, \quad F=5\,868.9, \quad \text{D.W.}=1.724$$

由参数估计结果$1-\hat{\delta}$=0.723 6，得$\hat{\delta}$=0.276 4。

最后得到长期货币流通需求模型的估计式

$$Y_t^e = -732.6 + 0.129\,2X_t + 26.97P_t \qquad (5.2.29)$$

估计结果表明，贷款额对中国货币流通量的影响，短期为 0.04，长期为 0.13，即贷款额每增加 1 亿元，短期货币流通需求量将增加 0.04 亿元，长期货币流通需求将增加 0.13 亿元；而反映物价水平的居民消费价格指数对中国货币流通量的影响，短期为 7.46，长期为 26.97，即价格指数每增加 1 个百分点，将导致短期货币流通需求量增加7.46亿元，长期货币流通需求增加26.97亿元。

尽管回归结果表明 D.W.=1.724，但不能据此判断自回归模型不存在自相关。拉格朗日乘数统计量 LM=0.637，小于 5%显著性水平下自由度为 1 的 χ^2 分布的临界值 $\chi^2_{0.05}(1)$=3.84，可判断模型已不存在一阶自相关。如果直接对下式作普通最小二乘回归：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 P_t + \mu_t$$

可得

$$Y_t = -385.9 + 0.112\,2X_t + 6.147\,6P_t$$

$$(-1.72) \quad (62.94) \quad (6.78)$$

$$R^2=0.996\,3, \quad \bar{R}^2=0.996\,1, \quad F=3\,682.2, \quad \text{D.W.}=0.017$$

由 D.W.值容易判断该模型随机干扰项具有序列相关性，因此，(5.2.28)式的设定更“正确”。

四、格兰杰因果关系检验

自回归分布滞后模型旨在揭示某变量的变化受其自身及其他变量过去行为的影响。然而，许多经济变量有着相互的影响关系。例如，GDP 的增长能够促进消费的增长，而反过来，消费的变化又是 GDP 变化的一个组成部分，因此，消费增加又能促进 GDP 的增加。现在的问题是：当两个变量间在时间上

有先导-滞后关系时，能否从统计上考察这种关系是单向的还是双向的呢？即主要是一个变量过去的行为在影响另一个变量的当前行为，还是双方的过去行为在相互影响着对方的当前行为？格兰杰(Granger)提出了一个简单的检验程序，习惯上称为**格兰杰因果关系检验**(Granger test of causality)。

对两变量 X 与 Y，格兰杰因果关系检验要求估计以下回归：

$$Y_t = \beta_0 + \sum_{i=1}^{m} \beta_i Y_{t-i} + \sum_{i=1}^{m} \alpha_i X_{t-i} \tag{5.2.30}$$

$$X_t = \delta_0 + \sum_{i=1}^{m} \delta_i X_{t-i} + \sum_{i=1}^{m} \lambda_i Y_{t-i} \tag{5.2.31}$$

可能存在有四种检验结果：

(1) X 对 Y 有单向影响，表现为(5.2.30)式 X 各滞后项前的参数整体不为零，而(5.2.31)式 Y 各滞后项前的参数整体为零；

(2) Y 对 X 有单向影响，表现为(5.2.31)式 Y 各滞后项前的参数整体不为零，而(5.2.30)式 X 各滞后项前的参数整体为零；

(3) Y 与 X 间存在双向影响，表现为 Y 与 X 各滞后项前的参数整体不为零；

(4) Y 与 X 间不存在影响，表现为 Y 与 X 各滞后项前的参数整体为零。

格兰杰检验是通过受约束的 F 检验完成的。如针对 X 不是 Y 的格兰杰原因这一假设，即针对(5.2.30)式中 X 滞后项前的参数整体为零的假设，分别做包含与不包含 X 滞后项的回归，记前者的残差平方和为 RSS_U，后者的残差平方和为 RSS_R；再计算 F 统计量：

$$F = \frac{(\text{RSS}_\text{R} - \text{RSS}_\text{U})/m}{\text{RSS}_\text{U}/(n-k)} \tag{5.2.32}$$

式中，m 为 X 的滞后项的个数，n 为样本容量，k 为包含可能存在的常数项及其他变量在内的无约束回归模型的待估参数的个数。

如果计算的 F 值大于给定显著性水平 α 下 F 分布的相应的临界值 $F_\alpha(m,n-k)$，则拒绝原假设，认为 X 是 Y 的格兰杰原因。

需要指出的是，格兰杰因果关系检验对于滞后期长度的选择有时很敏感。不同的滞后期可能会得到完全不同的检验结果。因此，一般而言，常进行不同滞后期长度的检验，以检验模型中随机干扰项不存在序列相关的滞后期长度来选取滞后期。

由于假设检验的零假设是不存在因果关系，在该假设下 F 统计量服从 F 分布，因此严格地说，该检验应该称为**格兰杰非因果关系检验**。

例 5.2.4

例 2.6.2 曾建立了中国居民总量消费函数，即通过回归模型考察了1978—2006 年间中国居民总量消费 Y 与总量可支配收入 X 的关系。显然，在建立居民总量消费函数时，我们认为可支配收入 X 的变动主要影响着消费支出 Y 的变动。但从宏观经济及其国民核算的角度看，消费的增加无疑也会拉动产出的增长，从而促进了可支配收入的增长，即居民消费与收入间可能是互为因果关系的。那么从中国 1978—2006 年的数据看，两者间的**格兰杰因果关系**又会是怎样的呢？

取 1 阶滞后，Eviews 软件给出的估计结果如表 5.2.3 所示。

表 5.2.3 X 与 Y 的格兰杰因果关系检验

Pairwise Granger Causality Tests
Sample: 1978 2006
Lags:1

Null Hypothesis:	Obs	F-Statistic	Prob.
Y does not Granger Cause X	28	6.347 08	0.018 5
X does not Granger Cause Y		15.111 8	0.000 7

由伴随概率知，在 5%的显著性水平下，既拒绝“X 不是 Y 的格兰杰原因”的假设，也拒绝“Y 不是 X 的格兰杰原因”的假设。因此，从 1 阶滞后的情况看，可支配收入 X 的增长与居民消费支出 Y 的增长互为格兰杰原因。从检验模型随机干扰项 1 阶序列相关的拉格朗日乘数检验看，以 Y 为被解释变量的模型的 LM=0.897，对应的伴随概率 P=0.343，表明在 5%的显著性水平下，该检验模型不存在序列相关性；但是，以 X 为被解释变量的模型的 LM=11.37，对应的伴随概率 P= 0.001，表明在 5%的显著性水平下，该检验模型存在严重的序列相关性。

表 5.2.4 给出了取 1～4 阶滞后的检验结果。可以看出，从 2 阶滞后期开始，检验模型都拒绝了“X 不是 Y 的格兰杰原因”的假设，而不拒绝“Y 不是 X 的格兰杰原因”的假设。当然，滞后阶数为 2 或 3 时，根据两类检验模型都不存在序列相关性，再由赤池信息准则，发现滞后 2 阶检验模型拥有较小的 AIC 值。据此，可判断可支配收入 X 是居民消费支出 Y 的格兰杰原因，而不是相反，即国民收入的增加更大程度地影响着消费的增加。

表 5.2.4 中国可支配收入 X 与居民消费支出 Y 的格兰杰因果关系检验

滞后长度	格兰杰因果性	F 检验的 P 值	LM(1)检验的 P 值	AIC 值	结论
1	$X \xrightarrow{\times} Y$	0.001	0.343	14.36	拒绝
	$Y \xrightarrow{\times} X$	0.019	0.001	16.87	拒绝

续表

滞后长度	格兰杰因果性	F 检验的 P 值	LM(1)检验的 P 值	AIC 值	结论
2	$X \xrightarrow{\times} Y$	0.012	0.734	14.46	拒绝
	$Y \xrightarrow{\times} X$	0.336	0.283	16.51	不拒绝
3	$X \xrightarrow{\times} Y$	0.025	0.143	14.54	拒绝
	$Y \xrightarrow{\times} X$	0.373	0.128	16.67	不拒绝
4	$X \xrightarrow{\times} Y$	0.029	0.841	14.56	拒绝
	$Y \xrightarrow{\times} X$	0.467	0.013	16.78	不拒绝

注：表中“ $\xrightarrow{\times}$ ”表示“箭头前的变量不是箭头后变量的格兰杰原因”。

*§ 5.3 模型设定偏误问题

到目前为止，经典计量经济学模型的回归分析，都集中在对模型的估计和对经典假设的相关检验方面，而较少关注模型的具体设定形式。如果模型通过了所有相关检验，就认为得到了一个“满意”的模型估计结果，从而可以进一步用于经济分析及预测。然而，如果我们设定了一个“错误的”或者说是“有偏误的”模型，即使所有的经典假设都满足，得到的估计结果也会与“实际”有偏误，这种偏误称为模型设定偏误。

一、模型设定偏误的类型

模型设定偏误主要有两大类：一类是关于解释变量选取的偏误，主要包括漏选相关变量和多选无关变量；另一类是关于模型函数形式选取的偏误。

1. 相关变量的遗漏(omitting relevant variable)

在建立模型时，由于人们认识上的偏差、理论分析的缺陷，或者是有关统计数据的限制，可能有意或无意地忽略了某些重要变量。例如，如果“正确”的模型为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu \tag{5.3.1}$$

而我们将模型设定为

$$Y = \alpha_0 + \alpha_1 X_1 + v \tag{5.3.2}$$

也就是说，设定模型时漏掉了一个相关的解释变量。这类错误称为**遗漏相关变量**。

由于“正确”模型可能包含有被解释变量 Y 与解释变量 X 的滞后项，即为自回归分布滞后模型，因此，遗漏相关变量可能表现为对 Y 或 X 滞后项的遗漏。这类模型设定偏误也称为**动态设定偏误**(dynamic mis-specification)。

2. 无关变量的误选(including irrelevant variable)

无关变量的误选是指在设定模型时，包括了无关解释变量。例如，如果

(5.3.1)式仍为“真”，但将模型设定为

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + v \tag{5.3.3}$$

也就是说，设定模型时，多选了一个无关解释变量。

3．错误的函数形式(wrong functional form)

错误的函数形式是指在设定模型时，选取了不正确的函数形式。最常见的就是当“真实”的函数形式为非线性时，却选取了线性的函数形式。例如，如果“真实”的回归函数为

$$Y = AX_1^{\beta_1} X_2^{\beta_2} \mathrm{e}^{\mu} \tag{5.3.4}$$

但却将模型设定为

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v \tag{5.3.5}$$

二、模型设定偏误的后果

当模型设定出现偏误时，模型估计结果也会与“实际”有偏差。这种偏差的性质和程度与模型设定偏误的类型密切相关。

1．遗漏相关变量偏误

采用遗漏相关变量的模型进行估计而带来的偏误称为**遗漏相关变量偏误**(omitting relevant variable bias)。设正确的模型为(5.3.1)式，而我们却对(5.3.2)式进行回归，X_1的参数估计为

$$\hat{\alpha}_1 = \frac{\sum x_{i1} y_i}{\sum x_{i1}^2} \tag{5.3.6}$$

将正确模型(5.3.1)式的离差形式

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \mu_i - \bar{\mu}$$

代入(5.3.6)式得

$$\begin{aligned} \hat{\alpha}_1 &= \frac{\sum x_{i1} y_i}{\sum x_{i1}^2} = \frac{\sum x_{i1}(\beta_1 x_{i1} + \beta_2 x_{i2} + \mu_i - \bar{\mu})}{\sum x_{i1}^2} \\ &= \beta_1 + \beta_2 \frac{\sum x_{i1} x_{i2}}{\sum x_{i1}^2} + \frac{\sum x_{i1}(\mu_i - \bar{\mu})}{\sum x_{i1}^2} \end{aligned} \tag{5.3.7}$$

(1) 如果漏掉的X_2与X_1相关，则(5.3.7)式中的第二项在小样本下求期望与大样本下求概率极限都不会为零，从而使得普通最小二乘估计量在小样本下是有偏的，在大样本下也是非一致的。

事实上，在正确模型为(5.3.1)式的情况下对(5.3.2)式进行回归，则(5.3.2)式的随机干扰项就包括了X_2，即$v = \beta_2 X_2 + \mu$，从而X_1与v是同期相关的。因此，如果$\beta_2 > 0$，且X_2与X_1正相关，则X_1与v正相关，导致X_1的参数被高估，而常数项被低估。

(2) 如果X_2与X_1不相关，则由(5.3.7)式易知α_1的估计满足无偏性与一致性，但这时α_0的估计却是有偏的。

(3) 随机干扰项的方差估计$\hat{\sigma}^2$也是有偏的。在同样的样本下，(5.3.2)式给出的样本残差与(5.3.1)式给出的样本残差也不相同，因此，由两组样本残差估计的随机干扰项的方差也会不同。如果(5.3.1)式是正确的估计，(5.3.2)式的估计则是有偏误的。

(4) $\hat{\alpha}_1$的方差是真估计量$\hat{\beta}_1$的方差的有偏估计。由(5.3.2)式与(5.3.1)式估计的X_1的参数的方差分别为

$$\mathrm{Var}(\hat{\alpha}_1)=\frac{\sigma^2}{\sum x_{i1}^2} \tag{5.3.8}$$

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1)&=\sigma^2\frac{\sum x_{i2}^2}{\sum x_{i1}^2\sum x_{i2}^2-(\sum x_{i1}x_{i2})^2}\\&=\frac{\sigma^2}{\sum x_{i1}^2(1-r_{x_1x_2}^2)}\end{aligned} \tag{5.3.9}$$

其中，$r_{x_1x_2}^2$为X_1与X_2的相关系数的平方。如果X_2与X_1相关，显然有$\mathrm{Var}(\hat{\alpha}_1)\neq\mathrm{Var}(\hat{\beta}_1)$，即使$X_2$与$X_1$不相关，由于由(5.3.2)式与(5.3.1)式估计的随机干扰项的方差不同，估计的X_1的参数的方差也会不同。

2. 包含无关变量偏误

采用包含无关解释变量的模型进行估计带来的偏误，称为**包含无关变量偏误**(including irrelevant variable bias)。

设正确的模型为(5.3.2)式，而我们却对(5.3.1)式进行估计。对于(5.3.1)式，如果$\beta_2=0$，则与(5.3.2)式相同，因此，可将(5.3.1)式视为以$\beta_2=0$为约束的正确模型(5.3.2)式的特殊形式。由于所有的经典假设都满足，因此对(5.3.1)式进行普通最小二乘估计，可得到无偏且一致的估计量。由于$\beta_2=0$，因此，$\mathrm{E}(\hat{\beta}_2)=0$。

尽管在包含无关变量的情况下，普通最小二乘估计量是无偏的，但却不具有最小方差性。事实上，对X_1前的参数的方差而言，正确模型(5.3.2)式与错误模型(5.3.1)式估计的方差分别由(5.3.8)式与(5.3.9)式给出。显然，当X_1与X_2完全线性无关时，两模型参数估计的方差相同，否则，包含无关变量的模型参数的方差大于正确模型参数估计的方差，即$\mathrm{Var}(\hat{\beta}_1)>\mathrm{Var}(\hat{\alpha}_1)$。

由此可见，在多选无关解释变量的情形下，普通最小二乘估计量仍是无偏且一致的，随机干扰项的方差σ^2也能被正确估计，但普通最小二乘估计量却往往是无效的。也就是说，包含无关变量的偏误主要表现为“错误”模型的普通最小二乘估计量的方差一般会大于“正确”模型相应参数估计量的方差。

PDG

3．错误函数形式的偏误

当选取了错误函数形式并对其进行估计时，带来的偏误称**错误函数形式偏误**(wrong functional form bias)。容易判断，这种偏误是全方位的。例如，如果“真实”的回归函数为(5.3.4)式给出的幂函数的形式，而在模型估计时设定的模型却为(5.3.5)式所示的线性形式。显然，模型(5.3.4)式中的参数 β_1 为弹性，而按(5.3.5)式估计出的 $\hat{\beta}_1$ 却是对一个单位 X_1 变化带来的 Y 相应变化的测量。两者具有完全不同的经济含义，估计结果一般也是不相同的。

三、模型设定偏误的检验

一旦模型设定有偏误，普通最小二乘估计可能带来不良后果。因此，对模型的设定偏误进行检验就显得非常重要。

1．检验是否含有无关变量

对于无关变量的误选检验比较简单，可用统计检验中的 t 检验与 F 检验完成。检验的基本思想是，如果模型中误选了无关变量，则其系数的真值应为零。因此，只需对无关变量系数的显著性进行检验即可。例如，对所选定的一个 k 元回归模型

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

如果我们怀疑其中第 j 个变量是与 Y 无关的变量，只需用通常的 t 检验去检验 β_j 的显著性即可。而要检验 X_2 与 X_3 是否同时应包括在模型中来，只需检验联合假设 H_0： $\beta_2 = \beta_3 = 0$ 即可，第三章已介绍了适用的 F 检验。

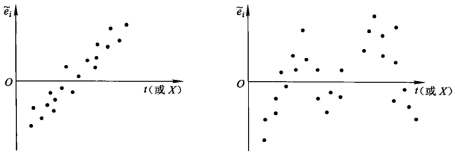
2．检验是否有相关变量的遗漏或函数形式设定偏误

在上面所列出的三种模型设定偏误中，遗漏相关变量与设定错误的函数形式的后果比多选不相关变量的情形要严重得多。不仅估计量有偏且不一致，而且随机干扰项的方差也往往被高估，从而使通常的推断程序变得无效，甚至参数的经济意义也可能不合理。而在多选不相关变量的情形下，后果仅是效率的损失。下面，我们着重介绍遗漏相关变量与设定错误的函数形式这两种模型设定偏误的检验。

(1) 残差图示法

对所设定的模型进行普通最小二乘回归，得到估计的残差序列 $\tilde{e}_t$ ，做出 $\tilde{e}_t$ 与时间 t 或某解释变量 X 的散点图，从图形考察估计的残差序列 $\tilde{e}_t$ 是否有规律地变动，来判断模型设定时是否遗漏了重要的解释变量或函数形式选取有偏误。

图 5.3.1 给出了残差序列随时间持续上升与呈现循环变化的两类图形。前者预示着模型设定时可能遗漏了随着时间的推移而持续上升的变量；后者则表明模型设定时可能遗漏了随着时间的推移而呈现循环变化的变量。



(a) 趋势变化　　(b) 循环变化

图 5.3.1　残差序列变化图

该方法曾在考察模型随机干扰项的异方差性与序列相关性时采用过。在那里我们已强调过，许多情形下，异方差或序列相关性往往是由于模型设定时漏掉了重要的解释变量而引起的，因此，当残差序列出现某种有规律的变化时，首先应考察模型中是否遗漏了某重要的解释变量。当我们肯定模型的设定是正确时，异方差或序列相关才是“真正”的异方差或序列相关。

当模型函数形式出现偏误时，残差序列也往往表现出某种有规律的变化特征。图 5.3.2 给出了一元回归模型中，真实模型呈幂函数形式，但却选取了线性函数进行回归的情形。在这种情形下，容易知道残差序列呈现先正、后负、再正的变化特征。

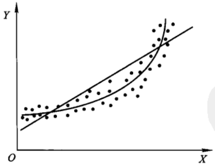


图 5.3.2　模型函数形式设定偏误时残差序列呈现正负交替变化

(2) 一般性设定偏误检验

残差图示法能够帮助我们初步判定在模型设定时是否遗漏了重要的解释变量，或者是否设定了有偏误的函数形式。但更准确更常用的判定方法是拉姆齐

(Ramsey)于 1969 年提出的所谓 **RESET 检验**(regression error specification test)。

我们仍假设正确模型为(5.3.1)式，却对(5.3.2)式进行估计。如果我们明确知道遗漏了一个相关变量 X_2，则问题变得相对简单。只需再估计(5.3.1)式，并检验变量 X_2 前的参数是否显著不为零即可。如果是显著的，就能判定(5.3.2)式的模型设定有误。

但问题是我们事先并不知道哪个变量被漏掉了，即无法确定 X_2 是什么。当然，如果能有一个替代变量 Z 来替代 X_2，我们就能进行上述检验。在拉姆齐的 RESET 检验中，采用(5.3.2)式中被解释变量 Y 的估计值 $\hat{Y}$ 的若干次幂来充当该“替代”变量。即先估计(5.3.2)式，得

$$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X_1$$

再用通过残差项 $\tilde{e}_t$ 与估计的 $\hat{Y}$ 的图形判断引入 $\hat{Y}$ 的若干次幂充当“替代”变量，进行普通最小二乘估计。如 $\tilde{e}_t$ 与 $\hat{Y}$ 的图形呈现曲线形变化时，回归模型可选为

$$Y = \beta_0 + \beta_1 X_1 + \gamma_1 \hat{Y}^2 + \gamma_2 \hat{Y}^3 + \mu \qquad (5.3.10)$$

再根据§3.5 介绍的增加解释变量的 F 检验来判断是否增加这些“替代”变量。当然，若仅增加一个“替代”变量，也可通过 t 检验来判断。

如果检验结果表明一个或若干个“替代”变量能够引入到模型中去，则说明模型设定时遗漏了相关变量。由于在进行(5.3.10)式的回归时，可引入若干个“替代”变量来判明是否有多于一个的变量被漏掉，因此，该方法被称为**一般性设定偏误检验**(test for general mis-specification)。

RESET 检验也可用来检验函数形式设定偏误的问题。例如，在一元回归模型中，假设真实的函数形式是非线性的，可用泰勒定理将其近似地表示为如下多项式的形式：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_1^3 + \cdots + \mu \qquad (5.3.11)$$

因此，如果我们用线性函数设定了模型，即意味着遗漏了相关变量 X_1^2, X_1^3，等等。所以，在一元回归中可以通过检验(5.3.11)式中 X_1 的各高次幂参数的显著性来判断是否将非线性模型误设成了线性模型。

对多元回归模型，非线性函数可能是关于若干个或全部解释变量的非线性，这时上述一元回归检验的程序已不适用，因为，模型中包含太多解释变量的高次幂及交叉项，容易导致自由度的损失以及出现多重共线性。这时仍可以按上面介绍的遗漏变量的程序进行检验。例如，我们对(5.3.1)式的二元线性模型进行估计，但却怀疑真实的函数形式是非线性的。这时，只需以(5.3.1)式估计出的 $\hat{Y}$ 的若干次幂为“替代”变量，进行类似于如下模型的估计：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma_1 \hat{Y}^2 + \gamma_2 \hat{Y}^3 + \mu \qquad (5.3.12)$$

再判断各“替代”变量的参数是否显著地不为零即可。这里采用 $\hat{Y}$ 的高次幂，

既避免了自由度的损失与多重共线性的问题，同时它已包含了解释变量的高次幂及交叉项提供的信息。

例 5.3.1

在§4.2 中国总量消费函数的例子中，估计了中国居民消费总量 Y 与总收入 X 的关系，但发现模型随机干扰项具有强烈的一阶自相关性。由于采用的是时间序列，Y 与 X 有着共同的随时间一致变动的趋势。由于建立的简单模型(4.2.23)式仅仅考虑了 X 对 Y 的影响，而未将时间趋势项独立引入模型，因此模型存在的强烈一阶自相关性可能就是遗漏了重要的相关变量造成的。换言之，仅仅刻画 Y 与 X 间因果关系的模型(4.2.23)式存在着模型设定偏误。下面对(4.2.23)式进行 RESET 检验。

首先，用原回归模型(4.2.23)式

$$\hat{Y}=2\ 091.3+0.437\ 5X$$

估计出消费总量序列 $\hat{Y}_t$。

其次，在原回归模型中加入新的解释变量 $\hat{Y}_t^2$ 后重新进行估计，得

$$\tilde{Y}_t=421.26+0.585\ 4X-0.000\ 008\ 6\hat{Y}_t^2$$

$$(1.24)\qquad(24.36)\qquad(-6.35)$$

$$R^2=0.995\ 3$$

原回归模型的可决系数为 $R^2=0.988\ 0$，由(3.6.17)式计算 F 统计量：

$$F=\frac{(0.995\ 3-0.988\ 0)/1}{(1-0.995\ 3)/(29-3)}=40.3$$

该值大于 5%显著性水平下、自由度为(1，26)的 F 分布的临界值 4.22，因此拒绝原模型与引入新变量的模型可决系数无显著差异的假设，表明原模型确实存在遗漏相关变量的设定偏误。

为了将 Y 与 X 随时间共同变化的时间趋势因素分离出来，模型引入了时间趋势项 T^2，并通过普通最小二乘法估计了二元回归模型得到(4.2.24)式：

$$\hat{Y}=3\ 328.1+0.176\ 2X+21.656T^2 \tag{4.2.24}$$

$$R^2=0.997\ 590$$

那么该模型是否就是一正确设定的模型呢？仍进行 RESET 检验。由该模型计算的消费总量序列记为 $\hat{\hat{Y}}_t$，并将它的平方项作为解释变量加入该二元模型中进行普通最小二乘估计得

$$\tilde{\tilde{Y}}_t=3\ 606.9+0.142\ 6X+22.70T^2-0.000\ 001\ 343\hat{\hat{Y}}_t^2$$

$$(7.34)\ (2.37)\quad (8.31)\qquad (0.62)$$

$$R^2=0.997\ 626$$

原二元回归模型(4.2.24)式的可决系数为 $R^2=0.997\ 590$，由(3.6.17)式计算 F 统计量：

$$F=\frac{(0.997\ 626-0.997\ 590)/1}{(1-0.997\ 626)/(29-4)}=0.38$$

该值小于 5%显著性水平下、自由度为(1，25)的 F 分布的临界值 4.24，表明加入时间趋势项 T^2 的总量消费模型已不存在设定偏误问题。同样地，通过再引入 $\hat{\hat{Y}}_t$ 的立方项后仍可验证原二元模型(4.2.24)式不存在设定偏误问题。

---

(3) 线性模型与双对数线性模型的选择

在设定模型时，一个较为棘手的问题是选取线性模型还是双对数线性模型。对于一元回归可通过变量的变化图形帮助确定。如例 2.6.2 中给出的居民总量可支配收入 X 与总量消费支出 Y，其图形是近似于指数函数的曲线(图 2.6.1)，因此对它们取对数后可近似地转化为直线，因此双对数线性模型可能更适合进行线性回归。

而对于多元回归，不同变量图形变化走势可能不同，比较难于判断。这时也无法直接通过判定系数的大小来辅助决策，因为在两类模型中被解释变量是不同的。为了在两类模型中进行比较，可用下面介绍的博克斯–考科斯(Box-Cox)变换进行。

第一步，计算被解释变量 Y 的样本几何均值：

$$\tilde{Y}=(Y_1Y_2\cdots Y_n)^{\frac{1}{n}}=\exp\left(\frac{1}{n}\sum\ln\ Y_i\right)$$

第二步，用得到的样本几何均值去除原被解释变量 Y，得到被解释变量的新序列 Y^*：

$$Y_i^*=\frac{Y_i}{\tilde{Y}}$$

第三步，用被解释变量的新序列 Y^* 替代原序列 Y，分别估计双对数线性模型与线性模型，这时得到的两个回归模型的残差平方和是可比的，从而可通过比较它们是否有显著差异来进行判断。扎瑞姆布卡(Zarembka，1968)提出的检验统计量为

$$\frac{1}{2}n\ln\frac{\mathrm{RSS}_2}{\mathrm{RSS}_1}$$

其中，RSS_2 与 RSS_1 分别为两个回归模型中对应的较大的残差平方和与较小的残差平方和，n 为样本容量。可以证明，该统计量在两个回归的残差平方

和无差异的假设下服从自由度为 1 的 χ^2 分布。如果该统计量的计算值大于给定显著性水平下的相应临界值，则拒绝两个回归的残差平方和无差异的原假设，从而应选择具有较小残差平方和的模型。

例 5.3.2

在§3.2 中国城镇居民人均消费函数的例子中，采用线性模型估计了 2006 年中国城镇居民人均消费支出 Y 与人均可支配收入 X_1 以及前一年人均消费支出 X_2 的关系，并得到可决系数 $R^2=0.975\,634$。如果采用双对数线性模型，普通最小二乘估计结果如下：

$$\ln\hat{Y}=-0.057+0.804\,2\ln X_1+0.176\,8\ln X_2$$

$$(-0.16)\qquad(7.45)\qquad(1.61)$$

$$R^2=0.961\,032\qquad F=345.2\qquad \text{D.W.}=1.871$$

可见，双对数线性模型的拟合优度比线性模型小一些。但我们不能就此简单地判断双对数线性模型劣于线性模型。下面进行 Box-Cox 变换。

容易计算原城镇居民人均消费支出样本的几何平均值为

$$\tilde{Y}=\exp(\frac{1}{n}\sum\ln(Y_i))=8\,139.44$$

以 $\tilde{Y}$ 去除 Y_i 得新的城镇居民人均消费支出序列：$Y_i^*=Y_i/\tilde{Y}$。以 Y_i^* 替代 Y_i，分别进行双对数线性模型与线性模型的回归，得

$$\ln\hat{Y}^*=-9.061+0.804\,3\ln X_1+0.176\,8\ln X_2,\qquad \text{RSS}_1=0.062\,944,$$

$$\hat{Y}^*=0.017\,6+0.000\,068X_1+0.000\,031X_2,\qquad \text{RSS}_2=0.069\,079$$

于是

$$\frac{1}{2}n\ln\left(\frac{\text{RSS}_2}{\text{RSS}_1}\right)=\frac{31}{2}\times\ln(1.097\,4)=1.442$$

该值小于 5%显著性水平下、自由度为 1 的 χ^2 分布的临界值 3.841，因此可判定双对数线性模与线性模型在拟合优度上无差异。

例 2.6.2 中国居民总量消费函数的建立中采用了如下非对数函数式

$$\hat{Y}=2\,091.29+0.437\,5X,\qquad R^2=0.987\,9$$

如果采用双对数线性模型可得如下估计结果：

$$\ln\hat{Y}=0.587\,3+0.880\,0\ln X,\qquad R^2=0.993\,0$$

可见双对数线性模型的 R^2 高于原线性模型。可通过 Box-Cox 变换进一步验证双对数线性模型确实“优”于原线性模型。

本章练习题

1. 回归模型中引入虚拟变量的作用是什么？有哪几种基本的引入方式？它们各适用于什么情况？

2. 在一项对北京某大学学生月消费支出的研究中，认为学生的消费支出除受其家庭的每月收入水平外，还受在学校中是否得到奖学金，来自农村还是城市，是经济发达地区还是欠发达地区，以及性别等因素的影响。试设定适当的模型，并导出如下情形下学生消费支出的平均水平：

(1) 来自欠发达农村地区的女生，未得到奖学金；

(2) 来自欠发达城市地区的男生，得到奖学金；

(3) 来自发达地区的农村女生，得到奖学金；

(4) 来自发达地区的城市男生，未得到奖学金。

3. 滞后变量模型有哪几种类型？分布滞后模型使用普通最小二乘方法存在哪些问题？

4. 产生模型设定偏误的主要原因是什么？模型设定偏误的后果以及检验方法有哪些？

5. 1970—1991 年美国制造业固定厂房设备投资 Y 和销售量 X 的相关数据如下表所示。

单位：10 亿美元

年份	厂房开支 Y	销售量 X	年份	厂房开支 Y	销售量 X
1970	36.99	52.805	1981	128.68	168.129
1971	33.60	55.906	1982	123.97	163.351
1972	35.42	63.027	1983	117.35	172.547
1973	42.35	72.931	1984	139.61	190.682
1974	52.48	84.790	1985	152.88	194.538
1975	53.66	86.589	1986	137.95	194.657
1976	68.53	98.797	1987	141.06	206.326
1977	67.48	113.201	1988	163.45	223.547
1978	78.13	126.905	1989	183.80	232.724
1979	95.13	143.936	1990	192.61	239.459
1980	112.60	154.391	1991	182.81	235.142

(1) 以 Y_t^* 代表理想的或长期的新建厂房设备企业开支，估计如下模型：

$$Y_t^* = \beta_0 + \beta_1 X_t + \mu_t$$

(2) 如果模型设定为 $Y_t^* = \beta_0 X_t^{\beta_2} e^{\mu_t}$，请用存量调整模型进行估计。同(1)中的结果相比，你会选择哪个模型？

(3) 以 X_t^* 代表理想的销售量，请估计如下模型：

$$Y_t = \beta_0 + \beta_1 X_t^* + \mu_t$$

与(1)中的模型相比，你认为哪个模型更适当一些？

6. 利用习题 5 所给的数据，试回答：

(1) 假定销售量对厂房设备支出有一个分布滞后效应，试用 4 期滞后和 2 次多项式去估计此分布滞后模型；

(2) 检验销量与厂房设备支出的格兰杰因果关系，使用直至 6 期为止的滞后并评述你的结果。

7. 如果真实的模型是 $Y_i = \beta_1 X_i + \mu_i$，但你却拟合了一个带截距项的模型 $Y_i = \alpha_0 + \alpha_1 X_i + v_i$，试评述这一设定误差的后果。

8. 在习题 7 中，假设真实的模型是带截距项的模型，而你却对过原点的模型进行了普通最小二乘回归。请评述这一模型误设的后果。

9. 在例 5.3.1 中，通过引入 $\hat{\hat{Y}}_t$ 的平方项后检验了二元模型（4.2.24）式不存在设定偏误，请进一步引入 $\hat{\hat{Y}}_t$ 的立方项检验二元模型（4.2.24）式是否仍不存在设定偏误问题。

10. 在第三章习题 13 中，假设有人不同意原幂函数模型是正确设定的模型，而下面的线性形式是正确设定的模型：

$$Y_i = \beta_0 + \beta_1 K_i + \beta_2 L_i + \mu_i$$

你将如何检验哪一个模型设定更正确？

11. 请通过 Box-Cox 变换检验例 2.6.2 中国居民总量消费函数的建立中，原线性模型与双对数线性模型哪一个更“优”。

第六章　联立方程计量经济学模型：理论与方法

6

联立方程计量经济学模型是相对于单方程计量经济学模型而言的。它以经济系统为研究对象，以揭示经济系统中各部分、各因素之间的数量关系和系统的数量特征为目标，用于经济系统的预测、分析和评价，是计量经济学模型的重要组成部分。它的理论与方法也是本书的重点内容之一。

§6.1　联立方程计量经济学模型的提出

联立方程计量经济学模型问题是从两方面提出来的。从研究对象的角度，为了满足实际研究对象的需要而建立联立方程计量经济学模型；从计量经济学理论方法的角度，为了估计联立方程计量经济学模型的需要而发展新的理论与方法。

一、经济研究中的联立方程计量经济学问题

单方程计量经济学模型，是用单一方程描述某一经济变量与影响该变量变化的诸因素之间的数量关系。所以，它适用于单一经济现象的研究，揭示其中的单向因果关系。但是，经济现象是极为复杂的，其中诸因素之间的关系，在很多情况下，不是单一方程描述的那种简单的单向因果关系，而是相互依存、互为因果的，这时，就必须用一组方程才能描述清楚。我们称这些经济现象为经济系统。

经济系统并没有严格的空间概念。国民经济是一个系统，一个地区的经济也是一个系统，甚至某一项经济活动也是一个系统。例如，我们进行商品购买决策，由于存在收入或预算的制约，在决定是否购买某种商品时，必须考虑到对其他商品的需求与其他商品的价格，这样，不同商品的需求量之间是互相影响、互为因果的。商品购买决策就是一个经济系统。

以一个由国内生产总值 (Y)、居民消费总额 (C)、投资总额 (I) 和政府消费额 (G) 等变量构成的简单的宏观经济系统为例。如果将政府消费额由系统外部给定，并对系统内部其他变量产生影响，就国内生产总值、居民消费总额、投资总额来讲，是互相影响并互为因果的。居民消费和投资当然取决于国内生

产总值，但反过来又影响国内生产总值。所以就无法用一个方程描述它们之间的关系，而需要建立一个由多个方程组成的方程系统。例如，可以建立如下的模型：

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{t1}, \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{t2}, \\ Y_t = C_t + I_t + G_t, \end{cases} \tag{6.1.1}$$

其中，第一个方程表示居民消费总额由国内生产总值决定；第二个方程表示投资总额由国内生产总值和前一年的国内生产总值共同决定；第三个方程表示国内生产总值由居民消费总额、投资总额和政府消费额共同决定，在假定进出口平衡的情况下，是一个恒等方程。这就是一个简单的描述宏观经济的联立方程计量经济学模型。实际经济研究中的联立方程计量经济学问题就是这样提出来的。

二、计量经济学方法中的联立方程问题

上面已经指出，经济系统中诸因素之间的关系必须用联立的计量经济学方程才能较好地描述。但是，如果联立方程计量经济学模型中的每个方程都可以看作一个单方程计量经济学模型，都可以用单方程计量经济学模型中讲述的方法加以研究，那么，从计量经济学理论与方法的角度，联立方程计量经济学模型的提出就没有什么新内容了。事实上，问题是存在的。主要表现于以下三个方面。

1. 随机解释变量问题

以上述简单的宏观经济模型为例。如果我们用单方程计量经济学模型方法估计其中的居民消费方程，从方程系统中可以判断，作为居民消费总额的解释变量的国内生产总值是一个随机变量。因为根据第三个方程，它是由居民消费与投资决定，而根据第一个和第二个方程，居民消费和投资都是随机变量，所以国内生产总值不是确定性变量。更进一步分析，居民消费 C 与随机干扰项 μ_1 相关，所以国内生产总值 Y 也与 μ_1 相关。如果直接利用普通最小二乘法估计居民消费方程，将得到关于 α_0 和 α_1 的有偏估计量，这一点在第四章中已经证明了。对于投资方程也是如此。

2. 损失变量信息问题

在一个经济系统中，变量之间或多或少地存在着某种关联。在估计联立方程系统中某一个随机方程参数时，必须考虑没有包含在该方程中的变量的数据信息。例如，居民消费方程中仅包含国内生产总值 Y，没有包含政府消费 G 和前期国内生产总值 Y_{t-1}，但是它们分别通过第二个和第三个方程对居民消费 C

产生影响。所以在估计居民消费方程的参数时，必须充分考虑政府消费G和前期国内生产总值Y_{t-1}的数据信息。而采用第二章的单方程计量经济学模型方法是无法实现这一点的。

3．损失方程之间的相关性信息问题

联立方程计量经济学模型系统中每个随机方程之间往往存在某种相关性，表现于不同方程随机干扰项之间，尤其在以时间序列数据作样本时，不同方程随机干扰项之间往往存在同期相关性，即在同一个样本点上，它们经常是相关的。原因为何？以上述宏观经济模型为例。如果经济景气对消费与投资具有虽不很显著但确实存在的影响，那么这种影响被分别包含在μ_1和μ_2中，导致在同一个样本点上μ_{t1}和μ_{t2}是相关的。这种相关性的存在意味着第二个方程包含有对第一个方程估计的有用信息；同样，第一个方程包含有对第二个方程估计的有用信息。如果单独把其中的一个方程抽取出来，用单方程模型方法进行估计，就割裂了两个方程之间的联系，没有利用全部的信息，造成估计的低效率。

通过以上分析，得到一个结论：必须发展新的估计方法估计联立方程计量经济学模型，以尽可能避免出现这些问题，这就从计量经济学理论方法上提出了联立方程问题。

§6.2　联立方程计量经济学模型的若干基本概念

在联立方程计量经济学模型中，有一些在单方程计量经济学模型中没有出现的概念，即使已经出现的概念，其内涵也发生了变化，所以搞清楚基本概念，是十分重要的。

一、变量

在联立方程计量经济学模型中，对于其中每个随机方程，其变量仍然有被解释变量与解释变量之分。但是对于模型系统而言，已经不能用被解释变量与解释变量来划分变量，正如上节所说的，同一个变量，在这个方程中作为被解释变量，在另一个方程中则可能作为解释变量。**对于联立方程计量经济学模型系统而言，将变量分为内生变量和外生变量两大类，外生变量与滞后内生变量又被统称为先决变量。**

1．内生变量

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Y满足

$$\mathrm{Cov}(Y_i, \mu_i) \neq 0$$

即

$$\mathrm{E}(Y_i\mu_i) \neq 0$$

因为

$$\begin{aligned}\mathrm{Cov}(Y_i,\mu_i) &= \mathrm{E}\{[Y_i - \mathrm{E}(Y_i)][\mu_i - \mathrm{E}(\mu_i)]\} \\ &= \mathrm{E}\{[Y_i - \mathrm{E}(Y_i)]\mu_i\} \\ &= \mathrm{E}(Y_i\mu_i) - \mathrm{E}(Y_i)\mathrm{E}(\mu_i) \\ &= \mathrm{E}(Y_i\mu_i)\end{aligned}$$

在 § 6.1 的宏观经济模型(6.1.1)式中，国内生产总值 (Y) 、居民消费总额 (C) 、投资总额 (I) 为内生变量。

2. 外生变量

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。

外生变量 X 一般满足

$$\mathrm{E}(X_i\mu_i) = 0$$

在 § 6.1 的宏观经济模型(6.1.1)式中，政府消费 G 是外生变量。

3. 先决变量(predetermined variable)

外生变量与滞后内生变量(lagged endogenous variable)统称为先决变量。

滞后内生变量是联立方程计量经济学模型中重要的不可缺少的一部分变量，用以反映经济系统的动态性与连续性。如果模型满足

$$\mathrm{E}(\mu_i\mu_{i-s}) = 0\,, \qquad s\neq 0$$

那么

$$\mathrm{E}(Y_{i-s}\mu_i) = 0\,, \qquad s\neq 0$$

在 § 6.1 的宏观经济模型中，前期国内生产总值 Y_{t-1} 为滞后内生变量，它与政府消费 G 一起构成先决变量。

在单方程计量经济学模型中，内生变量作为被解释变量，外生变量与滞后内生变量作为解释变量。而在联立方程计量经济学模型中，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。

二、结构式模型(structural model)

根据经济理论和行为规律建立的描述经济变量之间直接关系结构的计量经济学方程系统称为结构式模型。 § 6.1 的简单宏观经济模型：

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{t1}, \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{t2}, \qquad t=1,2,\cdots,n \\ Y_t = C_t + I_t + G_t, \end{cases} \tag{6.2.1}$$

就是一个结构式模型。

结构式模型中的每个方程都是**结构方程**(structural equation)，各个结构方程的参数称为**结构参数**(structural parameters or coefficient)。在结构方程中，解释变量中可以出现内生变量。**将一个内生变量表示为其他内生变量、先决变量和随机干扰项的函数形式，称为结构方程的正规形式。**结构方程的方程类型如下：

- 随机方程
  - 行为方程
  - 技术方程
  - 制度方程
  - 统计方程
- 恒等方程
  - 定义方程
  - 平衡方程
  - 经验方程

其中，行为方程描述经济系统中变量之间的行为关系，主要是因果关系，如用收入作为消费的解释变量建立的方程；技术方程描述由技术决定的变量之间的关系，如用总产值作为净产值的解释变量建立的方程；制度方程描述由制度决定的变量之间的关系，如用进口总额作为关税收入的解释变量建立的方程；统计方程描述由数据之间的相关性决定的变量之间的关系，如描述城镇居民收入与农村居民收入之间关系的方程。显然，在随机方程中，统计方程较多的结构式模型不是好的模型，应该尽可能地避免出现统计方程。定义方程与平衡方程之间是有区别的。前者是由经济学或经济统计学的定义决定的，如国内生产总值等于第一、二、三产业增加值之和；后者是由变量所代表的指标之间的平衡关系决定的，如政府消费等于消费总额减去居民消费。经验方程仅描述由经验得到的数据之间的确定性关系，没有什么实质性意义。所以在恒等方程中，经验方程较多的结构式模型不是好的模型，应该尽可能地避免出现经验方程。

习惯上用 $\boldsymbol{Y}$ 表示内生变量，$\boldsymbol{X}$ 表示先决变量，$\boldsymbol{\mu}$ 表示随机干扰项，$\boldsymbol{\beta}$ 表示内生变量的结构参数，$\boldsymbol{\gamma}$ 表示先决变量的结构参数，如果模型中有常数项，可以看作一个外生的虚变量 $\boldsymbol{X}_0$，它的观测值始终取 1，那么具有 g 个内生变量，k 个先决变量，g 个结构方程的模型称为完备的结构式模型。在完备的结构式模型中，独立的结构方程的数目等于内生变量的数目，每个内生变量都分别由一个方程来描述。一个完备的结构式模型可以写成

$$\boldsymbol{BY}+\boldsymbol{\Gamma X}=\boldsymbol{N} \tag{6.2.2}$$

或

$$(\boldsymbol{B\Gamma})\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{X}\end{pmatrix}=\boldsymbol{N} \tag{6.2.3}$$

其中

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\\ \vdots\\ \boldsymbol{Y}_g\end{pmatrix},\qquad \boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\\ \vdots\\ \boldsymbol{X}_k\end{pmatrix},\qquad \boldsymbol{N}=\begin{pmatrix}\boldsymbol{N}_1\\ \boldsymbol{N}_2\\ \vdots\\ \boldsymbol{N}_g\end{pmatrix}$$

用 n 表示样本容量，则

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\\ \vdots\\ \boldsymbol{Y}_g\end{pmatrix}=\begin{pmatrix}y_{11} & y_{12} & \cdots & y_{1n}\\ y_{21} & y_{22} & \cdots & y_{2n}\\ \vdots & \vdots & & \vdots\\ y_{g1} & y_{g2} & \cdots & y_{gn}\end{pmatrix}$$

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\\ \vdots\\ \boldsymbol{X}_k\end{pmatrix}=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1n}\\ x_{21} & x_{22} & \cdots & x_{2n}\\ \vdots & \vdots & & \vdots\\ x_{k1} & x_{k2} & \cdots & x_{kn}\end{pmatrix}$$

$$\boldsymbol{N}=\begin{pmatrix}\boldsymbol{N}_1\\ \boldsymbol{N}_2\\ \vdots\\ \boldsymbol{N}_g\end{pmatrix}=\begin{pmatrix}\mu_{11} & \mu_{12} & \cdots & \mu_{1n}\\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n}\\ \vdots & \vdots & & \vdots\\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gn}\end{pmatrix}$$

参数矩阵为

$$\boldsymbol{B}=\begin{pmatrix}\beta_{11} & \beta_{12} & \cdots & \beta_{1g}\\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g}\\ \vdots & \vdots & & \vdots\\ \beta_{g1} & \beta_{g2} & \cdots & \beta_{gg}\end{pmatrix}$$

$$\boldsymbol{\Gamma}=\begin{pmatrix}\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k}\\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k}\\ \vdots & \vdots & & \vdots\\ \gamma_{g1} & \gamma_{g2} & \cdots & \gamma_{gk}\end{pmatrix}$$

$\boldsymbol{B\Gamma}$ 为结构参数矩阵。

将(6.2.1)式表示的宏观经济模型写成矩阵方程(6.2.2)式的形式，其中各个矩阵为

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{C}_t\\ \boldsymbol{I}_t\\ \boldsymbol{Y}_t\end{pmatrix}=\begin{pmatrix}C_1 & C_2 & \cdots & C_n\\ I_1 & I_2 & \cdots & I_n\\ Y_1 & Y_2 & \cdots & Y_n\end{pmatrix}$$

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{1}\\ \boldsymbol{Y}_{t-1}\\ \boldsymbol{G}_t\end{pmatrix}=\begin{pmatrix}1 & 1 & \cdots & 1\\ Y_0 & Y_1 & \cdots & Y_{n-1}\\ G_1 & G_2 & \cdots & G_n\end{pmatrix}$$

$$\boldsymbol{N}=\begin{pmatrix}\boldsymbol{N}_1\\ \boldsymbol{N}_2\\ \boldsymbol{0}\end{pmatrix}=\begin{pmatrix}\mu_{11} & \mu_{12} & \cdots & \mu_{1n}\\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n}\\ 0 & 0 & \cdots & 0\end{pmatrix}$$

$$\boldsymbol{B\Gamma}=\begin{pmatrix}1 & 0 & -\alpha_1 & -\alpha_0 & 0 & 0\\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0\\ -1 & -1 & 1 & 0 & 0 & -1\end{pmatrix}$$

在一些教科书中，也采用(6.2.2)式、(6.2.3)式的转置形式表示结构式模型：

$$\boldsymbol{YB}+\boldsymbol{X\Gamma}=\boldsymbol{N} \tag{6.2.4}$$

$$(\boldsymbol{YX})\begin{pmatrix}\boldsymbol{B}\\ \boldsymbol{\Gamma}\end{pmatrix}=\boldsymbol{N} \tag{6.2.5}$$

注意，(6.2.4)式、(6.2.5)式中各个矩阵都是(6.2.2)式、(6.2.3)式中各个矩阵的转置，尽管在书写符号上未加区分。

三、简化式模型(reduced-form model)

将联立方程计量经济学模型的每个内生变量表示成所有先决变量和随机干扰项的函数，即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。显然，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，因此也不是我们研究的对象。但是，由于简化式模型中作为解释变量的变量中没有内生变量，可以采用普通最小二乘法估计每个方程的参数，所以它在联立方程计量经济学模型研究中具有重要的作用。简化式模型中每个方程称为**简化式方程**(reduced-form equation)，方程的参数称为**简化式参数**(reduced-form coefficient)。通常用 $\boldsymbol{\Pi}$ 表示简化式参数，于是简化式模型的矩阵形式为

$$\boldsymbol{Y}=\boldsymbol{\Pi X}+\boldsymbol{E} \tag{6.2.6}$$

其中

$$\boldsymbol{\Pi}=\begin{pmatrix}\pi_{11} & \pi_{12} & \cdots & \pi_{1k}\\ \pi_{21} & \pi_{22} & \cdots & \pi_{2k}\\ \vdots & \vdots & & \vdots\\ \pi_{g1} & \pi_{g2} & \cdots & \pi_{gk}\end{pmatrix}$$

$$\boldsymbol{E}=\begin{pmatrix}\boldsymbol{E}_1\\ \boldsymbol{E}_2\\ \vdots\\ \boldsymbol{E}_g\end{pmatrix}=\begin{pmatrix}\varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1n}\\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2n}\\ \vdots & \vdots & & \vdots\\ \varepsilon_{g1} & \varepsilon_{g2} & \cdots & \varepsilon_{gn}\end{pmatrix}$$

同样也可以用(6.2.6)式的转置形式表示简化式模型：

$$\boldsymbol{Y}=\boldsymbol{X\Pi}+\boldsymbol{E} \tag{6.2.7}$$

其中每个矩阵都是(6.2.6)式中同名矩阵的转置。

宏观经济模型(6.2.1)式的简化式模型为

$$\begin{cases}C_t=\pi_{10}+\pi_{11}Y_{t-1}+\pi_{12}G_t+\varepsilon_{t1},\\ I_t=\pi_{20}+\pi_{21}Y_{t-1}+\pi_{22}G_t+\varepsilon_{t2}\ , \qquad t=1,2,\cdots,n\\ Y_t=\pi_{30}+\pi_{31}Y_{t-1}+\pi_{32}G_t+\varepsilon_{t3},\end{cases} \tag{6.2.8}$$

四、参数关系体系

将(6.2.2)式作如下变换：

$$\boldsymbol{BY}=-\boldsymbol{\Gamma X}+\boldsymbol{N}$$

$$\boldsymbol{Y}=-\boldsymbol{B}^{-1}\boldsymbol{\Gamma X}+\boldsymbol{B}^{-1}\boldsymbol{N}$$

与(6.2.6)式比较，可以得到

$$\boldsymbol{\Pi}=-\boldsymbol{B}^{-1}\boldsymbol{\Gamma} \tag{6.2.9}$$

该式描述了简化式参数与结构式参数之间的关系，称为**参数关系体系**。

将(6.2.1)式结构式模型进行变量连续替代后得到

$$C_t=\frac{\alpha_0-\alpha_0\beta_1+\alpha_1\beta_2}{1-\alpha_1-\beta_1}+\frac{\alpha_1\beta_2}{1-\alpha_1-\beta_1}Y_{t-1}+\frac{\alpha_1}{1-\alpha_1-\beta_1}G_t+\frac{\mu_{t1}+\alpha_1\mu_{t2}-\beta_1\mu_{t1}}{1-\alpha_1-\beta_1}$$

$$I_t=\frac{\beta_0-\alpha_1\beta_0+\alpha_0\beta_1}{1-\alpha_1-\beta_1}+\frac{\beta_2-\alpha_1\beta_2}{1-\alpha_1-\beta_1}Y_{t-1}+\frac{\beta_1}{1-\alpha_1-\beta_1}G_t+\frac{\mu_{t2}-\alpha_1\mu_{t2}+\beta_1\mu_{t1}}{1-\alpha_1-\beta_1}$$

$$Y_t=\frac{\beta_0+\alpha_0}{1-\alpha_1-\beta_1}+\frac{\beta_2}{1-\alpha_1-\beta_1}Y_{t-1}+\frac{1}{1-\alpha_1-\beta_1}G_t+\frac{\mu_{t2}+\mu_{t1}}{1-\alpha_1-\beta_1}$$

与(6.2.8)式对照，得到简化式参数与结构式参数之间的关系体系为

$$\pi_{10}=\frac{\alpha_0-\alpha_0\beta_1+\alpha_1\beta_2}{1-\alpha_1-\beta_1}\,,\qquad \pi_{11}=\frac{\alpha_1\beta_2}{1-\alpha_1-\beta_1}\,,\qquad \pi_{12}=\frac{\alpha_1}{1-\alpha_1-\beta_1}$$

PDG

$$\pi_{20}=\frac{\beta_0-\alpha_1\beta_0+\alpha_0\beta_1}{1-\alpha_1-\beta_1}, \qquad \pi_{21}=\frac{\beta_2-\alpha_1\beta_2}{1-\alpha_1-\beta_1}, \qquad \pi_{22}=\frac{\beta_1}{1-\alpha_1-\beta_1}$$

$$\pi_{30}=\frac{\beta_0+\alpha_0}{1-\alpha_1-\beta_1}, \qquad \pi_{31}=\frac{\beta_2}{1-\alpha_1-\beta_1}, \qquad \pi_{32}=\frac{1}{1-\alpha_1-\beta_1}$$

利用参数关系体系，首先估计简化式参数，然后可以计算得到结构式参数。从参数关系体系还可以看出，简化式参数反映了先决变量对内生变量的直接与间接影响之和，这是简化式模型的另一个重要作用。例如，π_{21}表示Y_{t-1}对I_t的影响，即Y_{t-1}增加 1 个单位时对I_t的影响。根据

$$\pi_{21}=\frac{\beta_2-\alpha_1\beta_2}{1-\alpha_1-\beta_1}=\beta_2+\frac{\beta_1\beta_2}{1-\alpha_1-\beta_1}$$

这种影响被分成两部分，其中前一项β_2正是结构式方程中反映Y_{t-1}对I_t的直接影响的参数，后一项反映Y_{t-1}对I_t的间接影响，只有通过简化式模型才能得到。

§6.3 联立方程计量经济学模型的识别

联立方程计量经济学模型由多个方程组成，对方程之间的关系有严格的要求，否则模型就可能无法估计。所以在进行模型估计之前首先要判断它是否可以估计，这就是**模型的识别**。

一、识别的概念

我们先看一个例子。有如下 3 个方程构成的简单宏观经济模型：

$$\begin{cases} C_t=\alpha_0+\alpha_1 Y_t+\mu_{t1}, \\ I_t=\beta_0+\beta_1 Y_t+\mu_{t2}\ , \qquad t=1,2,\cdots,n \\ Y_t=C_t+I_t, \end{cases} \tag{6.3.1}$$

其中C为消费总额，包括居民消费和政府消费，在假定进出口平衡的情况下，国内生产总值为消费总额与投资总额之和。模型中消费总额与投资总额都用国内生产总值解释，在经济学上也是可以接受的。所以，如果该模型可以估计，不失为一个描述消费总额、投资总额和国内生产总值关系的总量宏观经济模型。

但是，分析发现，消费方程是包含C，Y和常数项的直接线性方程，而投资方程和国内生产总值方程的某种线性组合(消去I)所构成的新方程也是包含C,Y和常数项的直接线性方程。现在，问题出现了，当我们收集了C,Y的样本观测值并进行参数估计后，很难判断得到的是消费方程的参数估计量还是新组合方程的参数估计量。这时，我们只能认为原模型中的消费方程是不可估计的。这种情况被称为不可识别。

1. 识别的定义

在不同的教科书中，分别给出了识别的 3 种定义：

“如果联立方程计量经济学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。”

“如果联立方程计量经济学模型中某些方程的线性组合可以构成与某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计量经济学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。”

认真分析以上 3 种定义发现，应该以是否具有确定的统计形式作为识别的基本定义，即上述第 1 种；其他两种表述实际上是判断识别与否的方法。

什么是“统计形式”？即变量和方程关系式。什么是“具有确定的统计形式”？即模型系统中其他方程或所有方程的任意线性组合所构成的新的方程都不再具有这种统计形式。模型(6.3.1)式中的消费方程已经被证明不具有确定的统计形式，因为其他两个方程的线性组合形成的新方程与它的统计形式完全相同。如果某个结构方程不具有确定的统计形式，那么根据参数关系体系，在已知简化式模型参数估计值时，就不能得到该结构方程的确定的结构参数估计值。

2. 模型的识别

上述识别的定义是针对结构方程而言的。模型中每个需要估计其参数的随机方程都存在识别问题。**如果一个模型中的所有随机方程都是可以识别的，则认为该联立方程计量经济学模型系统是可以识别的**。反过来，如果一个模型系统中存在一个不可识别的随机方程，则认为该联立方程计量经济学模型系统是不可识别的。恒等方程由于不存在参数估计问题，所以也不存在识别问题。但是，必须注意，在判断随机方程的识别性问题时，应该将恒等方程考虑在内。例如，模型(6.3.1)式中正是恒等方程与投资方程的线性组合，构成了与消费方程具有相同统计形式的新方程，使得消费方程不可识别。

3. 恰好识别与过度识别

我们讲“某个随机方程，当给定有关变量的样本观测值，其参数具有确定的估计量”，包括两种情况：一是只有一组参数估计量；二是具有有限组参数估计量。如果某个随机方程具有一组参数估计量，称其为**恰好识别**(just identification)；如果某个随机方程具有多组参数估计量，称其为**过度识别**(over identification)。

为了更好地理解上述概念，我们通过模型(6.3.1)式及其改进形式逐步加以说明。

(1) 模型 1

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{t1}, \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{t2}\ , \qquad t=1,2,\cdots,n \\ Y_t = C_t + I_t, \end{cases}$$

已经判断消费方程不可识别。同样第 1 个与第 3 个方程的线性组合得到的新方程具有与投资方程相同的统计形式，所以投资方程也是不可识别的。于是，该模型系统不可识别。

该模型的简化式模型为

$$\begin{cases} C_t = \pi_{10} + \varepsilon_{t1}, \\ I_t = \pi_{20} + \varepsilon_{t2}, \qquad t=1,2,\cdots,n \\ Y_t = \pi_{30} + \varepsilon_{t3}, \end{cases}$$

参数关系体系为

$$\pi_{10} = \frac{\alpha_0 - \alpha_0\beta_1 + \alpha_1\beta_0}{1-\alpha_1-\beta_1}$$

$$\pi_{20} = \frac{\beta_0 - \alpha_1\beta_0 + \alpha_0\beta_1}{1-\alpha_1-\beta_1}$$

$$\pi_{30} = \frac{\beta_0 + \alpha_0}{1-\alpha_1-\beta_1}$$

由这 3 个方程组成的方程组中，剔除一个矛盾方程(方程 1 与方程 2 相加，右端等于方程 3 的右端，而左端并不一定相等，形成矛盾方程)，在已知 $\hat{\pi}_{10},\hat{\pi}_{20},\hat{\pi}_{30}$ 时，2 个方程不能求得 $\hat{\alpha}_0,\hat{\alpha}_1,\hat{\beta}_0,\hat{\beta}_1$ 的确定值。所以也证明消费方程与投资方程都是不可识别的。

(2) 模型 2

在模型 1 的投资方程中增加解释变量 Y_{t-1}，模型变为

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{t1}, \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{t2}\ , \qquad t=1,2,\cdots,n \\ Y_t = C_t + I_t, \end{cases} \tag{6.3.2}$$

这时，消费方程是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。但是，投资方程仍然是不可识别的，因为第 1 个，第 2 个与第 3 个方程的线性组合(消去 C)构成与它相同的统计形式。于是，该模型系统仍然不可识别。

该模型的简化式模型为

$$\begin{cases} C_t = \pi_{10} + \pi_{11} Y_{t-1} + \varepsilon_{t1}, \\ I_t = \pi_{20} + \pi_{21} Y_{t-1} + \varepsilon_{t2}\ , \qquad t=1,2,\cdots,n \\ Y_t = \pi_{30} + \pi_{31} Y_{t-1} + \varepsilon_{t3} \end{cases}$$

参数关系体系为

$$
\pi_{10}=\frac{\alpha_0-\alpha_0\beta_1+\alpha_1\beta_0}{1-\alpha_1-\beta_1}, \qquad \pi_{11}=\frac{\alpha_1\beta_2}{1-\alpha_1-\beta_1}
$$

$$
\pi_{20}=\frac{\beta_0-\alpha_1\beta_0+\alpha_0\beta_1}{1-\alpha_1-\beta_1}, \qquad \pi_{21}=\frac{\beta_2-\alpha_1\beta_2}{1-\alpha_1-\beta_1}
$$

$$
\pi_{30}=\frac{\beta_0+\alpha_0}{1-\alpha_1-\beta_1}, \qquad \pi_{31}=\frac{\beta_2}{1-\alpha_1-\beta_1}
$$

这 6 个方程组成的方程组中剔除 2 个矛盾方程，在已知$\hat{\pi}_{10},\hat{\pi}_{20},\hat{\pi}_{30},\hat{\pi}_{11}$，$\hat{\pi}_{21},\hat{\pi}_{31}$时，由 4 个方程不能求得所有 5 个结构参数的确定估计值，但是可以得到$\hat{\alpha}_0,\hat{\alpha}_1$的确定值。所以也证明消费方程可以识别，而投资方程都是不可识别的。而且，只能得到$\hat{\alpha}_0,\hat{\alpha}_1$的一组确定值，所以消费方程是恰好识别的方程。(读者自己求解上述方程组，验证这些结论。)

(3) 模型 3

在模型 2 的消费方程中增加解释变量C_{t-1}，模型变为

$$
\begin{cases}
C_t=\alpha_0+\alpha_1Y_t+\alpha_2C_{t-1}+\mu_{t1}, \\
I_t=\beta_0+\beta_1Y_t+\beta_2Y_{t-1}+\mu_{t2}, \qquad t=1,2,\cdots,n \\
Y_t=C_t+I_t,
\end{cases}
\tag{6.3.3}
$$

这时，消费方程仍然是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。而且，投资方程也是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。于是，该模型系统是可以识别的。

该模型的简化式模型为

$$
\begin{cases}
C_t=\pi_{10}+\pi_{11}Y_{t-1}+\pi_{12}C_{t-1}+\varepsilon_{t1}, \\
I_t=\pi_{20}+\pi_{21}Y_{t-1}+\pi_{22}C_{t-1}+\varepsilon_{t2}, \qquad t=1,2,\cdots,n \\
Y_t=\pi_{30}+\pi_{31}Y_{t-1}+\pi_{32}C_{t-1}+\varepsilon_{t3},
\end{cases}
$$

参数关系体系为

$$
\pi_{10}=\frac{\alpha_0-\alpha_0\beta_1+\alpha_1\beta_0}{1-\alpha_1-\beta_1}, \qquad \pi_{11}=\frac{\alpha_1\beta_2}{1-\alpha_1-\beta_1}, \qquad \pi_{12}=\frac{\alpha_2-\alpha_2\beta_1}{1-\alpha_1-\beta_1}
$$

$$
\pi_{20}=\frac{\beta_0-\alpha_1\beta_0+\alpha_0\beta_1}{1-\alpha_1-\beta_1}, \qquad \pi_{21}=\frac{\beta_2-\alpha_1\beta_2}{1-\alpha_1-\beta_1}, \qquad \pi_{22}=\frac{\alpha_2\beta_1}{1-\alpha_1-\beta_1}
$$

$$
\pi_{30}=\frac{\beta_0+\alpha_0}{1-\alpha_1-\beta_1}, \qquad \pi_{31}=\frac{\beta_2}{1-\alpha_1-\beta_1}, \qquad \pi_{32}=\frac{\alpha_2}{1-\alpha_1-\beta_1}
$$

这 9 个方程组成的方程组中剔除 3 个矛盾方程，在已知简化式参数估计值时，由 6 个方程能够求得所有 6 个结构参数的确定估计值。所以也证明消费方程和投资方程都是可以识别的。而且，只能得到所有 6 个结构参数的一组确定值，所以消费方程和投资方程都是恰好识别的方程。(读者自己求解上述方程组，验

证这些结论。）

(4) 模型 4

在模型 3 的消费方程中增加解释变量前一年的价格指数 P_{t-1}，模型变为

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \alpha_3 P_{t-1} + \mu_{t1}, \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{t2}, \qquad t=1,2,\cdots,n \\ Y_t = C_t + I_t, \end{cases} \tag{6.3.4}$$

这时，消费方程和投资方程仍然是可以识别的，因为任何方程的线性组合都不能构成与它们相同的统计形式。于是，该模型系统是可以识别的。

该模型的简化式模型为

$$\begin{cases} C_t = \pi_{10} + \pi_{11} Y_{t-1} + \pi_{12} C_{t-1} + \pi_{13} P_{t-1} + \varepsilon_{t1}, \\ I_t = \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} C_{t-1} + \pi_{23} P_{t-1} + \varepsilon_{t2}, \qquad t=1,2,\cdots,n \\ Y_t = \pi_{30} + \pi_{31} Y_{t-1} + \pi_{32} C_{t-1} + \pi_{33} P_{t-1} + \varepsilon_{t3}, \end{cases}$$

参数关系体系为

$$\begin{array}{lll} \pi_{10} = \dfrac{\alpha_0 - \alpha_0\beta_1 + \alpha_1\beta_0}{1-\alpha_1-\beta_1}, & \pi_{20} = \dfrac{\beta_0 - \alpha_1\beta_0 + \alpha_0\beta_1}{1-\alpha_1-\beta_1}, & \pi_{30} = \dfrac{\beta_0 + \alpha_0}{1-\alpha_1-\beta_1} \\ \pi_{11} = \dfrac{\alpha_1\beta_2}{1-\alpha_1-\beta_1}, & \pi_{21} = \dfrac{\beta_2 - \alpha_1\beta_2}{1-\alpha_1-\beta_1}, & \pi_{31} = \dfrac{\beta_2}{1-\alpha_1-\beta_1} \\ \pi_{12} = \dfrac{\alpha_2 - \alpha_2\beta_1}{1-\alpha_1-\beta_1}, & \pi_{22} = \dfrac{\alpha_2\beta_1}{1-\alpha_1-\beta_1}, & \pi_{32} = \dfrac{\alpha_2}{1-\alpha_1-\beta_1} \\ \pi_{13} = \dfrac{\alpha_3 - \alpha_3\beta_1}{1-\alpha_1-\beta_1}, & \pi_{23} = \dfrac{\alpha_3\beta_1}{1-\alpha_1-\beta_1}, & \pi_{33} = \dfrac{\alpha_3}{1-\alpha_1-\beta_1} \end{array}$$

这 12 个方程组成的方程组中剔除 4 个矛盾方程，在已知简化式参数估计值时，由 8 个方程能够求得所有 7 个结构参数的确定估计值。所以也证明消费方程和投资方程都是可以识别的。但是，求解结果表明，对于 $\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3$ 只能得到一组确定值，所以消费方程是恰好识别的方程；而对于 $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2$，能够得到多组确定值，所以投资方程是过度识别的方程。(读者自己求解上述方程组，验证这些结论。）

需要特别指出，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。模型 4 中的投资方程就是这种情况。

PDG

二、结构式识别条件

从识别的概念出发，完全可以对联立方程计量经济学模型的识别状态进行判断，实际中也是这样做的。但从理论的角度出发，人们总希望有一些规范的判断方法。这里首先介绍一种直接从待判断的结构方程出发的方法，称为**结构式条件**。

联立方程计量经济学模型的结构式(6.2.2)式

$$\boldsymbol{BY}+\boldsymbol{\Gamma X}=\boldsymbol{N}$$

中的第 i 个方程中包含 g_i 个内生变量(含被解释变量)和 k_i 个先决变量(含常数项)，模型系统中内生变量和先决变量的数目仍用 g 和 k 表示，矩阵 $\boldsymbol{B}_0\boldsymbol{\Gamma}_0$ 表示第 i 个方程中未包含的变量(包括内生变量和先决变量)在其他 $g-1$ 个方程中对应系数所组成的矩阵。于是，判断第 i 个结构方程识别状态的结构式条件为

如果 $R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)<g-1$，则第 i 个结构方程不可识别。

如果 $R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)=g-1$，则第 i 个结构方程可以识别，并且

　　如果 $k-k_i=g_i-1$，则第 i 个结构方程恰好识别；

　　如果 $k-k_i>g_i-1$，则第 i 个结构方程过度识别。

其中符号 R 表示矩阵的秩。一般将该条件的前一部分称为**秩条件**(rank condition)，用以判断结构方程是否识别；后一部分称为**阶条件**(order condition)，用以判断结构方程恰好识别或者过度识别。

例 6.3.1

现在以模型(6.3.4)式为例解释结构式条件的应用。模型为

$$\begin{cases}C_t=\alpha_0+\alpha_1Y_t+\alpha_2C_{t-1}+\alpha_3P_{t-1}+\mu_{t1}, & \\ I_t=\beta_0+\beta_1Y_t+\beta_2Y_{t-1}+\mu_{t2}, & t=1,2,\cdots,n \\ Y_t=C_t+I_t, & \end{cases}$$

结构参数矩阵为

$$\boldsymbol{B\Gamma}=\begin{pmatrix}1 & 0 & -\alpha_1 & -\alpha_0 & 0 & -\alpha_2 & -\alpha_3\\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0 & 0\\ -1 & -1 & 1 & 0 & 0 & 0 & 0\end{pmatrix}$$

首先判断第 1 个结构方程的识别状态。对于第 1 个方程，有

$$\boldsymbol{B}_0\boldsymbol{\Gamma}_0=\begin{pmatrix}1 & -\beta_2\\ -1 & 0\end{pmatrix}$$

$$R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)=2=g-1$$

所以，该方程可以识别。我们看到，矩阵 $\boldsymbol{B}_0\boldsymbol{\Gamma}_0$ 实际上就是矩阵 $\boldsymbol{B\Gamma}$ 除去第 1 个结构方程参数所在的行(第 1 行)和第 1 行中非 0 元素(对应于第 1 个结构方程包含的元素)所在的列之后剩下的元素按照原次序排列而得到的。先写出矩阵 $\boldsymbol{B\Gamma}$，然后再从中得到与所判断的方程对应的矩阵 $\boldsymbol{B}_0\boldsymbol{\Gamma}_0$，这样既简单，又不容易出错。又因为有

$$k-k_1=1=g_1-1$$

所以，第 1 个结构方程为恰好识别的结构方程。与我们上面的判断结论是一致的。

再看第 2 个结构方程，有

$$\boldsymbol{B}_0\boldsymbol{\Gamma}_0=\begin{pmatrix}1 & -\alpha_2 & -\alpha_3\\ -1 & 0 & 0\end{pmatrix}$$

$$R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)=2=g-1$$

所以，该方程可以识别。并且

$$k-k_2=2>g_2-1$$

所以，第 2 个结构方程为过度识别的结构方程。与我们上面的判断结论也是一致的。

第 3 个方程是平衡方程，不存在识别问题。

综合以上结果，该联立方程计量经济学模型是可以识别的。

例 6.3.2

再以(6.3.2)式模型为例。模型为

$$\begin{cases}C_t=\alpha_0+\alpha_1Y_t+\mu_{t1},\\ I_t=\beta_0+\beta_1Y_t+\beta_2Y_{t-1}+\mu_{t2}, & t=1,2,\cdots,n\\ Y_t=C_t+I_t,\end{cases}$$

结构参数矩阵为

$$\boldsymbol{B\Gamma}=\begin{pmatrix}1 & 0 & -\alpha_1 & -\alpha_0 & 0\\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2\\ -1 & -1 & 1 & 0 & 0\end{pmatrix}$$

首先判断第 1 个结构方程的识别状态。对于第 1 个方程，有

$$\boldsymbol{B}_0\boldsymbol{\Gamma}_0=\begin{pmatrix}1 & -\beta_2\\ -1 & 0\end{pmatrix}$$

$$R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)=2=g-1$$

所以，该方程可以识别。并且

$$k-k_1=1=g_1-1$$

所以，第 1 个结构方程为恰好识别的结构方程。

再看第 2 个结构方程，有

$$\boldsymbol{B}_0\boldsymbol{\Gamma}_0=\begin{pmatrix}1\\-1\end{pmatrix}$$

$$R(\boldsymbol{B}_0\boldsymbol{\Gamma}_0)=1<g-1$$

所以，该方程不可以识别。

综合以上结果，该联立方程计量经济学模型不可以识别。与上面的判断结论是一致的。

*三、简化式识别条件

如果已经知道联立方程计量经济学模型的简化式模型参数，那么可以通过对简化式模型的研究达到判断结构式模型是否识别的目的。对于简化式模型(6.2.6)式

$$\boldsymbol{Y}=\boldsymbol{\Pi X}+\boldsymbol{E}$$

简化式识别条件为

如果 $R(\boldsymbol{\Pi}_2)<g_i-1$，则第 i 个结构方程不可识别。

如果 $R(\boldsymbol{\Pi}_2)=g_i-1$，则第 i 个结构方程可以识别，并且

　　如果 $k-k_i=g_i-1$，则第 i 个结构方程恰好识别；

　　如果 $k-k_i>g_i-1$，则第 i 个结构方程过度识别。

其中 $\boldsymbol{\Pi}_2$ 是简化式参数矩阵 $\boldsymbol{\Pi}$ 中划去第 i 个结构方程所不包含的内生变量所对应的行和第 i 个结构方程中包含的先决变量所对应的列之后，剩下的参数按原次序组成的矩阵。至于为什么用 $\boldsymbol{\Pi}_2$ 而不用其他符号，是与它在矩阵 $\boldsymbol{\Pi}$ 中的分块位置有关。其他符号、变量的含义与结构式识别条件相同。一般也将该条件的前一部分称为秩条件，用以判断结构方程是否识别；后一部分称为阶条件，用以判断结构方程恰好识别或者过度识别。

例 6.3.3

某联立方程计量经济学模型，其结构式模型如下：

$$\begin{cases} Y_{i1}=\alpha_1 Y_{i2}+\alpha_2 X_{i1}+\alpha_3 X_{i2}+\mu_{i1}, \\ Y_{i2}=\beta_1 Y_{i3}+\beta_2 X_{i3}+\mu_{i2}, & i=1,2,\cdots,n \\ Y_{i3}=\gamma_1 Y_{i1}+\gamma_2 Y_{i2}+\gamma_3 X_{i3}+\mu_{i3}, \end{cases}$$

$k=3, g=3$, 已知其简化式模型参数矩阵为

$$\boldsymbol{\Pi}=\begin{pmatrix}4 & -2 & 3\\ 2 & -1 & 1\\ 2 & -1 & 0\end{pmatrix}$$

现在利用简化式条件判断结构式模型的识别状态。

对于第 1 个结构式方程，

$$k_1=2\,,\qquad g_1=2$$

$$\boldsymbol{\Pi}_2=\begin{pmatrix}3\\ 1\end{pmatrix}$$

因为

$$R(\boldsymbol{\Pi}_2)=1=g_1-1$$

所以该方程是可以识别的。又因为

$$k-k_1=1=g_1-1$$

所以该方程是恰好识别的。

对于第 2 个结构式方程，

$$k_2=1\,,\qquad g_2=2$$

$$\boldsymbol{\Pi}_2=\begin{pmatrix}2 & -1\\ 2 & -1\end{pmatrix}$$

因为

$$R(\boldsymbol{\Pi}_2)=1=g_2-1$$

所以该方程是可以识别的。又因为

$$k-k_2=2>g_2-1$$

所以该方程是过度识别的。

对于第 3 个结构式方程，

$$k_3=1\,,\qquad g_3=3$$

$$\boldsymbol{\Pi}_2=\begin{pmatrix}4 & -2\\ 2 & -1\\ 2 & -1\end{pmatrix}$$

因为

$$R(\boldsymbol{\Pi}_2)=1<g_3-1$$

所以该方程是不可以识别的。

综合上述结果，该联立方程计量经济学模型系统不可识别。

可以从数学上严格证明，简化式识别条件和结构式识别条件是等价的(有兴趣的读者可参阅《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)。

四、实际应用中的经验方法

当一个联立方程计量经济学模型系统中的方程数目比较多时，无论是从识别的概念出发，还是利用规范的结构式或联立方程计量经济学模型简化式识别条件，对模型进行识别，困难都是很大的，或者说是不可能的。因为一般实际联立方程计量经济学模型包含几百个、上千个方程是正常的。这就是理论与实际的脱节，理论上很严格的方法在实际中往往是无法应用的，在实际中应用的往往是一些经验方法。

关于联立方程计量经济学模型的识别问题，我们并不是等到理论模型已经建立了之后再像上面所介绍那样进行识别，而是在建立模型的过程中设法保证模型的可识别性。那么，在建立模型时就要遵循如下原则：

在建立某个结构方程时，要使该方程包含前面每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这个新引入的方程本身是可以识别的。只要前面每个方程都包含至少1个该方程所未包含的变量，并且互不相同，那么所有方程的任意线性组合都不能构成与该方程相同的统计形式。

在实际建模时，将每个方程所包含的变量记录在如表6.3.1所示的表式中，将是有帮助的。例如，在建立第4个方程时，必须包含变量1，2，3，4，5，6之外的至少一个变量；同时需要检查方程1，2，3是否都存在至少1个方程4所未包含的变量，且互不相同，这里可以认为方程1中的变量1，方程2中的变量4和5，方程3中的变量6满足要求。于是，所建立的方程4是可以识别的。

表6.3.1 变量记录表

	变量1	变量2	变量3	变量4	变量5	变量6	…
方程1	×	×		×			
方程2		×	×	×	×		
方程3	×		×	×		×	
方程4		×	×				
⋮							

§6.4 联立方程计量经济学模型的估计

一、概述

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**。显然，从模型估计的性质来讲，系统估计方法必然优于单方程方法，但从方法的复杂性来讲，单方程方法又优于系统估计方法。在实际中，单方程方法得到广泛的应用。

单方程估计方法主要解决的是联立方程计量经济学模型系统每个方程中的随机解释变量问题，同时尽可能地利用单个方程中没有包含的而在模型系统中包含的变量样本观测值的信息，但是没有考虑模型系统方程之间的相关性对单个方程参数估计量的影响。因此，也将单方程估计方法称为**有限信息估计方法**。

单方程估计方法按其方法原理又分为两类：一类以最小二乘为原理，如**间接最小二乘法**(Indirect Least Squares, ILS)、**二阶段最小二乘法**(Two Stage Least Squares，2SLS)、**工具变量法**(Instrumental Variable, IV)，我们称其为经典方法；一类不以最小二乘为原理，或者不直接从最小二乘原理出发，如以最大似然法为原理的**有限信息最大似然法**(Limited Information Maximum Likelihood, LIML)，以及仍然应用最小二乘原理，但并不以残差平方和最小为判断标准的**最小方差比方法**(Least Variable Ration, LVR)。

联立方程计量经济学模型的单方程估计方法不同于单方程计量经济学模型的估计方法，无论是研究对象还是方法本身都是不同的，不要将二者混淆。但是，单方程计量经济学模型估计方法的原理构成了联立方程计量经济学模型单方程估计方法的基础。

系统估计方法利用了模型系统提供的所有信息，包括方程之间的相关性信息。因此也将系统估计方法称为**完全信息估计方法**。

系统估计方法主要包括**三阶段最小二乘法**(Three Stage Least Squares, 3SLS)和**完全信息最大似然法**(Full Information Maximum Likelihood, FIML)。

本书只介绍几种简单常用的单方程估计方法。实际上，在下节我们会看到，在大量的联立方程计量经济学模型的应用研究中，仍然广泛应用普通最小二乘法进行模型的估计。

PDG

二、狭义的工具变量法(IV)

工具变量法是一类估计方法的统称，可以有各种不同的选择工具变量的方法。在这里仅指一种特定的工具变量而言，故称为“狭义的工具变量法”。

1. 工具变量的选取

对于联立方程计量经济学模型

$$\boldsymbol{BY}+\boldsymbol{\Gamma X}=\boldsymbol{N} \tag{6.4.1}$$

的每个结构方程，如第 1 个方程，可以写成如下形式：

$$\boldsymbol{Y}_1=\beta_{21}\boldsymbol{Y}_2+\beta_{31}\boldsymbol{Y}_3+\cdots+\beta_{g_11}\boldsymbol{Y}_{g_1}+\gamma_{11}\boldsymbol{X}_1+\gamma_{21}\boldsymbol{X}_2+\cdots+\gamma_{k_11}\boldsymbol{X}_{k_1}+\boldsymbol{N}_1 \tag{6.4.2}$$

该方程包含 g_1-1 个内生解释变量和 k_1 个先决解释变量。写成矩阵形式为

$$\boldsymbol{Y}_1=(\boldsymbol{Y}_0\quad\boldsymbol{X}_0)\begin{pmatrix}\boldsymbol{B}_0\\\boldsymbol{\Gamma}_0\end{pmatrix}+\boldsymbol{N}_1 \tag{6.4.3}$$

其中

$$\boldsymbol{Y}_0=(\boldsymbol{Y}_2\quad\boldsymbol{Y}_3\quad\cdots\quad\boldsymbol{Y}_{g_1})=\begin{pmatrix}Y_{12}&Y_{13}&\cdots&Y_{1g_1}\\Y_{22}&Y_{23}&\cdots&Y_{2g_1}\\\vdots&\vdots&&\vdots\\Y_{n2}&Y_{n3}&\cdots&Y_{ng_1}\end{pmatrix}$$

$$\boldsymbol{X}_0=(\boldsymbol{X}_1\quad\boldsymbol{X}_2\quad\cdots\quad\boldsymbol{X}_{k_1})=\begin{pmatrix}X_{11}&X_{12}&\cdots&X_{1k_1}\\X_{21}&X_{22}&\cdots&X_{2k_1}\\\vdots&\vdots&&\vdots\\X_{n1}&X_{n2}&\cdots&X_{nk_1}\end{pmatrix}$$

$$\boldsymbol{B}_0=\begin{pmatrix}\beta_{21}\\\beta_{31}\\\vdots\\\beta_{g_11}\end{pmatrix},\qquad\boldsymbol{\Gamma}_0=\begin{pmatrix}\gamma_{11}\\\gamma_{21}\\\vdots\\\gamma_{k_11}\end{pmatrix},\qquad\boldsymbol{Y}_1=\begin{pmatrix}Y_{11}\\Y_{21}\\\vdots\\Y_{n1}\end{pmatrix},\qquad\boldsymbol{N}_1=\begin{pmatrix}\mu_{11}\\\mu_{21}\\\vdots\\\mu_{n1}\end{pmatrix}$$

n 为样本容量，请读者注意，这里的 $\boldsymbol{B}_0,\boldsymbol{\Gamma}_0$ 的含义已不同于结构式识别条件中的 $\boldsymbol{B}_0,\boldsymbol{\Gamma}_0$。

欲估计结构方程(6.4.3)式，必须克服随机解释变量问题，有效的方法是工具变量法。在 § 4.4 中已经给出了工具变量的条件，在这里，自然就会想到，方程中没有包含的 $k-k_1$ 个先决变量基本满足工具变量的条件，可以选择它们作为方程中包含的 g_1-1 个内生解释变量的工具变量。如此选择工具变量的方法称为狭义的工具变量法。

如果结构方程(6.4.3)式是恰好识别的，即满足 $k-k_1=g_1-1$，那么，工具变

量的选择就很简单。如果结构方程(6.4.3)式是过度识别的，即满足$k-k_1>g_1-1$，那么，工具变量的选择就比较麻烦。而且参数估计结果有一定的任意性，因为每从$k-k_1$个没有包含在方程之中的先决变量中选出g_1-1个变量作为工具变量，就得到一组参数估计值。共计可能有$C_{k-k_1}^{g_1-1}$种不同的参数估计值。所以，一**般认为，这种工具变量法只适用于恰好识别的结构方程的估计**。

2. IV 参数估计量及其统计特性

选择$\boldsymbol{X}_0^*$作为$\boldsymbol{Y}_0$的工具变量，得到参数估计量为

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\Gamma}}_0\end{pmatrix}_{\mathrm{IV}}=((\boldsymbol{X}_0^*\quad\boldsymbol{X}_0)'\quad(\boldsymbol{Y}_0\quad\boldsymbol{X}_0))^{-1}(\boldsymbol{X}_0^*\quad\boldsymbol{X}_0)'\boldsymbol{Y}_1 \tag{6.4.4}$$

其中

$$\boldsymbol{X}_0^*=(\boldsymbol{X}_{k_1+1}\quad\boldsymbol{X}_{k_1+2}\quad\cdots\quad\boldsymbol{X}_k)=\begin{pmatrix}X_{1,k_1+1}&X_{1,k_1+2}&\cdots&X_{1k}\\X_{2,k_1+1}&X_{2,k_1+2}&\cdots&X_{2k}\\\vdots&\vdots&&\vdots\\X_{n,k_1+1}&X_{n,k_1+2}&\cdots&X_{nk}\end{pmatrix}$$

(6.4.4)式估计量的估计过程已经在§4.4 中介绍了，这里不再重复。

工具变量法参数估计量，正如在§4.4 中已经说明的，一般情况下，在小样本下是有偏的，但在大样本下是渐近无偏的。如果选取的工具变量与方程随机干扰项完全不相关，那么其参数估计量是无偏性估计量。

3. 参数估计量与工具变量的次序无关

对于恰好识别的结构方程，选择该方程中没有包含的$k-k_1$个先决变量作为方程中包含的g_1-1个内生解释变量的工具变量，虽然只能有一组选择，但在这一组中具体哪个先决变量作为哪个内生变量的工具变量，仍然具有任意性。但是这种任意性对参数估计量没有影响。为什么？

从§4.4 中知道，**工具变量法参数估计量是一个关于该参数估计量的正规方程组的解**。由该正规方程组的形成过程可以看出，**如果工具变量的次序不同，也就是工具变量被使用的先后次序不同，那么正规方程组中方程的次序将不相同**。但是由代数知识可知，在一个线性代数方程组中，**方程的次序不影响方程组的解**。所以，**只要选择的工具变量组中的变量是相同的，只能得到一种参数估计量，而与变量的次序无关**。这是一个重要的概念，请读者能够理解，在后续课程中还将多次用到这个概念。

三、间接最小二乘法(ILS)

联立方程计量经济学模型的结构方程中包含有内生解释变量，不能直接采

PDG

用普通最小二乘法估计其参数。但是对于简化式方程，正如在关于简化式模型概念介绍中提到的，可以采用普通最小二乘法直接估计其参数。于是就提出了间接最小二乘法：先对关于内生解释变量的简化式方程采用普通最小二乘法估计简化式参数，得到简化式参数估计量，然后通过参数关系体系，计算得到结构式参数的估计量。

间接最小二乘法只适用于恰好识别的结构方程的参数估计，因为只有恰好识别的结构方程，才能从参数关系体系中得到唯一一组结构参数的估计量。

1. 一个简单的例子

现有一个联立方程计量经济学模型，其结构式模型为

$$\begin{cases} Y_1 = \beta_{12}Y_2 + \gamma_{11}X_1 + \gamma_{12}X_2 + \mu_1 \\ Y_2 = \beta_{23}Y_3 + \gamma_{23}X_3 + \mu_2 \\ Y_3 = \beta_{31}Y_1 + \beta_{32}Y_2 + \gamma_{33}X_3 + \mu_3 \end{cases}$$

现欲估计第 1 个结构方程的参数，可以证明，该方程是恰好识别的，可以采用间接最小二乘法。该方程中有两个内生变量，相应的简化式方程为

$$\begin{cases} Y_1 = \pi_{11}X_1 + \pi_{12}X_2 + \pi_{13}X_3 + \varepsilon_1 \\ Y_2 = \pi_{21}X_1 + \pi_{22}X_2 + \pi_{23}X_3 + \varepsilon_2 \end{cases}$$

应用普通最小二乘法，在样本数据的支持下对每个简化式方程分别估计其参数，得到参数估计量 $\hat{\pi}_{ij}, i=1,2, j=1,2,3$ 。将简化式代入第 1 个结构方程，得到参数关系体系

$$\begin{cases} \pi_{11} - \beta_{12}\pi_{21} = \gamma_{11} \\ \pi_{12} - \beta_{12}\pi_{22} = \gamma_{12} \\ \pi_{13} - \beta_{12}\pi_{23} = 0 \end{cases}$$

由简化式参数估计量 $\hat{\pi}_{ij}, i=1,2, j=1,2,3$ ，计算得到结构参数估计值

$$\begin{cases} \hat{\beta}_{12} = \dfrac{\hat{\pi}_{13}}{\hat{\pi}_{23}} \\ \hat{\gamma}_{11} = \dfrac{\hat{\pi}_{13}\hat{\pi}_{21}}{\hat{\pi}_{23} - \hat{\pi}_{11}} \\ \hat{\gamma}_{12} = \dfrac{\hat{\pi}_{13}\hat{\pi}_{22}}{\hat{\pi}_{23} - \hat{\pi}_{12}} \end{cases}$$

2. 一般间接最小二乘法的估计过程

现在对结构方程(6.4.2)式的参数进行间接最小二乘估计。将(6.4.3)式改写成

$$\boldsymbol{Y}_1 - \boldsymbol{B}_0\boldsymbol{Y}_0 - \boldsymbol{\Gamma}_0\boldsymbol{X}_0 = \boldsymbol{N}_1$$

PDG

即

$$\begin{pmatrix} 1 & -\boldsymbol{B}_0 & -\boldsymbol{\Gamma}_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_0 \\ \boldsymbol{X}_0 \end{pmatrix} = \boldsymbol{N}_1 \tag{6.4.5}$$

$$\begin{pmatrix} \boldsymbol{B}_{00} & \boldsymbol{\Gamma}_{00} \end{pmatrix} \begin{pmatrix} \boldsymbol{Y}_{00} \\ \boldsymbol{X}_0 \end{pmatrix} = \boldsymbol{N}_1$$

其中

$$\boldsymbol{B}_{00} = \begin{pmatrix} 1 & -\boldsymbol{B}_0 \end{pmatrix}, \qquad \boldsymbol{\Gamma}_{00} = -\boldsymbol{\Gamma}_0, \qquad \boldsymbol{Y}_{00} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_0 \end{pmatrix}$$

内生变量的简化式模型为

$$\boldsymbol{Y}_{00} = \boldsymbol{\Pi}_{00}\boldsymbol{X} + \boldsymbol{E} \tag{6.4.6}$$

代入结构式模型，得到

$$\boldsymbol{B}_{00}\boldsymbol{\Pi}_{00}\boldsymbol{X} + \boldsymbol{\Gamma}_{00}\boldsymbol{X}_0 = 0$$

$$\boldsymbol{B}_{00}\boldsymbol{\Pi}_{00}\begin{pmatrix} \boldsymbol{X}_0 \\ \boldsymbol{X}_0^* \end{pmatrix} + \boldsymbol{\Gamma}_{00}\boldsymbol{X}_0 = 0$$

将 $\boldsymbol{\Pi}_{00}$ 分成两部分，一部分对应结构方程中包含的先决变量 $\boldsymbol{X}_0$，一部分对应结构方程中未包含的先决变量 $\boldsymbol{X}_0^*$，即

$$\boldsymbol{\Pi}_{00} = (\boldsymbol{\Pi}_{00}^1 \quad \boldsymbol{\Pi}_{00}^2)$$

于是有参数关系体系

$$\begin{cases} \boldsymbol{B}_{00}\boldsymbol{\Pi}_{00}^1 = \boldsymbol{\Gamma}_0 \\ \boldsymbol{B}_{00}\boldsymbol{\Pi}_{00}^2 = 0 \end{cases} \tag{6.4.7}$$

用普通最小二乘法估计简化式模型(6.4.6)式，得到 $\hat{\boldsymbol{\Pi}}_{00}$，代入参数关系体系(6.4.7)式,先由第 2 组方程计算得到 $\hat{\boldsymbol{B}}_{00}$，然后再代入第 1 组方程计算得到 $\hat{\boldsymbol{\Gamma}}_0$。于是得到了结构方程(6.4.2)式的结构参数估计量。

3．间接最小二乘法参数估计的统计性质

对于简化式模型应用普通最小二乘法得到的参数估计量具有线性性、无偏性、有效性。通过参数关系体系计算得到结构方程的结构参数估计量在小样本下是有偏的，在大样本下是渐近无偏的。

4．间接最小二乘法也是一种工具变量法

可以从数学上严格证明，采用间接最小二乘法估计结构方程(6.4.3)式等价于一种工具变量法，选择 $\boldsymbol{X}$ 作为 $(\boldsymbol{Y}_0,\boldsymbol{X}_0)$ 的工具变量，即用 $(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{k_1},\boldsymbol{X}_{k_1+1},\cdots,\boldsymbol{X}_k)$ 依次作为 $(\boldsymbol{Y}_2,\boldsymbol{Y}_3,\cdots,\boldsymbol{Y}_{g_1},\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{k_1})$ 的工具变量。请注意，这里对于结构方程中包含的先决变量也选择了其他先决变量作为工具变量。于是，

结构方程(6.4.3)式参数的间接最小二乘估计量可以写作

$$\begin{pmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{pmatrix}_{\mathrm{ILS}} = \left(X' \quad (Y_0 \quad X_0)\right)^{-1} X'Y_1 \tag{6.4.8}$$

这是一个重要的结论(关于它的数学证明，有兴趣的读者可以参考《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010)。

四、二阶段最小二乘法(2SLS)

狭义的工具变量法和间接最小二乘法一般只适用于联立方程计量经济学模型中恰好识别的结构方程的估计。但是，**在实际的联立方程计量经济学模型中，恰好识别的结构方程很少出现，一般情况下结构方程都是过度识别的。**因为实际的联立方程计量经济学模型一般包含较多数目的结构方程和先决变量，例如，包含 100 个方程，30 个先决变量的宏观经济模型不是大模型；而在每个结构方程中，如宏观经济模型中的生产方程、消费方程，一般仅包含 3~5 个变量，包括内生变量和先决变量。于是就出现了

$$k - k_i >> g_i - 1$$

的情况，所以结构方程大多是过度识别的。

二阶段最小二乘法是一种既适用于恰好识别的结构方程，又适用于过度识别的结构方程的单方程估计方法，由希尔（Theil）和巴斯曼尼（Basmann）分别于 1953 年和 1957 年各自独立提出，是一种应用最普遍的方法。

1. 二阶段最小二乘法

对于联立方程计量经济学模型(6.4.1)式中的第 1 个结构方程(6.4.3)式，由于内生解释变量 Y_0 是随机变量，不能直接采用普通最小二乘法。但是对于 Y_0 的简化式方程，即简化式模型

$$Y_0 = X\Pi_0 + E_0 \tag{6.4.9}$$

中的每个方程，不存在随机解释变量问题，可以直接采用普通最小二乘法估计其参数，并得到关于 Y_0 的估计值：

$$\hat{Y}_0 = X\hat{\Pi}_0 = X\left[(X'X)^{-1}X'Y_0\right] \tag{6.4.10}$$

这就是二阶段最小二乘法的第一阶段，即对简化式方程第一次使用普通最小二乘法。

用 Y_0 的估计量 $\hat{Y}_0$ 替换(6.4.3)式中的 Y_0，得到新的方程

$$Y_1 = (\hat{Y}_0 \quad X_0)\begin{pmatrix} B_0 \\ \Gamma_0 \end{pmatrix} + N_1 \tag{6.4.11}$$

显然，该方程中不存在随机解释变量问题，可以直接采用普通最小二乘法估计其参数，得到

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\Gamma}}_0\end{pmatrix}_{\text{2SLS}} = ((\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)' \quad (\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0))^{-1}(\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)'\boldsymbol{Y}_1 \qquad (6.4.12)$$

这就是二阶段最小二乘法的第二阶段，即对变换了的结构式方程使用普通最小二乘法。得到的参数估计量即为原结构方程(6.4.3)式参数的二阶段最小二乘估计量。

在应用二阶段最小二乘法的整个过程中，并没有涉及结构方程中内生解释变量和先决解释变量的数目，所以二阶段最小二乘法的应用与方程的识别状态无关，既适用于恰好识别的结构方程，又适用于过度识别的结构方程。

2．二阶段最小二乘估计量的统计性质

采用二阶段最小二乘法得到结构方程的结构参数估计量在小样本下是有偏的，在大样本下是渐近无偏的。下面我们将看到，由于二阶段最小二乘法也是一种工具变量法，它的估计量与工具变量法估计量是等价的，所以具有相同的统计性质。

3．二阶段最小二乘法也是一种工具变量法

如果我们不是用$\boldsymbol{Y}_0$的估计量$\hat{\boldsymbol{Y}}_0$替换(6.4.3)式中的$\boldsymbol{Y}_0$，而是用$\boldsymbol{Y}_0$的估计量$\hat{\boldsymbol{Y}}_0$作为(6.4.3)式中的$\boldsymbol{Y}_0$的工具变量，显然，因为$\hat{\boldsymbol{Y}}_0$是$\boldsymbol{X}$的线性组合，基本符合工具变量的条件。那么，按照工具变量方法的估计过程，应该得到如下的结构参数估计量：

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\Gamma}}_0\end{pmatrix} = ((\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)' \quad (\boldsymbol{Y}_0 \quad \boldsymbol{X}_0))^{-1}(\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)'\boldsymbol{Y}_1 \qquad (6.4.13)$$

将(6.4.12)式与(6.4.13)式进行比较发现，它们的区别仅仅是后者没有改变原方程中的解释变量$(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0)$。从数学上可以严格证明(6.4.12)式与(6.4.13)式表示的两组参数估计量是完全等价的，所以可以把二阶段最小二乘法也看作一种工具变量法，选择$\boldsymbol{Y}_0$的简化式方程的估计量$\hat{\boldsymbol{Y}}_0$作为结构式方程中$\boldsymbol{Y}_0$的工具变量。显然，工具变量的选取不同于上述的狭义工具变量法和间接最小二乘法，但都属于工具变量法。关于(6.4.12)式与(6.4.13)式等价性证明过程，有兴趣的读者可以参考《计量经济学学习指南与练习》，潘文卿，李子奈编著，高等教育出版社，2010。

五、对于恰好识别的结构方程，三种方法是等价的

上述三种单方程估计方法都适用于恰好识别的结构方程，对于同一个结构方程，选择不同的方法，应该得到相同的参数估计量。

对于(6.4.3)式所表示的结构方程，分别采用三种单方程估计方法得到的参

数估计量如下：

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\varGamma}}_0\end{pmatrix}_{\mathrm{IV}}=\left(\begin{pmatrix}\boldsymbol{X}_0^{*} & \boldsymbol{X}_0\end{pmatrix}'\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\right)^{-1}\begin{pmatrix}\boldsymbol{X}_0^{*} & \boldsymbol{X}_0\end{pmatrix}'\boldsymbol{Y}_1$$

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\varGamma}}_0\end{pmatrix}_{\mathrm{ILS}}=\left(\boldsymbol{X}'\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}_1$$

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\varGamma}}_0\end{pmatrix}_{\mathrm{2SLS}}=\left(\begin{pmatrix}\hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0\end{pmatrix}'\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\right)^{-1}\begin{pmatrix}\hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0\end{pmatrix}'\boldsymbol{Y}_1$$

可以看到，三种结果是用不同的工具变量法估计得到的，区别仅在于工具变量的选取不同。

比较狭义的工具变量法和间接最小二乘法的参数估计量(6.4.4)式与(6.4.8)式，它们选取了同样一组变量 $\boldsymbol{X}$ 作为结构方程中解释变量 $(\boldsymbol{Y}_0,\boldsymbol{X}_0)$ 的工具变量，只是次序不同。狭义的工具变量法用结构方程中未包含的先决变量 $\boldsymbol{X}_0^{*}$ 作为 $\boldsymbol{Y}_0$ 的工具变量，用结构方程中包含的先决变量 $\boldsymbol{X}_0$ 作为自己的工具变量；而间接最小二乘法则将先决变量 $\boldsymbol{X}$ 按自己的顺序作为 $(\boldsymbol{Y}_0,\boldsymbol{X}_0)$ 的工具变量，这就使得结构方程中包含的先决变量 $\boldsymbol{X}_0$ 也选择了其他先决变量作为工具变量，而不是自身。从前面的课程内容中已经知道，这两种不同的选取只影响正规方程组中方程的次序，并不影响方程组的解。所以狭义的工具变量法和间接最小二乘法的参数估计量是等价的。

比较二阶段最小二乘法和间接最小二乘法的参数估计量(6.4.13)式与(6.4.8)式。间接最小二乘法选取 $\boldsymbol{X}$ 作为结构方程中解释变量 $(\boldsymbol{Y}_0,\boldsymbol{X}_0)$ 的工具变量，二阶段最小二乘法选取 $\boldsymbol{X}$ 的线性组合

$$\hat{\boldsymbol{Y}}_0=\boldsymbol{X}\hat{\boldsymbol{\varPi}}_0=\boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_0]$$

作为结构方程中内生解释变量 $\boldsymbol{Y}_0$ 的工具变量，选取 $\boldsymbol{X}_0$ 作为自己的工具变量。这样使得关于二者参数估计量的正规方程组是不同的，分别为

$$\boldsymbol{X}'\boldsymbol{Y}_1=\left(\boldsymbol{X}'\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\right)\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\varGamma}}_0\end{pmatrix}_{\mathrm{ILS}}$$

$$\begin{pmatrix}\hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0\end{pmatrix}'\boldsymbol{Y}_1=\left(\begin{pmatrix}\hat{\boldsymbol{Y}}_0 & \boldsymbol{X}_0\end{pmatrix}'\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\right)\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\varGamma}}_0\end{pmatrix}_{\mathrm{2SLS}}$$

比较这两个正规方程组发现，后者可以由前者经过初等线性变换得到。而根据代数知识，初等线性变换不影响方程组的解。所以二阶段最小二乘法和间接最小二乘法的参数估计量是等价的。也可以对此进行严格证明。假设

$$\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\Gamma}}_0\end{pmatrix}_{\text{ILS}}=\begin{pmatrix}\hat{\boldsymbol{B}}_0\\ \hat{\boldsymbol{\Gamma}}_0\end{pmatrix}_{\text{2SLS}}$$

即

$$((\hat{\boldsymbol{Y}}_0\quad \boldsymbol{X}_0)'\quad(\boldsymbol{Y}_0\quad \boldsymbol{X}_0))^{-1}(\hat{\boldsymbol{Y}}_0\quad \boldsymbol{X}_0)'=(\boldsymbol{X}'\quad(\boldsymbol{Y}_0\quad \boldsymbol{X}_0))^{-1}\boldsymbol{X}'$$

两边同时左乘$\left(\boldsymbol{X}'\quad(\boldsymbol{Y}_0\quad \boldsymbol{X}_0)\right)$，有

$$\left(\boldsymbol{X}'\quad(\boldsymbol{Y}_0\quad \boldsymbol{X}_0)\right)\left(\left(\hat{\boldsymbol{Y}}_0\quad \boldsymbol{X}_0\right)'\quad\left(\boldsymbol{Y}_0\quad \boldsymbol{X}_0\right)\right)^{-1}\left(\hat{\boldsymbol{Y}}_0\quad \boldsymbol{X}_0\right)'=\boldsymbol{X}'$$

两边同时右乘$\left(\boldsymbol{Y}_0\quad \boldsymbol{X}_0\right)$，有

$$\left(\boldsymbol{X}'\quad(\boldsymbol{Y}_0\quad \boldsymbol{X}_0)\right)=\boldsymbol{X}'\left(\boldsymbol{Y}_0\quad \boldsymbol{X}_0\right)$$

该式显然成立。所以两种参数估计量是等价的假设成立。

结论是，**对于恰好识别的结构方程，狭义的工具变量法、间接最小二乘法和二阶段最小二乘法三种方法是等价的。**

六、简单宏观经济模型实例演示

例 6.4.1

下面建立一个包含 3 个方程的中国宏观经济模型，主要借此进行方法上的演示。

描写包含 3 个内生变量，即国内生产总值Y，居民消费总额C和投资总额I；3 个先决变量，即政府消费G(将净出口也包含其中，为了实现数据的平衡，该数据是按照$Y-C-I$计算出来的)，前期居民消费总额C_{t-1}和常数项。完备的结构式模型为

$$\begin{cases}C_t=\alpha_0+\alpha_1Y_t+\alpha_2C_{t-1}+\mu_{t1}\\ I_t=\beta_0+\beta_1Y_t+\mu_{t2}\\ Y_t=I_t+C_t+G_t\end{cases}\qquad t=1978,1979,\cdots,2007\qquad(6.4.14)$$

容易判断，消费方程是恰好识别的方程，投资方程是过度识别的方程，模型是可以识别的。现在对模型进行估计。样本观测值见表 6.4.1,数据来自《中国统计年鉴》(2008)。

表 6.4.1 中国宏观经济数据 单位：亿元

年份	Y	I	C	G	年份	Y	I	C	G
1978	3 605.6	1 377.9	1 759.1	468.6	1980	4 592.9	1 599.7	2 331.2	662
1979	4 092.6	1 478.9	2 011.5	602.2	1981	5 008.8	1 630.2	2 627.9	750.7

续表

年份	Y	I	C	G	年份	Y	I	C	G
1982	5 590	1 784.2	2 902.9	902.9	1995	63 216.9	25 470.1	28 369.7	9 377.1
1983	6 216.2	2 039	3 231.1	946.1	1996	74 163.6	28 784.9	33 955.9	11 422.8
1984	7 362.7	2 515.1	3 742	1 105.6	1997	81 658.5	29 968	36 921.5	14 769
1985	9 076.7	3 457.5	4 687.4	931.8	1998	86 531.6	31 314.2	39 229.3	15 988.1
1986	10 508.5	3 941.9	5 302.1	1 264.5	1999	91 125	32 951.5	41 920.4	16 253.1
1987	12 277.4	4 462	6 126.1	1 689.3	2000	98 749	34 842.8	45 854.6	18 051.6
1988	15 388.6	5 700.2	7 868.1	1 820.3	2001	108 972.4	39 769.4	49 213.2	19 989.8
1989	17 311.3	6 332.7	8 812.6	2 166	2002	120 350.3	45 565	52 571.3	22 214
1990	19 347.8	6 747	9 450.9	3 149.9	2003	136 398.8	55 963	56 834.4	23 601.4
1991	22 577.4	7 868	10 730.6	3 978.8	2004	160 280.4	69 168.4	63 833.5	27 278.5
1992	27 565.2	10 086.3	13 000.1	4 478.8	2005	188 692.1	80 646.3	71 217.5	36 828.3
1993	36 938.1	15 717.7	16 412.1	4 808.3	2006	221 651.3	94 402	80 476.9	46 772.4
1994	50 217.4	20 341.1	21 844.2	8 032.1	2007	263 242.5	111 417.4	93 317.2	58 507.9

1. 用狭义的工具变量法估计消费方程

选取消费方程中未包含的先决变量G作为内生解释变量Y的工具变量，得到结构参数的工具变量法估计量：

$$\hat{\alpha}_0 = 886.572\,5$$
$$\hat{\alpha}_1 = 0.085\,101$$
$$\hat{\alpha}_2 = 0.861\,701$$

2. 用间接最小二乘法估计消费方程

消费方程中包含的内生变量的简化式方程为

$$\begin{cases} C_t = \pi_{10} + \pi_{11}C_{t-1} + \pi_{12}G_t + \varepsilon_{t1} \\ Y_t = \pi_{20} + \pi_{21}C_{t-1} + \pi_{22}G_t + \varepsilon_{t2} \end{cases}$$

参数关系体系为

$$\begin{cases} \pi_{11} - \alpha_1\pi_{21} - \alpha_2 = 0 \\ \pi_{10} - \alpha_0 - \alpha_1\pi_{20} = 0 \\ \pi_{12} - \alpha_1\pi_{22} = 0 \end{cases}$$

用普通最小二乘法估计简化式方程，得到简化式参数估计量为

$$\hat{\pi}_{10} = 996.450\,7, \qquad \hat{\pi}_{20} = 1\,291.152$$
$$\hat{\pi}_{11} = 0.968\,812, \qquad \hat{\pi}_{21} = 1.258\,635$$
$$\hat{\pi}_{12} = 0.237\,809, \qquad \hat{\pi}_{22} = 2.794\,436$$

由参数关系体系计算得到结构参数间接最小二乘估计值为

$$\hat{\alpha}_1 = \frac{\hat{\pi}_{12}}{\hat{\pi}_{22}} = 0.085\,101$$
$$\hat{\alpha}_2 = \hat{\pi}_{11} - \hat{\alpha}_1\hat{\pi}_{21} = 0.861\,701$$
$$\hat{\alpha}_0 = \hat{\pi}_{10} - \hat{\alpha}_1\hat{\pi}_{20} = 886.572\,4$$

3. 用二阶段最小二乘法估计消费方程

二阶段最小二乘法的第一阶段是用普通最小二乘法估计内生解释变量的简化式方程，得到

$$\hat{Y}_t = 1\,291.152 + 1.258\,635 C_{t-1} + 2.794\,436 G_t$$

据此计算 $\hat{Y}_t$，替换结构方程中的 Y_t，再用普通最小二乘法估计变换了的结构式方程，得到消费方程的二阶段最小二乘参数估计量为

$$\hat{\alpha}_0 = 886.752\,5$$

$$\hat{\alpha}_1 = 0.085\,101$$

$$\hat{\alpha}_2 = 0.861\,701$$

比较上述消费方程的 3 种估计结果，证明这 3 种方法对于恰好识别的结构方程是等价的。估计量的差别只是很小的计算误差。

4. 用二阶段最小二乘法估计投资方程

投资方程是过度识别的结构方程，只能用二阶段最小二乘法估计。估计过程与上述二阶段最小二乘法估计消费方程的过程相同。得到投资方程的参数估计量为

$$\hat{\beta}_0 = -1\,455.814$$

$$\hat{\beta}_1 = 0.419\,774$$

至此，我们完成了该模型系统的估计。该例主要是为了演示估计方法，故未对模型估计中的其他问题进行讨论。

*七、主分量方法

主分量方法本身并不是联立方程计量经济学模型的估计方法，而是配合其他方法(如二阶段最小二乘法)使用于模型的估计过程之中。

1. 主分量方法的提出

二阶段最小二乘法是一种普遍适用的联立方程计量经济学模型的单方程估计方法，但是当它在实际模型估计中被应用时，立刻就会遇到不可逾越的困难。前面已经提过，一般实际应用的联立方程计量经济学模型，都具有相当多的结构方程数目和先决变量数目。假设一个宏观经济模型，包含 30 个先决变量，那么在二阶段最小二乘法的第一阶段——用普通最小二乘法估计简化式方程时，样本容量必须达到 100 左右，才能满足估计的需要，因为每个简化式方程都包含 30 个解释变量。这么大的样本容量实际上是达不到的，如何能够在较少样本的支持下完成对结构方程的二阶段最小二乘法估计？于是提出了主分量法。数学上的主分量方法早就成熟了，克洛伊可（Kloek）和梅内斯（Mennes）于 1960 年提出将它用于计量经济学模型的估计。

2. 主分量方法原理

所谓主分量方法，就是用较少数目的新变量 $\boldsymbol{Z}$ 重新表示原模型中较多数目的先决变量 $\boldsymbol{X}$ 的方法。例如，如果能够找到 5 个左右的新变量表示宏观经济模型中的 30 个先决变量，那么只需要 15 组以上的样本，就可以进行两阶段最小二乘法估计。显然，对充当主分量的变量是有严格要求的，主要有两条：一是它必须是先决变量 $\boldsymbol{X}$ 的线性组合，二是它们之间必须是正交的。前一条是保证主分量对先决变量 $\boldsymbol{X}$ 的代表性；后一条是保证主分量之间不出现共线性。

如何选择主分量？下面介绍一种简单适用的方法。

(1) 用两个主分量表示两个原变量

现在试图用两个主分量 Z_1 和 Z_2 表示两个原变量 X_1 和 X_2。显然 Z_1 和 Z_2 都应该是 X_1 和 X_2 的线性组合，即

$$Z_1 = a_{11}X_1 + a_{12}X_2$$
$$Z_2 = a_{21}X_1 + a_{22}X_2$$

也就是

$$\boldsymbol{Z} = \boldsymbol{XA} \tag{6.4.15}$$

其中

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} X_1 & X_2 \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

问题是如何选择矩阵 $\boldsymbol{A}$？首先选择 $\boldsymbol{a}_1$，它必须使得主分量 Z_1 具有最大的平方长度，即

$$\max(Z_1'Z_1) = \boldsymbol{a}_1'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_1 \tag{6.4.16}$$

并满足

$$\boldsymbol{a}_1'\boldsymbol{a}_1 = 1 \tag{6.4.17}$$

于是问题变为在(6.4.17)式的约束下求(6.4.16)式达到最大时的 $\boldsymbol{a}_1$。构造拉格朗日函数

$$P = \boldsymbol{a}_1'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_1 - \lambda_1(\boldsymbol{a}_1'\boldsymbol{a}_1 - 1)$$

使其对 $\boldsymbol{a}_1$ 的偏导数为 $\boldsymbol{0}$：

$$\frac{\partial P}{\partial \boldsymbol{a}_1} = 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_1 - 2\lambda_1\boldsymbol{a}_1 = \boldsymbol{0}$$

即

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_1 = \lambda_1\boldsymbol{a}_1$$

显然，λ_1 为 $\boldsymbol{X}'\boldsymbol{X}$ 的最大特征值，$\boldsymbol{a}_1$ 为对应的最大特征向量。由此求得了 $\boldsymbol{a}_1$，也就得到了第一个主分量 Z_1。

对于第二个主分量 Z_2，类似地有

$$\max(Z_2'Z_2)=\boldsymbol{a}_2'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_2$$

满足约束

$$\boldsymbol{a}_2'\boldsymbol{a}_2=1$$
$$Z_2'Z_1=0$$

后者是为了保证主分量之间不相关。该约束可进一步简化：

$$\boldsymbol{a}_2'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_1=0$$
$$\boldsymbol{a}_2'\lambda_1\boldsymbol{a}_1=0$$
$$\boldsymbol{a}_2'\boldsymbol{a}_1=0$$

构造拉格朗日函数

$$P=\boldsymbol{a}_2'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_2-\lambda_2(\boldsymbol{a}_2'\boldsymbol{a}_2-1)-\lambda_3(\boldsymbol{a}_2'\boldsymbol{a}_1)$$

使其对 $\boldsymbol{a}_2$ 的偏导数为 $\mathbf{0}$：

$$\frac{\partial P}{\partial \boldsymbol{a}_2}=2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_2-2\lambda_2\boldsymbol{a}_2-\lambda_3\boldsymbol{a}_1=\mathbf{0} \tag{6.4.18}$$

左乘 $\boldsymbol{a}_1'$ 得到

$$2\boldsymbol{a}_1'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_2-2\lambda_2\boldsymbol{a}_1'\boldsymbol{a}_2-\lambda_3\boldsymbol{a}_1'\boldsymbol{a}_1=0$$

式中前两项为 0，所以

$$\lambda_3\boldsymbol{a}_1'\boldsymbol{a}_1=0$$

因为 $\boldsymbol{a}_1'\boldsymbol{a}_1\neq 0$，所以 $\lambda_3=0$。于是(6.4.18)式变为

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{a}_2-\lambda_2\boldsymbol{a}_2=\mathbf{0}$$

λ_2 为 $\boldsymbol{X}'\boldsymbol{X}$ 的另一个最大特征值，$\boldsymbol{a}_2$ 为对应的特征向量。由此求得了 $\boldsymbol{a}_2$，也就得到了第二个主分量 Z_2。

由上可见，选择主分量就是求 $\boldsymbol{X}'\boldsymbol{X}$ 的特征值与特征向量。

(2) 用 k 个主分量表示 k 个原变量

对于 k 个先决变量 $\boldsymbol{X}$，如欲选择 k 个主分量，则只需要求得 $\boldsymbol{X}'\boldsymbol{X}$ 的 k 个特征值和对应的特征向量，于是

$$\boldsymbol{Z}=\boldsymbol{X}\boldsymbol{A}$$

其中

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_k\end{pmatrix}$$

并且有

$$\boldsymbol{Z}'\boldsymbol{Z}=\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}=\boldsymbol{\varLambda}=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k\end{pmatrix}$$

PDG

(3) 用 f 个主分量表示 k 个原变量

应用主分量的目的就是减少变量数目，所以一般要求用 f 个主分量表示 k 个原变量，那么其对应的 $\boldsymbol{A}=\begin{pmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_f\end{pmatrix}$ 应是 $\boldsymbol{X'X}$ 的前 f 个大的特征值所对应的特征向量。

(4) 在二阶段最小二乘法中主分量的选取

在应用二阶段最小二乘法估计联立方程计量经济学模型(6.4.1)式中的第 1 个结构方程(6.4.3)式

$$\boldsymbol{Y}_1=\begin{pmatrix}\boldsymbol{Y}_0 & \boldsymbol{X}_0\end{pmatrix}\begin{pmatrix}\boldsymbol{B}_0\\ \boldsymbol{\Gamma}_0\end{pmatrix}+\boldsymbol{N}_1$$

时，对于简化式方程

$$\boldsymbol{Y}_0=\boldsymbol{X\Pi}_0+\boldsymbol{E}_0=\begin{pmatrix}\boldsymbol{X}_0 & \boldsymbol{X}_0^*\end{pmatrix}\boldsymbol{\Pi}_0+\boldsymbol{E}_0 \tag{6.4.19}$$

来说，一般情况下，结构方程包含的先决解释变量 $\boldsymbol{X}_0$ 中变量的数目很有限，变量主要集中在结构方程未包含的先决变量 $\boldsymbol{X}_0^*$ 中。所以只需要选择主分量重新表示 $\boldsymbol{X}_0^*$，就可以有效地减少简化式方程中解释变量的数目，使得在有限样本的支持下模型得到估计。

如果选择 $\boldsymbol{X}_0^*$ 的 f 个主分量 $\boldsymbol{F}$，那么对新的简化式方程

$$\boldsymbol{Y}_0=\begin{pmatrix}\boldsymbol{X}_0 & \boldsymbol{F}\end{pmatrix}\boldsymbol{\Pi}_0^*+\boldsymbol{E}_0 \tag{6.4.20}$$

应用普通最小二乘法，得到参数估计量 $\hat{\boldsymbol{\Pi}}_0^*$，进而得到 $\hat{\boldsymbol{Y}}_0$，完成了第一阶段的估计任务。如果仍然希望得到原简化式模型参数的估计量 $\hat{\boldsymbol{\Pi}}_0$，可以将主分量 $\boldsymbol{F}$ 用 $\boldsymbol{X}_0^*$ 的线性组合代入(6.4.20)式，然后通过与(6.4.19)式对比，确定模型(6.4.19)式的参数的估计量 $\hat{\boldsymbol{\Pi}}_0$。

*八、*k* 级估计式

1．*k* 级估计式

k 级估计式本身不是一种估计方法，而是对各种方法得到的估计式的概括。对于联立方程计量经济学模型(6.4.1)式中的第 1 个结构方程

$$\boldsymbol{Y}_1=(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0)\begin{pmatrix}\boldsymbol{B}_0\\ \boldsymbol{\Gamma}_0\end{pmatrix}+\boldsymbol{N}_1$$

如果应用普通最小二乘法(这里讲的是“如果”，至于能否应用，后面将作出解释)进行估计，实际上等价于用($\boldsymbol{Y}_0 \quad \boldsymbol{X}_0$)作为自己工具变量，得到的参数估计式为

$$\begin{pmatrix} \hat{\boldsymbol{B}}_0 \\ \hat{\boldsymbol{\Gamma}}_0 \end{pmatrix}_{\text{OLS}} = ((\boldsymbol{Y}_0 \quad \boldsymbol{X}_0)'(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0))^{-1}(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0)'\boldsymbol{Y}_1$$

如果应用工具变量法、间接最小二乘法、二阶段最小二乘法进行估计，当结构方程是恰好识别时，估计结果是等价的；当结构方程是过度识别时，只能应用二阶段最小二乘法进行估计。所以以二阶段最小二乘法的估计式代表这 3 种经典方法，其估计式为

$$\begin{pmatrix} \hat{\boldsymbol{B}}_0 \\ \hat{\boldsymbol{\Gamma}}_0 \end{pmatrix}_{\text{2SLS}} = ((\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)'(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0))^{-1}(\hat{\boldsymbol{Y}}_0 \quad \boldsymbol{X}_0)'\boldsymbol{Y}_1$$

如果把参数估计式写成下列形式：

$$\begin{pmatrix} \hat{\boldsymbol{B}}_0 \\ \hat{\boldsymbol{\Gamma}}_0 \end{pmatrix} = ((\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0) \quad \boldsymbol{X}_0)'(\boldsymbol{Y}_0 \quad \boldsymbol{X}_0))^{-1}(\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0) \quad \boldsymbol{X}_0)'\boldsymbol{Y}_1 \tag{6.4.21}$$

显然，当

k=0 时，即为普通最小二乘法的估计式；

k=1 时，即为二阶段最小二乘法的估计式；

k 等于其他有限信息估计方法中的 λ 时，即为有限信息估计式。

故把(6.4.21)式称为 k 级估计式。

2. *k* 级估计式的性质

假设工具变量与随机干扰项不相关，即

$$\mathrm{P}\lim\left\{\frac{1}{n}\left[\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0)\right]\boldsymbol{N}_1\right\} = 0 \tag{6.4.22}$$

且先决变量与随机干扰项不相关，即

$$\mathrm{P}\lim\left[\frac{1}{n}(\boldsymbol{X}_0'\boldsymbol{N}_1)\right] = 0 \tag{6.4.23}$$

那么，容易证明 k 级估计式是一致性估计式。

但是，(6.4.22)式对 k 是有限制的，因为

$$\begin{aligned}
&\mathrm{P}\lim\left\{\frac{1}{n}\left[\boldsymbol{Y}_0 + k(\hat{\boldsymbol{Y}}_0 - \boldsymbol{Y}_0)\right]\boldsymbol{N}_1\right\} \\
&= \mathrm{P}\lim(1-k)\cdot P\lim\left[\frac{1}{n}(\boldsymbol{Y}_0'\boldsymbol{N}_1)\right] + \mathrm{P}\lim\ k\cdot \mathrm{P}\lim\left[\frac{1}{n}(\hat{\boldsymbol{Y}}_0'\boldsymbol{N}_1)\right] \\
&= \mathrm{P}\lim(1-k)\cdot \mathrm{P}\lim\left[\frac{1}{n}(\boldsymbol{Y}_0'\boldsymbol{N}_1)\right] + \\
&\quad \mathrm{P}\lim\ k\cdot \mathrm{P}\lim\left[\boldsymbol{Y}_0'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]\cdot \mathrm{P}\lim\left[\frac{1}{n}(\boldsymbol{X}'\boldsymbol{N}_1)\right] \\
&= \mathrm{P}\lim(1-k)\cdot \mathrm{P}\lim\left[\frac{1}{n}(\boldsymbol{Y}_0'\boldsymbol{N}_1)\right]
\end{aligned}$$

由于

$$\mathrm{P}\lim\left[\frac{1}{n}(\boldsymbol{Y}_0{}'\boldsymbol{N}_1)\right]\neq 0$$

所以，欲使(6.4.22)式成立，必须有

$$\mathrm{P}\lim(1-k)=0$$

也就是说，只有在二阶段最小二乘法或其他有限信息估计方法中，k 级估计式是一致性估计式，而在普通最小二乘法中，不具有一致性。

§6.5 联立方程计量经济学模型若干问题的讨论

联立方程计量经济学模型估计方法比较多，同样一个模型，可以选择几种不同的方法进行估计。估计方法的选择成为一个重要工作。在实际应用中，人们更多的是采用普通最小二乘法估计联立方程计量经济学模型，也需要给予合理的解释。既然是联立方程计量经济学模型，除了单方程计量经济学模型的检验方法仍然有效外，方程系统的性质也需要进行检验。这些是本节要专门讨论的问题。

一、估计方法的比较

选择模型估计方法，除了一些“硬约束”(例如，不可识别的模型不能估计，因此任何估计方法都不适用，过度识别的模型不能选择间接最小二乘法和工具变量法进行估计等)之外，有较大的灵活性。不同的估计方法，其参数估计量有不同的统计性质；模型的应用目的不同，对参数估计量的统计性质有不同的要求。二者的统一性，是选择估计方法最重要的准则。那么，对不同估计方法的统计特性进行比较研究显得尤其重要。

1. 大样本估计特性的比较

在大样本的情况下，各种参数估计方法的统计特性可以从数学上进行严格的证明，因而也可以将各种方法按照各个性质比较优劣。

(1) 按渐近无偏性比较优劣

除了普通最小二乘法方法外，所有方法的参数估计量都具有大样本下渐近无偏性。因而，除了普通最小二乘法方法最差外，其他方法无法比较优劣。

(2) 按渐近有效性比较优劣

按渐近有效性比较优劣，各种方法的排序如下：

① 普通最小二乘法　非一致性估计，未利用任何单方程外的信息；

② 工具变量法　利用了模型系统部分先决变量的数据信息；

③ 二阶段最小二乘法、有限信息最大似然法　利用了模型系统全部先决变量的数据信息；

④ 三阶段最小二乘法、完全信息最大似然法利用了模型系统全部先决变量的数据信息和结构方程相关性信息。

一般情况下，三阶段最小二乘法与完全信息最大似然法具有相同的渐近有效性。但是，在特殊情况下，例如，如果在开始估计之前已经知道方程系统随机干扰项的方差、协方差信息，完全信息最大似然法就可以充分利用这些信息，因而比三阶段最小二乘法更有效。

2. 小样本估计特性的 Monte Carlo 试验

参数估计量的大样本特性只是理论上的，实际上并没有“大样本”，所以对小样本估计特性进行比较更有实际意义。而在小样本的情况下，各种参数估计方法的统计特性无法从数学上进行严格的证明，因而提出了一种 Monte Carlo 试验方法。

(1) 小样本估计特性的 Monte Carlo 试验过程

Monte Carlo 试验按照下列过程进行。

第一步：利用随机数发生器产生随机干扰项分布的一组样本。

第二步：代入已经知道结构参数和先决变量观测值的结构模型中。

第三步：计算内生变量的样本观测值。

第四步：选用各种估计方法估计模型的结构参数。

上述步骤反复进行数百次，得到每一种估计方法的参数估计值的序列。

第五步：对每种估计方法的参数估计值序列进行统计分析。

第六步：与真实参数(即试验前已经知道的结构参数)进行比较，以判断各种估计方法的优劣。

(2) 小样本估计特性比较

将上述试验结果按照无偏性、有效性(即最小方差性)和最小均方差性对各种估计方法排序，结果如下：

① 无偏性

OLS	2SLS	3SLS LIML FIML

$$\longrightarrow$$

② 最小方差性

LIML	2SLS	FIML	OLS

$$\longrightarrow$$

③ 最小均方差性

OLS	LIML	2SLS	3SLS FIML

$$\longrightarrow$$

这里的方差按下式计算：

$$\mathrm{Var}=\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_i-\bar{\hat{\beta}})^2 \tag{6.5.1}$$

其中 N 为试验次数，$\hat{\beta}_i$ 为采用某种估计方法得到的第 i 次估计值，$\bar{\hat{\beta}}$ 为 N 次试验的平均值。均方差按下式计算：

$$\mathrm{MSE}=\mathrm{E}(\hat{\beta}-\beta)^2=\frac{1}{n}\sum_{i=1}^{N}(\hat{\beta}_i-\beta)^2 \tag{6.5.2}$$

注意，(6.5.1)式与(6.5.2)式是有原则区别的。前者是估计量的方差，反映估计量偏离均值的程度；后者反映估计量偏离真实值的程度。所以尽管普通最小二乘法具有最小方差性，但是由于它是有偏的，偏离真实值最为严重，所以它的最小均方差性仍然是最差的。

二、为什么普通最小二乘法被普遍采用

已有的关于联立方程计量经济学模型的研究成果，包括一些著名的、得到很好应用的模型，相当多地采用普通最小二乘法进行其参数估计。为什么？

1．小样本特性

上面已经讲到，在大样本的情况下，联立方程计量经济学模型各种参数估计方法的统计特性可以从数学上进行严格的证明，因而也有明确的优劣比较。但是参数估计量的大样本特性只是理论上的，实际上并没有“大样本”。而在小样本的情况下，各种参数估计方法的统计特性无法从数学上进行严格的证明，虽然通过 Monte Carlo 试验方法，以及其他统计分析方法，也可以得出普通最小二乘法并不是一种较好估计方法的一般性结论，但毕竟只是一个经验性结论。通过 Monte Carlo 试验方法，得到的反证也是有的。例如，米凯尔（Mikhail）于 1975 年在一些例子中，得出二阶段最小二乘法比三阶段最小二乘法更为有效的试验结果，与一般性结论不符合。所以，从理论上讲，在小样本情况下，各种估计方法的估计量都是有偏的。

2．充分利用样本数据信息

虽然除普通最小二乘法之外的其他估计方法可以部分或者全部地利用某个结构方程中未包含的先决变量的数据信息，从而提高参数估计量的统计性质。但是，在实际上它所付出的代价经常是牺牲了该方程所包含的变量的样本数据信息。为什么？以一个包含 10 个内生变量和 10 个方程的宏观经济模型为例。样本观测值只能来自统计资料，但由于历史的原因，生产方程可以得到 40 组样本观测值，而投资方程只能得到 20 组样本观测值。如果采用普通最小二乘法估计生产方程，就可以充分利用 40 组样本观测值的数据信息；如果采用普通最小

二乘法以外的方法估计生产方程，如二阶段最小二乘法，就只能利用生产方程20组样本观测值的数据信息，因为在它的第一阶段，必须按照模型系统所有先决变量中样本观测值最少的为标准确定样本容量。当然换取的是利用了生产方程中未包含的先决变量的20组样本观测值信息。这一得一失，很难判断得是否大于失。

3．确定性误差传递

所谓确定性误差，主要是结构方程的关系误差和外生变量的观测误差。对于普通最小二乘法，当估计某一个结构方程时，方程中没有包含的外生变量的观测误差和其他结构方程的关系误差对该方程的估计结果没有影响。如果采用二阶段最小二乘法，当估计某一个结构方程时，方程中没有包含的外生变量的观测误差对该方程的估计结果将生产影响，原因是在估计的第一阶段利用了所有外生变量的观测值信息。如果采用三阶段最小二乘法，方程中没有包含的外生变量的观测误差和其他结构方程的关系误差都将对每个结构方程的参数估计结果产生影响，原因是在估计的第一阶段利用了所有外生变量的观测值信息，在第三阶段利用了所有结构方程的关系信息。

4．样本容量不支持

由于实际的联立方程计量经济学模型中每个结构方程往往是过度识别的，适宜采用二阶段最小二乘法或三阶段最小二乘法，但是在其第一阶段要以所有先决变量作为解释变量，这就需要很大容量的样本。实际上是难以实现的。采用主分量方法等可以克服这个矛盾，但又带来方法的复杂性和新的误差。

5．实际模型的递推结构

联立方程计量经济学模型的主要应用领域是宏观经济模型，而宏观经济模型经常带有明显的递推结构。对于递推结构的联立方程计量经济学模型，是可以依次对每个结构方程采用普通最小二乘法的。

以上原因可以解释为什么相当多地采用普通最小二乘法进行联立方程计量经济学模型的参数估计。但是，必须承认，从理论上讲，普通最小二乘法用于联立方程计量经济学模型是不合适的；在实际上，应该尽可能创造条件，采用其他方法。事实上，进入20世纪70年代后，由于计算机技术的发展，以及数据统计技术的完善，普通最小二乘法已经不那么经常采用了。

三、联立方程计量经济学模型的检验

与单方程计量经济学模型一样，联立方程计量经济学模型在完成估计之后，也要进行检验。包括单方程检验和方程系统的检验。

凡是在单方程计量经济学模型中必须进行的各项检验，对于联立方程计量经济学模型中的结构方程，以及应用二阶段最小二乘法或三阶段最小二乘法过程中的简化式方程，都是适用的和需要的。在此不再重复。下面着重介绍模型系统的检验。

1. 拟合效果检验

对于联立方程计量经济学模型

$$\boldsymbol{YB}+\boldsymbol{X\Gamma}=\boldsymbol{N} \tag{6.5.3}$$

当结构参数估计量已经得到，并通过了对单个方程的检验之后，有

$$\hat{\boldsymbol{Y}}\hat{\boldsymbol{B}}+\boldsymbol{X}\hat{\boldsymbol{\Gamma}}=\boldsymbol{0} \tag{6.5.4}$$

将样本期的先决变量观测值代入(6.5.4)式，求解该方程组，即可得到内生变量的估计值$\hat{\boldsymbol{Y}}$。将估计值与实际观测值进行比较，据此判断模型系统的拟合效果。

如何求解方程组(6.5.4)？模型系统虽然是线性系统，但并不排除(6.5.4)式中存在非线性方程。这些方程所表现的变量之间的直接关系是非线性关系，但经过某种变换后以线性形式出现在模型中，例如用 Cobb–Dauglas 生产函数表示的生产方程。所以，对给定$\boldsymbol{X}$的值，求解内生变量的估计值$\hat{\boldsymbol{Y}}$常用的方法是迭代法。

常用的判断模型系统拟合效果的检验统计量是“均方百分比误差”，用 RMS 表示。其计算方法为

$$\mathrm{RMS}_i=\sqrt{\sum_{t=1}^{n}\frac{e_{it}^2}{n}}$$

$$e_{it}=\frac{y_{it}-\hat{y}_{it}}{y_{it}}$$

其中，RMS_i为第i个内生变量的“均方百分比误差”，n为样本容量。一般认为，在各种检验统计量中，RMS 具有更普遍的意义，对检验模型系统的总体拟合优度更为有效。

显然，当$\mathrm{RMS}_i=0$，表示第i个内生变量估计值与观测值完全拟合。一般地，在g个内生变量中，RMS<5%的变量数目占 70%以上，并且每个变量的 RMS 不大于 10%，则认为模型系统总体拟合效果较好。

2. 预测性能检验

建立联立方程计量经济学模型，一般要花费较长的时间，当模型建成后，样本期之后的时间截面上的内生变量实际观测值已经知道，这就有条件对模型系统进行预测检验。将该时间截面上的先决变量实际观测值代入模型，计算所有内生变量预测值，并计算其相对误差

$$\mathrm{RE}=\frac{y_{i0}-\hat{y}_{i0}}{y_{i0}}, \qquad i=1,2,\cdots,g$$

其中$y_{i0},\hat{y}_{i0}$分别为第i个内生变量的观测值与预测值，g为模型中内生变量数目。

同样，也没有绝对的标准。一般认为，$\mathrm{RE}<5\%$的变量数目占 70%以上，并且每个变量的相对误差不大于 10%，则认为模型系统总体预测性能较好。

有人会因此提出责难，认为这个标准太低了。例如，用模型来预测粮食产量，一般讲年实际增长率不会超过5%，而模型的预测误差允许达到5%，这样的预测还有什么意义呢？对于这类情况，应该在建立模型时加以考虑，如建立增量模型而不是总量模型。

3．方程间误差传递检验

由于联立方程计量经济学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。

如果关键路径上的方程数目为T,e_i为第i个方程的随机误差估计值，下列三个统计量都可以用来衡量关键路径上的误差水平。它们是

$$误差均值=\frac{1}{T}\sum_{i=1}^{T}e_i$$

$$均方根误差=\sqrt{\frac{1}{T}\sum_{i=1}^{T}e_i^2}$$

$$冯诺曼比=\frac{\sum_{i=2}^{T}(e_i-e_{i-1})^2}{\sum_{i=1}^{T}e_i^2}\frac{T}{T-1}$$

误差均值应用较少，因为存在正负相抵的问题，均方根误差和冯诺曼比应用较多，显然是越小越好。其中又以冯诺曼比对误差传递程度的检验功能最强，如果误差在方程之间没有传递，该比值为0。

4．样本点间误差传递检验

上述几种检验中构造的检验统计量都是在同一时间截面上计算其数值。在联立方程计量经济学模型系统中，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。由于存在滞后内生变量，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。

如果样本期为 $t=1,2,\cdots,n$，对于模型(6.5.4)式，给定 $t=1$ 时的所有先决变量的观测值，包括滞后内生变量，求解方程组，得到内生变量的预测值 $\hat{Y}_1$；对于 $t=2$，只外生给定外生变量的观测值，滞后内生变量则以前一时期的预测值代替，求解方程组，得到内生变量的预测值 $\hat{Y}_2$；如此逐年滚动预测，直至得到 $t=n$ 时的内生变量的预测值 $\hat{Y}_n$，并求出该滚动预测值与实际观测值的相对误差。另外，将 $t=n$ 时的所有先决变量的观测值，包括滞后内生变量的实际观测值，代入模型，求解方程组，得到内生变量的非滚动预测值 $\hat{Y}_n'$，并求出该非滚动预测值与实际观测值的相对误差。比较两种结果，二者的差异表明模型预测误差在不同的时间截面之间的传递。

从上述检验过程可以看出，滚动预测检验是较为严格有效的检验。

本章练习题

1．为什么要建立联立方程计量经济学模型？联立方程计量经济学模型适用于什么样的经济现象？

2．联立方程计量经济学模型的识别状况可以分为几类？其含义各是什么？

3．联立方程计量经济学模型的单方程估计有哪些主要的方法？其适用条件和统计性质各是什么？

4．一个有 2 个方程构成的简单商品供求模型如下：

$$供给方程：Q_t = \alpha_0 + \alpha_1 P_t + \mu_{t1}$$

$$需求方程：Q_t = \beta_0 + \beta_1 P_t + \mu_{t2}$$

其中，P 为均衡价格，Q_t 是供求平衡状态下的供给量或需求量。试从模型简化式与结构式关系体系回答下列问题：

(1) 该模型两个方程是否可识别？

(2) 如果对该模型需求函数增加消费者收入变量 Y_t，则两方程的识别状态有何变化？

(3) 如果再在上述模型的供给方程中引入新变量上期商品价格 P_{t-1}，则两方程的识别状态有何变化？

(4) 如果在需求函数中继续引入表示消费者财富的变量 W_t，则两方程的识别状态又有何变化？

5．对习题 4 联立模型的每种情况，按结构式识别条件进行识别。

6．某联立方程计量经济学模型有 3 个方程，3 个内生变量(Y_1，Y_2，Y_3)，3 个外生变量(X_1，X_2，X_3)和样本观测值始终为 1 的虚变量 C，样本容量为 n。其中第 2 个方程

$$Y_2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_3 + \alpha_3 X_3 + \mu_2$$

为恰好识别的结构方程。

(1) 写出用工具变量法估计该方程参数的正规方程组；

(2) 用间接最小二乘法估计该方程参数，也可以看成一种工具变量法，指出工具变量是如何选取的，并写出参数估计量的矩阵表达式；

(3) 用二阶段最小二乘法估计该方程参数，也可以看成一种工具变量法，指出 Y_3 的工具变量是什么，并写出参数估计量的矩阵表达式。

7. 下列为一个完备的联立方程计量经济学模型：

$$Y_t = \beta_0 + \beta_1 M_t + \gamma_1 C_t + \gamma_2 I_t + u_{t1}$$
$$M_t = \alpha_0 + \alpha_1 Y_t + \gamma_3 P_t + u_{t2}$$

其中，M 为货币供给量，Y 为国内生产总值，P 为价格总指数。C，I 分别为居民消费与投资。

(1) 指出模型的内生变量、外生变量、先决变量；

(2) 写出简化式模型，并导出结构式参数与简化式参数之间的关系；

(3) 用结构式条件确定模型的识别状态；

(4) 指出间接最小二乘法、工具变量法、二阶段最小二乘法中哪些可用于原模型第 1，2 个方程的参数估计。

8. 继续题 7，以如下中国的实际数据为资料，估计上述联立模型。要求恰好识别的方程按工具变量法与二阶段最小二乘法估计。

年份	货币与准货币 M_2 (亿元)	国内生产总值 GDP (亿元)	居民消费价格指数 P (1978=100)	居民消费 CONS (亿元)	固定资产投资 I (亿元)
1990	15 293.4	19 347.8	165.2	9 450.9	4 517
1991	19 349.9	22 577.4	170.8	10 730.6	5 594.5
1992	25 402.2	27 565.2	181.7	13 000.1	8 080.1
1993	34 879.8	36 938.1	208.5	16 412.1	13 072.3
1994	46 923.5	50 217.4	258.7	21 844.2	17 042.1
1995	60 750.5	63 216.9	302.9	28 369.7	20 019.3
1996	76 094.9	74 163.6	328.1	33 955.9	22 913.5
1997	90 995.3	81 658.5	337.3	36 921.5	24 941.1
1998	104 498.5	86 531.6	334.6	39 229.3	28 406.2
1999	119 897.9	91 125	329.9	41 920.4	29 854.7
2000	134 610.4	98 749	331.2	45 854.6	32 917.7
2001	158 301.9	108 972.4	333.5	49 213.2	37 213.5
2002	185 007	120 350.3	330.9	52 571.3	43 499.9
2003	221 222.8	136 398.8	334.8	56 834.4	55 566.6
2004	254 107	160 280.4	347.9	63 833.5	70 477.4
2005	298 755.7	188 692.1	354.2	71 217.5	88 773.6
2006	345 603.6	221 651.3	359.5	80 476.9	109 998.2
2007	403 442.2	263 242.5	376.7	93 317.2	137 323.9

第七章　扩展的单方程计量经济学模型

7

在第二章至第五章中讨论了经典的单方程计量经济学模型理论与方法。在这些章节中，讨论限于常参数、线性、揭示变量之间因果关系的单方程模型；被解释变量是连续的随机变量，其抽样是随机和不受限制的；在模型估计过程中或者只利用时间序列样本，或者只利用截面数据样本；主要依靠对经济理论和行为规律的理解确定模型的结构形式。在本章中，将讨论几种扩展模型，**主要包括将被解释变量抽样由完全随机扩展为受到限制的选择性样本模型，将被解释变量由连续的扩展为离散的离散选择模型，将单一种类的样本扩展为同时包含截面数据和时间序列数据的平行数据(panel data)模型等。**这些模型与方法，无论在计量经济学理论方面还是在实际应用方面，都具有重要意义。但是，这些模型都形成了各自丰富的内容体系，甚至是计量经济学的新分支学科，模型方法的数学过程较为复杂。这里只介绍其中最简单的模型，以了解这些模型理论与方法的概念与思路。

§7.1　选择性样本计量经济学模型

受限被解释变量(limited dependent variable)指被解释变量的观测值是连续的，但是受到某种限制，其抽样并非完全随机的，得到的观测值并不完全反映被解释变量的实际状态。选择性样本(selective sample)是受限被解释变量的主要形式，其样本观测值是在某种选择性限制的情况下抽取的。利用这样的样本观测值估计总体的参数，就是选择性样本计量经济学模型(selective sample model)要解决的问题。

选择性样本计量经济学模型的理论方法发展于 20 世纪 70 年代，赫克曼(James J. Heckman)做出了基础性的贡献，并因此获得了 2000 年诺贝尔经济学奖。近 10 年来，选择性样本模型已经成为微观计量经济学的主要内容之一，并且得到广泛的应用，特别是在劳动经济学、卫生经济学以及其他社会经济学领域。

一、经济生活中的选择性样本问题

经济生活中的选择性样本问题主要表现于以下常见的两类。

一类是“截断”(truncation)问题，即“掐头”或者“去尾”，也即不能从全部个体，而只能从一部分个体中随机抽取被解释变量的样本观测值，而这部分个体的观测值都大于或者小于某个确定值。例如，以居民收入为被解释变量建立居民收入模型。从理论上讲，居民收入样本数据应该从零到无穷大，但是由于客观条件所限，只能在收入处于某一数值以上或者某一数值以下的个体中取得样本观测值。再如，为了研究农民年贷款额的影响因素，对农村居民户进行随机抽调查。如果调查了 10 000 户，其中只有 6 000 户在一年内发生了贷款。如果仅以发生了贷款的 6 000 户的贷款额作为被解释变量观测值，显然是将其他没有发生贷款的 4 000 户“截断”掉了。从这样的样本数据出发，如果采用经典的方法估计总体回归模型的参数，显然是不合适的。

一类是“归并”(censoring)问题，即将被解释变量的处于某一范围的样本观测值都用一个相同的值代替。这类问题经常出现在“检查”、“调查”活动中，因此也称为“检查”问题，也被翻译为“删失”问题。例如，以居民对某种商品的需求量为被解释变量，建立需求函数模型。需求量的观测值是无法得到的，一般用实际购买量作为需求量的观测值。如果这种商品是限量购买的，正像我国过去长时期内所实行的那样，比如每户最多只能购买 100，那么得到的观测值将处于 0 与 100 之间，而且会有相当比例的观测值为 100。对于购买量小于 100 的个体，有理由认为这个购买量代表了他的需求量；但是对于购买量等于 100 的个体，他的需求量很可能是大于 100，所以这个购买量并不代表他的需求量。也就是说，凡是实际需求量大于 100 的，都用 100 作为样本观测值，等于是将大于 100 的观测值作了“归并”。再如，在上述的农民年贷款额影响因素分析模型中，如果以全部 10 000 户为样本，将其中没有发生贷款的 4 000 户的样本观测值设为 0，就发生了“归并”，是将贷款额小于等于 0 的值全部“归并”到 0。这类问题在微观经济活动调查中普遍存在。从这样的样本数据出发，如果采用经典的方法估计总体回归模型，显然也是不合适的。

二、“截断”问题的计量经济学模型

如果一个单方程计量经济学模型，只能从“掐头”或者“去尾”的连续区间随机抽取被解释变量的样本观测值，那么很显然，抽取每个样本观测值的概率以及抽取一组样本观测值的联合概率，与被解释变量的样本观测值不受限制的情况是不同的。如果能够知道在这种情况下抽取一组样本观测值的联合概率函数，那么就可以通过该函数极大化求得模型的参数估计量。这就是估计这类计量经济学模型的基本思路。

1. 截断分布

所谓“截断分布”，是完整分布的一部分，指“截断随机变量”的分布。

如果一个连续随机变量ξ的概率密度函数为$f(\xi)$，a为该随机变量分布范围内的一个常数，那么有

$$f(\xi|\xi>a)=\frac{f(\xi)}{P(\xi>a)} \tag{7.1.1}$$

这是由条件概率的定义导出的。

例如，如果ξ服从均匀分布$U(a,b)$，但是它只能在(c,b)内取得样本观测值，那么取得每个样本观测值的概率为

$$f(\xi|\xi>c)=\frac{f(\xi)}{P(\xi>c)}=\frac{\frac{1}{b-a}}{\int_c^b \frac{1}{b-a}\mathrm{d}\xi}=\frac{1}{b-c}$$

请读者注意，原来均匀分布随机变量的概率密度函数是$\frac{1}{b-a}$，而“截断随机变量”的概率密度函数是$\frac{1}{b-c}$，即在(c,b)内取得样本观测值的概率大于在(a,b)内取得样本观测值的概率。**这是截断问题的关键之处。**

如果ξ服从正态分布$N(\mu,\sigma)$，但是它只能在大于常数a的范围内取得样本观测值，那么取得每个样本观测值的概率为

$$\begin{aligned}f(\xi|\xi>a)&=\frac{f(\xi)}{P(\xi>a)}\\&=\frac{(2\pi\sigma^2)^{-1/2}\mathrm{e}^{-(\xi-\mu)^2/(2\sigma^2)}}{1-\Phi(\alpha)}\\&=\frac{\frac{1}{\sigma}\phi\left(\frac{\xi-\mu}{\sigma}\right)}{1-\Phi(\alpha)}\end{aligned} \tag{7.1.2}$$

其中，$\alpha=\frac{a-\mu}{\sigma}$，$\phi(\cdot)$是标准正态分布概率密度函数，$\Phi(\cdot)$是标准正态分布条件概率函数。显然，

$$P(\xi>a)=1-\Phi\left(\frac{a-\mu}{\sigma}\right)=1-\Phi(\bullet)$$

2．截断被解释变量数据计量经济学模型的最大似然估计

如果已经知道截断被解释变量的概率密度函数，自然会想到，可以采用最大似然法估计模型。对于模型

$$Y_i=\boldsymbol{X}_i\boldsymbol{\beta}+\mu_i \qquad \mu_i\sim N(0,\sigma^2) \tag{7.1.3}$$

有

$$Y_i|\boldsymbol{X}_i\sim N(\boldsymbol{X}_i\boldsymbol{\beta},\sigma^2)$$

如果Y_i只能在大于a的范围内取得观测值，从(7.1.2)式可以得到Y_i的概率密度函数为

$$f(Y_i)=\frac{\frac{1}{\sigma}\phi((Y_i-\boldsymbol{X}_i\boldsymbol{\beta})/\sigma)}{1-\Phi((a-\boldsymbol{X}_i\boldsymbol{\beta})/\sigma)}$$

于是(7.1.3)式的对数似然函数为

$$\begin{aligned}\ln L=-\frac{n}{2}(\ln(2\pi)+\ln\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{2}(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2\\-\sum_{i=1}^{n}\ln\left(1-\Phi\left(\frac{a-\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right)\right)\end{aligned}\tag{7.1.4}$$

该对数似然函数的极大化条件为

$$\frac{\partial\ln L}{\partial\begin{pmatrix}\boldsymbol{\beta}\\\sigma^2\end{pmatrix}}=\sum_{i=1}^{n}\begin{pmatrix}\left(\frac{Y_i-\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma^2}-\frac{\lambda_i}{\sigma}\right)\boldsymbol{X}_i\\-\frac{1}{2\sigma^2}+\frac{(Y_i-\boldsymbol{X}_i\boldsymbol{\beta})^2}{2\sigma^4}-\frac{\alpha_i\lambda_i}{2\sigma^2}\end{pmatrix}=\sum_{i=1}^{n}\boldsymbol{g}_i=\boldsymbol{0}\tag{7.1.5}$$

其中

$$\alpha_i=\left(a-\boldsymbol{X}_i\boldsymbol{\beta}\right)/\sigma$$
$$\lambda_i=\phi(\alpha_i)/(1-\Phi(\alpha_i))$$

求解(7.1.5)即可得到模型的参数估计量。当然，由于这是一个复杂的非线性问题，需要采用迭代方法求解(7.1.5)，例如牛顿法。当然，利用计量经济学软件可以很方便地实现模型的估计。

如果对模型进行再参数化，可以使得估计过程较为简单。以$a=0$为例，令$\boldsymbol{\gamma}=\boldsymbol{\beta}/\sigma$且$\theta=1/\sigma$，得到

$$\begin{aligned}\ln L=-\frac{n}{2}\ln(2\pi)+n\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\theta Y_i-\boldsymbol{X}_i\boldsymbol{\gamma})^2\\-\sum_{i=1}^{n}\ln\left(\Phi(\boldsymbol{X}_i\boldsymbol{\gamma})\right)\end{aligned}$$

这里利用了$1-\Phi(-t)=\Phi(t)$。对该对数似然函数极大化，求得到$\boldsymbol{\gamma}$和θ的估计量后再利用$\sigma=1/\theta$和$\boldsymbol{\beta}=\boldsymbol{\gamma}/\theta$求得原参数估计量。

以上只是介绍了方法的思路，了解这个思路，就可以在应用研究中正确地建立和估计模型。如果读者需要深入理解截断被解释变量数据计量经济学模型的理论方法，还需要参考其他计量经济学高级教科书。

下面简单介绍为什么不能采用普通最小二乘法估计截断被解释变量数据模型，可以不作为教学内容，供有兴趣的同学自学。

PDG

3．为什么截断被解释变量数据计量经济学模型不能采用普通最小二乘估计

对于截断被解释变量数据计量经济学模型，如果仍然把它看作为经典的线性模型，采用普通最小二乘法估计(7.1.3)式，会产生什么样的结果呢？

因为 Y_i 只能在大于 a 的范围内取得观测值，那么 Y_i 的条件均值为

$$\begin{aligned}\mathrm{E}(Y_i|Y_i>a)&=\int_a^{\infty}y_i\phi(Y_i|Y_i>a)\mathrm{d}Y_i\\&=\boldsymbol{X}_i\boldsymbol{\beta}+\sigma\frac{\phi((a-\boldsymbol{X}_i\boldsymbol{\beta})/\sigma)}{1-\Phi((a-\boldsymbol{X}_i\boldsymbol{\beta})/\sigma)}\end{aligned}\tag{7.1.6}$$

(7.1.6)式所示的条件均值是解释变量和待估参数的非线性函数。将(7.1.6)式记为

$$\mathrm{E}(Y_i|Y_i>a)=\boldsymbol{X}_i\boldsymbol{\beta}+\sigma\lambda(\alpha_i)\tag{7.1.7}$$

其中 $\alpha_i=\dfrac{\alpha-\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}$，于是有

$$\begin{aligned}\frac{\partial\mathrm{E}(Y_i|Y_i>a)}{\partial\boldsymbol{X}_i}&=\boldsymbol{\beta}+\sigma\left(\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha_i}\right)\frac{\partial\alpha_i}{\partial\boldsymbol{X}_i}\\&=\boldsymbol{\beta}+\sigma(\lambda_i^2-\alpha_i\lambda_i)\left(\frac{-\boldsymbol{\beta}}{\sigma}\right)\\&=\boldsymbol{\beta}(1-\lambda_{\mathbf{i}}^2+\alpha_i\lambda_i)\\&=\boldsymbol{\beta}(1-\delta(\alpha_{\mathbf{i}}))\end{aligned}\tag{7.1.8}$$

将(7.1.7)式写成

$$Y_i|Y_i>a=\mathrm{E}(Y_i|Y_i>a)+u_i=\boldsymbol{X}_i\boldsymbol{\beta}+\sigma\lambda(\alpha_i)+u_i\tag{7.1.9}$$

其中 u_i 是被解释变量观测值与条件期望值之差，具有零均值和异方差。其方差为

$$\mathrm{Var}(u_i)=\sigma^2(1-\lambda_i^2+\lambda_i\alpha_i)=\sigma^2(1-\delta_i)$$

对比(7.1.3)与(7.1.9)后发现，由于被解释变量数据的截断问题，使得原模型(7.1.3)变换为(7.1.9)式的模型。**如果采用普通最小二乘法直接估计(7.1.3)将会出现：第一，实际上忽略了一个非线性项 λ_i；第二，忽略了随机误差项实际上的异方差性**。这就造成参数估计量的偏误，被称为**“选择性偏误”**(selective bias)，而且如果不了解解释变量的分布，要估计该偏误的严重性也是很困难的。

例 7.1.1

为了建立某城市的城镇居民消费模型，经过对该城市的城镇居民消费行为的深入分析和经验检验，表明家庭人均收入(*X*)是决定家庭人均消费(*Y*)的唯一显著变量。在以工资收入为主要收入来源的城镇家庭中随机抽取 57 户样本，以他们的调查数据为样本观测值(见表 7.1.1)，建立该城镇居民消

PDG

费模型。

表 7.1.1 家庭人均月收入与月消费数据(元)

人均收入	人均消费	人均收入	人均消费	人均收入	人均消费
1 120	1 020	4 640	2 900	6 090	3 900
1 310	1 150	4 750	2 980	6 200	3 950
1 300	1 145	4 800	2 970	6 330	4 000
1 430	1 230	4 810	3 050	6 450	4 030
1 500	1 275	4 990	3 200	6 570	4 080
1 670	1 385	5 070	3 100	6 700	4 130
2 100	1 660	5 130	3 175	6 840	4 000
2 370	1 840	5 210	3 200	7 010	4 200
2 530	1 950	5 300	2 450	7 170	4 160
2 790	2 110	5 390	3 230	7 350	4 210
2 980	2 240	5 450	3 310	7 500	4 325
3 200	2 380	5 500	3 500	7 670	4 385
3 460	2 550	5 570	3 510	7 840	4 450
3 630	2 660	5 630	3 590	8 000	4 500
3 880	2 700	5 690	3 600	8 190	4 865
4 040	2 730	5 770	3 650	8 350	4 880
4 210	2 720	5 860	3 720	8 500	4 890
4 390	2 850	5 930	3 850	8 690	4 920
4 520	2 800	6 000	3 800	8 830	4 970

将该组样本看作不受任何限制下随机抽取的样本，采用普通最小二乘法估计模型，结果为

$$Y_i = 604.93 + 0.508\,3X_i \qquad i = 1,2,\cdots,57 \qquad R^2 = 0.977\,7$$

因为所有样本都是在以工资收入为主要收入来源的城镇家庭中抽取的，没有考虑到缺少稳定工资收入的低收入家庭和以财产收入为主的高收入家庭，显然样本具有选择性。将该组样本看作在消费水平大于 1 000 元、小于 5 000 元的特定人群中随机抽取的样本，重新采用最大似然法估计模型，结果为

$$Y_i = 556.70 + 0.519\,4X_i \qquad i = 1,2,\cdots,57 \qquad R^2 = 0.977\,5$$

同样的样本观测值，所建立的模型发生了变化，表示边际消费倾向的结构参数估计结果也发生了变化。

三、“归并”问题的计量经济学模型

1. 研究问题的思路

以一种简单的情况为例，讨论“归并”问题的计量经济学模型，即假设被解释变量服从正态分布，其样本观测值以 0 为界，凡小于 0 的都归并为 0，大于 0 的则取实际值。如果以 Y^*表示原始被解释变量，以 Y 表示归并后的被解释变量，那么则有

$$\begin{cases} Y=0, & Y^* \leqslant 0, \\ Y=Y^*, & Y^*>0, \end{cases} \quad 且\, Y^* \sim N(\mu,\sigma^2) \tag{7.1.10}$$

讨论这种简单的情况并不失一般性。如果样本观测值不是以 0 为界，而是以某一个数值 a 为界，则有

$$\begin{cases} Y=a, & Y^* \leqslant a, \\ Y=Y^*, & Y^*>a, \end{cases} \quad 且\, y^* \sim N(\mu,\sigma^2)$$

单方程线性“归并”问题的计量经济学模型为

$$\begin{cases} Y_{i*} = \boldsymbol{X}_i \boldsymbol{\beta} + \mu_i \quad \mu_i \sim N(0,\sigma^2) \\ Y_i = \max(Y_i^*, 0) \end{cases} \tag{7.1.11}$$

注意，这里实际观测到的被解释变量是 Y 而不是 Y^*。如果能够得到 Y_i 的概率密度函数，那么就可以方便地采用最大似然法估计模型，**这就是研究这类问题的思路**。

由于该模型是由托宾(Tobin)于 1958 年最早提出的，所以也称为 **Tobit 模型**。

2. “归并”变量的正态分布

由于原始被解释变量 Y^* 服从正态分布，根据(7.1.10)式，有

$$P(Y=0)=P(Y^* \leqslant 0)=\Phi\left(-\frac{\mu}{\sigma}\right)=1-\Phi\left(\frac{\mu}{\sigma}\right)$$

$$P(Y)=P(Y^*), \quad Y^*>0 \tag{7.1.12}$$

需要特别注意的是 $Y=0$ 的概率，它不是样本观测值取 0 的概率，而是样本观测值取小于等于 0 的值的概率。正是在这里，将归并样本和非归并样本严格区分开了。

3. 归并被解释变量数据计量经济学模型的最大似然估计

根据(7.1.12)式，可以得到所有样本的联合概率，即似然函数，然后很容易得到模型(7.1.11)的对数似然函数：

$$\ln L = \sum_{Y_i>0} -\frac{1}{2}\left(\ln(2\pi)+\ln\sigma^2+\frac{(Y_i - \boldsymbol{X}_i\boldsymbol{\beta})^2}{\sigma^2}\right) + \sum_{Y_i=0}\ln\left(1-\Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right)\right) \quad (7.1.13)$$

显然，(7.1.13)式由两部分组成，一部分对应于没有限制的观测值，是经典回归部分；一部分对应于受到限制的观测值。这是一个非标准的似然函数，它实际上是离散分布与连续分布的混合。

对(7.1.13)式极大化，就可以求得具有良好性质的参数估计量。同样，由于这是一个复杂的非线性问题，需要采用迭代方法求解，例如牛顿法。利用计量经济学软件可以很方便地实现模型的估计。

如果对(7.1.13)式进行再参数化，可使得估计过程更为简单，即令$\boldsymbol{\gamma} = \boldsymbol{\beta}/\sigma$ 和 $\theta = 1/\sigma$，得到

$$\ln L = \sum_{y_i>0} -\frac{1}{2}\left(\ln(2\pi) - \ln\theta^2 + (\theta Y_i - \boldsymbol{X}_i\boldsymbol{\gamma})^2\right) + \sum_{y_i=0}\ln\left(1-\Phi(\boldsymbol{X}_i)\boldsymbol{\gamma}\right) \quad (7.1.14)$$

对(7.1.14)式极大化，由于**黑塞**矩阵始终是负正定的，应用牛顿法求解时较为简单，且收敛速度较快，得到$\boldsymbol{\gamma}$和θ的估计量后再利用$\sigma = 1/\theta$和$\boldsymbol{\beta} = \boldsymbol{\gamma}/\theta$求得原参数估计量。

以上讨论的是左端归并的情况，如果出现右端归并，或者左右端同时归并的情况，原理是相同的，只是似然函数和对数似然函数的表述略有不同。在利用软件进行模型估计时，只需正确地输入归并状态，估计自动完成。

例 7.1.2

在例 7.1.1 中，如果增加了 3 个样本户，家庭人均收入分别为 1 080 元、1 040 元和 1 000 元，而家庭人均消费相同，都是 1 000 元。假设这 60 个样本不存在截断问题而存在左端归并问题，即家庭人均消费 1 000 元是≤1 000 元的归并。

将该组样本看作不受任何限制下随机抽取的样本，采用普通最小二乘法估计模型，结果为

$$Y_i = 571.83 + 0.513\,6X_i \qquad i = 1,2,\cdots,60 \qquad R^2 = 0.981\,1$$

将该组样本看作在消费水平为 1 000 元的归并样本，采用最大似然法重新估计模型，结果为

$$Y_i = 545.95 + 0.517\,8X_i \qquad i = 1,2,\cdots,60$$

同样的样本观测值，所建立的模型发生了变化，表示边际消费倾向的结构参数估计结果也发生了变化。

§7.2 二元离散选择模型

在经典计量经济学模型中，被解释变量通常被假定为连续变量。但是，经济分析中经常面临许多决策问题，或者称为选择问题，即人们必须在可供选择的几个方案中作出选择。这些可供选择的方案可以用离散的数据表示，例如，某一事件发生与否，分别用 1 和 0 表示；对某一建议持强烈反对、反对、中立、支持、强烈支持 5 种态度，可以分别用 0，1，2，3，4 表示。以这样的决策结果作为被解释变量建立的计量经济学模型，称为**离散被解释变量数据计量经济学模型**，或者称为**离散选择模型**(Discrete Choice Model, DCM)。如果被解释变量只能存在两种选择，称为**二元选择模型**(binary choice model)；如果被解释变量存在多种选择，称为**多元选择模型**(multiple choice model)。本节只介绍二元选择模型。

离散选择模型起源于费希纳(Fechner)于 1860 年进行的动物条件二元反射研究。1962 年，沃纳(Warner)首次将它应用于经济研究领域，用以研究公共交通工具和私人交通工具的选择问题。20 世纪 70 年代和 80 年代，离散选择模型被普遍应用于经济布局、企业定点、交通问题、就业问题、购买决策等经济决策领域的研究。从 1987 年出版的专著 Econometric Analysis of Discrete Choice(Börsch-Supan, Springer)所引用的文献可以看出，模型的估计方法主要发展于 20 世纪 80 年代初期。麦克法登(McFadden)因为在离散选择模型领域的贡献而获得 2000 年诺贝尔经济学奖。

一、二元离散选择模型的经济背景

实际经济生活中，人们经常遇到二元选择问题。

例如，公共交通工具和私人交通工具的选择问题。选择使用公共交通工具还是私人交通工具，取决于两类因素：一类是公共交通工具和私人交通工具所具有的属性，诸如速度、耗费时间、成本等；一类是决策个体所具有的属性，诸如职业、年龄、收入水平、健康状况等。从大量的统计中，可以发现选择结果与影响因素之间具有一定的因果关系。揭示这一因果关系并用于预测研究，对于制定交通工具发展规划无疑是十分重要的，这就需要建立计量经济学模型。

再如，对某种商品的购买决策问题。决定购买与否，取决于两类因素：一类是该商品本身所具有的属性，诸如性能、价格等；一类是消费者个体所具有的属性，诸如收入水平、对该商品的偏好程度等。从大量的统计中，可以发现选择结果与影响因素之间具有一定的因果关系。揭示这一因果关系并用于预测研究，对于生产厂家无疑是十分重要的，这也需要建立计量经济学模型。

又如，求职者对某种职业的选择问题。决定接受或者拒绝该职业，同样取决于两类因素：一类是该职业本身所具有的属性，诸如工作环境、工资水平、对求职者文化水平的要求等；一类是求职者个体所具有的属性，诸如年龄、文化水平、对职业的偏好等。从大量的统计中，可以发现选择结果与影响因素之间具有一定的因果关系。揭示这一因果关系并用于预测研究，对于用人单位如何适应就业市场，显然是十分有益的，这也需要建立计量经济学模型。

由此可见，二元选择问题在我们的经济生活中大量存在且应用广泛。

二、二元离散选择模型

1．原始模型

对于上述二元选择问题，可以建立如下计量经济学模型：

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \mu_i \tag{7.2.1}$$

其中$\boldsymbol{Y}_i$为观测值为 1 和 0 的决策被解释变量，$\boldsymbol{X}$为解释变量，包括选择对象所具有的属性和选择主体所具有的属性。在模型(7.2.1)式中因为$\mathrm{E}(\mu_i)=0$，所以$\mathrm{E}(Y_i)=\boldsymbol{X}_i\boldsymbol{\beta}$。令

$$p_i = P(Y_i = 1), \qquad 1 - p_i = P(Y_i = 0)$$

于是

$$\mathrm{E}(Y_i) = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = p_i$$

所以有

$$\mathrm{E}(Y_i) = P(Y_i = 1) = \boldsymbol{X}_i\boldsymbol{\beta}$$

对于该式右端的$\boldsymbol{X}_i\boldsymbol{\beta}$，并没有处于[0，1]范围内的限制，实际上很可能超出[0，1]的范围；而对于该式左端的$P(Y_i=1)$，则要求处于[0，1]范围内，于是上式产生了矛盾。另外，对于随机干扰项，有

$$\mu_i = \begin{cases} 1 - \boldsymbol{X}_i\boldsymbol{\beta}, & \text{当 } Y_i = 1, \text{其概率为 } \boldsymbol{X}_i\boldsymbol{\beta} \\ -\boldsymbol{X}_i\boldsymbol{\beta}, & \text{当 } Y_i = 0, \text{其概率为 } 1 - \boldsymbol{X}_i\boldsymbol{\beta} \end{cases}$$

显然，具有这种概率结构的随机干扰项具有异方差性。由于存在这两方面的问题，所以模型(7.2.1)式不能作为实际研究二元选择问题的模型。

2．效用模型

为了使二元选择问题的研究成为可能，我们必须首先建立随机效用模型。

以公共交通工具和私人交通工具的选择问题为例。如果某一个体选择公共交通工具，他的效用为U_i^1，上标表示选择结果，下标表示第 i 个个体。该效用是随机变量，并且由公共交通工具所具有的属性和决策个体所具有的属性解释。于是有

$$U_i^1 = \boldsymbol{X}_i\boldsymbol{\beta}^1 + \varepsilon_i^1 \tag{7.2.2}$$

类似地，如果某一个体选择私人交通工具，他的效用为U_i^0，该效用是随机变量，并且由私人交通工具所具有的属性和决策个体所具有的属性解释。于是有

$$U_i^0 = \boldsymbol{X}_i\boldsymbol{\beta}^0 + \varepsilon_i^0 \tag{7.2.3}$$

请注意，在模型(7.2.2)式和(7.2.3)式中，效用是不可观测的，我们能够得到的观测值仍然是选择结果，即 1 和 0。但是很显然，如果不可观测的$U_i^1 > U_i^0$，即对应于观测值 1，因为该个体选择公共交通工具的效用大于选择私人交通工具的效用，他当然会选择公共交通工具；相反，如果不可观测的$U_i^1 \leqslant U_i^0$，即对应于观测值 0，因为该个体选择公共交通工具的效用小于选择私人交通工具的效用，他当然会选择私人交通工具。

将(7.2.2)式与(7.2.3)式相减，得

$$U_i^1 - U_i^0 = \boldsymbol{X}_i(\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0) + (\varepsilon_i^1 - \varepsilon_i^0)$$

记为

$$Y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} \ + \mu_i^* \tag{7.2.4}$$

这就是我们要研究的二元选择模型。这是一个线性模型，其中$Y_i^*, \boldsymbol{X}_i, \boldsymbol{\beta}\ , \mu_i^*$分别为模型的被解释变量、解释变量、待估计参数和随机干扰项。

再来看个体选择$Y_i = 1$的概率。显然应该有

$$P(Y_i = 1) = P(Y_i^* > 0) = P(\mu_i^* > -\boldsymbol{X}_i\boldsymbol{\beta}) \tag{7.2.5}$$

3．最大似然估计

欲使得模型(7.2.4)式可以估计，就必须为μ_i^*选择一种特定的概率分布。**两种最常用的分布是标准正态分布和逻辑(logistic)分布，于是形成了两种最常用的二元选择模型——Probit 模型和 Logit 模型。**

无论是标准正态分布还是逻辑分布，由于它们是对称的，存在

$$F(-t) = 1 - F(t)$$

其中$F(t)$表示概率分布函数。于是(7.2.5)式可以改写为

$$\begin{aligned} P(Y_i = 1) &= P(Y_i^* > 0) = P(\mu_i^* > -\boldsymbol{X}_i\boldsymbol{\beta}) \\ &= 1 - P(\mu_i^* \leqslant -\boldsymbol{X}_i\boldsymbol{\beta}) \\ &= 1 - F(-\boldsymbol{X}_i\boldsymbol{\beta}) = F(\boldsymbol{X}_i\boldsymbol{\beta}) \end{aligned} \tag{7.2.6}$$

至此，可以得到模型(7.2.4)式的似然函数

$$P(Y_1, Y_2, \cdots, Y_n) = \prod_{Y_i=0}[1 - F(\boldsymbol{X}_i\boldsymbol{\beta})]\prod_{Y_i=1}F(\boldsymbol{X}_i\boldsymbol{\beta}) \tag{7.2.7}$$

即

$$L = \prod_{i=1}^{n}[F(\boldsymbol{X}_i\boldsymbol{\beta})]^{Y_i}[1 - F(\boldsymbol{X}_i\boldsymbol{\beta})]^{1-Y_i} \tag{7.2.8}$$

对数似然函数为

$$\ln L=\sum_{i=1}^{n}\left\{Y_i \ln F(\boldsymbol{X}_i\boldsymbol{\beta})+(1-Y_i)\ln[1-F(\boldsymbol{X}_i\boldsymbol{\beta})]\right\} \tag{7.2.9}$$

对数似然函数最大化的一阶条件为

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}}=\sum_{i=1}^{n}\left[\frac{Y_i f_i}{F_i}+(1-Y_i)\frac{-f_i}{1-F_i}\right]\boldsymbol{X}_i=\boldsymbol{0} \tag{7.2.10}$$

其中 f_i 表示概率密度函数。显然，在样本数据的支持下，如果知道(7.2.10)式中的概率分布函数和概率密度函数，求解该方程组，可以得到模型参数估计量。

三、二元 Probit 离散选择模型及其参数估计

Probit 模型是将标准正态分布作为(7.2.4)式中 μ_i^* 的概率分布而推导得到的。因为正态分布被认为是任何分布的自然的和首先的选择，于是二元 Probit 模型成为最常用的二元选择模型。标准正态分布的概率分布函数是

$$F(t)=\int_{-\infty}^{t}(2\pi)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \tag{7.2.11}$$

概率密度函数是

$$f(x)=(2\pi)^{-\frac{1}{2}}\exp\left(-\frac{x^2}{2}\right) \tag{7.2.12}$$

1．重复观测值不可以得到的情况下二元 Probit 离散选择模型的参数估计

在重复观测值不可以得到的情况下，(7.2.10)式写成

$$\begin{aligned}\frac{\partial \ln L}{\partial \boldsymbol{\beta}}&=\sum_{y_i=0}\frac{-f_i}{1-F_i}\boldsymbol{X}_i+\sum_{y_i=1}\frac{f_i}{F_i}\boldsymbol{X}_i\\&=\sum_{i=1}^{n}\left[\frac{q_i f(q_i\boldsymbol{X}_i\boldsymbol{\beta})}{F(q_i\boldsymbol{X}_i\boldsymbol{\beta})}\right]\boldsymbol{X}_i\\&=\sum_{i=1}^{n}\lambda_i\boldsymbol{X}_i\\&=\boldsymbol{0}\end{aligned} \tag{7.2.13}$$

其中 $$q_i=2Y_i-1$$

(7.2.13)式是关于 $\boldsymbol{\beta}$ 的非线性函数，不能直接求解，需采用完全信息最大似然法中所采用的迭代方法。

这里所谓“重复观测值不可以得到”，是指对每个决策者只有一个观测值。即使有多个观测值，也将其看作多个不同的决策者。

例 7.2.1

这里用一个简单的例子演示二元 Probit 离散选择模型及其参数估计。在一次选举中，由于候选人对高收入者有利，所以收入成为每个投票者表示同意或者反对的最主要影响因素。以投票者的态度(Y)作为被解释变量，以投票者的月收入(X)作为解释变量建立模型，同意者其观测值为 1，反对者其观测值为 0，样本数据见表 7.2.1。

原始模型为

$$Y_i = \alpha + \beta X_i + \mu_i, \qquad i = 1,2,\cdots,30$$

利用最简单的二元 Probit 离散选择模型参数估计软件(如 EViews 中的 Probit 估计)，估计结果为

$$\hat{\alpha} = -4.753\,9, \qquad \hat{\beta} = 0.003\,07, \qquad \ln L = -6.096\,2$$

计算得到模型(7.2.4)式的 $\hat{Y}_i^*$ 列于表 7.2.1 中。可见，虽然输入的是 Y 的观测值，但是作为估计对象的不是原始模型，而是模型(7.2.4)式。

按照方程

$$\text{YF}=1-\text{@CNORM}[-(-4.753\,9 + 0.003\,067{*}X)]$$

可以得到不同 X 值下的 Y 选择 1 的概率。例如，当 X=600 时，查标准正态分布表，对应于 2.913 7 的累积正态分布为 0.998 2；于是，Y 的预测值 YF=1–0.998 2=0.001 8，即对应于该个人，投赞成票的概率为 0 .001 8。

表 7.2.1　样本观测值及模拟值

X_i	Y_i	$\hat{Y}_i^*$ (Probit)	$\hat{Y}_i^*$ (Logit)	X_i	Y_i	$\hat{Y}_i^*$ (Probit)	$\hat{Y}_i^*$ (Logit)
100	0	−4.447 2	−7.602 9	1 600	0	0.153 3	0.262 2
200	0	−4.140 5	−7.078 6	1 700	1	0.460 0	0.786 5
300	0	−3.833 8	−6.554 3	1 800	0	0.766 7	1.310 8
400	0	−3.527 1	−6.029 9	1 900	1	1.073 4	1.835 2
500	0	−3.220 4	−5.505 6	2 000	1	1.380 1	2.359 5
600	0	−2.913 7	−4.981 2	2 100	1	1.686 8	2.883 9
700	0	−2.607 0	−4.456 9	2 200	1	1.993 5	3.408 2
800	0	−2.300 3	−3.932 6	2 300	1	2.300 2	3.932 5
900	0	−1.993 6	−3.408 2	2 400	1	2.606 9	4.456 9
1 000	0	−1.686 9	−2.883 8	2 500	1	2.913 6	4.981 2
1 100	0	−1.380 2	−2.359 5	2 600	1	3.220 3	5.505 6
1 200	0	−1.073 5	−1.835 2	2 700	1	3.527 0	6.029 9
1 300	1	−0.766 8	−1.310 9	2 800	1	3.833 7	6.554 2
1 400	0	−0.460 1	−0.786 5	2 900	1	4.140 4	7.078 6
1 500	1	−0.153 4	−0.262 2	3 000	1	4.447 1	7.602 9

2. 重复观测值可以得到的情况下二元 Probit 离散选择模型的参数估计

从理论上讲，“重复观测值可以得到”的情况是存在的，即对每个决策者有多个重复观测值。例如，观察某个人在外部条件不变的情况下对公共交通工具和私人交通工具的多次重复选择。在这种情况下，可以采用广义最小二乘法估计二元选择模型。

对第 i 个决策者重复观测 n_i 次，选择 $Y_i=1$ 的次数比例为 p_i，那么可以将 p_i 作为真实概率 P_i 的一个估计量。于是有

$$p_i = P_i + e_i = F(\boldsymbol{X}_i\boldsymbol{\beta}) + e_i \tag{7.2.14}$$

其中

$$\mathrm{E}(e_i)=0$$

$$\mathrm{Var}(e_i) = p_i\frac{1-p_i}{n_i}$$

对于标准正态分布的概率分布函数(7.2.11)式，定义“观测到的”**“概率单位”**(因此在一些教科书中将 Probit 模型译成“概率单位模型”)为

$$v_i = F^{-1}(p_i) = F^{-1}(P_i+e_i) \tag{7.2.15}$$

其中 F^{-1} 是标准正态分布的概率分布函数的反函数。用泰勒级数展开(7.2.15)式，只保留一阶项，则有

$$F^{-1}(P_i+e_i) = F^{-1}(P_i) + \frac{e_i}{f[F^{-1}(P_i)]} \tag{7.2.16}$$

于是(7.2.15)式可以改写为

$$v_i = F^{-1}(P_i) + u_i$$

其中

$$\mathrm{E}(u_i)=0$$

$$\mathrm{Var}(u_i) = \frac{P_i(1-P_i)}{n_i\{f[F^{-1}(P_i)]\}^2}$$

因为

$$F^{-1}(P_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

有

$$\begin{aligned} v_i &= \boldsymbol{X}_i\boldsymbol{\beta} + u_i \\ \boldsymbol{V} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U} \end{aligned} \tag{7.2.17}$$

采用广义最小二乘法估计(7.2.17)式，得到

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{V} \tag{7.2.18}$$

其中 $\boldsymbol{\Omega}$ 为 $\boldsymbol{U}$ 的方差-协方差矩阵。在实际估计过程中用它的估计量代替，即

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V} \tag{7.2.19}$$

而 $\hat{\boldsymbol{\Omega}}$ 则由 P_i 的估计量 p_i 构成，为了提高估计量的质量，可以采用迭代方法反

复求得 P_i 的估计量。

(7.2.19)式中 $\boldsymbol{V}$ 的观测值通过求解标准正态分布的概率分布函数

$$p_i = \int_{-\infty}^{v_i} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t$$

的反函数得到，而其中的 p_i 是实际观测得到的。为了使 p_i 的观测值比较可靠，一般要求对每个决策者都进行一定数量的次数(如 10 次左右)的观测。

四、二元 Logit 离散选择模型及其参数估计

Logit 模型是将逻辑分布作为(7.2.5)式中 μ_i^* 的概率分布而推导得到的。**Börsch-Supan 于 1987 年指出，如果选择是按照效用最大化而进行的，具有极限值的逻辑分布是较好的选择，这种情况下的二元选择模型应该采用 Logit 模型。**逻辑分布的概率分布函数是

$$F(t) = \frac{1}{1+\mathrm{e}^{-t}} \tag{7.2.20}$$

概率密度函数是

$$f(t) = \frac{\mathrm{e}^{-t}}{(1+\mathrm{e}^{-t})^2} \tag{7.2.21}$$

(7.2.20)式可以改写成

$$F(t) = \frac{\mathrm{e}^{t}}{1+\mathrm{e}^{t}} = \mathit{\Lambda}(t) \tag{7.2.22}$$

这里 $\mathit{\Lambda}$ 是通常用来表示逻辑分布的概率分布的符号。(7.2.21)式可以改写成

$$f(t) = \frac{\mathrm{e}^{t}}{(1+\mathrm{e}^{t})^2} = \mathit{\Lambda}(t)[1-\mathit{\Lambda}(t)] \tag{7.2.23}$$

1. 重复观测值不可以得到的情况下二元 Logit 离散选择模型的参数估计

在重复观测值不可以得到的情况下，将(7.2.22)式和(7.2.23)式代入(7.2.10)式，得到

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{n}\left[\frac{Y_i f_i}{F_i} + (1-Y_i)\frac{-f_i}{1-F_i}\right]\boldsymbol{X}_i \\ &= \sum_{i=1}^{n}[Y_i - \mathit{\Lambda}(\boldsymbol{X}_i\boldsymbol{\beta})]\boldsymbol{X}_i = \boldsymbol{0} \end{aligned} \tag{7.2.24}$$

(7.2.24)式是关于 $\boldsymbol{\beta}$ 的非线性函数，不能直接求解，需采用完全信息最大似然法中所采用的迭代方法。

同样，这里所谓“重复观测值不可以得到”，是指对每个决策者只有一个观测值。

对于例 7.2.1，利用最简单的二元 Logit 离散选择模型参数估计软件(如 EViews 中的 Logit 估计)，估计结果为

$$\hat{\alpha}=-8.127\,3,\qquad \hat{\beta}=0.005\,24,\qquad \ln L=-6.259\,9$$

计算得到模型(7.2.4)式的Y_i^*列于表 7.2.1 中。同样可见，虽然输入的是 Y 的观测值，但是作为估计对象的不是原始模型，而是模型(7.2.4)式。

2．重复观测值可以得到的情况下二元 Logit 离散选择模型的参数估计

在重复观测值可以得到的情况下，同样可以采用广义最小二乘法估计二元 Logit 选择模型。

由(7.2.20)式可以得到

$$\frac{F(t)}{1-F(t)}=\mathrm{e}^t \tag{7.2.25}$$

同样地，对第 i 个决策者重复观测 n_i 次，选择 $Y_i=1$ 的次数比例为 p_i，那么可以将 p_i 作为真实概率 P_i 的一个估计量。于是有

$$p_i=P_i+e_i=F(\boldsymbol{X}_i\boldsymbol{\beta})+e_i \tag{7.2.26}$$

其中

$$\mathrm{E}(e_i)=0$$

$$\mathrm{Var}(e_i)=\frac{p_i(1-p_i)}{n_i}$$

用样本重复观测得到的 p_i 构成“**成败比例**” $\frac{p_i}{1-p_i}$ (因此在一些教科书中将 Logit 模型译成“对数成败比例模型”)，取对数并进行泰勒展开，有

$$\ln\frac{p_i}{1-p_i}\approx\ln\frac{P_i}{1-P_i}+\frac{e_i}{P_i(1-P_i)} \tag{7.2.27}$$

在(7.2.25)式中，用 P_i 代替 $F(t)$，再用 $\boldsymbol{X}_i\boldsymbol{\beta}$ 代入 t，然后代入(7.2.27)式，得到

$$\ln\frac{p_i}{1-p_i}\approx\ln e^{\boldsymbol{X}_i\boldsymbol{\beta}}+u_i=\boldsymbol{X}_i\boldsymbol{\beta}+u_i \tag{7.2.28}$$

令 $v_i=\ln\frac{p_i}{1-p_i}$，则有

$$\begin{aligned} v_i&=\boldsymbol{X}_i\boldsymbol{\beta}+u_i\\ \boldsymbol{V}&=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{U}\end{aligned} \tag{7.2.29}$$

采用广义最小二乘法估计(7.2.29)式，得到

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{V} \tag{7.2.30}$$

其中 $\hat{\boldsymbol{\Omega}}$ 由 P_i 的估计量 p_i 构成。同样地，为了提高估计量的质量，可以采用迭代方法反复求得 P_i 的估计量。$\boldsymbol{V}$ 的观测值不需要求解概率分布函数的反函数，而是由实际观测得到的 p_i 直接计算得到。

五、一个实际例题

例 7.2.2

某商业银行从历史贷款客户中随机抽取 78 个样本，根据设计的指标体系分别计算它们的“商业信用支持度”(XY)和“市场竞争地位等级”(SC)，对它们贷款的结果(JG)采用二元离散变量，1 表示贷款成功，0 表示贷款失败。样本观测值见表 7.2.2。目的是研究 JG 与 XY，SC 之间的关系，并为正确贷款决策提供支持。

表 7.2.2　样本观测值

JG	XY	SC	JGF	JG	XY	SC	JGF	JG	XY	SC	JGF
0	125.0	−2	0.000 0	0	1500	−2	0.000 0	0	54.00	−1	0.000 0
0	599.0	−2	0.000 0	0	96.00	0	0.000 0	1	42.00	2	1.000 0
0	100.0	−2	0.000 0	1	−8.000	0	1.000 0	0	42.00	0	0.020 9
0	160.0	−2	0.000 0	0	375.0	−2	0.000 0	1	18.00	2	1.000 0
0	46.00	−2	0.000 0	0	42.00	−1	6.5×10^{-13}	0	80.00	1	6.4×10^{-12}
0	80.00	−2	0.000 0	1	5.000	2	1.000 0	1	−5.000	0	1.000 0
0	133.0	−2	0.000 0	0	172.0	−2	0.000 0	0	326.0	2	0.000 0
0	350.0	−1	0.000 0	1	−8.000	0	1.000 0	0	261.0	1	0.000 0
1	23.00	0	0.997 9	0	89.00	−2	0.000 0	1	−2.000	−1	0.999 9
0	60.00	−2	0.000 0	0	128.0	−2	0.000 0	0	14.00	−2	3.9×10^{-7}
0	70.00	−1	0.000 0	1	6.000	0	1.000 0	1	22.00	0	0.999 1
1	−8.000	0	1.000 0	0	150.0	−1	0.000 0	0	113.0	1	0.000 0
0	400.0	−2	0.000 0	1	54.00	2	1.000 0	1	42.00	1	0.998 7
0	72.00	0	0.000 0	0	28.00	−2	0.000 0	1	57.00	2	0.999 9
0	120.0	−1	0.000 0	1	25.00	0	0.990 6	0	146.0	0	0.000 0
1	40.00	1	0.999 8	1	23.00	0	0.997 9	1	15.00	0	1.000 0
1	35.00	1	0.999 9	1	14.00	0	1.000 0	0	26.00	−2	4.4×10^{-16}
1	26.00	1	1.000 0	0	49.00	−1	0.000 0	0	89.00	−2	0.000 0
1	15.00	−1	0.447 2	0	14.00	−1	0.549 8	1	5.000	1	1.000 0
0	69.00	−1	0.000 0	0	61.00	0	2.1×10^{-12}	1	−9.000	−1	1.000 0
0	107.0	1	0.000 0	1	40.00	2	1.000 0	1	4.000	1	1.000 0
1	29.00	1	1.000 0	0	30.00	−2	0.000 0	0	54.00	−2	0.000 0
1	2.000	1	1.000 0	0	112.0	−1	0.000 0	1	32.00	1	1.000 0
1	37.00	1	0.999 9	0	78.00	−2	0.000 0	0	54.00	0	1.4×10^{-7}
0	53.00	−1	0.000 0	1	0.000	0	1.000 0	0	131.0	−2	0.000 0
0	194.0	0	0.000 0	0	131.0	−2	0.000 0	1	15.00	0	1.000 0

PDG

1. 估计模型

采用 EViews 中的 Probit 模型估计方法，以 JG 为被解释变量，常数项，XY 和 SC 为解释变量，得到如表 7.2.3 所示的输出结果。

表 7.2.3　模型估计输出结果

```
Dependent Variable: JG
Method: ML - Binary Probit
Date: 10/06/04    Time: 23:25
Sample: 1 78
Included observations: 78
Convergence achieved after 13 iterations
Covariance matrix computed using second derivatives
==========================================================================
Variable              Coefficient   Std. Error      z-Statistic      Prob.
==========================================================================
C                        8.797358   7.544042          1.166133      0.2436
XY                      -0.257882   0.228894         -1.126645      0.2599
SC                       5.061789   4.458465          1.135321      0.2562
==========================================================================
Mean dependent var       0.410256      S.D. dependent var           0.495064
S.E. of regression       0.090067      Akaike info criterion        0.118973
Sum squared resid        0.608402      Schwarz criterion            0.209616
Log likelihood          -1.639954      Hannan-Quinn criter.         0.155259
Restr. log likelihood   -52.80224      Avg. log likelihood         -0.021025
LR statistic (2 df)      102.3246      McFadden R-squared           0.968942
Probability(LR stat)     0.000000
==========================================================================
Obs with Dep=0              46           Total obs                     78
Obs with Dep=1              32
==========================================================================
```

用回归方程表示如下：

JGF=1-@CNORM[-(8.797 358 375-0.257 881 662 4*XY+5.061 788 664*SC)]

2. 模拟与预测

该方程表示，当 XY 和 SC 已知时，代入方程，可以计算贷款成功的概率 JGF。例如，将表 7.2.2 中第 1 个样本观测值 XY=125，SC=–2 代入方程右边，计算括号内的值为 33.561 4。查标准正态分布表，对应于 33.561 4 的累积正态分布为 1.0。于是，JG 的预测值 JGF=1–1.0=0，即对应于该客户，贷款成功的概率为 0。将表 7.2.2 中第 19 个样本观测值 XY=15，SC=–1 代入方程右边，计算括号内的值为 0.132 655 2。查标准正态分布表，对应于 0.132 655 2 的累积正态分布为 0.551 7。于是，JG 的预测值 JGF=1–0.551 7 =0.448 3，即对应于该客户，贷款成功的概率为 0.448 3。

PDG

如果有一个新客户，根据客户资料，计算的“商业信用支持度”(XY)和“市场竞争地位等级”(SC)，代入模型，就可以得到贷款成功的概率，以此决定是否给予贷款。

六、二元离散选择模型的检验

经过估计的二元离散选择模型是否是一个好的模型？类似于经典的单方程模型，需要进行检验。主要的检验包括拟合优度检验、总体显著性检验、变量显著性检验、预测(回代)效果检验、异方差性检验和省略变量检验等。其中变量显著性检验的原理及检验统计量与经典单方程模型相同，而异方差性检验和省略变量检验的原理及检验统计量比较复杂，这里只简单介绍拟合优度检验、总体显著性检验和预测(回代)效果检验。

需要说明的是，由于经典单方程计量经济学模型主要采用以最小二乘原理为基础的模型估计方法，其检验统计量大多是基于残差平方和而构建的，例如拟合优度检验的 R^2 统计量、总体显著性检验的 F 统计量、变量显著性检验 t 或 z 统计量、约束回归检验的 F 统计量。而包括离散选择模型在内的非经典计量经济学模型主要采用以最大似然原理为基础的模型估计方法，所以其检验统计量大多是基于似然函数值而构建的，例如 Wald 统计量、LR 统计量、LM 统计量。

1. 拟合优度检验

设 L_0 为模型中所有解释变量的系数都为 0 时的似然函数值，显然有

$$\ln L_0 = n(P\ln P + (1-P)\ln(1-P))$$

其中 P 为样本观测值中被解释变量等于 1 的比例，n 为样本数目。设 L 为模型估计得到的似然函数值，构造一个统计量：

$$R^2 = 1 - \frac{\ln L}{\ln L_0}$$

显然，如果模型完全不拟合样本观测值，L 等于 L_0，则有 $R^2 = 0$；如果模型完全拟合样本观测值，L 等于 1，则有 $R^2 = 1$。所以 R^2 可以作为检验模型拟合优度的统计量，R^2 越接近于 1，模型的拟合效果越好。

在例 7.2.2 的模型估计输出结果表 7.2.3 中可以发现，当模型中所有解释变量的系数都为 0 时的对数似然函数值 $\ln L_0 = -52.802\,24$，模型估计得到的似然函数值 $\ln L = -1.639\,954$，计算得到 $R^2 = 0.968\,942$。因此可以判断这是一个拟合效果较好的模型。(在表中统计量 R^2 被称为“McFadden R-squared”)

2. 总体显著性检验

总体显著性检验的 0 假设为：$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$，备择假设为：解

释变量的系数不全为 0。构造一个似然比(likelihood ratio, LR)统计量：

$$\mathrm{LR}=-2(\ln L_0-\ln L)\sim\chi^2(k)$$

其中 L_0 为模型满足 0 假设(所有解释变量的系数都为 0)时的似然函数值，L 为模型估计得到的似然函数值。直观上看，如果LR 较大，表明 L_0 与 L 之间的差较大，倾向于拒绝 0 假设而接受模型总体显著的备择假设。

对于例 7.2.2 的模型，由 $\ln L_0=-52.802\,24$ 和 $\ln L=-1.639\,954$，计算得到 $\mathrm{LR}=102.324\,6$。由 χ^2 分布表查得：$\chi^2_{0.01}(2)=9.21$，可见，在 0.01 的显著水平上，该模型拒绝总体不显著的 0 假设。

3. 回代效果检验

当二元离散选择模型被估计后，将所有样本的解释变量观测值代入模型，计算得到每个样本的被解释变量选择 1 的概率，与每个样本被解释变量的实际观测值进行比较，以判断模型的预测(回代)效果，这也是一种实际有效的模型检验方法。

对于例 7.2.2 的模型，表 7.2.2 中“JGF”列即为模型计算得到的每个样本的被解释变量选择 1 的概率。从中可见，除了 2 个样本外，所有样本都通过了回代检验。没有通过回代检验的 2 个样本中，1 个样本的选择结果为 1，回代算得的选择 1 的概率为 0.447 2；另 1 个样本的选择结果为 0，回代算得的选择 1 的概率为 0.549 8。如何看待回代结果呢？这是一个与临界值有关的问题。

通常有多种方法确定临界值。一是“朴素方法”，即以 0.50 为临界值。该方法适合于全部样本中选择 1 和选择 0 的样本数目相当的情况。在例 7.2.2 的模型中，选择 1 和选择 0 的样本数目分别为 32 和 46，差异较大，不适合采用该方法。二是“先验方法”，即以全部样本中选择 1 的样本所占的比例为临界值。例如在例 7.2.2 的模型中，选择 1 的样本的比例为 0.41。但是，该方法适合于以全部个体作为样本的情况，而例 7.2.2 中的 78 个样本仅是贷款客户的极少部分，所以也不适合采用该方法。三是“最优方法”，即以“犯第一类错误最小”为原则确定临界值的方法。例如在例 7.2.2 的模型中，如果以 0.50 为临界值，则有 2 个样本发生“弃真”，即犯第一类错误；如果以 0.41 为临界值，则发生“弃真”的样本只有 1 个。所以以 0.41 作为临界值比较合适。

§7.3 平行数据计量经济学模型

所谓“平行数据”，也被翻译成“面板数据”，指在时间序列上取多个截面，在这些截面上同时选取样本观测值所构成的样本数据。平行数据计量经济学模型是近 20 年来计量经济学理论方法的重要发展之一，已经形成了与截面数据模

型相对应的完整的模型体系，具有很好的应用价值。本节将介绍**固定影响平行数据模型**(panel data model with fixed-effect)，与之相对应的是**随机影响平行数据模型**(panel data model with random-effects)，一般将它放在高级课程中。在固定影响平行数据模型中，本节仅介绍最简单的变截距模型和变系数模型，关于平行数据的动态模型、联立方程模型、选择性样本模型、离散选择模型等，以及平行数据的时间序列分析，本节将不予涉及。有兴趣的读者可以参考 Analysis of Panel Data(Cheng Hsiao)等专门书籍。

一、平行数据模型概述

1．经济分析中的平行数据问题

在经济分析中，尤其是通过建立计量经济学模型所进行的经济分析中，经常发现，只利用截面数据或者只利用时间序列数据不能满足分析目的的需要。

例如，如果分析生产成本问题，只利用截面数据，即选择同一截面上不同规模的企业数据作为样本观测值，可以分析成本与企业规模的关系，但是不能分析技术进步对成本的影响；只利用时间序列数据，即选择同一企业在不同时间上的数据作为样本观测值，可以分析成本与技术进步的关系，但是不能分析企业规模对成本的影响。如果采用平行数据，即在不同时间上选择不同规模的企业数据作为样本观测值，无疑既可以分析成本与技术进步的关系，也可以分析成本与企业规模的关系。

再如，分析目前我国的结构性失业问题，它既受到各地区产业结构的影响，也受到国家在各个时期的宏观政策的影响。只利用截面数据，即选择同一时间上不同省市的数据作为样本观测值，可以分析各省市不同的产业结构对结构性失业的影响，但是不能分析国家的宏观政策对各省市结构性失业的影响；只利用时间序列数据，即选择同一省市或者全国在不同时间上的数据作为样本观测值，可以分析国家的宏观政策对结构性失业的影响，但是不能分析不同的产业结构对结构性失业的影响。如果采用平行数据，即在不同的时间上选择不同省市的数据作为样本观测值，无疑既可以分析不同的产业结构对结构性失业的影响，也可以分析国家的宏观政策对结构性失业的影响。

由于平行数据计量经济学模型综合应用了截面数据和时间序列数据，相比于单独利用截面数据或者时间序列数据的模型，在减少模型的设定偏误，以及减少模型的估计偏误方面，都具有很多优点，因此它在目前的经济分析中成为仅次于经典单方程模型且被广泛采用的模型类型。

2．平行数据模型简介

单方程平行数据模型的一般形式为

$$y_{it}=\alpha_i+\boldsymbol{x}_{it}\boldsymbol{\beta}_i+u_{it}\,,\qquad i=1,\cdots,n\,,\qquad t=1,\cdots,T \tag{7.3.1}$$

其中 $\boldsymbol{X}_{it}$ 为 $1\times K$ 向量，$\boldsymbol{\beta}_i$ 为 $K\times 1$ 向量，K 为解释变量的数目。按照规范的表示，这里的 $\boldsymbol{X}_{it}$ 和 $\boldsymbol{\beta}_i$ 应该写成矩阵 $\boldsymbol{X}_{it}$ 和 $\boldsymbol{B}_i$，而且有

$$\boldsymbol{X}_{it}=(X_{it1},X_{it2},\cdots,X_{itK}) \qquad \boldsymbol{B}_i=(\beta_{i1},\beta_{i2},\cdots,\beta_{iK})'$$

但是为了简化书写，在本节中采用(7.3.1)式的表示。误差项 u_{it} 服从均值为零，方差为 σ_u^2 的正态分布。

模型(7.3.1)式常用的有如下三种情形：

情形 1：　$\alpha_i=\alpha_j\,,\beta_i=\beta_j$

情形 2：　$\alpha_i\neq\alpha_j\,,\beta_i=\beta_j$

情形 3：　$\alpha_i\neq\alpha_j\,,\beta_i\neq\beta_j$

对于情形 1，在横截面上无个体影响，也无结构变化，则普通最小二乘估计给出了 α 和 β 的一致有效估计，相当于将多个时期的截面数据放在一起作为样本数据。对于情形 2，这时的模型称为**变截距模型**(panel data model with variable intercept)，在横截面上的个体影响不同，个体影响表现为模型中被忽略的反映个体差异的变量的影响，又分为**固定影响**(fixed-effect)和**随机影响**(random-effect)两种情况。对于情形 3，这时的模型称为**变系数模型**(panel data model with variable coefficient)，除存在个体影响外，在横截面上还存在变化的经济结构，因而结构参数在不同横截面单位上是不同的，也分为**固定影响和随机影响**两种情况。

典型的平行数据是横截面单位较多而时期较少的数据。这样，该技术主要集中于横截面的变化，或异方差上(因为截面数据容易产生异方差，这在经典线性模型中已经讨论了)。

如果变化并不反映在横截面个体之间，而是反映在不同的截面之间，即在时间序列上，分析方法是完全相同的。

二、模型的设定

由于可以构造和检验比以往单独用横截面数据或时间序列数据更现实的行为方程模型，这就大大地丰富了计量经济学的经验研究。但平行数据包括二维的数据(横截面和时间)，如果模型设定不正确，将造成较大的偏差，估计结果与实际将相差甚远。所以，在建立平行数据模型时必须控制不可观察的个体和(或)时间的特征以避免模型设定的偏差并改进参数估计的有效性。

如果可获得的数据来自简单可控制的实验，则可以应用标准统计方法。不幸的是，多数平行数据来自经济活动的复杂过程。这样，若假设经济变量 y 在每个时点上都是由参数化的概率分布函数 $\mathrm{P}(y\mid\beta)$ (β 为参数)生成的，实际上是不现实的。忽视这种在横截面或时间上参数的本质上的差异可能会导致参数估

计不是一致估计或估计出的参数值无意义。例如，考虑模型(7.3.1)，参数 α_i 和 β_i 在不同的横截面样本点(即同一横截面的不同个体样本点)上不同，在不同时间上相同。这样，在不同的横截面样本点上，y 的抽样分布是不同的。但在同一横截面样本点上，y 在不同时间上的抽样分布是相同的。此时，若对该平行数据建立模型

$$y_{it}=\alpha+x_{it}\beta+u_{it}, \qquad i=1,\cdots,n, \qquad t=1,\cdots,T$$

则参数的最小二乘估计将不可能是一致估计，且估计值无任何意义。当模型(7.3.1)中不同横截面的参数相同，即 $\beta_i=\beta$ 时，也有同样的问题。

于是，研究平行数据的第一步是检验刻画被解释变量 y 的参数是否在所有横截面样本点和时间上都是常数，即检验所研究的问题属于上述 3 种情况中的哪一种，以便确定模型的形式。广泛使用的检验是**协变分析检验**(analysis of covariance)，也称 F 检验，主要检验以下两个假设。

假设 1：斜率在不同的横截面样本点和时间上都相同，但截距不相同。

$$H_1: \quad y_{it}=\alpha_i+x_{it}\beta+u_{it} \tag{7.3.2}$$

假设 2：截距和斜率在不同的横截面样本点和时间上都相同。

$$H_2: \quad y_{it}=\alpha+x_{it}\beta+u_{it} \tag{7.3.3}$$

显然，如果接受了假设 2，则没有必要进行进一步的检验。如果拒绝了假设 2，就应该检验假设 1，判断斜率是否都相等。如果假设 1 被拒绝，就应该采用模型(7.3.1)。

该检验的原理就是本书多元线性回归模型中介绍的参数约束检验。模型(7.3.3)就是对模型(7.3.1)施加了参数 α 和 β 在不同的截面个体上都相同的约束；模型(7.3.2)就是对模型(7.3.1)施加了参数 β 在不同的截面个体上都相同的约束。所以读者可以根据参数约束检验的原理自己构造检验假设2和假设1的统计量，并完成该模型设定检验，其中重要的是需要采用相应的估计方法首先估计模型(7.3.3)、(7.3.2)和(7.3.1)，得到它们各自的残差平方和，用以计算检验统计量的值。

下面首先介绍用以进行假设检验的 F 统计量的计算方法。

记

$$\begin{aligned}\bar{y}_i&=\frac{1}{T}\sum_{t=1}^{T}y_{it}\\ \bar{x}_i&=\frac{1}{T}\sum_{t=1}^{T}x_{it}\end{aligned} \tag{7.3.4}$$

模型(7.3.1)参数的普通最小二乘估计为

$$\hat{\beta}_i=W_{xx,i}^{-1}W_{xy,i}$$

PDG

$$\hat{\alpha}_i = \overline{y}_i - \overline{x}_i \hat{\beta}_i \tag{7.3.5}$$

称之为**群内估计**，其中

$$W_{xx,i} = \sum_{t=1}^{T} (x_{it} - \overline{x}_i)'(x_{it} - \overline{x}_i)$$

$$W_{xy,i} = \sum_{t=1}^{T} (x_{it} - \overline{x}_i)'(y_{it} - \overline{y}_i)$$

$$W_{yy,i} = \sum_{t=1}^{T} (y_{it} - \overline{y}_i)^2 \tag{7.3.6}$$

第 i 群的残差平方和是 $\mathrm{RSS}_i = W_{yy,i} - W'_{xy,i} W^{-1}_{xx,i} W_{xy,i}$，模型(7.3.1)的残差平方和为

$$S_1 = \sum_{i=1}^{n} \mathrm{RSS}_i \tag{7.3.7}$$

模型(7.3.2)参数的普通最小二乘估计为

$$\hat{\beta}_w = W_{xx}^{-1} W_{xy}$$

$$\hat{\alpha}_i = \overline{y}_i - \overline{x}_i \hat{\beta}_w \tag{7.3.8}$$

其中

$$W_{xx} = \sum_{i=1}^{n} W_{xx,i}\text{，}\quad W_{xy} = \sum_{i=1}^{n} W_{xy,i} \tag{7.3.9}$$

令

$$W_{yy} = \sum_{i=1}^{n} W_{yy,i} \tag{7.3.10}$$

模型(7.3.2)的残差平方和为

$$S_2 = W_{yy} - W'_{xy} W_{xx}^{-1} W_{xy} \tag{7.3.11}$$

模型(7.3.3)参数的普通最小二乘估计为

$$\hat{\beta} = T_{xx}^{-1} T_{xy}$$

$$\hat{\alpha} = \overline{y} - \overline{x} \hat{\beta} \tag{7.3.12}$$

其中

$$T_{xx} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \overline{x})'(x_{it} - \overline{x})$$

$$T_{xy} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \overline{x})'(y_{it} - \overline{y})$$

$$T_{yy} = \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \overline{y})^2 \tag{7.3.13}$$

$$\overline{y} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it}\text{，}\qquad \overline{x} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it} \tag{7.3.14}$$

模型(7.3.3)的残差平方和为

PDG

$$S_3 = T_{yy} - T'_{xy} T_{xx}^{-1} T_{xy} \tag{7.3.15}$$

由此可以得到下列结论：

(1) $S_1 / \sigma_u^2 \sim \chi^2[n(T-K-1)]$；

(2) 在 H_2 下，$S_3 / \sigma_u^2 \sim \chi^2[nT-(K+1)]$ 和 $(S_3 - S_1)/\sigma_u^2 \sim \chi^2[(n-1)(K+1)]$；

(3) $(S_3 - S_1)/\sigma_u^2$ 与 S_1 / σ_u^2 独立。

所以，得到检验 H_2 的 F 统计量：

$$\begin{aligned} F_2 &= \frac{(S_3 - S_1)/[(n-1)(K+1)]}{S_1/[nT-n(K+1)]} \\ &\sim F[(n-1)(K+1), n(T-K-1)] \end{aligned} \tag{7.3.16}$$

同时得到下列结论：

(1) 在 H_1 下，$S_2 / \sigma_u^2 \sim \chi^2[n(T-1)-K]$ 和 $(S_2 - S_1)/\sigma_u^2 \sim \chi^2[(n-1)K]$；

(2) $(S_2 - S_1)/\sigma_u^2$ 与 S_1 / σ_u^2 独立。

所以，得到检验 H_1 的 F 统计量：

$$\begin{aligned} F_1 &= \frac{(S_2 - S_1)/[(n-1)K]}{S_1/[nT-n(K+1)]} \\ &\sim F[(n-1)K, n(T-K-1)] \end{aligned} \tag{7.3.17}$$

给定显著性水平，查 F 分布表，得到临界值，与由计算得到的 F 统计量数值进行比较，即可得到拒绝或者接受假设的结论。需要说明的是，在本书推荐的 EViews 软件中，该检验不能自动完成。

三、固定影响变截距模型

变截距模型是应用最广泛的一种平行数据模型，可表示为

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}, \qquad i = 1, \cdots, n, \qquad t = 1, \cdots, T$$

其中 $\boldsymbol{x}_{it}$ 为 $1 \times K$ 向量，$\boldsymbol{\beta}$ 为 $K \times 1$ 向量；α_i 为个体影响，即为模型中被忽略的反映个体差异变量的影响；u_{it} 为随机干扰项，即为模型中被忽略的随横截面和时间变化的因素的影响，假设其均值为零，方差为 σ_u^2，并假定 u_{it} 与 $\boldsymbol{x}_{it}$ 不相关。如果横截面的个体影响可以用常数项 α_i 的差别来说明，这样 α_i 是一个待估未知参数，称为**固定影响变截距模型**。如果横截面的个体影响可以用不变的常数项和变化的随机项之和 $\alpha_0 + \varepsilon_i$ 的差别来说明，称为**随机影响变截距模型**。

1. 固定影响模型：最小二乘虚拟变量模型及其参数估计

对于固定影响变截距模型。变截距 α_i 是一个待估未知参数。令 $\boldsymbol{y}_i$ 和 $\boldsymbol{X}_i$ 是第 i 个个体的 T 个观测值向量和矩阵，并令 $\boldsymbol{u}_i$ 是随机干扰项 $T \times 1$ 向量，(7.3.2)

式可写成

$$\boldsymbol{y}_i = \boldsymbol{e}\alpha_i + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i, \qquad i = 1, \cdots, n \tag{7.3.18}$$

其中
$$\boldsymbol{y}_i = \begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{Ti} \end{pmatrix}_{T\times 1}, \quad \boldsymbol{e} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}_{K\times 1}$$

$$\boldsymbol{u}_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{Ti} \end{pmatrix}, \quad \boldsymbol{X}_i = \begin{pmatrix} x_{1i1} & x_{1i2} & \cdots & x_{1iK} \\ x_{2i1} & x_{2i2} & \cdots & x_{2iK} \\ \vdots & \vdots & & \vdots \\ x_{Ti1} & x_{Ti2} & \cdots & x_{TiK} \end{pmatrix}_{T\times K}$$

(7.3.18)也可写成

$$\boldsymbol{y} = (\boldsymbol{d}_1 \ \ \boldsymbol{d}_2 \ \ \cdots \ \ \boldsymbol{d}_n \ \ \boldsymbol{X})\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{u} \tag{7.3.19}$$

其中

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{nT\times 1}, \quad (\boldsymbol{d}_1 \ \ \boldsymbol{d}_2 \ \ \cdots \ \ \boldsymbol{d}_n) = \begin{pmatrix} e & 0 & \cdots & 0 \\ 0 & e & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e \end{pmatrix}_{nT\times n}$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix}_{nT\times K}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{n\times 1}, \quad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_n \end{pmatrix}_{nT\times 1}$$

其中 $\boldsymbol{d}_i$ 是代表第 i 个单位的虚拟变量。

令 $\boldsymbol{D} = (\boldsymbol{d}_1 \ \ \boldsymbol{d}_2 \ \ \cdots \ \ \boldsymbol{d}_n)$，则(7.3.19)式等价于

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{7.3.20}$$

该模型通常被称为**最小二乘虚拟变量模型**(Least-Squares Dummy-Variable, LSDV)，有时也称之为**协方差分析模型**(analysis-of-covariance model)(解释变量既有定量的，也有定性的)。如果 n 充分小，此模型可以当作具有 $n+K$ 个参数的多元回归模型，参数可由普通最小二乘法进行估计。

当 n 很大，甚至成千上万时，普通最小二乘法的计算可能超过任何计算机的存储容量。此时，可用下列分块回归的方法进行计算。

令 $\boldsymbol{Q}=\boldsymbol{I}_T-\frac{1}{T}\boldsymbol{ee}'$，因为 $\boldsymbol{I}_T\boldsymbol{e}=\frac{1}{T}\boldsymbol{ee}'\boldsymbol{e}$，所以 $\boldsymbol{Qe}=\boldsymbol{0}$，则由(7.3.18)式有

$$\boldsymbol{Q}\boldsymbol{y}_i=\boldsymbol{Qe}\alpha_i+\boldsymbol{QX}_i\boldsymbol{\beta}+\boldsymbol{Qu}_i=\boldsymbol{QX}_i\boldsymbol{\beta}+\boldsymbol{Qu}_i \tag{7.3.21}$$

于是

$$\boldsymbol{X}_i'\boldsymbol{Q}\boldsymbol{y}_i=\boldsymbol{X}_i'\boldsymbol{QX}_i\boldsymbol{\beta}+\boldsymbol{X}_i'\boldsymbol{Qu}_i$$

$$\sum_i\boldsymbol{X}_i'\boldsymbol{Q}\boldsymbol{y}_i=\left(\sum_i\boldsymbol{X}_i'\boldsymbol{QX}_i\right)\boldsymbol{\beta}+\sum_i\boldsymbol{X}_i'\boldsymbol{Qu}_i$$

$$\hat{\boldsymbol{\beta}}_{CV}=\left(\sum_{i=1}^n\boldsymbol{X}_i'\boldsymbol{QX}_i\right)^{-1}\left(\sum_{i=1}^n\boldsymbol{X}_i'\boldsymbol{Q}\boldsymbol{y}_i\right) \tag{7.3.22}$$

由于模型(7.3.20)也称为协方差分析模型，所以参数 $\boldsymbol{\beta}$ 的最小二乘虚拟变量模型估计也叫做协方差估计。$\boldsymbol{\beta}$ 的协方差估计是无偏的，且当 n 或 T 趋于无穷大时为一致估计。它的协方差矩阵为

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{CV})=\sigma_u^2\left(\sum_{i=1}^n\boldsymbol{X}_i'\boldsymbol{QX}_i\right)^{-1} \tag{7.3.23}$$

截距的估计为

$$\hat{\alpha}_i=\bar{y}_i-\bar{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}}_{CV} \tag{7.3.24}$$

$$\mathrm{Var}(\hat{\alpha}_i)=\frac{\sigma_u^2}{T}+\bar{\boldsymbol{X}}_i\mathrm{Var}(\hat{\boldsymbol{\beta}}_{CV})\bar{\boldsymbol{X}}_i' \tag{7.3.25}$$

截距的估计是无偏估计，且仅当 T 趋于无穷大时为一致估计。

方差 $\boldsymbol{\sigma}_u^2$ 的估计量为

$$s^2=\sum_{i=1}^n\sum_{t=1}^T\frac{(y_{it}-\hat{\alpha}_i-\boldsymbol{x}_{it}\hat{\boldsymbol{\beta}}_{CV})^2}{nT-n-K} \tag{7.3.26}$$

可以利用 F 检验来检验 $\alpha_i=\alpha_j$ 的假设，在该假设下

$$F=\frac{(R_u^2-R_p^2)/(n-1)}{(1-R_u^2)/(nT-n-K)} \tag{7.3.27}$$

服从 $F(n-1,nT-n-K)$，其中 R^2 为判定系数，下标 u 表示非约束模型，而 p 表示约束模型。

从以上估计过程可以看到，当 n 很大，参数 $\boldsymbol{\beta}$ 的数目并没有变化，而参数 $\boldsymbol{\alpha}$ 的数目等于 n，使得最小二乘虚拟变量模型估计无法进行。所以，所谓分块估计，就是首先设法消去参数 $\boldsymbol{\alpha}$，估计参数 $\boldsymbol{\beta}$；然后再在每个截面个体上利用变量的观测值和参数 $\boldsymbol{\beta}$ 的估计值，计算参数 $\boldsymbol{\alpha}$ 的估计量。

2. 固定影响变截距模型实例

例 7.3.1

建立我国城镇居民储蓄模型。

(1) 模型设定

以城镇居民家庭人均年储蓄额为被解释变量，以城镇居民家庭人均年可支配收入(其他因素经过检验表明不显著)为被解释变量。因为我国不同地区居民平均收入水平差距较大，为了将地区之间的影响引入模型，采用平行数据作为样本数据。选择北京、贵州、辽宁、吉林、新疆、安徽、山东、广东、山西、湖南、青海、上海 12 个地区从 1992 年到 1996 年的共 60 组数据，并消除了价格因素。利用上述检验方法，计算得到

$$S_1 = 1\,385\,549, \qquad S_2 = 1\,744\,353, \qquad S_3 = 4\,519\,840$$

$$F_1 = 0.85, \qquad F_2 = 3.70$$

查 F 分布表，给定 10%的显著性水平，得到临界值：

$$F(22,36) = 1.85, \qquad F(11,36) = 2.07$$

由于 $F_2 > 1.85$，所以拒绝 H_2；由于 $F_1 < 2.07$，所以接受 H_1。因此模型应该采用第二种形式，为变截距模型。具体形式为

$$S_{it} = \alpha_i + \beta I_{it} + \mu_{it}, \qquad i = 1,2,\cdots,12, \qquad t = 1992,\cdots,1996$$

其中 S, I 分别表示人均年储蓄额和人均年可支配收入。

(2) 固定影响变截距模型的估计

将模型假定为固定影响变截距模型，引入虚变量使之成为(7.3.19)式的形式。对模型进行估计，本例采用 EViews 软件，当然采用其他软件也是一样的。估计结果为

$$\hat{\beta} = 0.521\,9$$

$$\hat{\alpha}_1 = -577.351\,0, \qquad \hat{\alpha}_2 = -1\,029.709, \qquad \hat{\alpha}_3 = -775.157\,3$$

$$\hat{\alpha}_4 = -720.057\,3, \qquad \hat{\alpha}_5 = -877.626\,4, \qquad \hat{\alpha}_6 = -1\,004.523$$

$$\hat{\alpha}_7 = -989.062\,5, \qquad \hat{\alpha}_8 = -1\,628.885, \qquad \hat{\alpha}_9 = -687.383\,2$$

$$\hat{\alpha}_{10} = -1\,144.25\,1, \qquad \hat{\alpha}_{11} = -789.093\,1, \qquad \hat{\alpha}_{12} = -1\,191.264$$

(3) 结论

从模型可以看出，对所选择的 12 个地区，虽然有相同的储蓄倾向，但是实际储蓄水平有较大的差异。如果分地区分别建立以时间序列数据为样本的模型，不能得到不同地区具有相同储蓄倾向的结论；如果以不同地区的截面数据为样本建立模型，则不能考察不同地区的不同储蓄水平。

四、固定影响变系数模型

变截距模型中的截距变化反映了方程中未出现的变量对被解释变量的影响，或是随着截面个体而变，或是随着时间而变。但有时，变化的经济结构或不同的社会经济背景因素使得响应参数(也称结构参数)也随着时间或横截面个体不同而变化。

当数据不支持不变响应参数模型，且变量之间关系的设定也很恰当时，就必须考虑在时间或横截面上系数变化的变系数模型。系数随横截面上个体而改变的模型为

$$y_{it} = X_{it}\boldsymbol{\beta}_i + u_{it}, \qquad i=1,\cdots,n, \qquad t=1,\cdots,T \tag{7.3.28}$$

其中 X_{it} 和 $\boldsymbol{\beta}_i$ 是解释变量和参数向量，也可写成

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{u}_i \tag{7.3.29}$$

其中

$$\boldsymbol{y}_i = \begin{pmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{Ti} \end{pmatrix}_{T\times 1}, \qquad \boldsymbol{X}_i = \begin{pmatrix} x_{1i1} & x_{1i2} & \cdots & x_{1iK} \\ x_{2i1} & x_{2i2} & \cdots & x_{2iK} \\ \vdots & \vdots & & \vdots \\ x_{Ti1} & x_{Ti2} & \cdots & x_{TiK} \end{pmatrix}_{T\times K}$$

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iK} \end{pmatrix}, \qquad \boldsymbol{u}_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{Ti} \end{pmatrix}$$

当将 $\boldsymbol{\beta}_i$ 视为固定的不同的常数时，称为**固定影响变系数模型**，可写成

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{7.3.30}$$

其中

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}_{nT\times 1}, \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{X}_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{X}_n \end{pmatrix}_{nT\times nK}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix}_{nK\times 1}, \qquad \boldsymbol{u} = \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_n \end{pmatrix}_{nT\times 1}$$

显然，如果随机干扰项在不同横截面个体之间不相关，即 $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_j') = \boldsymbol{0}, i \neq j$ 且 $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i') = \sigma_i^2\boldsymbol{I}$，上述模型的参数估计极为简单，即以每个截面个体的时间序

列数据为样本，采用经典单方程计量经济学模型的估计方法分别估计其参数。即使采用广义最小二乘法估计同时得到$\boldsymbol{\beta}=(\boldsymbol{\beta}_1' \quad \cdots \quad \boldsymbol{\beta}_n')'$的广义最小二乘法估计量，也是与在每个横截面个体上$\boldsymbol{\beta}_i$的经典单方程估计一样。

如果随机干扰项在不同横截面个体之间的协方差不为零，即$\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_j') \neq \boldsymbol{0}$，则$\boldsymbol{\beta}=(\boldsymbol{\beta}_1' \quad \cdots \quad \boldsymbol{\beta}_n')'$的广义最小二乘法估计比在每个横截面个体上$\beta_i$的经典单方程估计更有效。

记$\boldsymbol{\Omega}_{ij}=\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_j')$，则

$$\boldsymbol{V}=\begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} & \cdots & \boldsymbol{\Omega}_{1n} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} & \cdots & \boldsymbol{\Omega}_{2n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Omega}_{n1} & \boldsymbol{\Omega}_{n2} & \cdots & \boldsymbol{\Omega}_{nn} \end{pmatrix}_{nT\times nT} \tag{7.3.31}$$

参数的广义最小二乘法估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}=(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} \tag{7.3.32}$$

如何得到协方差矩阵的估计量呢？一种可行的方法是：首先采用经典单方程计量经济学模型的估计方法分别估计每个横截面个体上的$\boldsymbol{\beta}_i$，计算残差估计值，以此构造协方差矩阵的估计量，这类似于经典单方程计量经济学模型的广义最小二乘法估计。

这里顺便解释一个问题：在时间序列上有多少个时点才能称为“平行数据”？例如，31个省(市、区)的20年的数据当然是一组“平行数据”，那么31个省(市、区)的3年的数据能否构成一组“平行数据”呢？首先要区分“平行数据”和“平行数据模型”。凡是1组以上的截面数据，例如31个省(市、区)的3年的数据就可以称为“平行数据”，但是这样的数据并不能采用“平行数据模型”的理论方法来分析。从上述的固定影响变系数模型的估计中可以看到，如果随机干扰项在不同横截面个体之间不相关，即以每个截面个体的时间序列数据为样本，采用经典单方程模型的估计方法分别估计其参数，那么要求时间序列必须足够长，才能实现模型的估计。例如，如果模型中有4个待估参数，那么按照经典模型对样本容量的要求，时间序列的时点至少要大于待估参数数目的3倍(即12)。所以，如果时间序列的时点太少，作为“平行数据模型”的第一步，即模型设定检验，就无法实现，更谈不上建立“平行数据模型”了。

本章练习题

1. 在经典计量经济学模型中，通常选择哪些类型的数据作为样本数据？对被解释变量

样本数据有哪些假定?

2．某一截面数据计量经济学模型 $y_i = \boldsymbol{\beta}'\boldsymbol{X}_i + \mu_i$，被解释变量服从正态分布，其样本观测值为 $y_1, y_2, \cdots, y_n$，其中 y_1, y_2, y_3 取相同值 a，其他观测值均大于 a。分别将该组样本看作未受限制的随机抽取样本、以 a 为截断点的选择性样本、以 a 为归并点的选择性样本，分别采用最大似然法估计模型。

(1) 写出 3 种情况下的对数似然函数表达式。

(2) 比较 3 种情况下的对数似然函数值的大小，并加以简单证明。

3．令 Y 表示一个学生在一所大学是否在第 4 年后能免试推荐攻读硕士学位的虚拟变量。设 X_1 与 X_2 分别是其入学时的考试成绩以及大学前两年各门必修课的平均成绩，X_3 是其在第三学年每周学习的小时数。假设利用 420 个学生的数据得到如下的 Logit 模型：

$$P_i = \mathrm{E}(Y=1) = \frac{1}{1+\mathrm{e}^{-(-1.1+0.002X_1+0.007X_2+0.02X_3)}}$$

假设 X_1 与 X_2 固定在 85 分的水平上，计算每周花 40 小时与花 20 小时学习的学生在推荐攻读硕士学位概率上的估计差异。

4．对重复观测数据(分组数据)，试证明以“成败比例”为特征的 Logit 模型

$$\ln\frac{p_i}{1-p_i} \approx \ln\frac{P_i}{1-P_i} + \frac{e_i}{P_i(1-P_i)}$$

中误差项的方差为

$$\mathrm{Var}\left[\frac{e_i}{P_i(1-P_i)}\right] = \frac{1}{N_i P_i (1-P_i)}$$

其中已知

$$\mathrm{Var}(e_i) = \frac{P_i(1-P_i)}{N_i}$$

5．在申请出国读学位的 16 名学生中有如下 GRE 数量与词汇成绩，其中 9 位学生获得入学准入。请根据下表中资料估计 Logit 模型与 Probit 模型。

学生编号	数量成绩 Q	词汇成绩 V	是否准入 Y(1=准，0=不准)	学生编号	数量成绩 Q	词汇成绩 V	是否准入 Y(1=准，0=不准)
1	760	550	1	9	520	660	1
2	600	350	0	10	800	250	0
3	720	320	0	11	670	480	0
4	710	630	1	12	670	520	1
5	530	430	1	13	780	710	1
6	650	570	0	14	520	450	0
7	800	500	1	15	680	590	1
8	650	680	1	16	500	380	0

6．下表列出了美国、加拿大、英国在 1980—1999 年的失业率 Y 以及对制造业的补助 X 的相关数据资料。考虑如下模型：

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \mu_{it}$$

(1) 根据上述回归模型分别估计这三个国家 Y 关于 X 的回归方程；

(2) 将三个国家的数据合并成一个大样本，按上述模型估计一个总的回归方程；

(3) 估计变截距固定影响模型；

(4) 根据上述三类回归方程的估计结果，判断哪类模型更好一些。

年份	美　国		加拿大		英　国	
	补助 X/(美元/小时)	失业率 Y/%	补助 X/(美元/小时)	失业率 Y/%	补助 X/(美元/小时)	失业率 Y/%
1980	55.6	7.1	49	7.2	43.7	7.0
1981	61.1	7.6	54.1	7.3	44.1	10.5
1982	67.0	9.7	59.6	10.6	42.2	11.3
1983	68.8	9.6	63.9	11.5	39.0	11.8
1984	71.2	7.5	64.3	10.9	37.2	11.7
1985	75.1	7.2	63.5	10.2	39.0	11.2
1986	78.5	7.0	63.3	9.2	47.8	11.2
1987	80.7	6.2	68.0	8.4	60.2	10.3
1988	64.0	5.5	76.0	7.3	68.3	8.6
1989	86.6	5.3	84.1	7.0	67.7	7.2
1990	90.8	5.6	91.5	7.7	81.7	6.9
1991	95.6	6.8	100.1	9.8	90.5	8.8
1992	100.0	7.5	100.0	10.6	100.0	10.1
1993	102.7	6.9	95.5	10.7	88.7	10.5
1994	105.6	6.1	91.7	9.4	92.3	9.7
1995	107.9	5.6	93.3	8.5	95.9	8.7
1996	109.3	5.4	93.1	8.7	95.6	8.2
1997	111.4	4.9	94.4	8.2	103.3	7.0
1998	117.3	4.5	90.6	7.5	109.8	6.3
1999	123.2	4.9	91.9	5.7	112.2	6.1

7．继续习题 6，请用普通最小二乘法与广义最小二乘法估计固定影响变系数模型，并对两种估计方法所得估计结果进行比较。

第八章　时间序列计量经济学模型

8

在第一章中已提到，经济分析中所用的三大类重要数据中，时间序列数据是最常见，也是最重要的一类数据。因此，对时间序列数据的分析也就成了计量经济分析最为重要的内容之一。迄今为止，对时间序列的分析是通过建立以因果关系为基础的结构模型进行的。而无论是单方程计量经济学模型还是联立方程计量经济学模型，这种分析背后有一个隐含的假设，即这些数据是**平稳的**(stationary)。否则的话，通常的 t，F 等假设检验则不可信。在经典回归分析中，通过假设样本观测点趋于无穷时，解释变量 X 任何时刻观测值的方差趋于有界常数，给出了 X 平稳性的一个重要条件。这样，既为大样本下的统计推断奠定了基础，也使得所考察的时间序列更靠近平稳性这一假设。

涉及时间序列数据的另一问题是**虚假回归**(spurious regression)或**伪回归**，即如果有两列时间序列数据表现出一致的变化趋势(非平稳的)，即使它们之间没有任何经济关系，若进行回归也可表现出较高的可决系数。在现实经济生活中，实际的时间序列数据往往是非平稳的，而且主要的经济变量，如消费、收入往往表现为一致地上升或下降。这样，仍然通过前面的因果关系模型进行分析，一般不会得到有意义的结果。时间序列分析模型方法就是在这样的情况下，以通过揭示时间序列自身的变化规律为主线而发展起来的全新的计量经济学方法论。时间序列分析已组成现代计量经济学的重要内容，并广泛应用于经济分析与预测当中。格兰杰和恩格尔因为在该领域的突出贡献而获得 2003 年诺贝尔经济学奖。

§8.1　时间序列的平稳性及其检验

一、时间序列数据的平稳性

时间序列分析中首先遇到的问题是关于时间序列数据的平稳性问题。假定某个时间序列是由某一**随机过程**(stochastic process)生成的，即假定时间序列 $\{X_t\}(t=1,2,\cdots)$ 的每个数值都从一个概率分布中随机得到，如果 X_t 满足下列条件：

(1) 均值$\mathrm{E}(X_t)=\mu$，与时间t无关的常数；

(2) 方差$\mathrm{Var}(X_t)=\sigma^2$，与时间$t$无关的常数；

(3) 协方差$\mathrm{Cov}(X_t X_{t+k})=\gamma_k$，只与时期间隔$k$有关，与时间$t$无关的常数。

则称该随机时间序列是（宽）平稳的，而该随机过程是一个**平稳随机过程**(stationary stochastic process)。

例 8.1.1

最简单的随机时间序列X_t是一个具有零均值同方差的独立分布序列：

$$X_t=\mu_t\,,\qquad \mu_t\sim N(0,\sigma^2) \tag{8.1.1}$$

该序列常被称为是一个**白噪声**(white noise)。由于X_t具有相同的均值与方差，且协方差为零，因此由定义知一个白噪声序列是平稳的。

例 8.1.2

另一个简单的随机时间列序被称为**随机游走**(random walk)，该序列由如下随机过程生成：

$$X_t=X_{t-1}+\mu_t \tag{8.1.2}$$

这里，μ_t是一个白噪声。

容易知道该序列有相同的均值$\mathrm{E}(X_t)=\mathrm{E}(X_{t-1})$。为了检验该序列是否具有相同的方差，可假设$X_t$的初值为$X_0$，则易知

$$X_1=X_0+\mu_1$$

$$X_2=X_1+\mu_2=X_0+\mu_1+\mu_2$$

…………

$$X_t=X_0+\mu_1+\mu_2+\cdots+\mu_t$$

假定初始值X_0为常数，μ_t是一个白噪声，因此$\mathrm{Var}(X_t)=t\sigma^2$，即$X_t$的方差与时间$t$有关而非常数，故它是非平稳序列。

然而，对X_t取**一阶差分**(first difference)

$$\Delta X_t=X_t-X_{t-1}=\mu_t \tag{8.1.3}$$

由于μ_t是一个白噪声，则序列$\{\Delta X_t\}$是平稳的。后面将会看到，如果一个时间序列是非平稳的，它常常可通过取差分的方法形成平稳序列。

事实上，随机游走(8.1.2)式是下面称之为**1 阶自回归 AR(1)过程**的特例

$$X_t=\phi X_{t-1}+\mu_t \tag{8.1.4}$$

不难验证，$|\phi|>1$时，该随机过程生成的时间序列是发散的，表现为持续上升($\phi>1$)或持续下降($\phi<-1$)，因此是非平稳的；$\phi=1$时，是一个随机游走

过程，也是非平稳的。在第二节中将证明，只有当$-1<\phi<1$时，该随机过程才是平稳的。

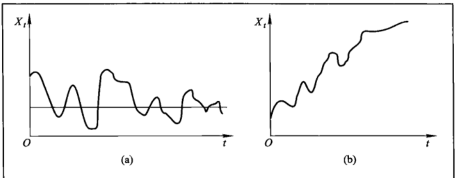
(8.1.4)式又是如下k阶自回归AR(k)过程的特例：

$$X_t=\phi_1 X_{t-1}+\phi_2 X_{t-2}+\cdots+\phi_k X_{t-k}+\mu_t \tag{8.1.5}$$

该随机过程平稳性条件也将在第二节中介绍。

二、平稳性的图示判断

给出一个随机时间序列，首先可通过该序列的时间路径图来粗略地判断它是否是平稳的。平稳时间序列(图 8.1.1(a))在图形上往往表现出一种围绕其均值不断波动的过程；而非平稳时间序列(图 8.1.1(b))则往往表现出在不同的时间段具有不同的均值(如持续上升或持续下降)。



(a) (b)

图 8.1.1 平稳时间序列与非平稳时间序列图

然而，这种直观的图示也常产生误导，因此需要进行进一步的判别。通常的做法是检验样本自相关函数及其图形。首先定义随机时间序列的**自相关函数**(autocorrelation function, ACF)如下：

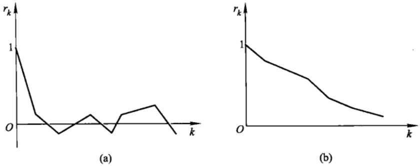
$$\rho_k=\frac{\gamma_k}{\gamma_0} \tag{8.1.6}$$

分子是时间序列滞后k期的协方差，分母是方差，因此自相关函数是关于滞后期k的递减函数。

由于实际上对一个随机过程只有一个实现(样本)，因此，只能计算**样本自相关函数**(sample autocorrelation function)，也称为样本自相关系数。一个时间序列的样本自相关函数定义为

$$r_k=\frac{\sum_{t=1}^{n-k}\left(X_t-\bar{X}\right)\left(X_{t+k}-\bar{X}\right)}{\sum_{t=1}^{n}\left(X_t-\bar{X}\right)^2}, \qquad k=1,2,3,\cdots \tag{8.1.7}$$

易知，随着 k 的增加，样本自相关函数下降且趋于零。但从下降速度来看，平稳序列要比非平稳序列快得多。图 8.1.2 给出了图 8.1.1 中平稳时间序列(a)与非平稳时间序列(b)的样本自相关函数图。



(a)　　(b)

图 8.1.2　平稳时间序列与非平稳时间序列样本相关图

确定样本自相关函数某一数值 r_k 是否足够接近于 0 是非常有用的，因为它可检验对应的自相关函数 ρ_k 的真值是否为 0 的假设。巴特雷特(Bartlett)曾证明，如果时间序列由白噪声过程生成，则对所有的 $k>0$，样本自相关系数近似地服从均值为 0、方差为 $1/n$ 的正态分布，其中 n 为样本数。

也可检验对所有的 $k>0$，自相关系数都为 0 的联合假设，这可通过如下 Q_{LB} 统计量进行：

$$Q_{\mathrm{LB}}=n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k} \tag{8.1.8}$$

该统计量近似地服从自由度为 m 的 χ^2 分布(m 为滞后长度)。因此，如果计算的 Q 值大于显著性水平为 α 的临界值，则有 $1-\alpha$ 的把握拒绝所有 ρ_k ($k>0$)同时为 0 的假设。

例 8.1.3

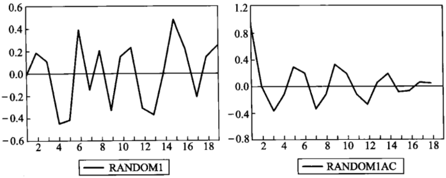
表 8.1.1 中序列 Random1 是通过一个随机过程(随机函数)生成的有 19 个样本的随机时间序列。容易验证该样本序列的均值为 0，方差为 0.078 9。从图形看(图 8.1.3)，它在其样本均值 0 附近上下波动，且样本自相关系数迅速下降到 0，随后在 0 附近波动且逐渐收敛于 0。由于该序列由一个随机过程生成，可以认为不存在序列相关性，因此该序列为一个白噪声。根据巴特雷特曾证明的，该序列的自相关系数应遵从均值为 0、方差为 1/19 的正态分布，因此任一 ρ_k ($k>0$)的 95%的置信区间都将是[−0.449 7,　0.449 7]。可以

看出 $k>0$ 时，r_k 的值确实落在了该区间内，因此可以接受 $\rho_k\,(k>0)$ 为 0 的假设。同样地，从 Q_{LB} 统计量的计算值看，滞后 17 期的计算值为 26.38，未超过 5%显著性水平的临界值 27.58，因此可以接受所有的自相关系数 $\rho_k\,(k>0)$ 都为 0 的假设。因此，该随机过程是一个平稳过程。

序列 Random2 是由(8.1.2)式生成的一个随机游走时间序列样本(图 8.1.4)，其中第 0 项取值为 0，随机项是由 Random1 表示的白噪声。图形表示出该序列具有相同的均值，但从样本自相关图看，虽然自相关系数迅速下降到 0，但随着时间的推移，则在 0 附近波动且呈发散趋势。样本自相关系数显示 $r_1=0.48$，落在了区间[–0.449 7, 0.449 7]之外，因此在 5%的显著性水平下拒绝 ρ_1 的真值为 0 的假设。该随机游走序列是非平稳的。

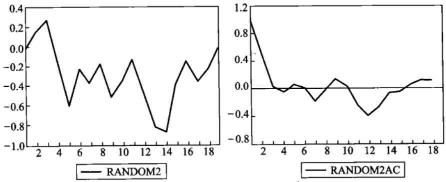
表 8.1.1 一个纯随机序列与随机游走序列的检验

序号	Random1	自相关系数 $r_k\ (k=0,1,\cdots,17)$	Q_{LB}	Random2	自相关系数 $r_k\ (k=0,1,\cdots,17)$	Q_{LB}
1	–0.031	1.000		–0.031	1.000	
2	0.188	–0.051	0.059	0.157	0.480	5.116
3	0.108	–0.393	3.679	0.264	0.018	5.123
4	–0.455	–0.147	4.216	–0.191	–0.069	5.241
5	–0.426	0.280	6.300	–0.616	0.028	5.261
6	0.387	0.187	7.297	–0.229	–0.016	5.269
7	–0.156	–0.363	11.332	–0.385	–0.219	6.745
8	0.204	–0.148	12.058	–0.181	–0.063	6.876
9	–0.340	0.315	15.646	–0.521	0.126	7.454
10	0.157	0.194	17.153	–0.364	0.024	7.477
11	0.228	–0.139	18.010	–0.136	–0.249	10.229
12	–0.315	–0.297	22.414	–0.451	–0.404	18.389
13	–0.377	0.034	22.481	–0.828	–0.284	22.994
14	–0.056	0.165	24.288	–0.884	–0.088	23.514
15	0.478	–0.105	25.162	–0.406	–0.066	23.866
16	0.244	–0.094	26.036	–0.162	0.037	24.004
17	–0.215	0.039	26.240	–0.377	0.105	25.483
18	0.141	0.027	26.381	–0.236	0.093	27.198
19	0.236			0.000		



(a) (b)

图 8.1.3 纯随机序列 Random1 样本图及其样本自相关系数图



(a) (b)

图 8.1.4 随机游走序列 Random2 样本图及其样本自相关系数图

例 8.1.4

检验经居民消费价格指数缩减的中国支出法 GDP 时间序列的平稳性。

在 § 2.6 中的表 2.6.3 中，曾列出了 1978—2006 年间中国支出法 GDP 与居民消费价格指数 CPI（1990=100）的时间序列，并由它们推算出了经 CPI 缩减的实际支出法 GDP 的时间序列 GDPC。图 8.1.5 显示，GDPC 表现出了一个持续上升的过程，即在不同的时间段上，其均值是不同的，因此可初步判断是非平稳的。而且从它们的样本自相关系数的变化看，也是缓

慢下降的，再次表明它们的非平稳性。从滞后 28 期的 Q_{LB} 统计量看，计算值为 231.68，超过了显著性水平为 5%时的临界值 32.67。因此进一步否定了这两个时间序列的自相关系数在滞后一期之后的值全部为 0 的假设。这样，我们得出的结论是 1978—2006 年中国实际 GDP 时间序列是非平稳序列。

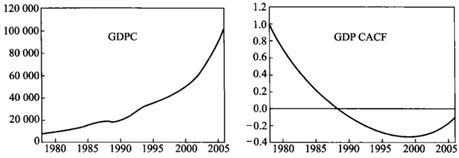


图 8.1.5　1978—2006 年中国实际 GDP 时间序列及其样本自相关图

例 8.1.5

检验例 2.6.2 中关于居民总量消费 Y 与总量可支配收入 X 这两个时间序列的平稳性。

在本章开始时已指出，对时间序列运用传统的回归技术进行回归，是建立在时间序列平稳性这一基本假定基础之上的。因此，在建立总量消费函数的回归方程之前，应对这两个序列的平稳性进行检验。

从图形(图 8.1.6)上容易得出居民总量消费 Y 与总量可支配收入是非平稳的这一结论。从滞后 28 期的 Q_{LB} 统计量看，这两个序列的统计量计算值分别为 240.8 与 267.1，均超过了显著性水平为 5%的临界值 32.67，否定了它们的自相关系数在滞后一期之后的值全部为 0 的假设，再次表明它们的非平稳性。就此来说，运用传统的回归方法建立它们的回归方程是无实际意义的。不过，我们将在第三节中看到，如果两个非平稳时间序列是协整的，则传统的回归结果却是有意义的，而这两个变量的对数序列恰是协整的。

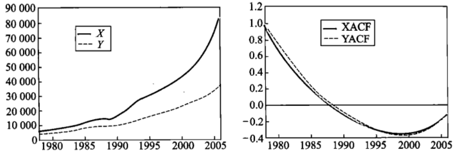


图 8.1.6 1978—2006 年总量消费 *Y* 与总量可支配收入 *X* 时间序列及其样本自相关图

三、平稳性的单位根检验

对时间序列的平稳性除了通过图形直观判断外，运用统计量进行统计检验则是更为准确与重要的。**单位根检验**(unit root test)是统计检验中普遍应用的一种检验方法。

1. DF 检验

我们已知道，随机游走序列

$$X_t = X_{t-1} + \mu_t$$

是非平稳的，其中 μ_t 是白噪声。而该序列可看成是随机模型

$$X_t = \rho X_{t-1} + \mu_t \tag{8.1.9}$$

中参数 $\rho=1$ 时的情形。也就是说，对(8.1.9)式作回归，如果确实发现 $\rho=1$，则称随机变量 X_t 有一个**单位根**。显然，一个有单位根的时间序列就是随机游走序列，而随机游走序列是非平稳的。因此，要判断某时间序列是否是平稳的，可通过(8.1.9)式判断它是否有单位根。这就是时间序列平稳性的单位根检验。

(8.1.9)式可变形成差分形式

$$\begin{aligned}\Delta X_t &= (\rho-1)X_{t-1} + \mu_t \\ &= \delta X_{t-1} + \mu_t\end{aligned} \tag{8.1.10}$$

检验(8.1.9)式是否存在单位根 $\rho=1$，也可通过(8.1.10)式判断是否有 $\delta=0$。

一般地，检验一个时间序列 X_t 的平稳性，可通过检验带有截距项的一阶自回归模型

$$X_t = \alpha + \rho X_{t-1} + \mu_t \tag{8.1.11}$$

中的参数 ρ 是否小于 1，或者说检验其等价变形式

$$\Delta X_t = \alpha + \delta X_{t-1} + \mu_t \tag{8.1.12}$$

中的参数 δ 是否小于 0。

在第二节中我们将证明，(8.1.11)式中的参数 ρ 大于或等于 1 时，时间序列 X_t 是非平稳的，对应于(8.1.12)式，则是 δ 大于或等于 0。因此，针对(8.1.12)式，是在备择假设 H_1：$\delta<0$ 下检验零假设 H_0：$\delta=0$。这可通过普通最小二乘法下的 t 检验完成。

然而，在零假设(序列非平稳)下，即使在大样本下统计量也是有偏误的(向下偏倚)，通常的 t 检验无法使用。迪基(Dickey)和福勒(Fuller)于 1976 年提出了这一情形下 t 统计量服从的分布(这时的 t 统计量也称为 τ 统计量)，即 DF 分布(见表 8.1.2)。因此，检验仍采用普通最小二乘法估计(8.1.12)式，计算 t 统计量的值，并与 DF 分布表中给定显著性水平下的临界值比较。如果 t 统计量的值小于临界值(左尾单侧检验)，这意味着 δ 足够小，则拒绝零假设 H_0：$\delta=0$，认为时间序列不存在单位根，是平稳的。

表 8.1.2 DF 分布临界值表

显著性水平	样本容量					t 分布临界值 ($n=+\infty$)
	25	50	100	500		
0.01	–3.75	–3.58	–3.51	–3.44	–3.43	–2.33
0.05	3.00	–2.93	–2.89	–2.87	–2.86	–1.65
0.10	2.63	–2.60	–2.58	–2.57	–2.57	–1.28

2. ADF 检验

在上述使用(8.1.12)式对时间序列进行平稳性检验中，实际上假定了时间序列是由具有白噪声随机干扰项的一阶自回归过程 AR(1)生成的。但在实际检验中，时间序列可能由更高阶的自回归过程生成，或者随机干扰项并非是白噪声，这样用普通最小二乘法进行估计得到的 t 统计量的渐近分布会受到无关参数的干扰，导致 DF 检验无效。另外，如果时间序列包含有明显的随时间变化的某种趋势(如上升或下降)，则 DF 检验必须保证能够除去这些趋势，否则时间趋势成分会进入随机干扰项。这两种情况都偏离了随机干扰项为白噪声的情形，统计量的渐近分布随之改变。

为了保证 DF 检验中随机干扰项的白噪声特性，迪基和福勒对 DF 检验进行了扩充，形成了 **ADF 检验**(augment Dickey–Fuller test)。ADF 检验是通过下面三个模型完成的：

模型 1： $$\Delta X_t = \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \tag{8.1.13}$$

模型 2： $$\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \tag{8.1.14}$$

模型 3： $$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=1}^{m} \beta_i \Delta X_{t-i} + \varepsilon_t \tag{8.1.15}$$

模型 3 中的 t 是时间变量，代表了时间序列随时间变化的某种趋势(如果有的话)。零假设都是 H_0： $\delta=0$，即存在一个单位根。模型 1 与另两个模型的差别在于是否包含常数项和趋势项。

实际检验时从模型 3 开始，然后模型 2，最后是模型 1。何时检验拒绝零假设，即原序列不存在单位根，为平稳序列，何时可停止检验。否则，就要继续检验，直到检验完模型 1 为止。检验原理与 DF 检验相同，只是对模型 1，2，3 进行检验时，有各自相应的临界值表。表 8.1.3 给出了三个模型所使用的 ADF 分布临界值表。

表 8.1.3 不同模型使用的 ADF 分布临界值表

模型	统计量	样本容量	0.01	0.025	0.05	0.10
1	τ_δ	25	–2.66	–2.26	–1.95	–1.60
		50	–2.62	–2.25	–1.95	–1.61
		100	–2.60	–2.24	–1.95	–1.61
		250	–2.58	–2.23	–1.95	–1.61
		500	–2.58	–2.23	–1.95	–1.61
		>500	–2.58	–2.23	–1.95	–1.61
2	τ_δ	25	–3.75	–3.33	–3.00	–2.62
		50	–3.58	–3.22	–2.93	–2.60
		100	–3.51	–3.17	–2.89	–2.58
		250	–3.46	–3.14	–2.88	–2.57
		500	–3.44	–3.13	–2.87	–2.57
		>500	–3.43	–3.12	–2.86	–2.57
		25	3.41	2.97	2.61	2.20
		50	3.28	2.89	2.56	2.18
		100	3.22	2.86	2.54	2.17
		250	3.19	2.84	2.53	2.16
		500	3.18	2.83	2.52	2.16
		>500	3.18	2.83	2.52	2.16
3	τ_δ	25	–4.38	–3.95	–3.60	–3.24
		50	–4.15	–3.80	–3.50	–3.18

续表

模型	统计量	样本容量	0.01	0.025	0.05	0.10
3	τ_δ	100	–4.04	–3.73	–3.45	–3.15
		250	–3.99	–3.69	–3.43	–3.13
		500	–3.98	–3.68	–3.42	–3.13
		>500	–3.96	–3.66	–3.41	–3.12
	τ_α	25	4.05	3.59	3.20	2.77
		50	3.87	3.47	3.14	2.75
		100	3.78	3.42	3.11	2.73
		250	3.74	3.39	3.09	2.73
		500	3.72	3.38	3.08	2.72
		>500	3.71	3.38	3.08	2.72
	τ_β	25	3.74	3.25	2.85	2.39
		50	3.60	3.18	2.81	2.38
		100	3.53	3.14	2.79	2.38
		250	3.49	3.12	2.79	2.38
		500	3.48	3.11	2.78	2.38
		>500	3.46	3.11	2.78	2.38

从模型 3 转到模型 2 时，需要检验β=0 是否成立，不幸的是这里β的 t 统计量仍然不服从正态分布；从模型 2 转到模型 1 时，需要检验α=0 是否成立，这里α的 t 统计量也不服从正态分布。不过无论是哪种模型，无论真实的数据是一个单位根过程还是一个平稳过程，滞后项的 t 统计量都是服从正态分布的。

一个简单的检验是同时估计出上述三个模型的适当形式，然后通过 ADF 临界值表检验零假设 H_0： δ=0。只要其中有一个模型的检验结果拒绝了零假设，就可以认为时间序列是平稳的。当三个模型的检验结果都不能拒绝零假设时，则认为时间序列是非平稳的。这里所谓模型适当的形式就是在每个模型中选取适当的滞后差分项，以使模型的残差项是一个白噪声(主要保证不存在自相关)。

例 8.1.6

检验 1978—2006 年间中国实际支出法国内生产总值 GDPC 时间序列的平稳性。

经过尝试，模型 3 取了 1 阶滞后：

$$\widehat{\Delta \mathrm{GDPC}}_t = -223.7 - 101.2\mathrm{T} + 0.093\mathrm{GDPC}_{t-1} + 0.733\Delta\mathrm{GDPC}_{t-1}$$

$$(-0.55)\quad(-1.23)\quad(1.97)\qquad\qquad(3.53)$$

通过拉格朗日乘数检验(Lagrange multiplier test，LM)对随机干扰项的自相关性进行检验，LM(1)=1.22，LM(2)=1.56，可见不存在自相关性，因此该模型的设定是正确的。

从 $GDPC_{t-1}$ 的参数估计值看，其 t 统计量的值大于临界值(单尾)，不能拒绝存在单位根的零假设。同时，由于时间项 T 的 t 统计量也小于 ADF 分布表中的临界值(双尾)，因此不能拒绝不存在趋势项的零假设。需进一步检验模型 2。

经试验，模型 2 中滞后项取 1 阶：

$$\Delta\widehat{GDPC}_t = -504.8 + 0.041GDPC_{t-1} + 0.882\Delta GDPC_{t-1}$$

(−1.47)　(1.90)　　(5.15)

LM(1)=0.83　　LM(2)=0.91

由于模型残差项不存在自相关性，因此该模型的设定是正确的。从 $GDPC_{t-1}$ 的参数值看，其 t 统计量为正值，大于临界值，不能拒绝存在单位根的零假设。同时，由于常数项的 t 统计量也小于 ADF 分布表中的临界值，因此不能拒绝不存常数项的零假设。需进一步检验模型 1。

经试验，模型 1 中滞后项取 1 阶：

$$\Delta\widehat{GDPC}_t = 0.020GDPC_{t-1} + 0.985\Delta GDPC_{t-1}$$

(1.20)　　(6.18)

LM(1)=0.12　　LM(2)=0.17

由于模型残差不存在自相关性，因此模型的设定是正确的。从 $GDPC_{t-1}$ 的参数值看，其 t 统计量为正值，大于临界值，不能拒绝存在单位根的零假设。

至此，可断定中国实际支出法 GDP 时间序列是非平稳的。

例 8.1.7

检验例 2.6.2 中关于居民总量消费 Y 与总量可支配收入 X 这两个时间序列的平稳性。

对总量消费 Y 来说，经过尝试，三个模型的适当形式分别为

模型 3：

$$\Delta\hat{Y}_t = -4.516 - 32.75T + 0.118Y_{t-1}$$

（−0.03）（−0.92）（3.48）

LM（1）=2.19　　LM（2）=4.09

模型 2：

$$\Delta\hat{Y}_t = -55.86 + 0.088Y_{t-1}$$

（−0.38）（9.88）

LM（1）=2.47　　LM（2）=4.21

模型 1：

$$\Delta\hat{Y}_t = 0.085Y_{t-1}$$

（18.90）

LM（1）=2.43　　　　LM（2）=4.24

三个模型中 Y_{t-1} 参数估计值的 t 统计量的值均大于各自的临界值，因此不能拒绝存在单位根的零假设，即总量消费序列 Y 是非平稳的。

对于总量可支配收入 X 时间序列来说，三个模型的适当形式为

模型 3：

$$\Delta\hat{X}_t = -205.97 - 65.22T + 0.081X_{t-1} + 0.725\Delta X_{t-1}$$

(−0.47)　(−0.75)　(1.59)　(3.51)

LM(1)=0.14　　　　LM(2)=0.94

模型 2：

$$\Delta\hat{X}_t = -412.83 + 0.046X_{t-1} + 0.799\Delta X_{t-1}$$

(−1.20)　(2.09)　(4.43)

LM(1)=0.08　　　　LM(2)=0.64

模型 1：

$$\Delta\hat{X}_t = 0.030X_{t-1} + 0.864\Delta X_{t-1}$$

(1.70)　(4.97)

LM(1)=0　　　　LM(2)= 0.281

三个模型中 X_{t-1} 的参数估计值的 t 统计量的值均比 ADF 临界值表中各自的临界值大，不能拒绝该时间序列存在单位根的假设，因此可判断人均居民消费序列是非平稳的。

四、单整、趋势平稳与差分平稳随机过程

1. 单整

随机游走序列

$$X_t = X_{t-1} + \mu_t$$

经差分后等价地变形为

$$\Delta X_t = X_t - X_{t-1} = \mu_t \tag{8.1.16}$$

由于 μ_t 是一个白噪声，因此差分后的序列$\{\Delta X_t\}$是平稳的。

如果一个时间序列经过一次差分变成平稳的，就称原序列是 **1 阶单整**(integrated of 1)序列，记为 $I(1)$ 。一般地，如果一个时间序列经过 d 次差分后变成平稳序列，则称原序列是 ***d* 阶单整**(integrated of *d*)序列，记为 $I(d)$ 。显然，$I(0)$ 代表平稳时间序列。

现实生活中，只有少数经济指标的时间序列表现为平稳的，如利率，而大多数指标的时间序列是非平稳的，如一些存量指标常常是 2 阶单整的，以不变价格表示的流量指标，如消费额、收入等常表现为 1 阶单整。大多数非平稳的时间序列一般可通过一次或多次差分变为平稳的。但也有一些时间序列，无论经过多少次差分，都不能变为平稳的。这种序列被称为**非单整的**(nonintegrated)。

例 8.1.8

检验例 8.1.4 中中国实际支出法 GDP 的单整性。

经过试算，发现中国实际支出法 GDP 是 2 阶单整的，适当的检验模型为

$$\Delta^3\widehat{\mathrm{GDPC}}_t = -0.535\Delta^2\mathrm{GDPC}_{t-1}$$
$$(-2.88)$$
$$R^2=0.244\,6,\quad \mathrm{LM}(1)=0.00,\quad \mathrm{LM}(2)=0.00$$

例 8.1.9

检验例 8.1.5 中中国居民总量消费 Y 与总量可支配收入 X 的单整性。

经过试算，发现中国总量可支配收入 X 是 2 阶单整的，适当的检验模型为

$$\Delta^3\hat{X}_t = -0.773\Delta^2 X_{t-1}$$
$$(-3.91)$$
$$R^2=0.377\,7,\quad \mathrm{LM}(1)=0.00,\quad \mathrm{LM}(2)= 0.00$$

同样地，居民总量消费 Y 也是 2 阶单整的，适当的检验模型为

$$\Delta^3\hat{Y}_t = -1.047\Delta^2 Y_{t-1}$$
$$(-5.02)$$
$$R^2=0.501\,3,\quad \mathrm{LM}(1)=0.00,\quad \mathrm{LM}(2)= 0.00$$

2. 趋势平稳与差分平稳随机过程

前面已指出，一些非平稳的经济时间序列往往表现出共同的变化趋势，而这些序列间本身不一定有直接的关联关系，这时对这些数据进行回归，尽管有较高的 R^2，但其结果是没有任何实际意义的。这种现象我们称之为**虚假回归**。例如用中国的劳动力时间序列与美国的 GDP 时间序列作回归，会得到较高的 R^2，但不能认为两者有直接的关联关系，而只不过它们有共同的趋势罢了，这种回归结果我们认为是虚假的。

为了避免这种虚假回归的产生，通常的做法是引入作为趋势变量的时间，这样包含有时间趋势变量的回归，可以消除这种趋势性的影响。然而这种做法，只有当趋势性变量是**确定性的**而非**随机性的**，才会是有效的。换言之，一个包含有某种确定性趋势的非平稳时间序列，可以通过引入表示这一确定性趋势的趋势变量，而将确定性趋势分离出来。

考虑如下的含有 1 阶自回归的随机过程：

$$X_t = \alpha + \beta t + \rho X_{t-1} + \mu_t \tag{8.1.17}$$

其中 μ_t 是一个白噪声，t 为一个时间趋势。

如果 $\rho=1$，$\beta=0$，则(8.1.17)式成为一个带位移的随机游走过程：

$$X_t = \alpha + X_{t-1} + \mu_t \tag{8.1.18}$$

根据 α 的正负，X_t 表现出明显的上升或下降趋势。这种趋势称为**随机性趋势**(stochastic trend)。

如果 $\rho=0$，$\beta \neq 0$，则(8.1.17)式成为一个带时间趋势的随机变化过程：

$$X_t = \alpha + \beta t + \mu_t \tag{8.1.19}$$

根据 β 的正负，X_t 表现出明显的上升或下降趋势。这种趋势称为**确定性趋势**(deterministic trend)。

如果 $\rho=1$，$\beta \neq 0$，则 X_t 包含确定性与随机性两种趋势。

判断一个非平稳的时间序列，它的趋势是随机性的还是确定性的，可通过 ADF 检验中所用的第 3 个模型(8.1.15)式进行。该模型中已引入了表示确定性趋势的时间变量 t，即分离出了确定性趋势的影响。因此，如果检验结果表明所给时间序列有单位根，且时间变量前的参数显著为零，则该序列显示出随机性趋势；如果没有单位根，且时间变量前的参数显著地异于零，则该序列显示出确定性趋势。

随机性趋势可通过差分的方法消除，如(8.1.18)式可通过差分变换为 $\Delta X_t = \alpha + \mu_t$。该时间序列 X_t 称为**差分平稳过程**(difference stationary process)；而确定性趋势无法通过差分的方法消除，只能通过除去趋势项消除，如(8.1.19)式可通过除去 βt 变换为 $X_t - \beta t = \alpha + \mu_t$，其中 μ_t 是平稳的，因此 X_t 称为**趋势平稳过程**(trend stationary process)。

最后需要说明的是，趋势平稳过程代表了一个时间序列长期稳定的变化过程，因而用于进行长期预测是更为可靠的。

§ 8.2　随机时间序列分析模型

在讨论了平稳时间序列的重要性之后，接下来的一个实际问题就是如何建立一个平稳时间序列的模型，以及如何利用所建的模型进行预测。与经典回归分析不同的是，我们这里所建立的时间序列模型主要不是以不同变量间的因果关系为基础，而是寻找时间序列自身的变化规律。同样地，在预测一个时间序列未来的变化时，我们不再使用一组与之有因果关系的其他变量，而只是用该序列的过去行为来预测未来。

一、时间序列模型的基本概念及其适用性

1．时间序列模型的基本概念

时间序列模型是指仅用它的过去值及随机扰动项所建立起来的模型，其一般形式为

$$X_t = F(X_{t-1}, X_{t-2}, \cdots, \mu_t) \tag{8.2.1}$$

建立具体的时间序列模型，需解决如下三个问题：模型的具体形式、时序变量的滞后期以及随机扰动项的结构。例如，取线性方程、1 期滞后以及白噪声随机扰动项($\mu_t = \varepsilon_t$)，模型将是一个**1 阶自回归过程 AR(1)**:

$$X_t = \varphi X_{t-1} + \varepsilon_t \tag{8.2.2}$$

这里，ε_t 特指白噪声。

一般的 ***p*** **阶自回归过程 AR(*p*)**是

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \mu_t \tag{8.2.3}$$

如果随机扰动项 μ_t 是一个白噪声($\mu_t = \varepsilon_t$)，则称(8.2.3)式为一个纯 **AR(*p*)过程**(pure AR(*p*) process)，记为

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \tag{8.2.4}$$

如果 μ_t 不是一个白噪声，通常认为它是一个 q 阶的**移动平均**(moving average)**过程 MA(*q*)**:

$$\mu_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{8.2.5}$$

(8.2.5)式给出了一个纯 **MA(*q*)过程**(pure MA(*p*) process)。将(8.2.3)式与(8.2.5)式结合，得到一个一般的**自回归移动平均**(autoregressive moving average)**过程 ARMA(*p*,*q*)**:

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{8.2.6}$$

(8.2.6)式表明，一个随机时间序列可以通过一个自回归移动平均过程生成，即该序列可以由其自身的过去或滞后值以及随机干扰项来解释。如果该序列是平稳的，即它的行为并不会随着时间的推移而变化，那么我们就可以通过该序列过去的行为来预测未来。这也正是随机时间序列分析模型的优势所在。

2．时间序列分析模型的适用性

迄今为止，对一个时间序列 X_t 的变动进行解释或预测，是通过某个单一方程回归模型或联立方程回归模型进行的，由于它们以因果关系为基础，且具有一定的模型结构，因此也常称为**结构式模型**。然而，如果 X_t 波动的主要原因可能是我们无法解释的因素，如气候、消费者偏好的变化，则利用结构式模型来解释 X_t 的变动就比较困难或不可能，因为要取得相应的量化数据，并建立令人

满意的回归模型是很困难的。有时，即使能估计出一个较为满意的因果关系回归方程，但由于对某些解释变量未来值的预测本身就非常困难，甚至比预测被解释变量的未来值更困难，这时因果关系的回归模型及其预测技术就不适用了。

在这些情况下，我们采用另一条预测途径：通过时间序列的历史数据，得出关于其过去行为的有关结论，进而对时间序列未来行为进行推断。例如，时间序列过去是否有明显的增长趋势，如果增长趋势在过去的行为中占主导地位，能否认为它也会在未来的行为里占主导地位，或者时间序列显示出循环周期性行为，我们能否利用过去的这种行为来外推它的未来走向。随机时间序列分析模型，就是要通过序列过去的变化特征来预测未来的变化趋势。

使用时间序列分析模型的另一个原因在于，如果经济理论正确地阐释了现实经济结构，则这一结构可以写成类似于(8.2.6)式的时间序列分析模型的形式。例如，对于如下最简单的宏观经济模型：

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_t$$
$$Y_t = C_t + I_t$$

这里，C，I，Y 分别表示消费、投资与国民收入。C 与 Y 作为内生变量，它们的运动是由作为外生变量的 I 的运动及随机干扰项 μ_t 的变化决定的。上述模型可作变形如下：

$$C_t = \frac{\alpha_2}{1-\alpha_1} C_{t-1} + \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1}{1-\alpha_1} I_t + \frac{1}{1-\alpha_1} \mu_t$$
$$Y_t = \frac{\alpha_2}{1-\alpha_1} Y_{t-1} + \frac{\alpha_0}{1-\alpha_1} + \frac{1}{1-\alpha_1} I_t - \frac{\alpha_2}{1-\alpha_1} I_{t-1} + \frac{1}{1-\alpha_1} \mu_t$$

两个方程等式右边除去第一项外的剩余部分可看成一个综合性的随机干扰项，其特征依赖于投资项 I 的行为。如果 I 是一个白噪声，则消费序列 C 就成为一个 1 阶自回归过程 AR(1)，而收入序列 Y 就成为一个(1，1)阶的自回归移动平均过程 ARMA(1,1)。

二、随机时间序列模型的平稳性条件

自回归移动平均模型(ARMA)是随机时间序列分析模型的普遍形式，自回归模型(AR)和移动平均模型(MA)是它的特殊情况。关于这几类模型的研究，是时间序列分析的重点内容，主要包括模型的平稳性分析、模型的识别和模型的估计。

1．AR(p)模型的平稳性条件

随机时间序列模型作为随机过程的描述，它的平稳性与该随机过程的平稳性是等价的，因此，可通过它所生成的随机时间序列的平稳性来判断。如果一

个 p 阶自回归模型 AR(p)生成的时间序列是平稳的，就说该 AR(p)模型是平稳的，否则，就说该 AR(p)模型是非平稳的。

考虑 p 阶自回归模型 AR(p)：

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \tag{8.2.4}$$

引入**滞后算子**(lag operator)$\boldsymbol{L}$：

$$LX_t = X_{t-1}\text{，}\quad L^2 X_t = X_{t-2}\text{，}\quad \cdots\text{，}\quad L^p X_t = X_{t-p}$$

(8.2.4)式变换为

$$(1-\varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)X_t = \varepsilon_t$$

记 $\varPhi(L) = (1-\varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)$，则称多项式方程

$$\varPhi(z) = (1-\varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p) = 0$$

为 AR(p)的**特征方程**(characteristic equation)。可以证明，如果该特征方程的所有根在单位圆外(根的模大于 1)，则 AR(p)模型是平稳的。

例 8.2.1

AR(1)模型的平稳性条件。对 1 阶自回归模型 AR(1)：

$$X_t = \varphi X_{t-1} + \varepsilon_t$$

方程两边平方再求数学期望，得到 X_t 的方差

$$\mathrm{E}(X_t^2) = \varphi^2 \mathrm{E}(X_{t-1}^2) + \mathrm{E}(\varepsilon_t^2) + 2\varphi \mathrm{E}(X_{t-1}\varepsilon_t)$$

由于 X_t 仅与 ε_t 相关，因此，$\mathrm{E}(X_{t-1}\varepsilon_t)$ =0。如果该模型稳定，则有 $\mathrm{E}(X_t^2) = \mathrm{E}(X_{t-1}^2)$，从而上式可变换为

$$\gamma_0 = \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$$

在平稳条件下，该方差是一个非负的常数，从而有 $|\varphi|$ <1。而 AR(1)的特征方程

$$\varPhi(z) = 1 - \varphi z = 0$$

的根为

$$z = 1/\varphi$$

AR(1)稳定，即 $|\varphi|$<1，意味着特征根大于 1。

例 8.2.2

AR(2)模型的平稳性。对 AR(2)模型

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t \tag{8.2.7}$$

方程两边同乘以 X_t，再取期望得

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \mathrm{E}(X_t\varepsilon_t)$$

由(8.2.7)式知

$$\mathrm{E}(X_t\varepsilon_t) = \varphi_1\mathrm{E}(X_{t-1}\varepsilon_t) + \varphi_2\mathrm{E}(X_{t-2}\varepsilon_t) + \mathrm{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$$

于是

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \sigma_\varepsilon^2 \tag{8.2.8}$$

同样地，由(8.2.7)式还可得到

$$\begin{aligned} \gamma_1 &= \varphi_1\gamma_0 + \varphi_2\gamma_1 \\ \gamma_2 &= \varphi_1\gamma_1 + \varphi_2\gamma_0 \end{aligned} \tag{8.2.9}$$

于是方差为

$$\gamma_0 = \frac{(1-\varphi_2)\sigma_\varepsilon^2}{(1+\varphi_2)(1-\varphi_1-\varphi_2)(1+\varphi_1-\varphi_2)}$$

由平稳性的定义，该方差必须是一个正常数，于是有

$$\varphi_1+\varphi_2<1, \qquad \varphi_2-\varphi_1<1, \qquad |\varphi_2|<1$$

这就是 AR(2)的平稳性条件，或称为平稳域，它是一个顶点分别为(–2,–1)，(2,–1)，(0,1)的三角形(图 8.2.1)。

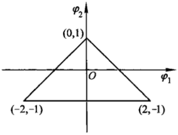


图 8.2.1 AR(2)模型的平稳域

AR(2)模型(8.2.7)式对应的特征方程 $1-\varphi_1 z-\varphi_2 z^2=0$ 的两个根 z_1，z_2 满足

$$z_1 z_2 = -\frac{1}{\varphi_2}, \qquad z_1+z_2 = -\frac{\varphi_1}{\varphi_2}$$

解出 φ_1，φ_2：

$$\varphi_1 = \frac{z_1+z_2}{z_1 z_2}, \qquad \varphi_2 = -\frac{1}{z_1 z_2}$$

由 AR(1)的平稳性，$|\varphi_2|=\dfrac{1}{|z_1||z_2|}<1$，则至少有一个根的模大于 1，不妨设 $|z_1|>1$，有

$$\varphi_1+\varphi_2=\frac{z_1+z_2}{z_1z_2}-\frac{1}{z_1z_2}=1-\left(1-\frac{1}{z_1}\right)\left(1-\frac{1}{z_2}\right)<1$$

$$\left(1-\frac{1}{z_1}\right)\left(1-\frac{1}{z_2}\right)>0$$

于是$|z_2|>1$。由 $\varphi_2-\varphi_1<1$ 可推出同样的结果。

对高阶自回归模型 AR(p)来说，多数情况下没有必要直接计算其特征方程的特征根，但有一些有用的规则可用来检验高阶自回归模型的稳定性：

(1) AR(p)模型稳定的必要条件是

$$\varphi_1+\varphi_2+\cdots+\varphi_p<1$$

(2) 由于 $\varphi_i(i=1,2,\cdots,p)$ 可正可负，AR(p)模型稳定的充分条件是

$$|\varphi_1|+|\varphi_2|+\cdots+|\varphi_p|<1$$

2．MR(q)模型的平稳性

对于移动平均模型 MR(q)：

$$X_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q} \tag{8.2.10}$$

其中 ε_t 是一个白噪声，于是

$$\mathrm{E}(X_t)=\mathrm{E}(\varepsilon_t)-\theta_1\mathrm{E}(\varepsilon_{t-1})-\cdots-\theta_q\mathrm{E}(\varepsilon_{t-q})=0$$

$$\gamma_0=\mathrm{Var}\left(X_t\right)=(1+\theta_1^2+\cdots+\theta_q^2)\sigma_\varepsilon^2$$

$$\gamma_1=\mathrm{Cov}(X_t,X_{t-1})=(-\theta_1+\theta_1\theta_2+\theta_2\theta_3+\cdots+\theta_{q-1}\theta_q)\sigma_\varepsilon^2$$

…………

$$\gamma_{q-1}=\mathrm{Cov}(X_t,X_{t-q+1})=(-\theta_{q-1}+\theta_1\theta_q)\sigma_\varepsilon^2$$

$$\gamma_q=\mathrm{Cov}(X_t,X_{t-q})=-\theta_q\sigma_\varepsilon^2$$

当滞后期大于 q 时，X_t 的自协方差系数为 0。因此，有限阶移动平均模型总是平稳的。

3．ARMA(p,q)模型的平稳性

由于 ARMA(p,q)模型是 AR(p)模型与 MA(q)模型的组合：

$$X_t=\varphi_1X_{t-1}+\cdots+\varphi_pX_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$$

而 MA(q)模型总是平稳的，因此 ARMA(p,q)模型的平稳性取决于 AR(p)部分的平稳性。当 AR(p)部分平稳时，则该 ARMA(p,q)模型是平稳的；否则，不是平稳的。

由于随机时间序列总是由某个随机过程或随机模型生成的，因此一个平稳的时间序列总可以找到生成它的平稳的随机过程或模型。同时，由§8.1 的内容可知，一个非平稳的随机时间序列可以通过差分的方法将它变换为平稳的，对差分后平稳的时间序列也可找出对应的平稳随机过程或模型。因此，如果我们将一个非平稳时间序列通过 d 次差分，将它变为平稳的，然后用一个平稳的 ARMA(p,q) 模型作为它的生成模型，则我们就说该原始时间序列是一个**自回归单整移动平均**(autoregressive integrated moving average)**时间序列**，记为 ARIMA(p,d,q)。例如，一个 ARIMA(2,1,2)时间序列在它成为平稳序列之前先得差分一次，然后用一个 ARMA(2,2)模型作为它的生成模型。当然，一个 ARIMA(p,0,0)过程表示了一个纯 AR(p)平稳过程；一个 ARIMA(0,0, q)表示一个纯 MA(q)平稳过程。

三、随机时间序列模型的识别

所谓随机时间序列模型的识别，就是对于一个平稳的随机时间序列，找出生成它的合适的随机过程或模型，即判断该时间序列是遵循纯 AR 过程，还是遵循纯 MA 过程或 ARMA 过程。所使用的工具主要是时间序列的自相关函数(Autocorrelation Function，ACF)及偏自相关函数(Partial Autocorrelation Function，PACF)。

1. AR(p)过程

(1) 自相关函数 ACF

1 阶自回归模型 AR(1)：

$$X_t = \varphi X_{t-1} + \varepsilon_t$$

的 k 阶滞后自协方差为

$$\gamma_k = \mathrm{E}\left[X_{t-k}(\varphi X_{t-1} + \varepsilon_t)\right] = \varphi\gamma_{k-1} = \varphi^k \gamma_0 \qquad (k = 1,2,\cdots)$$

因此，AR(1)模型的自相关函数为

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi^k \qquad (k = 1,2,\cdots)$$

由 AR(1)的稳定性知$|\varphi|<1$，因此，$k \to +\infty$ 时，ρ_k 呈指数型衰减，直到零。这种现象称为**拖尾**或称 AR(1)有**无穷记忆**(infinite memory)。注意，$\varphi<0$ 时，衰减呈振荡状。

在例 8.2.2 中，我们曾得到过 2 阶自回归模型 AR(2)：

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$$

的方差 γ_0 以及滞后 1 期与 2 期的自协方差 γ_1, γ_2。类似地可写出一般的 k 期滞后自协方差

$$\gamma_k = \mathrm{E}\left[X_{t-k}(\varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t)\right] = \varphi_1\gamma_{k-1} + \varphi_2\gamma_{k-2} \qquad (k = 2,3,\cdots)$$

于是 AR(2)的 k 阶自相关函数为

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} \qquad (k = 2,3,\cdots)$$

PDG

其中 $\rho_1=\varphi_1/(1-\varphi_2)$，$\rho_0=1$。

同样地，如果AR(2)是平稳的，则由 $\varphi_1+\varphi_2<1$ 知 $|\rho_k|$ 衰减趋于零，呈拖尾状。至于衰减的形式，要看AR(2)特征根的实虚性，若为实根，则呈单调或振荡型衰减；若为虚根，则呈正弦波型衰减。

一般地，p 阶自回归模型AR(p)：

$$X_t=\varphi_1X_{t-1}+\varphi_2X_{t-2}+\cdots+\varphi_pX_{t-p}+\varepsilon_t$$

的 k 期滞后方差为

$$\begin{aligned}\gamma_k&=\mathrm{E}\left[X_{t-k}(\varphi_1X_{t-1}+\varphi_2X_{t-2}+\cdots+\varphi_pX_{t-p}+\varepsilon_t)\right]\\&=\varphi_1\gamma_{k-1}+\varphi_2\gamma_{k-2}+\cdots+\varphi_p\gamma_{k-p}\end{aligned}\tag{8.2.11}$$

从而有自相关函数

$$\rho_k=\varphi_1\rho_{k-1}+\varphi_2\rho_{k-2}+\cdots+\varphi_p\rho_{k-p}\tag{8.2.12}$$

可见，无论 k 有多大，ρ_k 的计算均与其1到 p 阶滞后的自相关函数有关，因此呈拖尾状。如果AR(p)是平稳的，则 $|\rho_k|$ 递减且趋于零。

事实上，自相关函数(8.2.12)式是一个 p 阶差分方程，其通解为 $\rho_k=\sum_{i=1}^{p}C_i\lambda_i^k$，其中 $z_i=1/\lambda_i$ 是AR(p)特征方程 $\Phi(z)=0$ 的特征根，由AR(p)平稳的条件知，$|z_i|>1$ 或 $|\lambda_i|<1$，因此，当 λ_i 均为实数根时，ρ_k 呈几何型衰减(单调或振荡)；当存在虚数根时，则一对共轭复根构成通解中的一个阻尼正弦波项，ρ_k 呈正弦波型衰减。

(2) 偏自相关函数PACF

自相关函数ACF(k)给出了 X_t 与 X_{t-k} 的总体相关性，但总体相关性可能掩盖了变量间完全不同的隐含关系。例如，在AR(1)随机过程中，X_t 与 X_{t-2} 间有相关性可能主要是由于它们各自与 X_{t-1} 间的相关性带来的，$\rho_2=\varphi^2=\rho_1^2=\mathrm{E}(X_tX_{t-1})\mathrm{E}(X_{t-1}X_{t-2})$，即自相关函数中包含了这种所有的“间接”相关。与之相反，X_t 与 X_{t-k} 间的**偏自相关函数**则是消除了中间变量 $X_{t-1},\cdots,X_{t-k+1}$ 带来的间接相关后的直接相关性，它是在已知序列值 $X_{t-1},\cdots,X_{t-k+1}$ 的条件下，X_t 与 X_{t-k} 间关系的度量。因此，在AR(1)中，从 X_t 中去掉 X_{t-1} 的影响，则只剩下随机干扰项 ε_t，显然它与 X_{t-2} 无关，因此我们说 X_t 与 X_{t-2} 的偏自相关函数为零，记为 $\rho_2^*=\mathrm{Corr}(\varepsilon_t,X_{t-2})=0$。同样地，在AR($p$)过程中，对所有的 $k>p$，X_t 与 X_{t-k} 间的偏自相关函数为零。可见AR(p)的一个主要特征是，$k>p$ 时，$\rho_k^*=\mathrm{Corr}(\varepsilon_t,X_{t-k})=0$，即 ρ_k^* 在 p 以后是截尾的。

由AR(p)过程的特征可以得到对随机时间序列的识别原则：若 X_t 的偏自相关函数在 p 以后截尾，即 $k>p$ 时，ρ_k^*=0，而它的自相关函数 ρ_k 是拖尾的，则

此序列是自回归 AR(p)序列。

需指出的是，在实际识别时，由于样本偏自相关函数r_k^*是总体偏自相关函数ρ_k^*的一个估计，由于样本的随机性，当$k>p$时，r_k^*不会全为 0，而是在 0 的上下波动。但可以证明，当$k>p$时，r_k^*服从渐近正态分布

$$r_k^* \sim N\left(0, \frac{1}{n}\right)$$

式中 n 表示样本容量。因此，如果计算的r_k^*满足$r_k^* < \frac{2}{\sqrt{n}}$，我们就有 95.5%的把握判断原时间序列在$k>p$之后截尾。

2．MA(q)过程

对 MA(1)过程

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

可容易地写出它的自协方差系数：

$$\begin{aligned} \gamma_0 &= (1+\theta^2)\sigma_\varepsilon^2 \\ \gamma_1 &= -\theta^2 \sigma_\varepsilon^2 \\ \gamma_2 &= \gamma_3 = \cdots = 0 \end{aligned} \tag{8.2.13}$$

于是，MA(1)过程的自相关函数为

$$\begin{aligned} \rho_1 &= \frac{-\theta}{(1+\theta^2)} \\ \rho_2 &= \rho_3 = \cdots = 0 \end{aligned} \tag{8.2.14}$$

可见，当$k>1$时，$\rho_k=0$，即X_t与X_{t-k}不相关，MA(1)自相关函数是截尾的。

MA(1)过程可以等价地写成ε_t关于无穷序列 X_t，X_{t-1}，…的线性组合的形式：

$$\varepsilon_t = X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \cdots$$

或

$$X_t = \theta X_{t-1} + \theta^2 X_{t-2} + \cdots + \varepsilon_t \tag{8.2.15}$$

(8.2.15)式是一个 AR(∞) 过程，它的偏自相关函数非截尾但却趋于零，因此 MA(1) 的偏自相关函数是拖尾但却趋于零的。

注意，(8.2.15)式只有当$|\theta|<1$时才有意义，否则意味着距X_t越远的X值，对X_t的影响越大，显然不符合常理。因此，我们把$|\theta|<1$称为 MA(1)的**可逆性条件**(invertibility condition)或**可逆域**。

一般地，q 阶移动平均过程 MA(q)：

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

的自协方差系数为

$$r_k = \mathrm{E}(X_t X_{t-k}) = \begin{cases} \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2), & 当 k=0 \\ \sigma_\varepsilon^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q), & 当 1 \leqslant k \leqslant q \\ 0, & 当 k>q \end{cases}$$

相应的自相关函数为

$$\rho_k = \frac{r_k}{r_0} = \begin{cases} 1, & 当 k=0 \\ (-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)/(1+\theta_1^2+\cdots+\theta_q^2), & 当 1 \leqslant k \leqslant q \\ 0, & 当 k>q \end{cases} \quad (8.2.16)$$

可见，当 $k>q$ 时，X_t 与 X_{t+k} 不相关，即存在截尾现象。因此，当 $k>q$ 时，$\rho_k=0$ 是 MA(q)的一个特征。换句话说，可以根据自相关函数是否从某一点开始一直为 0 来判断 MA(q)模型的阶。

与 MA(1)相仿，可以验证 MA(q)过程的偏自相关函数是拖尾但趋于零的。因此，可得 MA(q)模型的识别规则：**若随机时间序列的自相关函数截尾，即自 q 以后 $\rho_k=0(k>q)$，而它的偏自相关函数是拖尾的，则此序列是 q 阶滑动平均 MA(q)序列。**

同样需要注意的是，在实际识别时，由于样本自相关函数 r_k 是总体自相关函数 ρ_k 的一个估计，由于样本的随机性，当 $k>q$ 时，r_k 不会全为 0，而是在 0 的上下波动。但可以证明，当 $k>q$ 时，r_k 服从渐近正态分布

$$r_k \sim N\left(0, \frac{1}{n}\right)$$

式中 n 表示样本容量。因此，如果计算的 r_k 满足 $r_k < \dfrac{2}{\sqrt{n}}$，我们就有 95.5%的把握判断原时间序列在 $k>q$ 之后截尾。

3. ARMA(p, q)过程

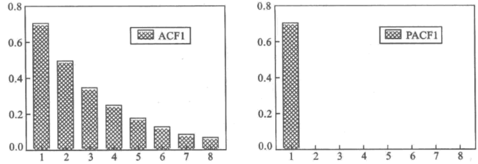
ARMA(p,q)的自相关函数，可以看作 MA(q)的自相关函数和 AR(p)的自相关函数的混合物。当 p=0 时，它具有截尾性质；当 q=0 时，它具有拖尾性质；当 p，q 都不为 0 时，它具有拖尾性质。

从识别上看，通常，ARMA(p,q)过程的偏自相关函数可能在 p 阶滞后前有几项明显的尖柱(spikes)，但从 p 阶滞后项开始逐渐趋向于零；而它的 ACF 则是在 q 阶滞后前有几项明显的尖柱，从 q 阶滞后项开始逐渐趋向于零。

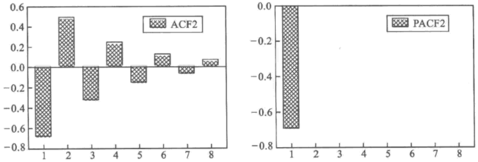
表 8.2.1 列出了 ARMA(p,q)模型 ACF 与 PACF 的变动特征，图 8.2.2 给出了若干示意图(左边是每个模型的 ACF，右边是每个模型的 PACF)。

表 8.2.1　ARMA(*p*,*q*)模型的 ACF 与 PACF 理论模式

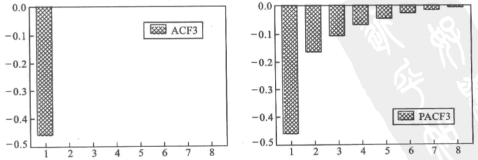
模　　型	ACF	PACF
白噪声	$\rho_k=0$	$\rho_k=0$
AR(p)	衰减趋于零(几何型或振荡型)	p 阶后截尾： $\rho_k^*=0$， $k>q$
MA(q)	q 阶后截尾： $\rho_k=0$， $k>q$	衰减趋于零(几何型或振荡型)
ARMA(p,q)	q 阶后衰减趋于零(几何型或振荡型)	p 阶后衰减趋于零(几何型或振荡型)



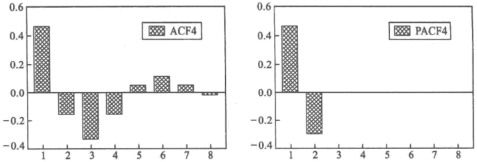
模型1：$X_t=0.7X_{t-1}+\varepsilon_t$



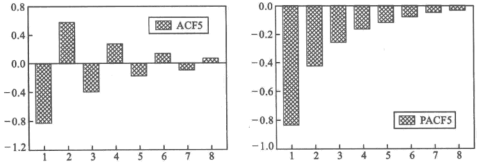
模型2：$X_t=-0.7X_{t-1}+\varepsilon_t$



模型3：$X_t=\varepsilon_t-0.7\varepsilon_{t-1}$



模型4：$X_t = 0.7X_{t-1} - 0.49X_{t-2} + \varepsilon_t$



模型5：$X_t = -0.7X_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-1}$

图 8.2.2 ARMA(p,q)模型的 ACF 与 PACF 理论模式

四、随机时间序列模型的估计

经过模型识别，确定了时间序列分析模型的模型结构和阶数，接着就可以对模型进行参数估计。AR(p)，MA(q)，ARMA(p,q)模型的估计方法较多，大体上分为 3 类：最小二乘估计、矩估计和利用自相关函数的直接估计。下面有选择地加以介绍。

1. AR(p)模型的 Yule Walker 方程估计

在前面讨论 AR(p)模型的识别问题时，我们曾得到

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \cdots + \varphi_p\rho_{k-p}$$

利用 $\rho_k = \rho_{-k}$，得到如下方程组：

$$\rho_1 = \varphi_1 + \varphi_2\rho_1 + \cdots + \varphi_p\rho_{p-1}$$

$$\rho_2 = \varphi_1\rho_1 + \varphi_2 + \cdots + \varphi_p\rho_{p-2}$$

…………

$$\rho_p = \varphi_1\rho_{p-1} + \varphi_2\rho_{p-1} + \cdots + \varphi_p\rho_{p-k}$$

称此方程组为 Yule Walker 方程组。该方程组建立了 AR(p)模型的模型参数 $\varphi_1,\varphi_2,\cdots,\varphi_p$ 与自相关函数 $\rho_1,\rho_2,\cdots,\rho_p$ 的关系。利用实际时间序列提供的信息，首先求得自相关函数的估计值 $\hat{\rho}_1,\hat{\rho}_2,\cdots,\hat{\rho}_p$，然后利用 Yule Walker 方程组，求解模型参数的估计值 $\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$。结果如下：

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \hat{\rho}_0 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix} \tag{8.2.17}$$

由于 $\varepsilon_t = X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p}$，于是

$$\sigma_\varepsilon^2 = E(\varepsilon_t^2) = \gamma_0 - 2\sum_{j=1}^{p}\varphi_j\gamma_j + \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i}$$

由(8.2.11)式容易证明

$$\sum_{j=1}^{p}\varphi_j\gamma_j = \sum_{j=1}^{p}\varphi_j\sum_{i=1}^{p}\varphi_i\gamma_{j-i} = \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i}$$

于是

$$\sigma_\varepsilon^2 = \gamma_0 - \sum_{i,j=1}^{p}\varphi_i\varphi_j\gamma_{j-i}$$

从而可得 σ_ε^2 的估计值

$$\hat{\sigma}_\varepsilon^2 = \hat{\gamma}_0 - \sum_{i,j=1}^{p}\hat{\varphi}_i\hat{\varphi}_j\hat{\gamma}_{j-i} \tag{8.2.18}$$

在具体计算时，总体自相关函数的估计值 $\hat{\rho}_k$ 可用样本自相关函数 r_k 替代。

2. MA(q)模型的矩估计

将 MA(q)模型的自协方差函数中的各个量用估计量代替，得到

$$\hat{\gamma}_k = \begin{cases} \hat{\sigma}_\varepsilon^2(1+\hat{\theta}_1^2+\hat{\theta}_2^2+\cdots+\hat{\theta}_q^2), & \text{当}k=0 \\ \hat{\sigma}_\varepsilon^2(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q), & \text{当}1\leqslant k\leqslant q \\ 0, & \text{当}k>q \end{cases} \tag{8.2.19}$$

利用实际时间序列提供的信息，首先求得自协方差函数的估计值，于是(8.2.19)式是一个包含 q+1 个待估参数 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_q,\hat{\sigma}_\varepsilon^2$ 的非线性方程组，可以用直接法或迭代法求解。常用的迭代方法有线性迭代法和牛顿-拉夫森(Newton-Raphson)迭代法。

(1) MA(1)模型的直接算法

对于 MA(1)模型，(8.2.19)式相应地写成

$$\begin{cases}\hat{\gamma}_0=\hat{\sigma}_\varepsilon^2(1+\hat{\theta}_1^2)\\ \hat{\gamma}_1=-\hat{\sigma}_\varepsilon^2\hat{\theta}_1\end{cases}$$

于是
$$\hat{\theta}_1=-\frac{\hat{\gamma}_1}{\hat{\sigma}_\varepsilon^2}$$

从而有
$$\hat{\sigma}_\varepsilon^4-\hat{\gamma}_0\hat{\sigma}_\varepsilon^2+\hat{\gamma}_1^2=0$$

于是有解

$$\hat{\sigma}_\varepsilon^2=\frac{\hat{\gamma}_0\pm\sqrt{\hat{\gamma}_0^2-4\hat{\gamma}_1^2}}{2}=\frac{\hat{\gamma}_0}{2}(1\pm\sqrt{1-4\hat{\rho}_1^2})$$

$$\hat{\theta}_1=-\hat{\gamma}_1\Big/\hat{\sigma}_\varepsilon^2=-2\hat{\rho}_1\Big/(1\pm\sqrt{1-4\hat{\rho}_1^2})$$

由于参数估计有两组解，可根据可逆性条件$|\theta_1|<1$来判断选取一组。

(2)　MA(q)模型的迭代算法

对于$q>1$的MA(q)模型，一般用迭代算法估计参数。

由(8.2.19)式得

$$\begin{cases}\hat{\sigma}_\varepsilon^2=\dfrac{\hat{\gamma}_0}{1+\hat{\theta}_1^2+\cdots+\hat{\theta}_q^2}\\ \hat{\theta}_k=-\left(\dfrac{\hat{\gamma}_k}{\hat{\sigma}_\varepsilon^2}-\hat{\theta}_1\hat{\theta}_{k+1}-\hat{\theta}_2\hat{\theta}_{k+2}-\cdots-\hat{\theta}_{q-k}\hat{\theta}_q\right)\end{cases}\tag{8.2.20}$$

第一步，给出$\hat{\sigma}_\varepsilon^2,\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_q$的一组初值，如

$$\hat{\sigma}_\varepsilon^2(0)=\hat{\gamma}_0,\qquad \hat{\theta}_1(0)=\hat{\theta}_2(0)=\cdots\hat{\theta}_q(0)=0$$

代入(8.2.20)式，计算出第一次迭代值

$$\hat{\sigma}_\varepsilon^2(1)=\hat{\gamma}_0,\qquad \hat{\theta}_k(1)=-\hat{\gamma}_k/\hat{\gamma}_0$$

第二步，将第一次迭代值代入(8.2.20)式，计算出第二次迭代值

$$\hat{\sigma}_\varepsilon^2(2)=\frac{\hat{\gamma}_0}{1+\hat{\theta}_1^2(1)+\cdots+\hat{\theta}_q^2(1)}$$

$$\hat{\theta}_k(2)=-\frac{\hat{\gamma}_k}{\hat{\gamma}_0-\hat{\theta}_1(1)\hat{\theta}_{k+1}(1)-\cdots-\hat{\theta}_{q-k}(1)\hat{\theta}_q(1)}$$

由此反复迭代下去，直到第m步的迭代值与第$m-1$步的迭代值相差不大时(满足一定的精度)，便停止迭代，并用第m步的迭代结果作为(8.2.20)的近似解。

3．ARMA(p,q)模型的矩估计

在ARMA(p,q)中共有$(p+q+1)$个待估参数，$\varphi_1,\varphi_2,\cdots,\varphi_p$与$\theta_1,\theta_2,\cdots,\theta_q$以及$\sigma_\varepsilon^2$，其估计量计算步骤及公式如下：

第一步，估计$\varphi_1,\varphi_2,\cdots,\varphi_p$。计算公式为

$$\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{pmatrix}=\begin{pmatrix}\hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p+1}\\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p}\\ \vdots & \vdots & & \vdots\\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \cdots & \hat{\rho}_q\end{pmatrix}^{-1}\begin{pmatrix}\hat{\rho}_{q+1}\\ \hat{\rho}_{q+2}\\ \vdots\\ \hat{\rho}_{q+p}\end{pmatrix}$$

其中$\hat{\rho}_k$是总体自相关函数的估计值，可用样本自相关函数r_k代替。

第二步，改写模型，求$\theta_1,\theta_2,\cdots,\theta_q$以及$\sigma_\varepsilon^2$的估计值。将模型

$$X_t=\varphi_1X_{t-1}+\varphi_2X_{t-2}+\cdots+\varphi_pX_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q}$$

改写为

$$X_t-\varphi_1X_{t-1}-\varphi_2X_{t-2}-\cdots-\varphi_pX_{t-p}=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q}\qquad(8.2.21)$$

令

$$\tilde{X}_t=X_t-\hat{\varphi}_1X_{t-1}-\hat{\varphi}_2X_{t-2}-\cdots-\hat{\varphi}_pX_{t-p}$$

于是(8.2.21)式可以写成

$$\tilde{X}_t=\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q}$$

构成一个 MA 模型。根据估计 MA 模型的方法，可以得到$\theta_1,\theta_2,\cdots,\theta_q$以及$\sigma_\varepsilon^2$的估计值。

4．AR(*p*)的最小二乘估计

假设模型(8.2.4)式的参数估计值$\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$已经得到，即有

$$X_t=\hat{\varphi}_1X_{t-1}+\hat{\varphi}_2X_{t-2}+\cdots+\hat{\varphi}_pX_{t-p}+\hat{\varepsilon}_t$$

残差的平方和为

$$S(\hat{\varphi})=\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2=\sum_{t=p+1}^{n}(X_t-\hat{\varphi}_1X_{t-1}-\hat{\varphi}_2X_{t-2}-\cdots-\hat{\varphi}_pX_{t-p})^2\qquad(8.2.22)$$

根据最小二乘原理，所要求的参数估计值$\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$应该使得(8.2.22)式达到极小。所以它们应该是下列方程组的解：

$$\frac{\partial S}{\partial\hat{\varphi}_j}=0$$

即

$$\sum_{t=p+1}^{n}(X_t-\hat{\varphi}_1X_{t-1}-\hat{\varphi}_2X_{t-2}-\cdots-\hat{\varphi}_pX_{t-p})X_{t-j}=0,\qquad j=1,2,\cdots,p\qquad(8.2.23)$$

解该方程组，就可得到待估参数的估计值。

为了与 AR(*p*)模型的 Yule Walker 方程估计进行比较，将(8.2.23)式改写成

$$\frac{\hat{\varphi}_1}{n}\sum_{t=p+1}^{n}X_{t-1}X_{t-j}+\frac{\hat{\varphi}_2}{n}\sum_{t=p+1}^{n}X_{t-2}X_{t-j}+\cdots+\frac{\hat{\varphi}_p}{n}\sum_{t=p+1}^{n}X_{t-p}X_{t-j}=\frac{1}{n}\sum_{t=p+1}^{n}X_tX_{t-j}$$

$$j=1,2,\cdots,p$$

由自协方差函数的定义，并将自协方差函数的估计值

$$\hat{\gamma}_k=\frac{1}{n}\sum_{t=p+1}^{n-k}X_{t+k}X_t$$

代入，上式表示的方程组即为

$$\hat{\varphi}_1\hat{\gamma}_{j-1}+\hat{\varphi}_2\hat{\gamma}_{j-2}+\cdots+\hat{\varphi}_p\hat{\gamma}_{j-p}=\hat{\gamma}_j\,,\qquad j=1,2,\cdots,p$$

或

$$\hat{\varphi}_1r_{j-1}+\hat{\varphi}_2r_{j-2}+\cdots+\hat{\varphi}_pr_{j-p}=r_j\,,\qquad j=1,2,\cdots,p$$

解该方程组，得到

$$\begin{pmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{pmatrix}=\begin{pmatrix}r_0 & r_1 & \cdots & r_{p-1}\\ r_1 & r_0 & \cdots & r_{p-2}\\ \vdots & \vdots & & \vdots\\ r_{p-1} & r_{p-2} & \cdots & r_0\end{pmatrix}^{-1}\begin{pmatrix}r_1\\ r_2\\ \vdots\\ r_p\end{pmatrix}$$

即为参数的最小二乘估计。与(8.2.17)式的估计值比较发现，当 n 足够大时，二者是相似的。σ_ε^2 的估计值为

$$\hat{\sigma}_\varepsilon^2=\frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\varepsilon}_t^2=\frac{S(\hat{\varphi})}{n-p}$$

需要说明的是，在上述模型的平稳性、识别与估计的讨论中，ARMA(p,q)模型中均未包含常数项。如果包含常数项，该常数项并不影响模型的原有性质，因为通过适当的变形，可将包含常数项的模型转换为不含常数项的模型。下面以一般的 ARMA(p,q)模型为例说明。

对含有常数项的模型

$$X_t=\alpha+\varphi_1X_{t-1}+\cdots+\varphi_pX_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$$

方程两边同减去 $\dfrac{\alpha}{1-\varphi_1-\cdots-\varphi_p}$，则可得到

$$x_t=\varphi_1x_{t-1}+\cdots+\varphi_px_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q}$$

其中，$x_i=X_i-\dfrac{\alpha}{1-\varphi_1-\cdots-\varphi_p},i=t,t-1,\cdots,t-p$。

五、随机时间序列模型的检验

时间序列模型的识别与估计过程往往是同步进行的。由于在实际识别 ARMA(p,q)模型时，滞后项阶数的选择并不是一件容易的事，因此模型在识别与估计之后还需进行检验。由于 ARMA(p,q)模型的识别与估计是在假设随机干扰项 ε_t 是一个白噪声的基础上进行的，因此，如果估计的模型确认正确的话，

残差 $\hat{\varepsilon}_t$ 应代表一个白噪声序列。如果通过所估计的模型计算的样本残差不代表一个白噪声，则说明模型的识别与估计有误，需重新识别与估计。在实际检验时，主要检验残差序列是否存在自相关。可用§8.1 提出的 Q_{LB} 统计量进行 χ^2 检验。因此，在给定显著性水平下，可计算不同滞后期的 Q_{LB} 值，通过与 χ^2 分布表中的相应临界值比较，来检验是否拒绝残差序列为白噪声的假设，若某一个 Q_{LB} 大于相应临界值，则应拒绝所估计的模型，需重新识别与估计。

另外一个遇到的问题是，在实际识别 ARMA(p,q)模型时，需多次反复尝试，有可能存在不止一组(p,q)值都能通过识别检验。显然，增加 p 与 q 的阶数，可增加拟合优度，但却同时降低了自由度。因此，对可能的适当的模型，存在着模型的“简洁性”与模型的拟合优度的权衡选择问题。常用的模型选择的判别标准有第三章提到的赤池信息准则(AIC)与施瓦茨准则(SC)。在选择可能的模型时，AIC 与 SC 越小越好，显然如果添加的滞后项没有解释能力，则对残差平方和 RSS 值的减小没有多大帮助，却增加待估参数的个数，因此使得 AIC 或 SC 的值增加。需注意的是在不同模型间进行比较时，必须选取相同的时间段。

例 8.2.3

中国实际支出法 GDP 的 ARMA(p,q)模型估计。

在§8.1 中已知，中国实际支出法国内生产总值 GDPC 是非平稳的，但它的 2 阶差分是平稳的，即 GDPC 是 $I(2)$ 时间序列。因此，可以对经过 2 阶差分后的 GDP 建立适当的 ARMA(p,q)模型。记 GDPC 经 2 阶差分后的新序列为 GDPC2，图 8.2.3 给出了该新序列的样本自相关函数图与偏自相关函数图及相应滞后期的函数值。可以看出，样本自相关函数与偏自相关函数图形都是在滞后 1 期时迅速趋于 0，而且在 $k \geqslant 1$ 以后，样本自相关函数值 r_k 与偏自相关函数值 r_k^* 都落在了 95%的置信区间[−0.3772, 0.3772]的内部，因此在 5%的显著性水平下不拒绝从滞后 1 期开始 $\rho_k=0$，$\rho_k^*=0$ 的假设。据此，可认为 GDPC2 是一个白噪声，从而可建立 GDPC2 的纯 MA(0)模型，或建立 GDPC 的 ARIMA(0,2,0)模型：

$$\mathrm{GDPC2}_t=\varepsilon_t \text{或} \Delta^2\mathrm{GDPC}_t=\varepsilon_t$$

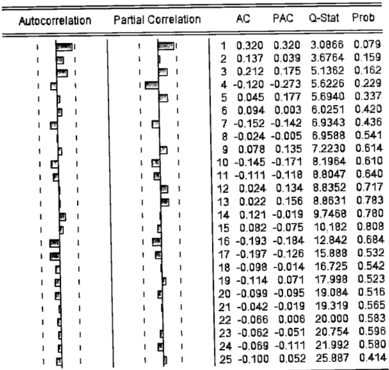
当然，由于在滞后 1 期时，$r_1=r_1^*=0.32$，接近于 5%显著性水平下的临界值 0.37，也可考虑建立 ARMA(1,1)模型，或建立纯 AR(1)模型，或建立纯 MA(1)模型。

若建立纯 AR(1)模型时，由 Yule Walker 方程可得

$$\hat{\varphi}=1^{-1}(0.320)=0.320$$

从而有

$$\widehat{\mathrm{GDPC2}}_t=0.320\mathrm{GDPC2}_{t-1}$$



	AC	PAC	Q-Stat	Prob
1	0.320	0.320	3.0866	0.079
2	0.137	0.039	3.6764	0.159
3	0.212	0.175	5.1362	0.162
4	-0.120	-0.273	5.6226	0.229
5	0.045	0.177	5.6940	0.337
6	0.094	0.003	6.0251	0.420
7	-0.152	-0.142	6.9343	0.436
8	-0.024	-0.005	6.9588	0.541
9	0.078	0.135	7.2230	0.614
10	-0.145	-0.171	8.1964	0.610
11	-0.111	-0.118	8.8047	0.640
12	0.024	0.134	8.8352	0.717
13	0.022	0.156	8.8631	0.783
14	0.121	-0.019	9.7468	0.780
15	0.082	-0.075	10.182	0.808
16	-0.193	-0.184	12.842	0.684
17	-0.197	-0.126	15.888	0.532
18	-0.098	-0.014	16.725	0.542
19	-0.114	0.071	17.998	0.523
20	-0.099	-0.095	19.084	0.516
21	-0.042	-0.019	19.319	0.565
22	-0.066	0.006	20.000	0.583
23	-0.062	-0.051	20.754	0.596
24	-0.069	-0.111	21.992	0.580
25	-0.100	0.052	25.887	0.414

图 8.2.3 GDPC2 序列的样本自相关函数图与偏自相关函数图

用普通最小二乘法回归的结果为

$$\widehat{\mathrm{GDPC2}}_t = 0.465\mathrm{GDPC2}_{t-1}$$

$$(2.50)$$

$$R^2 = 0.008,\ \mathrm{D.W.} = 2.1$$

有时，用普通最小二乘法回归时，也可加入常数项。本例中加入常数项的回归式为

$$\widehat{\mathrm{GDPC2}}_t = 364.9 + 0.340\mathrm{GDPC2}_{t-1}$$

$$(1.69)\quad(1.75)$$

$$R^2 = 0.113,\ \mathrm{D.W.} = 2.1$$

建立纯 MA(1)模型的普通最小二乘估计为

$$\widehat{\mathrm{GDPC2}}_t = \varepsilon_t + 0.485\varepsilon_{t-1}$$

$$(2.83)$$

$$R^2 = 0.011，\ \mathrm{D.W.} = 2.02$$

在建立 ARMA(1，1)模型时，发现 GDPC2 的 1 期滞后项的参数估计值超过 1，不符合常理，因此不再建立 ARMA(1，1)模型。

最后，我们通过上述 ARIMA 模型对中国实际支出法 GDP 进行向后 1 期预测。

由纯 MA(0)模型可知

$$\begin{aligned}\Delta^2\mathrm{GDPC}_t &= \Delta\mathrm{GDPC}_t - \Delta\mathrm{GDPC}_{t-1}\\ &= (\mathrm{GDPC}_t - \mathrm{GDPC}_{t-1}) - (\mathrm{GDPC}_{t-1} - \mathrm{GDPC}_{t-2})\\ &= \mathrm{GDPC}_t - 2\mathrm{GDPC}_{t-1} + \mathrm{GDPC}_{t-2} = \varepsilon_t\end{aligned}$$

即

$$\mathrm{GDPC}_t = 2\mathrm{GDPC}_{t-1} - \mathrm{GDPC}_{t-2} + \varepsilon_t$$

由于 ε_t 表示预测期的随机干扰项，是未知的，可假设为 0，于是 2007 年中国实际支出法 GDP 的预测值为

$$\widehat{\mathrm{GDPC}}_{2007} = 2\times 101\,617.5 - 88\,001.2 = 115\,233.8 \text{（亿元）}$$

由纯 AR(1)模型可知

$$\Delta^2\mathrm{GDPC}_t = \varphi\Delta^2\mathrm{GDPC}_{t-1} + \varepsilon_t$$

由该式容易推出

$$\mathrm{GDPC}_t = (2+\varphi)\mathrm{GDPC}_{t-1} - (1+2\varphi)\mathrm{GDPC}_{t-2} + \varphi\mathrm{GDPC}_{t-3} + \varepsilon_t$$

于是 2007 年中国实际支出法 GDP 的预测值为

$$\begin{aligned}\widehat{\mathrm{GDPC}}_{2007} &= (2+0.320)\times 101\,617.5 - (1+2\times 0.320)88\,001.2 + 0.320\times 76\,095.7\\ &=115\,781.3\text{（亿元）}\end{aligned}$$

由纯 MA(1)模型可知

$$\Delta^2\mathrm{GDPC}_t = \varepsilon_t + 0.485\varepsilon_{t-1}$$

于是 2007 年中国实际支出法 GDP 的预测式为

$$\widehat{\mathrm{GDPC}}_{2007} = 2\mathrm{GDPC}_{2006} - \mathrm{GDPC}_{2005} + 0.485\hat{\varepsilon}_{2006}$$

其中，$\hat{\varepsilon}_{2006}$ 为模型滞后 1 期的相应残差项的估计值，它的计算比较复杂。一般可由软件直接给出 MA 模型的外推预测，Eviews 软件给出的 2007 年 $\Delta^2\mathrm{GDPC}_{2007}$ 的外推预测结果为 253.7，于是 2007 年中国实际支出法 GDP 的预测值为

$$\widehat{\mathrm{GDPC}}_{2007} = 2\mathrm{GDPC}_{2006} - \mathrm{GDPC}_{2005} + 253.7 = 115\,487.5\text{（亿元）}$$

2007 年中国名义支出法 GDP 为 263 242.5 亿元，以 1990 年为基准的居民消费价格指数为 228.1，由此可推得当年中国实际支出法 GDP 为 115 408.6 亿元，据此所计算的上述 ARIMA 模型的预测误差见表 8.2.2。

表 8.2.2 中国实际支出法 GDP 外推预测 单位：亿元

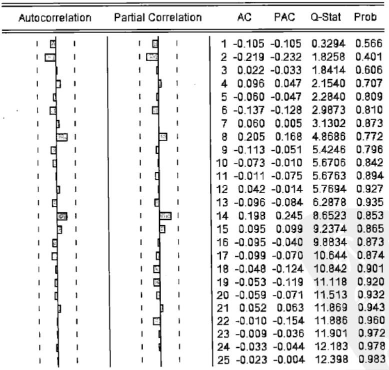
	实际值	MA(0)模型预测值	AR(1)模型预测值	MA(1)模型预测值
115 408.6	相对误差（%）	115 233.8	115 781.3	115 487.5
		–0.15	0.32	0.07

例 8.2.4

中国居民总量消费的 ARMA(p,q)模型

在§8.1 中曾指出，例 2.6.2 中经居民消费价格指数 CPI 调整的中国居民总量消费（Y）与总量可支配收入（X）这两个时间序列是非平稳的，因此不宜直接建立它们的回归方程。由于它们都是 $I(2)$ 时间序列，因此可以建立它们的 ARIMA(p,d,q)模型。这里只建立居民总量消费 Y 的随机时间序列模型。

中国实际居民总量消费序列 Y 经过 2 次差分后的新序列记为 $Y2$，其自相关函数、偏自相关函数图形及有关的函数值见图 8.2.4。



```
Autocorrelation    Partial Correlation        AC     PAC    Q-Stat   Prob

                                        1 -0.105 -0.105   0.3294  0.566
                                        2 -0.219 -0.232   1.8258  0.401
                                        3  0.022 -0.033   1.8414  0.606
                                        4  0.096  0.047   2.1540  0.707
                                        5 -0.060 -0.047   2.2840  0.809
                                        6 -0.137 -0.128   2.9873  0.810
                                        7  0.060  0.005   3.1302  0.873
                                        8  0.205  0.168   4.8686  0.772
                                        9 -0.113 -0.051   5.4246  0.796
                                       10 -0.073 -0.010   5.6706  0.842
                                       11 -0.011 -0.075   5.6763  0.894
                                       12  0.042 -0.014   5.7694  0.927
                                       13 -0.096 -0.084   6.2878  0.935
                                       14  0.198  0.245   8.6523  0.853
                                       15  0.095  0.099   9.2374  0.865
                                       16 -0.095 -0.040   9.8834  0.873
                                       17 -0.099 -0.070   10.644  0.874
                                       18 -0.048 -0.124   10.842  0.901
                                       19 -0.053 -0.119   11.118  0.920
                                       20 -0.059 -0.071   11.513  0.932
                                       21  0.052  0.063   11.869  0.943
                                       22 -0.010 -0.154   11.886  0.960
                                       23 -0.009 -0.036   11.901  0.972
                                       24 -0.033 -0.044   12.183  0.978
                                       25 -0.023 -0.004   12.398  0.983
```

图 8.2.4 $Y2$ 序列的样本自相关函数图与偏自相关函数图

在 5%的显著性水平下，容易验证该序列本身就接近于一个白噪声，因此可建立零阶 MA(0)模型：

$$Y2_t = \varepsilon_t$$

于是，中国居民总量消费 Y 的 ARIMA 模型为

$$Y_t = 2Y_{t-1} - Y_{t-2} + \varepsilon_t$$

由该模型可对 2007 年实际居民总量消费进行外推预测：

$$\hat{Y}_{2007} = 2 \times 36\,811.6 - 33\,214.0 = 40\,409.2 \text{（亿元）}$$

2007 年，中国名义居民消费总量为 93 317.2 亿元，以 1990 年为基准的居民消费价格指数为 228.1，由此可推得 2007 年中国实际居民消费总量为 40 910.7 亿元，可见上述 ARIMA 模型的预测相对误差为–1.2%，该误差比例 2.6.2 中一元回归模型的预测误差 7.1%小得多。

§8.3 协整与误差修正模型

前面已提到，经典回归模型是建立在平稳数据变量基础上的。对于非平稳变量，不能使用经典回归模型，否则会出现虚假回归等诸多问题。由于许多经济变量是非平稳的，这就给经典的回归分析方法带来了很大限制。例如，在前面讨论中国居民消费总支出与可支配总收入变量的例子中，由于它们是非平稳的，就此直接建立回归模型，其结果的可信程度将有所降低。从例 8.2.4 中已经看到，就对 2007 年的预测效果看，例 2.6.2 所给出的传统的因果关系回归模型确实比 ARMA 模型差，因果关系模型预测的相对误差为 7.1%，ARMA 模型的反为–1.2%。当然，正像例 4.2.1 所讨论的那样，例 2.6.2 所给出的因果关系模型既存在着设定偏误的问题，还存在着序列相关的问题，因此预测误差肯定会大一些。可以验证，如果用模型设定更为“正确”的（4.2.24）式进行预测，2007 年居民实际总量消费为 39 624.4 亿元，相对误差减小到–3.1%；如果再用进一步消除了序列相关的模型（4.2.25）式进行预测，得到 2007 年居民实际总量消费为 39 889.2 亿元，相对误差进一步减小到–2.5%。可见这时因果关系模型的效果并不“差”。其原因在于从经济理论上说，可支配收入决定着居民的消费支出水平，而且它们之间存在着长期的稳定关系，即它们之间是协整的(cointegration)。本节中将会看到，具有协整关系的经济变量间具有长期的稳定关系，因此是可以使用经典回归模型方法建立回归模型的。

一、长期均衡关系与协整

经济理论指出，某些经济变量间确实存在着长期均衡关系。这种均衡关系

意味着经济系统不存在破坏均衡的内在机制。如果变量在某时期受到干扰后偏离其长期均衡点，则均衡机制将会在下一期进行调整以使其重新回到均衡状态。

假设 X 与 Y 间的长期“均衡关系”由下面的(8.3.1)式描述

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t \tag{8.3.1}$$

式中 μ_t 是随机干扰项。该均衡关系意味着给定 X 的一个值，Y 相应的均衡值也随之确定为 $\alpha_0 + \alpha_1 X$。在 $t-1$ 期末，存在下述三种情形之一：

(1) Y 等于它的均衡值

$$Y_{t-1} = \alpha_0 + \alpha_1 X_{t-1} \tag{8.3.2}$$

(2) Y 小于它的均衡值

$$Y_{t-1} < \alpha_0 + \alpha_1 X_{t-1}$$

(3) Y 大于它的均衡值

$$Y_{t-1} > \alpha_0 + \alpha_1 X_{t-1}$$

在时期 t，假设 X 有一个变化量 ΔX_t，如果变量 X 与 Y 在时期 t 与 $t-1$ 期末仍满足它们间的长期均衡关系，则 Y 的相应变化量 ΔY_t 由下式给出：

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t \tag{8.3.3}$$

式中，$v_t = \mu_t - \mu_{t-1}$。然而情况往往并非如此。如果 $t-1$ 期末，发生了上述第二种情况，即 Y 的值小于其均衡值，则 Y 的变化往往会比第一种情形下 Y 的变化 ΔY_t 大一些；反之，如果 Y 的值大于其均衡值，则 Y 的变化往往会小于第一种情形下的 ΔY_t。

可见，如果(8.3.1)式正确地提示了 X 与 Y 间的长期稳定的“均衡关系”，则意味着 Y 对其均衡点的偏离从本质上说是“临时性”的。因此，一个重要的假设就是随机干扰项 μ_t 必须是平稳序列。显然，如果 μ_t 有随机性趋势(上升或下降)，则会导致 Y 对其均衡点的任何偏离都会被长期累积下来而不能被消除。

(8.3.1)式中的随机干扰项 μ_t 也被称为**非均衡误差**(disequilibrium error)，它是变量 X 与 Y 的一个线性组合：

$$\mu_t = Y_t - \alpha_0 - \alpha_1 X_t \tag{8.3.4}$$

因此，如果(8.3.1)式所揭示的 X 与 Y 间的长期均衡关系正确的话，(8.3.4)式表述的非均衡误差应是一个平稳时间序列，并且具有零均值，即 μ_t 是具有零均值的 $I(0)$ 序列。

正像前面所指出的，许多经济变量是非平稳的，即它们是 1 阶或更高阶的单整时间序列。但从这里我们已看到，非平稳的时间序列，它们的线性组合也可能成为平稳的。如假设(8.3.1)式中的 X 与 Y 是 $I(1)$ 序列，如果该式所表述的它们间的长期均衡关系成立的话，则意味着由非均衡误差(8.3.4)式给出的线性组合是 $I(0)$ 序列。这时我们称变量 X 与 Y 是**协整的**。

一般地，如果序列 $X_{1t}, X_{2t}, \cdots, X_{kt}$ 都是 d 阶单整的，存在向量 $\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \cdots, \alpha_k)$，使得 $\boldsymbol{Z}_t=\boldsymbol{\alpha X}_t' \sim I(d-b)$，其中，$b>0, \boldsymbol{X}_t=(X_{1t}, X_{2t}, \cdots, X_{kt})'$，则认为序列 $X_{1t}, X_{2t}, \cdots, X_{kt}$ 是 (d,b) 阶协整，记为 $\boldsymbol{X}_t \sim \mathrm{CI}(d,b)$，$\boldsymbol{\alpha}$ 为**协整向量**(cointegrated vector)。

由此可见，**如果两个变量都是单整变量，只有当它们的单整阶相同时，才可能协整**；如果它们的单整阶不相同，就不可能协整。

三个以上的变量，如果具有不同的单整阶数，有可能经过线性组合构成低阶单整变量。例如，如果存在

$$W_t \sim I(1), \qquad V_t \sim I(2), \qquad U_t \sim I(2)$$

并且

$$P_t = aV_t + bU_t \sim I(1)$$
$$Q_t = cW_t + eP_t \sim I(0)$$

那么认为

$$V_t, U_t \sim \mathrm{CI}(2,1)$$
$$W_t, P_t \sim \mathrm{CI}(1,1)$$

从协整的定义可以看出，(d,d) 阶协整是一类非常重要的协整关系，它的经济意义在于：两个变量，虽然它们具有各自的长期波动规律，但是如果它们是 (d,d) 阶协整的，则它们之间存在着一个长期稳定的比例关系。例如，前面提到的中国居民总量消费 Y 和总量可支配收入 X，它们各自都是 2 阶单整序列，并且我们将会看到，它们取对数后的序列 $\ln Y$ 与 $\ln X$ 各自都是 1 阶单整的，而且 $\ln Y$ 与 $\ln X$ 也是(1,1)阶协整的，这就说明它们的对数序列间存在着一个长期稳定的比例关系。从计量经济学模型的意义上讲，建立如下居民总量消费函数模型：

$$\ln Y_t = \alpha_0 + \alpha_1 \ln X_t + \mu_t$$

变量的选择是合理的，随机误差项也一定是一个白噪声，而且模型参数有合理的经济解释。这也解释了尽管这两个时间序列是非平稳的，但却可以用经典的回归分析方法建立双对数因果关系回归模型的原因。

从这里，我们已经初步认识到，检验变量之间的协整关系，在建立计量经济学模型中是非常重要的。而且，从变量之间是否具有协整关系出发选择模型的变量，其数据基础是牢固的，且其统计性质是优良的。

二、协整的检验

1. 两变量的 Engle-Granger 检验

在时间序列分析中，最令人关注的一种协整关系是（1，1）阶协整。为了

检验两个均呈现 1 阶单整的变量 Y_t, X_t 是否为协整，恩格尔和格兰杰于 1987 年提出两步检验法，也称为 **EG 检验**。

第一步，用普通最小二乘法估计方程(8.3.1)并计算非均衡误差，得到

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_t$$
$$e_t = Y_t - \hat{Y}_t$$

称为**协整回归**(cointegrating)或**静态回归**(static regression)。

第二步，检验 e_t 的单整性。如果 e_t 为稳定序列 $I(0)$，则认为变量 Y_t, X_t 为(1,1)阶协整；否则，认为变量 X_t，Y_t 不存在协整关系。

检验 e_t 的单整性的方法即是§8.1 中使用的 DF 检验或者 ADF 检验。由于协整回归中已含有截距项，则检验模型中无须再用截距项；如果协整回归中还含有趋势项，则检验模型中也无须再用时间趋势项。使用模型 1：

$$\Delta e_t = \delta\, e_{t-1} + \sum_{i=1}^{p} \theta_i \Delta e_{t-i} + \varepsilon_t$$

进行检验时，拒绝零假设 $H_0 : \delta = 0$，意味着残差项 e_t 是平稳序列，从而说明 X 与 Y 是协整的。

一个需要注意的问题是，这里的 DF 检验或 ADF 检验是针对协整回归计算出的残差项 e_t 而非真正的非均衡误差 μ_t 进行的。而普通最小二乘法采用了残差最小平方和原理，因此估计量 δ 往往是向下偏倚的，这样将导致拒绝零假设的机会比实际情形大。于是对 e_t 平稳性检验的 DF 与 ADF 临界值应该比正常的 DF 与 ADF 临界值还要小。麦金农(Mackinnon，1991)通过模拟试验给出了协整检验的临界值，表 8.3.1 是双变量情形下不同样本容量的临界值。

表 8.3.1 双变量协整 ADF 检验临界值

样本容量	显著性水平		
	0.01	0.05	0.10
25	−4.37	−3.59	−3.22
50	−4.12	−3.46	−3.13
100	−4.01	−3.39	−3.09
∞	−3.90	−3.33	−3.05

例 8.3.1

对表 2.6.3 中经居民消费价格指数调整后的 1978—2006 年中国居民总量消费 Y 与总量可支配收入 X 的数据，检验它们取对数的序列 $\ln Y$ 与 $\ln X$ 间的协整关系。

PDG

对于 lnY 与 lnX，容易验证它们均是 $I(1)$序列，适当的检验模型如下：

$$\Delta^2 \ln \hat{Y}_t = 0.059 - 0.741\Delta \ln Y_{t-1}$$

$$(3.55)\quad(-3.89)$$

$$\mathrm{LM}(1)=1.358 \qquad \mathrm{LM}(2)=1.571$$

$$\Delta^2 \ln \hat{X}_t = 0.071 - 0.784\Delta \ln X_{t-1}$$

$$(3.58)\quad(-3.97)$$

$$\mathrm{LM}(1)=1.062 \qquad \mathrm{LM}(2)=2.022$$

在 5%的显著性水平下，ADF 检验的临界值为–2.97。

对 lnY 与 lnX 做如下协整回归

$$\ln \hat{Y}_t = 0.587 + 0.880 \ln X_t$$

$$(4.11)\quad(61.89)$$

$$R^2=0.993\,0 \qquad \mathrm{D.W}=0.415$$

通过对该式计算的残差序列 e_t 进行 ADF 检验，得适当的检验模型为

$$\Delta\hat{e}_t = -0.631e_{t-1} + 0.337\Delta e_{t-1} + 0.298\Delta e_{t-2} + 0.390\Delta e_{t-3} + 0.494\Delta e_{t-4}$$

$$(-3.69)\quad(1.78)\qquad(1.58)\qquad(2.14)\qquad(2.58)$$

$$\mathrm{LM}(1)=1.89 \qquad \mathrm{LM}(2)=2.99$$

由附录表六容易算得，在 5%的显著性水平下，协整的 ADF 检验临界值为 $-3.3377-\dfrac{5.967}{29}-\dfrac{8.98}{29^2}=-3.554$，$e_{t-1}$ 前参数的 t 值为–3.69，因此拒绝存在单位根的假设，表明残差项是稳定的。据此判断，中国居民总量消费的对数序列 lnY 与可支配总收入的对数序列 lnX 是（1,1）阶协整的，说明了这两个变量的对数序列间存在长期稳定的“均衡”关系。

2. 多变量协整关系的检验

多变量协整关系的检验要比双变量复杂一些，主要原因在于协整变量间可能存在多种稳定的线性组合。假设有 4 个 $I(1)$ 变量 Z，X，Y，W，它们有如下的长期均衡关系：

$$Z_t = \alpha_0 + \alpha_1 W_t + \alpha_2 X_t + \alpha_3 Y_t + \mu_t \tag{8.3.5}$$

其中，非均衡误差项 μ_t 应是 $I(0)$ 序列：

$$\mu_t = Z_t - \alpha_0 - \alpha_1 W_t - \alpha_2 X_t - \alpha_3 Y_t \tag{8.3.6}$$

然而，如果 Z 与 W，X 与 Y 间分别存在长期均衡关系：

$$Z_t = \beta_0 + \beta_1 W_t + v_{1t}$$

$$X_t = \gamma_0 + \gamma_1 Y_t + v_{2t}$$

则非均衡误差项 v_{1t}，v_{2t} 一定是稳定序列 $I(0)$。于是它们的任意线性组合也是

稳定的。例如，

$$v_t = v_{1t} + v_{2t} = Z_t - \beta_0 - \gamma_0 - \beta_1 W_t + X_t - \gamma_1 Y_t \tag{8.3.7}$$

一定是$I(0)$序列。由于v_t像(8.3.6)式中的μ_t一样，也是Z，X，Y，W四个变量的线性组合，由此(8.3.7)式也成为这 4 个变量的另一个稳定线性组合。$(1,-\alpha_0,-\alpha_1,-\alpha_2,-\alpha_3)$是对应于(8.3.6)式的协整向量，$(1,-\beta_0-\gamma_0,-\beta_1,1,-\gamma_1)$是对应于(8.3.7)式的协整向量。

对于多变量的协整检验过程，基本与双变量情形相同，即需要检验变量是否具有同阶单整性，以及是否存在稳定的线性组合。后者需通过设置一个变量为被解释变量，其他变量为解释变量，进行普通最小二乘估计并检验残差序列是否平稳。如果不平稳，则需更换被解释变量，进行同样的普通最小二乘估计及相应的残差项检验。当所有的变量都被作为被解释变量检验之后，仍不能得到平稳的残差项序列，则认为这些变量间不存在(d,d)阶协整。

同样地，检验残差项是否平稳的 DF 与 ADF 检验临界值要比通常的 DF 与 ADF 检验临界值小，而且该临界值还受到所检验变量的个数的影响。表 8.3.2 给出了麦金农(1991)通过模拟试验得到的不同变量协整检验的临界值(其一般计算表式见附录)。

表 8.3.2　多变量协整检验 ADF 临界值

样本	变量数=3 显著性水平			变量数=4 显著性水平			变量数=6 显著性水平		
容量	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
25	–4.92	–4.1	–3.71	–5.43	–4.56	–4.15	–6.36	–5.41	–4.96
50	–4.59	–3.92	–3.58	–5.02	–4.32	–3.98	–5.78	–5.05	–4.69
100	–4.44	–3.83	–3.51	–4.83	–4.21	–3.89	–5.51	–4.88	–4.56
	–4.30	–3.74	–3.45	–4.65	–4.1	–3.81	–5.24	–4.7	–4.42

约翰森（Johansen）于 1988 年，以及与居斯利斯（Juselius）一起于 1990 年提出了一种基于向量自回归模型的多重协整检验方法，通常称为 Johansen 检验，或 JJ 检验，是一种进行多重协整检验的较好方法。这种方法将在高级课程中介绍。

三、误差修正模型

1. 误差修正模型

前面已经提到，对于非稳定时间序列，可通过差分的方法将其化为稳定序列，然后才可建立经典的回归分析模型。例如当我们建立人均消费水平(Y)与人均可支配收入(X)之间的回归模型

$$Y_t = \alpha_0 + \alpha_1 X_t + \mu_t \tag{8.3.1}$$

时，如果Y与X具有共同的向上或向下的变化趋势，为了避免虚假回归，通常需要通过差分的方法建立消除变量的共同变化趋势，使之成为稳定序列，再建立差分回归模型

$$\Delta Y_t = \alpha_1 \Delta X_t + v_t \tag{8.3.2$'$}$$

式中，$v_t = \mu_t - \mu_{t-1}$。

然而，这种做法会引起两个问题：一是，如果X与Y间存在着长期稳定的均衡关系(8.3.1)式，且误差项μ_t不存在序列相关，则差分(8.3.2′)式中的v_t是一个一阶移动平均时间序列，因而是序列相关的；二是，如果采用(8.3.2)式的差分形式进行估计，则关于变量水平值的重要信息将被忽略。这时模型只表达了X与Y间的短期关系，而没有揭示它们间的长期关系。因为，从长期均衡的观点看，Y在第t期的变化不仅取决于X本身的变化，还取决于X与Y在$t-1$期末的状态，尤其是X与Y在$t-1$期的不平衡程度。

另外，使用差分变量也往往会得出不能令人满意的回归方程。例如，使用(8.3.2′)式回归时，很少出现截距项显著为零的情况，即我们常常会得到如下形式的方程：

$$\Delta Y_t = \hat{\alpha}_0 + \hat{\alpha}_1 \Delta X_t + v_t\,, \qquad \hat{\alpha}_0 \neq 0 \tag{8.3.8}$$

这样在X保持不变时，如果模型存在静态均衡(static equilibrium)，Y也会保持它的长期均衡值不变。但如果使用(8.3.8)式，即使X保持不变，Y也会处于长期上升($\hat{\alpha}_0 > 0$)或下降($\hat{\alpha}_0 < 0$)的过程中，这意味着X与Y间不存在静态均衡。这与大多数具有静态均衡的经济理论假说不相符。很明显，如果收入保持稳定，我们就不能期望消费支出永远不停地变化。

可见，简单差分不一定能解决非平稳时间序列所遇到的全部问题，因此，误差修正模型便应运而生。

误差修正模型(Error Correction Model，ECM)是一种具有特定形式的计量经济学模型，它的主要形式是由大卫德森（Davidson），亨格瑞（Hendry），斯巴（Srba）和耶（Yeo）于 1978 年提出的，称为 DHSY 模型。为了便于理解，我们通过一个具体的模型来介绍它的结构。

假设两个变量X与Y的长期均衡关系如(8.3.1)式所示，由于现实经济中X与Y很少处在均衡点上，因此我们实际观测到的只是X与Y间的短期的或非均衡的关系，假设具有如下(1, 1)阶分布滞后形式：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \delta Y_{t-1} + \mu_t \tag{8.3.9}$$

该模型显示出第t期的Y值，不仅与X的变化有关，而且与$t-1$期X与Y的状态值有关。

由于变量可能是非平稳的，因此不能直接运用普通最小二乘法。对(8.3.9)式适当变形得

$$\begin{aligned}\Delta Y_t &= \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} - (1-\delta) Y_{t-1} + \mu_t \\ &= \beta_1 \Delta X_t - (1-\delta)\left(Y_{t-1} - \frac{\beta_0}{1-\delta} - \frac{\beta_1 + \beta_2}{1-\delta} X_{t-1} \right) + \mu_t\end{aligned}$$

或

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \mu_t \tag{8.3.10}$$

其中 $\lambda = 1-\delta$，$\alpha_0 = \beta_0/(1-\delta)$，$\alpha_1 = (\beta_1 + \beta_2)/(1-\delta)$

如果将(8.3.10)中的参数α_0，α_1与(8.3.1)式中的相应参数视为相等，则(8.3.10)式中括号内的项就是$t-1$期的非均衡误差项。于是(8.3.10)式表明Y的变化决定于X的变化以及前一时期的非均衡程度。同时，(8.3.10)式也弥补了简单差分(8.3.2)式的不足，因为该式含有用X，Y水平值表示的前期非均衡程度。因此，Y的值已对前期的非均衡程度做出了修正。(8.3.10)式称为**一阶误差修正模型**(first-order error correction model)。

模型(8.3.10)可以写成：

$$\Delta Y_t = \beta_1 \Delta X_t - \lambda \cdot \mathrm{ecm}_{t-1} + \mu_t \tag{8.3.11}$$

其中 ecm 表示**误差修正项**。由(8.3.11)可知，一般情况下$|\delta|<1$，所以有$0<\lambda<1$。我们可以据此分析 ecm 的修正作用：若 t-1 时刻Y大于其长期均衡解$\alpha_0 + \alpha_1 X$，ecm 为正，则$-\lambda \cdot \mathrm{ecm}$为负，使得$\Delta Y_t$减少；若 t-1 时刻Y小于其长期均衡解$\alpha_0 + \alpha_1 X$，ecm 为负，$-\lambda \cdot \mathrm{ecm}$为正，使得$\Delta Y_t$增大。这体现了长期非均衡误差对$Y_t$的控制。

需要注意的是，在实际分析中，变量常以对数的形式出现，其主要原因在于变量对数的差分近似地等于该变量的变化率，而经济变量的变化率常常是稳定序列，因此适合于包含在经典回归方程中。于是长期均衡模型(8.3.1)中的α_1可视为Y关于X的**长期弹性**(long-run elasticity)，而短期非均衡模型(8.3.9)中的β_1可视为Y关于X的**短期弹性**(short-run elasticity)。

更复杂的误差修正模型可依照一阶误差修正模型类似地建立。如具有季度数据的变量，可在短期非均衡模型(8.3.9)中引入更多的滞后项。引入二阶滞后项的模型为

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \mu_t \tag{8.3.12}$$

经过适当的恒等变形，可得如下误差修正模型：

$$\Delta Y_t = -\delta_2 \Delta Y_{t-1} + \beta_1 \Delta X_t - \beta_3 \Delta X_{t-1} - \lambda (Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1}) + \mu_t \tag{8.3.13}$$

其中 $\lambda = 1 - \delta_1 - \delta_2$，$\alpha_0 = \beta_0 / \lambda$，$\alpha_1 = (\beta_1 + \beta_2 + \beta_3)/\lambda$

同样地，引入三阶滞后项的误差修正模型与(8.3.13)式相仿，只不过模型中多出差分滞后项ΔY_{t-2}，ΔX_{t-2}。

多变量的误差修正模型也可类似地建立。例如三个变量如果存在如下长期均衡关系：

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_t \tag{8.3.14}$$

则其一阶非均衡关系可写成

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \gamma_1 Z_t + \gamma_2 Z_{t-1} + \delta Y_{t-1} + \mu_t \tag{8.3.15}$$

于是它的一个误差修正模型为

$$\Delta Y_t = \beta_1 \Delta X_t + \gamma_1 \Delta Z_t - \lambda(Y_{t-1} - \alpha_0 - \alpha_1 X_{t-1} - \alpha_2 Z_{t-1}) + \mu_t \tag{8.3.16}$$

其中 $\lambda = 1-\delta$，$\alpha_0 = \beta_0/\lambda$，$\alpha_1 = (\beta_1+\beta_2)/\lambda$，$\alpha_2 = (\gamma_1+\gamma_2)/\lambda$

2. 误差修正模型的建立

(1) 格兰杰表述定理

误差修正模型有许多明显的优点，如一阶差分项的使用消除了变量可能存在的趋势因素，从而避免了虚假回归问题；一阶差分项的使用也消除模型可能存在的多重共线性问题；误差修正项的引入保证了变量水平值的信息没有被忽视；由于误差修正项本身的平稳性，使得该模型可以用经典的回归方法进行估计，尤其是模型中差分项可以使用通常的 t 检验与 F 检验来进行选取，等等。于是，一个重要的问题就是是否变量间的关系都可以通过误差修正模型来表述。就此问题，恩格尔与格兰杰于 1987 年提出了著名的格兰杰**表述定理**(Granger representation theorem)：

如果变量 X 与 Y 是协整的，则它们间的短期非均衡关系总能由一个误差修正模型表述，即

$$\Delta Y_t = \text{lagged}(\Delta Y, \Delta X) - \lambda \cdot \text{ecm}_{t-1} + \mu_t, \qquad 0<\lambda<1 \tag{8.3.17}$$

其中，ecm_t 是非均衡误差项或者说成是长期均衡偏差项，λ 是短期调整参数。

对于上述(1,1)阶自回归分布滞后模型(8.3.9)式，如果

$$Y_t \sim I(1), \qquad X_t \sim I(1)$$

那么，(8.3.10)式左边 $\Delta Y_t \sim I(0)$，右边的 $\Delta X_t \sim I(0)$，于是，只有 Y 与 X 协整，才能保证右边也是 $I(0)$。因此，建立误差修正模型，需要首先对变量进行协整分析，以发现变量之间的协整关系，即长期均衡关系，并以这种关系构成误差修正项。然后建立短期模型，将误差修正项看作一个解释变量，连同其他反映短期波动的解释变量一起，建立短期模型，即误差修正模型。

注意，由于(8.3.17)式中没有明确指出 ΔY 与 ΔX 的滞后项数，因此，可以是多个；同时，由于一阶差分项是 $I(0)$ 变量，因此模型中也允许使用 X 的非滞后差分项 ΔX_t。

格兰杰表述定理可类似地推广到多个变量的情形中去。

(2) Engle-Granger(E-G)两步法

由协整与误差修正模型的关系，可以得到误差修正模型建立的 E-G 两步法：

第一步，进行协整回归(普通最小二乘法)，检验变量间的协整关系，估计协整向量(长期均衡关系参数)；

第二步，若协整性存在，则以第一步求到的残差作为非均衡误差项加入到误差修正模型中，并用普通最小二乘法估计相应参数。

需要注意的是，在进行变量间的协整检验时，如有必要可在协整回归式中加入趋势项，这时，对残差项的稳定性检验就无须再设趋势项。另外，第二步中变量差分滞后项的多少，可以残差项序列是否存在自相关性来判断。如果存在自相关，则应加入变量差分的滞后项。

(3) 直接估计法

也可以采用打开误差修整模型中非均衡误差项括号的方法直接用普通最小二乘法估计模型。但仍须事先对变量间的协整关系进行检验。例如对双变量误差修正模型(8.3.10)式，可打开非均衡误差项的括号直接估计下式：

$$\Delta Y_t = \lambda\alpha_0 + \beta_1\Delta X_t - \lambda Y_{t-1} + \lambda\alpha_1 X_{t-1} + \mu_t$$

这时短期弹性与长期弹性可一并获得。需注意的是，用不同方法建立的误差修正模型结果也往往不一样。

例 8.3.2

建立中国居民总量消费 Y 的误差修正模型。

例 8.3.1 中验证了中国居民总量消费（Y）与可支配总收入（X）的对数序列间呈协整关系。下面尝试建立它们的误差修正模型。

以 $\ln Y$ 关于 $\ln X$ 的协整回归中稳定残差序列 e_t 作为误差修正项，可建立如下误差修正模型

$$\Delta\ln\hat{Y}_t = 0.507\Delta\ln X_t + 0.405\Delta\ln Y_{t-1} - 0.209e_{t-1} \qquad (8.3.18)$$

$$\begin{array}{ccc} (5.46) & (3.78) & (-2.04) \end{array}$$

$$R^2 = 0.394 \qquad \text{LM}(1)=0.10 \qquad \text{LM}(2)=0.82$$

由例 8.3.1 中的协整回归式可得 $\ln Y$ 关于 $\ln X$ 的长期弹性为 0.880；由（8.3.18）式可得 $\ln Y$ 关于 $\ln X$ 的短期弹性 0.507。

最后，给出误差修正模型的预测情况。

由例 8.3.1 中的协整回归式计算 2006 年关于长期均衡点的偏差：

$$e_{2006} = \ln Y_{2006} - 0.587 - 0.880\ln X_{2006} = \ln 36\,811.6 - 0.587 - 0.880\ln 85\,624.1$$

$$= -0.068\,23$$

由（8.3.18）式预测 2007 年的短期波动

$$\Delta\ln\hat{Y}_{2007}=0.507\Delta\ln X_{2007}+0.405\Delta\ln Y_{2006}-0.209e_{2006}$$

$$=0.507\times(\ln(95\ 405.7)-\ln(85\ 624.1))+0.405\times(\ln(36\ 811.6)-\ln(33\ 214.0))-0.209\times(-0.068\ 23)=0.110\ 753$$

于是

$$\ln\hat{Y}_{2007}=\ln Y_{2006}+\Delta\ln Y_{2007}$$

$$=\ln 36\ 811.6+0.110\ 753=10.62\ 432$$

$$\hat{Y}_{2007}=\exp(\ln\hat{Y}_{2007})=e^{10.624\ 32}=41\ 122.9$$

例 8.2.4 中已计算出 2007 年中国实际居民消费总量为 40 910.7 亿元，可见这里误差修正模型预测的相对误差为 0.52%。这一误差既小于例 8.2.4 中 ARMA 模型的预测误差，更小于例 2.6.2 中传统回归模型的预测误差。

本章练习题

1．设时间序列$\{X_t\}$由$X_t=\delta_0+\delta_1 t+\varepsilon_t$生成，如果$\varepsilon_t$是一个具有零均值、同方差、不序列相关的白噪声，问：(1) X_t是平稳时间序列吗？(2) $X_t-E(X_t)$是平稳时间序列吗？

2．设时间序列X_t由下面随机过程生成：$X_t=Z_t+\varepsilon_t$，其中ε_t为一个均值为 0，方差为σ_ε^2的白噪声序列；Z_t是一个均值为 0，方差为σ_z^2，协方差恒为常数a的平稳时间序列；ε_t与Z_t不相关。

(1) 求X_t的期望与方差，它们与时间t有关吗？

(2) 求协方差$\mathrm{Cov}(X_t,X_{t+k})$，并指出$X_t$是否是平稳的。

(3) 证明：X_t的自相关函数为$\rho_k=a/(\sigma_z^2+\sigma_\varepsilon^2)$。

3．假设两时间序列X_t与Y_t都是随机游走序列。证明：如果X_t与Y_t是协整的，则X_t与Y_{t-1}也是协整的。

4．假设两个时间序列X_t与Y_t都是$I(1)$序列，但存在某个不为 0 的β，使$Y_t-\beta X_t$是$I(0)$。证明：对于任何$\delta\neq\beta$，组合$Y_t-\delta X_t$一定是$I(1)$的。

5．假设两个时间序列X_t与Y_t满足$Y_t=\beta X_t+\varepsilon_{1t}$与$\Delta X_t=\alpha\Delta X_{t-1}+\varepsilon_{2t}$，其中，$\beta\neq 0$，$|\alpha|<1$，且$\varepsilon_{1t}$与$\varepsilon_{2t}$分别是两$I(0)$序列。证明：从这两个方程可以推出一个如下形式的误差修正模型：

$$\Delta Y_t=\alpha_1\Delta X_{t-1}+\delta(Y_{t-1}-\beta X_{t-1})+\varepsilon_t$$

其中，$\alpha_1=\beta\alpha$，$\delta=-1$，$\varepsilon_t=\varepsilon_{1t}+\beta\varepsilon_{2t}$

6．检验例 8.3.1 中经居民消费价格指数调整后的中国居民总量消费Y与总量可支配收入X的对数序列$\ln Y$与$\ln X$是非平稳的。

7．教材例 2.6.2 曾给出了 1978—2006 年中国居民消费价格指数 CPI(1990 年=100)。

（1）作出时间序列 CPI 的样本相关图，并通过图形判断该时间序列的平稳性。

（2）对 CPI 序列进行单位根检验，以进一步明确它们的平稳性。

（3）检验 CPI 的单整性；

（4）尝试建立 CPI 的 ARIMA 模型。

8．观察中国货物进出口数据发现两者间有很强的同步性，由于中国的加工贸易占总贸易量的一半左右，一种观点认为中国的货物进口很大程度上受货物出口波动的影响。下表给出了 1978—2007 年中国货物进、出口额的自然对数序列 LM，LX。

年　份	LX	LM	年　份	LX	LM
1978	4.579 9	4.690 4	1993	6.821 5	6.946 6
1979	4.917 1	5.054 3	1994	7.098 5	7.052 8
1980	5.199 6	5.299 3	1995	7.305 1	7.186 0
1981	5.394 1	5.394 5	1996	7.320 2	7.235 8
1982	5.408 1	5.262 2	1997	7.510 9	7.261 0
1983	5.404 0	5.365 5	1998	7.515 9	7.245 9
1984	5.566 1	5.613 5	1999	7.575 2	7.412 8
1985	5.611 3	6.046 2	2000	7.820 8	7.719 1
1986	5.734 6	6.061 7	2001	7.886 5	7.797 9
1987	5.977 4	6.068 7	2002	8.088 3	7.990 1
1988	6.163 7	6.314 8	2003	8.385 3	8.325 5
1989	6.264 2	6.382 5	2004	8.688 3	8.632 7
1990	6.431 2	6.279 5	2005	8.938 5	8.794 7
1991	6.578 0	6.458 2	2006	9.178 8	8.976 5
1992	6.744 5	6.692 0	2007	9.407 4	9.165 3

（1）对 LX 与 LM 序列进行单位根检验，检验它们的平稳性。

（2）检验 LX 与 LM 的单整性；

（3）检验 LX 与 LM 的协整性；

（4）如果 LX 与 LM 是协整的，请估计 LX 关于 LM 的误差修正模型。

第九章　计量经济学应用模型

9

在第一章中已经介绍过，计量经济学模型主要用于结构分析、经济预测、政策评价、检验与发展经济理论，这是从计量经济学模型的功能角度来讲的。从计量经济学模型的应用领域来讲，在宏观经济和微观经济的各个领域，可以说计量经济学模型无所不在，并且其应用范围也由经济领域扩展到劳动、卫生、教育、人口、家庭等社会领域。从计量经济学模型的类型来讲，各种经典的和非经典的模型都得到广泛应用，并且在不断发展。

在计量经济学模型被广泛应用并不断发展的同时，对它的质疑和否定也从未间断过。存在这种现象，一方面是出于对计量经济学模型性质的不同认识，以及对它的哲学基础、经济学基础和统计学基础的不同理解和解释；另一方面是出于计量经济学模型实际应用中的问题和错误，没有正确地建立和应用计量经济学模型，结果是授人以柄。

学习计量经济学的目的，一方面是发展计量经济学，另一方面是应用计量经济学模型，后者更为重要。作为本书的最后一章，其目的是帮助读者正确地建立和应用计量经济学模型。本章分为 4 节：第一节主要讨论计量经济学应用模型类型的设定，强调模型类型对被解释变量数据类型的依赖性；第二节讨论计量经济学应用中总体回归模型的设定，着重阐述“从一般到简单”的模型设定原则；第三节讨论计量经济学应用模型中变量之间函数关系的设定，主张理论与经验并重；第四节讨论计量经济学应用模型中变量性质的设定，强调变量性质设定的相对性。

§9.1　计量经济学应用模型类型设定

作为建立计量经济学应用模型的第一步，就是针对研究对象和研究目的，选择适当类型的模型，称之为模型类型设定。如果模型类型设定错误，一切就都错了。

一、问题的提出

经过近 80 年的发展，计量经济学已经形成了门类齐全的模型体系。在这

一体系中，主要有参数模型和非参数模型，单方程模型和联立方程模型，截面数据模型、时间序列数据模型和平行数据模型；即便在截面数据单方程参数模型中，还包括经典模型、选择性样本模型、计数数据(count data)模型、离散选择模型、持续时间数据(duration data)模型等多种类型。我们已经在前面的章节中介绍过其中一些模型，而有些在本书中还没有接触到。那么，面对一个实际经济问题，在如此众多的模型类型中，应该选择哪种模型，自然是首先要明确的问题。

例 9.1.1

某单位对我国农户借贷需求进行了较为广泛的调查，采集了 16 个省(自治区)的 72 个县(市)、440 多个村庄的 5 100 家农户的数据。其中，在一年中发生借贷行为的农户占 55.3%(包括向亲友借贷)共 2 820 户，其余 2 280 户没有发生借贷。现已经收集了该 5 100 户中每户的一年中发生的借贷额、家庭总收入、总支出、总收入中农业生产经营收入所占比例、总支出中生产性支出所占的比例、户主受教育程度、户主健康状况、家庭人口数等近 100 项数据。

为了对农户借贷行为进行因素分析，即建立以农户借贷额为被解释变量、各种影响因素为解释变量的农户借贷因素分析模型，不同的研究者建立了下面 4 种不同类型的计量经济学模型。

(1) 仅利用 2 820 户发生借贷的农户为样本，即以他们的借贷额为被解释变量，各种影响因素为解释变量，建立经典的回归模型。

(2) 为了充分利用没有发生借贷的农户的信息，于是利用 5 100 户为样本，即以他们的借贷额为被解释变量，其中没有发生借贷的农户借贷额为 0，以各种影响因素为解释变量，建立经典的回归模型。

(3) 进一步分析发现，不应该将没有发生借贷的农户的借贷额统统视为 0，而应该视为小于等于 0(≤0)，于是利用 5 100 户为样本，建立归并数据模型(Tobit 模型)。

(4) 更进一步分析发现，不应该将没有发生借贷的农户的借贷额统统视为小于等于 0，因为其中一部分农户有借贷需求，只是因为各种原因(例如提出借贷被拒绝，担心借不到而不敢提出借贷要求等)而没有发生实际借贷。所以，应该按照赫克曼(Heckman)两步法建立模型，即首先利用全部样本信息建立借贷是否发生的二元选择模型，然后再利用 2 820 户发生借贷的农户为样本，建立借贷额的因素分析回归模型。

显然，最后一种模型是正确的，其他都是不正确的。

例 9.1.1 中面临的问题，属于截面数据单方程计量经济学应用模型类型的选择，也是实际计量经济学应用研究中遇到的最多的问题，那么，应该根据什么来设定应用模型的类型？

例 9.1.2

为了研究我国城镇居民的食品需求及其与各个影响因素之间的关系，在城镇居民中随机抽取了 5 000 户作为样本。以每户的人均月食品需求量 q_i 为被解释变量，以人均月收入 I_i 与购买食品的平均价格 p_i 为解释变量。因为不同收入水平的家庭购买食品的场所、时间不同，所以调查显示，他们购买食品的平均价格是不同的。另外，其他商品的价格对食品需求应该产生影响，考虑采用扣除食品后的消费品零售价格指数作为解释变量，但由于获得数据的困难，同时考虑到该价格指数对于不同家庭的差异不大，所以没有将它引入模型。于是建立了如下单方程食品需求函数模型：

$$\ln q_i = \alpha + \beta \ln p_i + \gamma \ln I_i + \mu_i \qquad i = 1, 2, \cdots, 5000$$

该模型受到了尖锐的批评：有人认为首先在模型类型设定上就存在问题，应该建立包括各类商品需求量的联立方程模型，而不应该选择单方程模型。显然，这些批评是正确的。

例 9.1.2 提出的问题是关于单方程模型和联立方程模型之间的选择问题，属于模型类型设定的范畴，同样是计量经济学应用研究中经常遇到的，那么，应该根据什么来进行选择？

例 9.1.3

在一篇研究我国工业资本配置效率的论文中，按照杰弗里 • 瓦格勒(Jeffrey Wurgler)在 2000 年提出的资本配置效率模型，以投资增长率为被解释变量，以利润增长率为解释变量，忽略其他因素。选择我国 39 个工业行业 1991 年至 1999 年共 9 年的 351 组数据为样本，建立平行数据模型。显然，如果反映资本配置效率的参数 β 接近于 1，说明资本配置效率比较高。

(1) 为了进行国际比较，建立了如下**截距和系数都不变的模型**：

$$\ln \frac{I_{it}}{I_{i,t-1}} = \alpha + \beta \ln \frac{V_{it}}{V_{i,t-1}} + \mu_{it}$$

利用 39 个行业 9 年的 351 个样本，采用普通最小二乘估计。估计结果表明，我国工业资本配置效率不仅低于发达国家，也低于大多数发展中国家。

(2) 为了定量刻画我国每年的资本配置效率，建立了如下**截距和系数都随时间变化的模型**：

$$\ln\frac{I_{it}}{I_{i,t-1}}=\alpha_t+\beta_t\ln\frac{V_{it}}{V_{i,t-1}}+\mu_{it}$$

分别用每年的行业截面数据，采用普通最小二乘估计。从估计结果可以看出，我国资本配置效率呈逐年下滑趋势。

(3) 为了分析我国工业行业的成长性，又建立了如下**截距和系数都随行业变化的模型**：

$$\ln\frac{I_{it}}{I_{i,t-1}}=\alpha_i+\beta_i\ln\frac{V_{it}}{V_{i,t-1}}+\mu_{it}$$

分别以每个行业的时间序列数据为样本观测值，进行普通最小二乘估计，得到每个行业的资本配置效率估计量。从估计结果选择其中最具发展潜力的 5 个行业。

该项研究也受到了尖锐的批评：对于同样一组样本观测值，根据研究目的的需要，建立了 3 个不同类型的模型，显然是不正确的。正确的反映该数据生成过程的模型只能是 1 个，不可能是 3 个。

例 9.1.3 同样属于应用模型类型设定问题，在不同类型的平行数据模型中，应该选择哪类模型？这个问题的答案是清楚的，在本书 § 7.3 中的“模型设定”小节中首先介绍的就是模型设定检验，通过严格的统计检验，确定应该建立什么类型的模型。但是，在已有的平行数据模型的应用研究中，很少注意这个问题，经常是根据研究目的的需要设定模型类型。

以上几例从不同的角度指出了计量经济学应用模型类型设定问题的普遍性和重要性。下面将分别讨论一些指导模型类型设定的原则，仍然采用举例式进行讨论，希望读者能够从中得到一些启发。

二、单方程应用模型类型对被解释变量数据类型的依赖性

在经济、社会问题研究中，当研究对象确定之后，表征该经济、社会活动结果的数据自然地被确定了。例如，研究我国经济增长的影响因素以及各个因素对增长的贡献，那么表征经济增长结果的 GDP 时间序列自然地成为模型研究的对象；研究学生在本科 4 年内不及格的课程门数与什么因素有关，那么表征不及格门数的计数数据 0，1，2…自然地成为模型研究的对象；研究农户的借贷方式由哪些因素决定，那么表征农户向各种正规金融和非正规金融机构借贷的选择结果的离散选择数据 0，1，2…自然地成为模型研究的对象。计量经济学应用研究的第一步，就是根据表征所要研究的经济、社会活动结果的数据

类型确定应该建立什么类型的计量经济学模型，即根据作为被解释变量观测值的数据类型设定应用模型的类型。在这一步骤中，**被解释变量观测值数据的类型决定了计量经济学模型的类型。**

作为被解释变量观测值的数据分为三类：截面数据(cross-sectional data)、时间序列数据(time-series data)和平行数据(panel data，也译为面板数据、综列数据)。

1．经典截面数据模型

对于截面数据，只有当数据是在截面总体中由随机抽样得到的样本观测值，并且变量具有连续的随机分布时，才能够将模型类型设定为经典的计量经济学模型。经典计量经济学模型的数学基础建立在随机抽样的截面数据之上。但是，在实际的经验实证研究中，面对的截面数据经常是非随机抽样得到的，或者是离散的，如果仍然采用经典计量经济学的模型设定，错误就不可避免了。事实上，20 世纪 70 年代以来，针对这些类型数据的模型已经得到发展并建立了坚实的数学基础。本书第二章至第五章详细介绍了经典截面数据模型的理论方法。

2．选择性样本模型

如果被解释变量的样本观测值并不是在截面总体中由随机抽样得到的，那么经典截面数据模型不再适用。例如，在例 9.1.1 中，如果只利用 2 820 户发生借贷的农户为样本，建立经典的回归模型，被称为“截断数据”(truncation data)，就是错误的。这类数据在实际经济分析中十分常见，特别在微观经济社会问题研究中大量存在。人们抽取的样本经常是“掐头”或者“去尾”的。对于这类数据，因为抽取每个样本的概率发生了变化，如果仍然采用经典计量经济学模型，其估计结果就产生了“选择性偏误”，应该建立截断数据模型，正如 § 7.1 中所讨论的。

例如，如果我们分析学生的学习成绩与相关影响因素之间的关系，学习成绩的最高分为 100，最低分为 0。处于 0 与 100 之间的得分，是学习成绩的真实反映；而表现为 100 分和 0 分的学生，实际学习成绩是不同的，所以应该将 100 分看为大于等于 100 分的归并，将 0 分看为小于等于 0 分的归并。这类数据被称为“归并数据”(censored data)。它们在经济分析中也是常见的，例如受到供给限制条件下的商品需求量、尚处于失业状态下的失业时间。类似地，因为抽取处于归并点的每个样本的概率发生了变化，如果仍然采用经典计量经济学模型，其估计结果也会产生“选择性偏误”，应该建立归并数据模型，同样如 § 7.1 中所讨论的。

3．离散选择模型

如果被解释变量的样本观测值并不是连续的，而是离散的，并且以此表征选择结果，那么经典截面数据模型也不再适用。人们几乎每时每刻都面临着这

类问题。选择结果受哪些因素的影响？各个因素的影响程度有多大？当然可以通过建立计量经济学模型来分析，但是应该建立专门的离散选择模型。例如，在§7.2中所讨论的二元离散选择模型。或者如上所述，将各种正规金融和非正规金融机构分为商业银行、农村信用社、互助金融、地下钱庄和亲友互借5类，那么表征农户向各种正规金融和非正规金融机构借贷的选择结果的离散选择数据0，1，2，3，4自然地成为模型的被解释变量观测值，就应该建立多元离散选择模型。

4．计数数据模型

人们经常要研究表现为计数数据的社会、经济活动结果受哪些因素的影响。例如，汽车一个月内发生事故的次数、学生本科4年内不及格的课程门数、大学毕业生参加工作后的前5年内调换工作的次数、个人一年内到医院就诊的次数。这些数据都是离散的非负整数，在随机抽取的一组样本中，零元素和绝对值较小的数据出现得较为频繁，重复抽样的正态分布假设不再适用。显然，对于这样的问题，不可以建立以正态性假设为基础的经典计量经济学模型，应该建立专门发展的计数数据模型。例如 Gilbert(1979)提出的泊松回归模型，Hausman,Hall 和 Griliches(1984)提出的负二项回归模型。

5．持续时间数据模型

以某项活动持续时间作为研究对象，例如研究失业持续时间与影响因素之间的关系，在这类问题中，仅从数据方面看存在两个问题：一是失业已经持续的时间并不是失业持续时间的真实反映，不能作为失业持续时间的观测值；二是取得部分解释变量的样本观测值存在困难，因为它们在持续时间内是变化的。毫无疑问，持续时间数据问题也不能建立经典的计量经济学模型。

6．时间序列分析模型

对于时间序列数据，正如第八章所讨论的，经典计量经济学模型只能建立在平稳时间序列基础之上，因为只有对满足渐近不相关的协方差平稳序列，才可以适用基于截面数据的统计推断方法，建立时间序列模型。协方差平稳性和渐近不相关性为时间序列分析适用大数定律和中心极限定理创造了条件，替代了截面数据分析中的随机抽样假定。否则，数据的时间序列性破坏了随机抽样假定，取消了样本点之间的独立性，样本点将发生序列相关。如果序列相关性不能足够快地趋于零，那么在统计推断中发挥关键作用的大数定律、中心极限定理等极限法则就缺乏应用基础。很可惜，实际的时间序列很少是平稳的。由于宏观经济仍然是我国学者进行经验实证研究的主要领域，而宏观时间序列大多是非平稳的，于是出现了大量的错误。只有经济行为上存在长期均衡关系，在数据上存在协整关系的非平稳时间序列，才能够建立经典的结构模型。

7. 平行数据模型

在截面数据和时间序列数据中存在的问题也同时存在于平行数据中，并且对平行数据还提出了模型设定的专门问题，例如变截距和变系数问题、随机影响和固定影响问题等，已经发展形成了一套完整的模型方法体系。必须依据新的模型方法体系设定总体理论模型类型，才能进行可靠的经验实证研究。

三、单方程模型和联立方程模型的选择对经济行为的依赖性

计量经济学应用模型应该是对研究对象的经济行为的客观描述。如果研究对象是相对独立的经济活动，其中存在着清晰的单向因果关系，那么可以将该应用模型设定为单方程模型。例如，研究城镇居民家庭的消费，通过消费行为分析发现，其消费主要取决于家庭收入、资产存量和社会保障制度，它们之间存在单向因果关系，因为消费对家庭收入、资产存量和社会保障制度并不产生影响。这样，我们就可以将城镇居民家庭消费模型设定为单方程模型。但是，如果研究对象并不是相对独立的经济活动，而是属于一个经济系统，在该经济系统的变量之间存在着复杂的互为因果关系，那么就应该将应用模型设定为联立方程模型。例如，研究我国的居民总消费，通过行为分析发现，虽然居民总消费水平主要取决于 GDP 的水平，但是它也受到投资水平的间接影响，而且更重要的是，居民总消费水平也反过来影响 GDP 的水平。这样，一个单方程模型就不能够完全描述居民总消费行为，必须建立一个包括居民总消费、投资和 GDP 的联立方程模型，正如 § 6.1 所讨论的。

由此可见，**在计量经济学应用研究中，单方程模型和联立方程模型的选择对经济行为具有依赖性**。下面以需求函数模型为例，进一步讨论经济行为在计量经济学应用模型类型设定中的作用，同时借此了解经济学中的需求理论和需求函数理论，因为需求函数模型是计量经济学应用研究的一个重要领域。

例 9.1.4

扩展的线性支出系统需求函数模型的推导。

根据描述需求行为的需求理论，人们对各种商品的需求量是在预算约束下，由效用函数在效用最大化下导出的。效用是商品需求量的函数，效用最大是目标，预算是约束。例如，对直接效用函数

$$U = u(q_1, q_2, \cdots, q_n) \tag{9.1.1}$$

在预算

$$\sum_{i=1}^{n} q_i p_i = I$$

约束下对效用极大化，得到的商品需求量组合为最优商品组合，该组合中的商品需求量 q_i 是收入 I 和价格 p_i 的函数，就是需求函数。

需求函数的推导过程如下。构造如下的拉格朗日函数：

$$L(q_1,q_2,\cdots,q_n,\lambda)=u(q_1,q_2,\cdots,q_n)+\lambda(I-\sum_{i=1}^{n}q_ip_i) \qquad (9.1.2)$$

最优商品组合必须满足一阶极值条件：

$$\begin{cases}\dfrac{\partial L}{\partial q_i}=\dfrac{\partial u}{\partial q_i}-\lambda p_i=0 & i=1,2,\cdots,n\\ \dfrac{\partial L}{\partial \lambda}=I-\sum\limits_{i=1}^{n}q_ip_i=0\end{cases}$$

求解该方程组即可得到所求的各种商品的需求函数。

下面介绍线性支出系统(Linear Expenditure System，LES)需求函数模型和扩展的线性支出系统(Extend Linear Expenditure System，ELES)需求函数模型。克莱因(Klein)和鲁宾(Rubin)于 1947 年提出了如下形式的**直接效用函数**

$$U=\sum_{i=1}^{n}u_i(q_i)=\sum_{i=1}^{n}b_i\ln(q_i-r_i) \qquad (9.1.3)$$

其中 r_i 为对第 i 种商品的基本需求量，b_i 为边际预算份额。该效用函数认为，效用具有可加性，即总效用为各种商品的效用之和；而各种商品的效用取决于实际需求量与基本需求量之差。(9.1.3)式在预算

$$\sum_{i=1}^{n}q_ip_i=V$$

约束下极大化，即构造如下的拉格朗日函数：

$$L(q_1,q_2,\cdots,q_n,\lambda)=\sum_{i=1}^{n}b_i\ln(q_i-r_i)+\lambda(V-\sum_{i=1}^{n}q_ip_i) \qquad (9.1.4)$$

由极值条件得到如下方程组：

$$\begin{cases}\dfrac{\partial L}{\partial q_i}=\dfrac{b_i}{q_i-r_i}-\lambda\cdot p_i=0 & i=1,2,\cdots,n\\ \dfrac{\partial L}{\partial \lambda}=\sum\limits_{i=1}^{n}q_ip_i-V=0\end{cases} \qquad (9.1.5)$$

求解该方程组即得到英国计量经济学家斯通(R.Stone)于 1954 年提出的线性支出系统需求函数：

$$q_i=r_i+\frac{b_i}{p_i}(V-\sum_j p_jr_j) \qquad i=1,2,\cdots,n \qquad (9.1.6)$$

PDG

线性支出系统需求函数(以下简称 LES)的经济意义十分清楚。**对第 i 种商品的需求量等于两部分之和。第一部分为基本需求量，即维持基本生活所必需的；第二部分为总预算扣除对所有商品的基本需求支出后剩余部分中愿意用于对第 i 种商品的需求，与消费者的偏好有关。**该模型系统是一个由 n 个方程组成的联立方程模型。LES 在估计上存在困难。(9.1.6)式中待估参数为基本需求量 r_i 和边际预算份额 b_i。但是，由于总预算 V 是对所有商品的需求支出之和，是内生变量，无法外生给出，使得模型难以估计。所以 LES 并没有被实际应用。

为克服 LES 在估计上的困难，1973 年柳弛(Liuch)对 LES 作了两点修改，提出了扩展的线性支出系统需求函数模型(以下简称 ELES)。这两点修改是：以收入 I 代替预算 V；将 b_i 的概念由边际预算份额改为边际消费倾向。于是模型表达式为

$$q_i = r_i + \frac{b_i}{p_i}(I - \sum_j p_j r_j) \qquad i = 1,2,\cdots,n \qquad (9.1.7)$$

其中待估参数为基本需求量 r_i 和边际消费倾向 b_i。按照它们的经济意义，应该有

$$r_i > 0, \quad 0 \leqslant b_i < 1, \quad \sum_i b_i \leqslant 1$$

由收入和价格的样本观测值可以对模型进行估计。ELES 是一类经济意义清楚、具有广泛应用价值的需求函数模型，属于联立方程模型。

根据例 9.1.4 中对需求行为的分析和需求函数的推导，可以发现，即使仅研究某一种商品的需求，也应该建立联立方程模型。很显然，对于例 9.1.2，尽管研究的对象是仅仅城镇居民的食品需求，并不是所有各类商品的需求，但是由于各类商品的需求是通过预算(收入)相联系的，它们之间是紧密相关的，所以建立孤立的食品需求单方程模型是不可取的。

§9.2 计量经济学应用模型总体回归模型设定

在计量经济学应用研究中设定了应用模型的类型后，接下来的工作是设定总体回归模型。只有设定了正确的总体回归模型，才能通过严格的数学过程和统计推断，得到正确的研究结果。因此，它决定了应用研究的成败。正如 §1.2 中提到的，总体回归模型设定包括选择变量，确定变量之间的函数关系，以及设定待估参数的期望值。本节采用总体回归模型设定为标题，将涉及总体设定

的一些基本概念和原则，在内容上则主要讨论应用模型的变量选择问题，即选择哪些变量进入模型。至于变量之间函数关系的设定，将在§9.3中专门讨论。

一、问题的提出及其重要性

目前，计量经济学应用研究中存在的问题很多，错误也比较普遍。重要原因之一，是缺少对计量经济学模型方法论基础的研究和理解。作为一种方法论，它的哲学基础、经济学基础、数学基础和统计学基础还没有受到足够的重视。计量经济学模型方法论基础集中体现于总体回归模型的设定。

任何科学研究，无论是自然科学还是社会科学，都是试图回答：如何从经历到的过去、特殊、局部，推论到没有经历到的未来、一般、整体。它们也都遵循以下过程：首先是关于偶然的、个别的、特殊的现象的观察；其次是从偶然的、个别的、特殊的现象的观察中，提出假说，或者是理论，或者是模型，这些假说是关于必然、一般、普遍现象而言的；然后需要对假说进行检验，检验方法一般包括实验的方法、预测的方法和回归的方法；最后是发现关于必然、一般、普遍的规律。经济学研究也是如此，不同于自然科学的是，它在推论过程的提出假说阶段中，根据是否引入价值判断，有规范研究和实证研究之分。如前所说，计量经济学模型是一种主流的实证经济研究方法论。

计量经济学模型的总体设定，就是上述从观察到的样本出发，提出关于总体假设的过程，并用计量经济学模型的形式加以表述。以经典单方程线性计量经济学模型为例：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \qquad \mu_i \sim N(0,\sigma^2) \qquad i = 1, 2\cdots, n \qquad (9.2.1)$$

给定任何被解释变量Y，要对其进行完全的解释，需考虑所有对其有直接影响的因素集$\boldsymbol{\Omega}$。按照与被解释变量关联关系的恒常性和显著性两个维度，对$\boldsymbol{\Omega}$进行分解，将显著的恒常性因素集$\boldsymbol{X}$作为解释变量。这里的“恒常性”，或者覆盖所有的截面个体，或者覆盖时间序列的所有时点。计量经济学应用研究的任务就是找到被解释变量与恒常性因素之间的关联关系，即所谓的经济规律。对于显著的偶然因素，通过数据诊断发现存在这些因素的“奇异点”，然后通过技术手段消除其影响。但对于非显著因素，无论是恒常性还是偶然性的，尽管它们的单独影响可以忽略不计，却不能简单忽略掉无数非显著因素的影响。格林(W.H.Greene)在2003年时指出，没有什么模型可以期望它能处理经济现实的无数偶然因素，因此在经验模型中纳入随机要素是必须的，被解释变量的观测值不仅要考虑已经了解清楚的变量，也要考虑来自人们并不了解的偶然性和无数微弱因素的影响。

因此，总体回归模型设定的主要任务是确定影响被解释变量Y的显著恒常因素集$\boldsymbol{X}$，以及确定被解释变量Y与$\boldsymbol{X}$之间的关系形式和关系参数。在本节，

主要讨论前者，而将后者放到§9.3专门讨论。

在目前的计量经济学应用研究中，大量的问题和错误存在于总体回归模型设定中：或者按照研究者的研究目的选择模型的解释变量；或者照搬某种经济理论，简单地按照理论的提示选择模型的解释变量；或者根据变量观测值显示的统计关系选择模型的解释变量。凡此种种，都是片面的甚至是错误的。本节将针对这些问题和错误，提出总体回归模型设定的若干原则。

二、计量经济学模型总体设定的“一般性”原则

1．总体回归模型设定的“研究目的导向”及其问题

任何应用研究都有特定的研究目的，例如分析某两个经济变量之间的关系，或者评价某项经济政策的效果。于是，按照特定的研究目的进行计量经济学模型总体模型的设定，成为计量经济学应用研究的普遍现象和最严重的问题。

例9.2.1

为了分析我国经济增长中各个投入要素的贡献，需要建立总量生产函数模型。根据对我国经济增长行为的深入分析，认为可以将所有投入要素分为4类：制度、技术、资本、劳动。以 ZD 表示制度变迁指数，JS 表示技术进步指数，ZB 表示资本投入指数，LD 表示劳动投入指数，如果采用 Cobb-Dauglas(也称 C-D)生产函数模型形式，以4类要素的时间序列数据为样本观测值，建立如下模型：

$$\ln \mathrm{GDP}_t = \beta_0 + \beta_1 \ln ZD_t + \beta_2 \ln JS_t + \beta_3 \ln ZB_t + \beta_4 \ln LD_t + \mu_t$$
$$t = 1, 2, \cdots, T \tag{9.2.2}$$

该总体回归模型设定符合“一般性”原则。因为模型中包括了所有对被解释变量产生影响的变量。

但是，在一篇专门研究我国制度变迁与经济增长的关系的论文中，以 GDP 为被解释变量，仅以制度变迁指数作为解释变量，建立了一元对数线性模型：

$$\ln \mathrm{GDP}_t = \alpha_0 + \alpha_1 \ln ZD_t + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.3}$$

估计结果显示，制度变迁对于 GDP 的弹性系数为2.1。于是得到结论：我国制度变迁指数每提高1%，国内生产总值将增长2.1%。

在另外一篇专门研究我国资本投入对经济增长贡献的论文中，以 GDP 为被解释变量，仅以固定资产原值(代表资本投入指数)作为解释变量，建立了一元对数线性模型：

$$\ln \mathrm{GDP}_t = \gamma_0 + \gamma_1 \ln ZB_t + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.4}$$

模型估计显示，拟合优度达到 0.95。于是得到结论：我国的资本投入可以很好地解释 GDP 的增长，所以其他投入要素，包括劳动、技术等对经济增长贡献几乎为 0。

例 9.2.1 中的模型(9.2.3)和(9.2.4)是典型的“研究目的导向”，其研究结论显然是十分荒谬的。而类似的应用研究到处可见，包括在权威的经济学刊物上。例如，在一篇研究我国证券市场发展对宏观经济影响的论文中，为了分析证券市场发展对财政收入的影响，作者以我国财政收入为被解释变量，以股票融资额为唯一解释变量，建立了一元线性模型。估计结果显示，股票融资额增加 1 亿元，财政收入将增加 4.729 亿元。再如，在一篇研究我国财政支农支出对缩小城乡收入差距的影响的论文中，将我国财政支农支出作为唯一的解释变量，以城乡收入差距作为被解释变量，建立了一元线性模型。估计结果显示，随着财政支农支出的增加，城乡收入差距竟然在不断扩大！

2．总体回归模型设定的“一般性”原则

计量经济学模型总体设定，必须遵循“唯一性”原则。对于同一个作为研究对象的被解释变量，它和所有影响因素之间只能存在一种客观的正确的关系。或者说，对于一组被解释变量样本观测值，只能由一种客观的数据生成过程生成。所以，正确的总体模型只能是一个。不同的研究者、不同的研究目的、不同的数据选择方法、不同的数据集，会对模型的约化和简化过程产生影响，会使得最终的应用模型有所不同。但是，作为研究起点的总体模型必须是唯一的。而这个具有“唯一性”的模型，必须最具“一般性”。

计量经济学模型总体设定，必须遵循“一般性”的原则，即作为建模起点的总体模型必须能够包容所有经过约化得到的“简洁”的模型。具体讲，它应该包含所有对被解释变量产生影响的变量，尽管其中的某些变量会因为显著性不高或者不满足正交性条件等原因在后来的约化过程中被排除。在计量经济学模型发展的历史上，曾经倡导过“从简单到一般”的建模思路，那是由于历史的局限，已经被“从一般到简单”的建模思路所取代。

3．为什么必须遵循“一般性”原则？

可以从逻辑学、经济学和统计学三个方面加以解释。

从逻辑学上讲，计量经济学模型方法是一种经验实证的方法，它是建立在证伪和证实的不对称性的逻辑学基础之上的。一旦总体模型被设定，利用样本数据进行的经验检验只能发现已经包含其中的哪些变量是不显著的，而不能发现没有包含其中的显著变量。例如对于模型(9.2.2)，如果劳动投入对于 GDP 的影响不显著，那么可以通过经验发现并加以剔除；但是对于模型(9.2.3)或(9.2.4)，

经验检验则不可能告诉你模型缺少了哪些显著的解释变量。

从经济学上讲，总体回归模型必须反映现实的经济活动，而现实经济活动中变量之间的关系是复杂的。一些经济学理论经常采用简洁的语言，揭示两个变量之间的关系。例如，需求法则指出“需求量随着价格的上升而下降”。但是它的前提是“其他因素不变”。而在现实经济活动中，所有因素都在变化，如果仅用价格作为需求量的唯一解释变量建立并估计模型，是很难合理地检验需求法则的，因为收入也是变化的。

从统计学上讲，只有首先建立最一般的模型，才能保证模型的随机扰动项满足“源生性”和基本假使。如果省略了显著的变量，例如模型(9.2.3)或(9.2.4)，其随机扰动项中包括了省略的变量对被解释变量的影响，破坏了基本假设，在此基础上进行的模型估计和推断都是无效的。

4．不易发现的违背“一般性”原则的实例

模型(9.2.3)或(9.2.4)的错误是容易发现的。为了加深读者对“一般性”原则的理解，下面列举几个不易发现的违背“一般性”原则的实例。

例 9.2.2

在某权威刊物发表的一篇关于中国和印度不平衡发展的比较研究的论文中，着重分析不平衡增长对贫困的影响。为了分析产业间的增长不平衡对贫困的影响，以贫困率为被解释变量，以人均 GDP 和三大产业在 GDP 中的份额为解释变量，建立了一组回归模型；为了分析居民收入增长不平衡对贫困的影响，以贫困率为被解释变量，以农村居民平均收入增长率、城市居民平均收入增长率，以及人口流动效应为解释变量，建立了另外一组回归模型。

读者肯定会问，既然贫困率受到产业间的增长不平衡和居民收入增长不平衡的共同影响，为什么不建立一组包括两方面因素的模型而要分别建立模型？分别建立的模型的估计结果和统计推断有意义吗？贫困率除了受到不平衡增长的影响外，制度因素、政策因素等也有重要影响，为什么在模型中未予考虑？

例 9.2.3

在一篇关于居民社会信任水平的影响因素分析的论文中，通过二元 Probit 模型分析居民的社会信任水平与各个影响因素之间的关系。作者利用实际调查的微观数据，以二元离散变量表示居民的社会信任水平，如果受访者表示社会上大多数人可以信任，该变量赋值为 1；反之为 0。

论文严肃科学地分析了居民社会信任水平的影响因素，将其分为三类：个人因素，例如性别、年龄、受教育程度、收入水平、就业情况、宗教信仰；社区因素，例如在本市居住的时间、日常语言类型；社会因素，例如是否参加社会团体、对政府的评价、对媒体的评价。论文首先选择个人因素作为解释变量建立模型，估计其参数；然后“控制”个人因素，引入社区因素作为解释变量建立模型，估计其参数；最后“控制”个人因素和社区因素，引入社会因素作为解释变量建立模型，估计其参数。

一个显而易见的问题是，为什么不直接建立一个最“一般”的包括所有影响因素的总体模型？既然影响因素包括三类，那么以其中某一类作为解释变量建立模型，其参数估计结果有意义吗？

---

5. 什么情况下可以不遵循“一般性”原则？

在进行了上述讨论后，读者肯定会提出一个问题：发表于国内外权威刊物的计量经济学应用研究论文，建立模型的通常程序是“从简单到复杂”，即开始设定简单的模型，包括较少的解释变量，经过估计后，如果发现拟合效果不好，再增加解释变量，直到满意为止。难道它们都是错误的？

首先，必须明确这种“从简单到复杂”的模型设定思路，不符合计量经济学模型方法论的逻辑学、经济学和统计学基础，容易走上实用主义的歧途，不值得提倡。

其次，应该承认在特定的情况下，这些研究的结论是可以成立的，即如果所有显著的影响因素(解释变量)在行为上是独立的，在统计上是不相关的，那么“简单”模型的结论是能够成立的。回到例 9.2.1，如果制度变迁指数、技术进步指数、资本投入指数和劳动投入指数是完全独立的，那么根据本书第三章介绍的理论，模型(9.2.3)中参数 α_1 和模型(9.2.2)中的参数 β_1 的估计结果应该是相同的，模型(9.2.4)中参数 γ_1 和模型(9.2.2)中的参数 β_3 的估计结果应该是相同的。同样，在例 9.2.2 中，如果能够证明产业增长不平衡、居民收入增长不平衡是相互独立的，那么模型的结果是可靠的，可是在论文中并未见到这样的证明。在例 9.2.3 中，如果能够证明个人因素、社区因素、社会因素是相互独立的，那么模型的结果是可靠的，在论文中也未见到这样的证明。

三、计量经济学模型总体设定的“现实性”原则

1. 总体回归模型设定的“先验理论导向”及其问题

20 世纪 30 年代至 70 年代发展的经典计量经济学模型，经济理论在其总体模型设定中起导向作用。计量经济学根据已有的经济理论进行总体模型的设定，

将模型估计和模型检验看成是自己的主要任务。经济理论可以被认为是嵌入计量经济学模型的，相对经验数据而言具有先验性。克莱因(L.R. Klein，1974)指出，经济理论能够提出一些用数学形式表达，然后再从计量经济学观点加以检验的假设，但是必须指出，学院式的经济理论仅仅是建立假设的来源之一。但是，在经典计量经济学模型的应用研究中，直接依据经济学理论设定总体模型的现象十分普遍，因此**经典计量经济学模型通常被认为是先验理论导向。**

问题在于，**能否以先验的经济学理论作为计量经济学模型总体设定的导向？答案是否定的**。因为在它们之间，至少存在以下几个障碍。

第一，正统经济学以经济人假设和理性选择为其理论体系的基石，任何一种理论都建立在决策主体是理性的和决策行为是最优的基础之上。而计量经济学模型总体设定的目的，是建立能够描述人们实际观察到的经济活动之中蕴藏着的一般规律的总体模型，毫无疑问，实际经济活动既不是“理性”的，也不是“最优”的。

第二，正统经济学理论强调“简单”，认为只有简单的理论才能够揭示本质。而计量经济学模型恰恰相反，它强调“一般”，必须将经济活动所涉及的所有因素包含其中。所以，即使经济学理论是正确的，也不能据此设定计量经济学模型，因为它舍弃了太多显著的因素。

第三，对于同一个研究对象，不同的研究者依据不同的先验理论，就会设定不同的模型。例如，以居民消费为研究对象，分别依据绝对收入消费理论、相对收入消费理论、持久收入消费理论、生命周期消费理论以及合理预期消费理论，就会选择不同的解释变量和不同的函数形式，设定不同的居民消费总体模型。

2. 总体回归模型设定的“现实性”的原则

通俗地讲，经济学理论所揭示的是理想的经济世界，而计量经济学模型描述的是现实的经济世界。所以，计量经济学应用研究的总体回归模型设定必须遵循“现实性”的原则。

所谓“现实性”的原则，就是客观地分析研究对象的现实行为，从中发现变量之间的因果关系，并由此设定总体回归模型。下面通过两个实例进行解释。

例 9.2.4

在一篇实证研究我国货币-产出非对称影响关系的论文中，作者采用多元 STAR 模型研究我国货币-产出关系，模型系统中包括的变量有表示产出的实际工业产出指数、表示货币的 M1 和表示价格的消费价格指数，而货币流通速度被合理地省略了。显然，如此选择模型系统的变量，所依据的是经典的货币需求理论。

PDG

那么需要讨论的是，经典的货币需求理论是否反映我国的实际？以此作为描述货币-产出关系的总体模型设定的依据是否可靠？按照经济学中经典的货币需求理论，货币需求系统仅包含货币需求量、经济活动总量、价格和货币流通速度，按照理论导向，该论文中设定的总体模型是正确的。但是，人们都知道，经济学在它的发展过程中，出现了许多货币需求理论，那么不同的研究者依据不同的理论就可以设定不同的总体模型。更为重要的是，论文研究的是我国的货币与产出之间的关系，而在我国的货币需求系统中，相互有关联的因素远不只是论文所涉及的 3 个，特别是政策因素对我国的货币需求量的影响是不可忽略的。

例 9.2.5

在一篇关于人民币汇率的均衡、失调、波动与调整的论文中，作者通过理论分析和经验检验，得到描述实际汇率(REER)与相对供给(SU)、相对需求(DE)之间长期均衡关系的模型：$\mathrm{REER} = -0.33\mathrm{SU} + 0.26\mathrm{DE} + 4.59$，并在此基础上建立了反映短期变化之间关系的误差修正模型。显然，经典的汇率决定理论在该模型总体设定中起了导向作用。

人们同样会问：经典的汇率决定理论是否反映我国的实际？以此作为选择模型变量的依据是否可靠？虽然上述模型描述了实际汇率与相对供给、相对需求之间长期的均衡关系，但其他因素的影响却被忽略了，至少我国的汇率政策对实际汇率是存在显著影响的。

对计量经济学模型总体设定先验理论导向的批评和提倡“现实性”原则，并不意味着完全否定经济学理论在模型设定中的作用。描述理想经济世界的经济学理论可以指导我们正确分析现实经济世界的经济行为关系；简洁的经济学理论至少揭示了“一般”经济系统中的一部分经济关系。经济学理论将作用于经济行为关系分析，而不是直接作用于模型总体设定。

四、计量经济学模型总体设定的“统计检验必要性”原则

1. 总体回归模型设定的“数据关系导向”及其问题

在第二次世界大战结束后的 20 多年中，由于当时主流的经济理论，特别是宏观经济理论与现实经济活动之间较好的一致性，以先验理论为导向的经典计量经济学模型得到了迅速的扩张和广泛的应用。但是，经典模型对 20 世纪 70 年代经济衰退和滞胀的预测和政策分析的失效，引来了著名的“卢卡斯批判”。卢卡斯批判从表面上看是对结构模型和模型结构不变性的批判，而实质上是对

模型总体设定先验理论导向的批判。基于截面数据的经典模型面临先验理论与经济现实的脱节，而被迫更多地转而依赖数据关系和统计分析。这就直接导致了计量经济学总体模型设定转向“数据关系导向”。

同时，基于时间序列数据的计量经济学模型由于存在非平稳性和序列相关性，其统计分析理论方法得到了迅速发展，一方面为模型总体设定提供了强大的工具，另一方面又将模型设定引入仅仅依赖数据的歧途。

数据的时间序列性破坏了计量经济学模型的随机抽样假定，取消了样本点之间的独立性，样本点之间将发生序列相关。如果序列相关性不能足够快地趋于零，在统计推断中发挥关键作用的大数定律、中心极限定理等极限法则缺乏应用基础。所以，只有对满足渐近不相关的协方差平稳序列，才可以适用基于截面数据的统计推断方法，建立时间序列模型。这样，协方差平稳性和渐近不相关性在时间序列分析中扮演了一个非常重要的角色，为时间序列分析适用大数定律和中心极限定理创造了条件，替代了截面数据分析中的随机抽样假定。

但是经济现实中的随机过程都很难符合这些条件。在不适用大数定律和中心极限定理的情况下，经典模型的计量分析常会产生欺骗性的结论。对包含非平稳随机变量的模型的谬误回归，引出两个问题，一是是否可以统计确定具有恒常关系的非平稳随机变量之间的模型；二是如何处理非平稳随机过程，为适用统计方法建立模型创造条件。于是，对时间序列的非平稳性的识别与处理，即单位根检验，以及在非平稳随机过程之间建立恒常的数据关系，即协整检验，成为模型总体设定的主要任务，正如本书第八章所讨论的。

这样，带来的新的问题是，计量分析的理论基础——经济行为理论反而被忽略了。时间序列的数据协整关系是结果，而不是原因；由于经济现实的系统关联性，满足统计协整关系的变量很多，但是可以纳入基于经济行为建立的动态均衡模型的变量并不多。因此，协整关系检验是模型总体设定的必要条件，但不是充分条件。必须在经济行为分析，即经济系统中变量之间动力学关系分析的基础之上，才能有效发挥协整检验的作用。

例 9.2.6

举一个比较极端的例子。在一项关于我国城镇居民收入的研究中，作者为了检验城镇居民收入对农村居民消费存在影响，对城镇居民人均收入(CZSR)和农村居民人均消费(NCXF)两个时间序列数据进行了严格的统计分析。首先进行单位根检验，发现它们都是 2 阶单整序列。然后进行格兰杰因果关系检验，发现在 5%的显著性水平上，城镇居民人均收入是农村居民人均消费的格兰杰原因。最后进行协整检验，发现它们之间存在(2，2)阶协整。于是得到了描述二者之间长期均衡关系的模型，并进行模型估计

得到：

$$\mathrm{NCXF}_t = 558.07 + 0.181\,7\mathrm{CZSR}_t$$

根据模型指出，城镇居民人均收入每提高 100 元，可以使得农村居民人均消费提高 18.17 元。

这个结论显然是错误的，但是所有统计检验却是严格的。问题在哪里？对农村居民消费行为进行分析，不难发现农村居民收入是最主要的影响因素。而将农村居民人均收入引入模型，容易发现，城镇居民收入并不显著。

2. 模型总体设定的“统计检验必要性”原则

类似于例 9.2.6 的情况在计量经济学应用模型研究中并不少见。尤其是它们采用了先进的统计分析方法，使之更具有隐蔽性。因此，需要正确认识、对待统计分析在计量经济学模型总体设定中的作用。

如果经济时间序列在经济行为上存在直接因果关系，那么它们在统计上一定存在协整关系，一定能够通过统计检验，包括因果关系检验和协整检验。但是反过来，在统计上能够通过因果关系检验和协整检验的经济时间序列，在经济行为上并不一定存在直接因果关系。由于人们认识的局限，在经济行为分析中发现的因果关系并不一定都是正确的，所以在经济行为分析的基础上进行统计检验是完全必要的，以此达到“去伪存真”的效果。从这个意义上，单位根检验、因果关系检验和协整检验理论，给出了总体回归模型设定的有效工具。这就是计量经济学模型总体设定的“统计检验必要性”原则。

五、计量经济学模型总体设定的“经济主体动力学关系导向”原则

对计量经济学模型总体设定的讨论，必须首先明确两个问题。第一，要确定的不是经济主体内在的本质意义的属性，而是经济主体之间的关系意义的属性；第二，要确定的是主体之间的动力学关系，不是作为主体经济活动结果的经济变量之间的数据关系。这就是计量经济学模型总体设定的“经济主体动力学关系导向”原则。

而事实上，无论先验理论导向，还是数据关系导向，计量经济学模型总体设定所忽视的正是经济主体之间的动力学关系。计量经济学模型分析的目的不是为了确定在主体关系意义上无所指的经济变量之间的关系。经济变量及相关数据是经济主体活动的结果，脱离主体互动关系建构的变量，不过是纯粹的数字。从关系论的角度看，主体的任何行为，都应在主体和其身处的环境之间寻找原因。正如自然科学的动力学研究一样，物体运动状态发生变化的根本原因是物体的环境与物体之间的作用力。同样地，经济主体发生任何行为，都必然

由主体与其身处的环境之间的作用引起。

经济主体与其身处的环境之间的动力学过程，是真正的数据生成过程。与经济主体的特定动力学过程相关的数据，将为相应动力学关系的描述提供经验基础。以经济主体与环境之间的动力学关系分析为基础和前提，基于该动力学过程生成的数据，以数据统计分析为必要条件，验证确定的经济主体与环境的互动关系，正是计量经济学总体模型所要界定的因果关系。只有动力学关系的理论分析，没有基于统计相关性的经验支持，是无法确认这样的动力学关系的。同样，只有数据关系的统计分析，没有具有良好的动力学关系理论框架，也会使统计分析误入歧途。正是在这个意义上，基于主体动力学关系的计量经济学模型总体设定，可以实现先验理论导向和数据关系导向的综合。

可以用图 9.2.1 清晰地描述先验的经济理论、数据的统计分析、经济主体的动力学关系与计量经济学总体模型之间的关系。在这里，先验的经济理论并不直接作为总体模型设定的导向，而是指导经济主体的动力学关系分析；数据的统计分析也不直接作为总体模型设定的导向，而是对经济主体的动力学关系进行检验；而对总体模型设定起直接导向作用的，是经济主体的动力学关系。

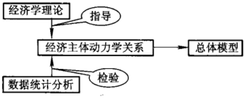


图 9.2.1 理论、数据、动力学关系与总体模型

以经济主体的动力学关系为导向设定的总体回归模型，毫无疑问满足上述的“现实性”原则和“统计检验必要性”原则，但是它是否满足“一般性”原则，仍然需要检验。检验的准则就是总体模型随机扰动项的源生性和正态性。如果模型设定正确，随机扰动项所包含的仅仅是非显著因素的影响，这样的随机扰动是源生的。只要保证随机扰动项的源生性，它所包含的因素满足独立性，以及对随机扰动的影响均匀小的条件，根据中心极限定理，这样的随机扰动项服从正态分布。所以，在动力学关系导向的计量经济学模型总体设定中，中心极限定理仍然居于十分重要的地位。

六、案例——消费理论与消费函数模型

消费理论是宏观经济学理论的重要内容，旨在研究消费行为。关于消费行为的研究，即消费理论，一直受到高度重视，出现了各种消费理论。这里的消

费指消费总量，而不是对具体商品或服务的消费需求，这是它有别于需求理论的主要之点。它的研究对象可以是一个国家、一个群体，甚至一个个体，但一定是研究对象的总消费。消费函数模型是关于研究对象的总消费与影响因素，主要是可支配的总收入之间关系的数学表达式，它也是计量经济学应用模型中一个重要的组成部分。按照不同的消费理论，可以建立不同的消费函数模型。

本节以此为案例，一方面试图加深读者对总体回归模型“先验理论导向”的认识，以及强调遵循“现实性”原则的重要性；另一方面也借此介绍这一重要的计量经济学模型应用研究领域。

1. 绝对收入假设消费函数模型

(1) 绝对收入假设消费函数模型

凯恩斯(Keynesian)认为，消费是由收入唯一决定的，消费与收入之间存在着稳定的函数关系。随着收入的增加，消费将增加，但消费的增长低于收入的增长，即边际消费倾向递减。根据这一理论假设，按照总体回归模型“先验理论导向”，可以建立如下消费函数模型：

$$C_t = \alpha + \beta Y_t + \mu_t \quad t = 1, 2, \cdots, T \tag{9.2.5}$$

其中C表示消费额，Y表示收入，α, β为待估参数。从经济意义上讲，α为自发性消费，β为边际消费倾向，于是有$0 < \beta < 1$，$\alpha > 0$。模型(9.2.5)可以很方便地采用单方程模型的估计方法估计其参数。

(2) 关于绝对收入假设消费函数模型的讨论

模型(9.2.5)表达了凯恩斯所说的消费是由收入唯一决定的假设，但是由于边际消费倾向β为常数，并没有真正反映边际消费倾向递减的规律。在一般的教科书上，以(9.2.5)式满足

$$0 < \frac{\partial C}{\partial Y} < 1 \text{，} \quad \frac{\partial C}{\partial Y} < \frac{C}{Y}$$

为由，认为模型反映了边际消费倾向递减规律。实际上，建立变参数模型，即假设

$$\beta = \beta_0 + \beta_1 Y_t$$

其中$\beta_1 < 0$，代入(9.2.5)式得

$$C_t = \alpha + \beta_0 Y_t + \beta_1 Y_t^2 + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.6}$$

上式可以较好地反映边际消费倾向递减规律。(9.2.6)式仍然可以很方便地采用单方程模型的估计方法估计其参数。

2. 相对收入假设消费函数模型

(1)“示范性”假设消费函数模型

绝对收入假设消费函数模型认为消费者的消费行为是独立的，不受周围环境的影响。这种消费行为假设是不符合客观实际的。**杜伊森贝里(Duesenberry)**

认为，消费者的消费行为不仅受自身收入的影响，也受周围人的消费水平的影响。例如，若周围人的消费水平较高，即使某个消费者的收入水平较低，也企图接近周围人的消费水平，于是他的边际消费倾向就会比较高。这种现象被称为消费的“示范性”。

由消费的“示范性”，个人的平均消费倾向不仅与收入有关，而且与个人所处的群体的收入分布有关，在收入分布中处于低收入的个人，往往有较高的消费倾向，即

$$\frac{C_i}{Y_i}=\alpha_0+\alpha_1\frac{\overline{Y}_i}{Y_i} \tag{9.2.7}$$

其中 $\overline{Y}_i$ 为该消费者所处的群体的平均收入水平。从(9.2.7)式可以看出，当 $\alpha_0,\alpha_1,\overline{Y}_i$ 一定时，对于较低的 Y_i，其 C_i/Y_i 较高。这就是“示范性”的作用。(9.2.7)式的计量形态可表示为

$$C_i=\alpha_0 Y_i+\alpha_1\overline{Y}_i+\mu_i \qquad i=1,2,\cdots,n \tag{9.2.8}$$

其中待估参数 $0<\alpha_0<1$，反映个人的边际消费倾向；$0<\alpha_1<1$，反映群体平均收入水平对个体消费的影响。该模型可以很方便地采用单方程模型的估计方法估计其参数。但是，样本必须取自不同的群体，否则不能反映“示范性”对消费的影响。

(2)“不可逆性”假设消费函数模型

绝对收入假设消费函数模型认为消费者的消费行为只由当前收入水平决定，与历史上曾经发生的消费活动无关。这种消费行为假设也是不符合客观实际的。**杜伊森贝里认为，消费者的消费支出水平不仅受当前收入的影响，也受自己历史上曾经实现的消费水平的影响**。例如，若历史上曾经达到较高的消费水平，即使当前的收入水平较低，也企图接近历史上曾经达到的消费水平，于是当前的边际消费倾向就会比较高。这种现象被称为消费的“不可逆性”。

由消费的“不可逆性”，当前的平均消费倾向不仅与收入有关，而且与所曾经达到的消费水平，即曾经达到的最高收入水平有关，当前收入低于曾经达到的最高收入时，往往有较高的消费倾向，即

$$\frac{C_t}{Y_t}=\alpha_0+\alpha_1\frac{Y_0}{Y_t} \tag{9.2.9}$$

其中 Y_0 为该消费者曾经达到的最高收入水平。从(9.2.9)式可以看出，当 α_0,α_1,Y_0 一定时，对于较低的 Y_t，其 C_t/Y_t 较高。这就是“不可逆性”的作用。(9.2.9)式的计量形态可表示为

$$C_t=\alpha_0 Y_t+\alpha_1 Y_0+\mu_t \qquad t=1,2,\cdots,T \tag{9.2.10}$$

其中待估参数 $0<\alpha_0<1$，反映当前的边际消费倾向；$0<\alpha_1<1$，反映曾经达到

的最高收入水平对当前消费的影响。一般情况下，收入具有随时间递增的趋势，所以可以用前一个时期的收入代替曾经达到的最高收入。于是模型(9.2.10)可以改写为

$$C_t = \alpha_0 Y_t + \alpha_1 Y_{t-1} + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.11}$$

该模型也可以很方便地采用单方程模型的估计方法估计其参数。

3．生命周期假设消费函数模型

莫迪利亚尼(Modigliani)，布拉姆帕格(Brumberg)和安东(Ando)于 1954 年提出，消费者现期消费不仅与现期收入有关，而且与消费者以后各期收入的期望值、开始时的资产数量和年龄有关。消费者一生中消费支出流量的现值要等于一生中各期收入流量的现值。所以，消费者的预算约束为

$$\sum_{t=1}^{T} \frac{C_t}{(1+r)^{t-1}} = \sum_{t=1}^{T} \frac{Y_t}{(1+r)^{t-1}}$$

其中 r 为贴现率。在预算约束下，消费者总希望将自己一生的全部收入在消费支出中进行最优分配，使得效用函数 $U(C_1, C_2, \cdots, C_T)$ 达到最大。于是推导消费函数问题就变成下列拉格朗日函数的极值问题：

$$L(C_1, C_2, \cdots, C_T, \lambda) = U(C_1, C_2, \cdots, C_T) + \lambda\left(\sum_{t=1}^{T} \frac{Y_t}{(1+r)^{t-1}} - \sum_{t=1}^{T} \frac{C_t}{(1+r)^{t-1}}\right) \tag{9.2.12}$$

(9.2.12)式的极值条件为

$$\begin{cases} \dfrac{\partial L}{\partial C_t} = \dfrac{\partial U}{\partial C_t} - \dfrac{\lambda}{(1+r)^{t-1}} = 0 \qquad t = 1, 2, \cdots, T \\ \dfrac{\partial L}{\partial \lambda} = \sum\limits_{t=1}^{T} \dfrac{Y_t}{(1+r)^{t-1}} - \sum\limits_{t=1}^{T} \dfrac{C_t}{(1+r)^{t-1}} = 0 \end{cases}$$

求解该方程组，即可得到最优消费的消费函数为

$$C_t = c_t(Y_1, Y_2, \cdots, Y_T, r)$$

表明消费是各个时期的收入和贴现率的函数。

一般近似地用下列函数描述生命周期假设消费函数模型：

$$C_t = \alpha_1 Y_t + \alpha_2 A_t + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.13}$$

其中 A_t 为时刻 t 的资产存量，待估参数 $0 < \alpha_1 < 1$，反映当前的边际消费倾向；$0 < \alpha_2 < 1$，反映消费者已经积累的财富对当前消费的影响。模型(9.2.13) 可以很方便地采用单方程模型的估计方法估计其参数。

4．持久收入假设消费函数模型

弗里德曼(Friedman)于 1957 年提出了消费的持久收入假设，它是对凯恩斯的绝对收入假设的修正与补充。分析消费者的消费行为发现，在消费中有一部分是经常的必须保证的基本消费，另一部分是非经常的额外消费；而收入也可

以分成两部分，一部分是可以预料到的长久性的、带有常规性的持久收入，另一部分是非连续性的、带有偶然性的瞬时收入，即

$$Y_t = Y_t^p + Y_t^t$$

$$C_t = C_t^p + C_t^t$$

其中 Y_t, Y_t^p, Y_t^t 分别为实际收入、持久收入和瞬时收入；C_t, C_t^p, C_t^t 分别为实际消费、持久消费和瞬时消费。持久消费由持久收入决定，瞬时消费由瞬时收入决定。于是持久收入假设消费函数模型的一种计量形态是

$$C_t = \alpha_0 + \alpha_1 Y_t^p + \alpha_2 Y_t^t + \mu_t \qquad t = 1, 2, \cdots, T \qquad (9.2.14)$$

估计(9.2.14)式的参数的困难在于样本观测值的选取，因为能够得到的是实际收入，而不是持久收入和瞬时收入。弗里德曼建议，对于时间序列数据，时刻 t 的持久收入可以表示为各期实际收入的加权和：

$$Y_t^p = \lambda Y_t + \lambda(1-\lambda)Y_{t-1} + \lambda(1-\lambda)^2 Y_{t-2} + \cdots \qquad 0 < \lambda < 1$$

即 $Y_t^p - Y_{t-1}^p = \lambda(Y_t - Y_{t-1}^p)$。在实际应用时，首先给定一个 λ 值，计算每年的持久收入观测值，再由此计算瞬时收入观测值，然后估计模型(9.2.14)。反复修改 λ 值，直至取得满意的拟合结果。

5. 合理预期的消费函数模型

理性预期理论认为，人们可以对原因变量进行预期，然后根据原因变量的预期值对结果变量进行预测。于是在消费函数研究中，假设第 t 期的消费是收入预期值 Y_t^e 的函数，即

$$C_t = \alpha + \beta Y_t^e \qquad (9.2.15)$$

表示**消费者按收入预期决定自己的消费计划和实现消费**。而收入预期值 Y_t^e 是现期实际收入与前一期预期收入的加权和：

$$Y_t^e = (1-\lambda)Y_t + \lambda Y_{t-1}^e = (1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots)$$

代入(9.2.15)式得到

$$C_t = \alpha + \beta\ (1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots)$$

$$C_{t-1} = \alpha + \beta\ (1-\lambda)(Y_{t-1} + \lambda Y_{t-2} + \lambda^2 Y_{t-3} + \cdots)$$

$$C_t - \lambda C_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)Y_t$$

于是可以将合理预期的消费函数模型的计量形态表示为

$$C_t = \alpha(1-\lambda) + \lambda C_{t-1} + \beta(1-\lambda)Y_t + \mu_t \qquad t = 1, 2, \cdots, T \qquad (9.2.16)$$

模型(9.2.16) 可以很方便地采用单方程线性模型的估计方法估计其参数。

6. 适应预期的消费函数模型

适应预期理论认为，人们可以根据原因变量的实际值对结果变量进行预期，但是实际上往往达不到预期的结果，因而需要对结果变量的预期值进行调

整。于是，在消费函数研究中，假设第 t 期的消费预期值 C_t^e 是收入的函数，即

$$C_t^e = \alpha + \beta Y \tag{9.2.17}$$

表示**消费者按收入决定自己的消费预期**。而由于种种原因，实际消费与消费预期值之间存在如下关系：

$$C_t - C_{t-1} = \lambda(C_t^e - C_{t-1})$$

其中，λ 为调整系数。可以将该式写成

$$C_t^e = \frac{1}{\lambda}C_t + \frac{\lambda-1}{\lambda}C_{t-1}$$

代入(9.2.17)式即可求得消费函数模型，其计量形态为

$$C_t = \lambda\alpha + (1-\lambda)C_{t-1} + \lambda\beta Y_t + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.2.18}$$

可以很容易地估计该模型。

7. 一点启示

从以上已经介绍的消费理论和消费函数模型可以看到，不同的理论假设导出不同的模型。如果我们试图建立中国的消费函数模型，不同的研究者，依据不同的消费理论，就可以设定不同的消费函数模型。如果仅仅试图通过研究检验每种消费理论是否适合于我国，也许是有意义的。如果研究的目的是为了揭示我国的消费行为，揭示影响消费的各个因素对消费的实际影响，那么不同的研究者就会得到不同的结论，这样的研究是没有意义的。先验的经济理论，可以指导我们分析实际的经济行为关系，但不能直接作为总体回归模型设定的导向。

§9.3 计量经济学应用模型函数关系设定

在确定了计量经济学应用模型的模型类型、模型中应该包含的解释变量后，模型设定的下一项任务就是设定解释变量与被解释变量之间的函数关系。它们之间是直接线性关系还是非线性关系？如果是非线性关系，是否可以通过简单变换化为线性关系？这也属于总体回归模型设定的内容之一，是选择估计方法进行模型估计的前提之一。

一、模型的关系类型

对于一个单方程计量经济学模型，设 Y 为被解释变量，$X = (X_1, X_2, \cdots, X_k)$ 为解释变量向量，它是直接影响变量 Y 的 k 个变量。模型的一般形式为

$$f(Y, \mathbf{X}, \theta, \beta) = \mu \tag{9.3.1}$$

一般情况下，被解释变量和解释变量是可以分离的，于是(9.3.1)式可以写成

$$h(Y_i,\theta)=g(X_i,\beta)+\mu_i \qquad i=1,2,\cdots,n \tag{9.3.2}$$

其中 $h(\cdot)$ 和 $g(\cdot)$ 是非线性函数，θ 和 β 是参数，μ_i 是随机扰动项。通过被解释变量的参数变换，常见的单方程非线性计量经济学模型表示为

$$Y_i=m(\boldsymbol{X}_i,\beta)+\mu_i \qquad i=1,2,\cdots,n \tag{9.3.3}$$

如果(9.3.3)式可以通过简单的变换化为线性的，或者被解释变量和解释变量原本就呈现直接线性关系，模型就是我们所熟悉的形式：

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\cdots+\beta_kX_{ki}+\mu_i \qquad \mu_i\sim N(0,\sigma^2) \quad i=1,2\cdots,n \tag{9.3.4}$$

在实际计量经济学应用研究中，开始就将总体回归模型设定为(9.3.4)式的情况并不多见，因为实际经济活动中解释变量与被解释变量之间呈现直接线性关系的并不多。例如下面将要介绍的 C-D 生产函数模型、CES 生产函数模型都是(9.3.3)式的非线性模型，但是，它们大都可以通过简单的变换，例如变量置换、函数变换、级数变换等，变换成为(9.3.4)式的线性模型。

本节讨论的计量经济学应用模型变量函数关系的设定，主要是针对(9.3.3)式而言的。

二、模型关系误设的后果

计量经济学应用模型的关系误设的后果主要表现于经济学和统计学两个方面。其经济学后果是显而易见的。若反映被解释变量和解释变量之间关系的参数在经济含义上和具体数值上都具有严重的偏误，即使后面的统计推断再严密，其结论也是没有意义的。

1．“源生的”随机扰动项

统计学后果主要表现于随机扰动项。从经济学意义上，$X=(X_1,X_2,\cdots,X_k)$ 包含了所有对 Y 具有显著影响的因素，$m(X_i,\beta)$ 表达了这些因素与 Y_i 之间的动力学关系，生成了 Y_i 的条件期望值。但是，无数不显著因素的影响对于生成 Y_i 的观测值是不可忽略的，“不显著”不是“没有影响”。从统计学意义上，在 $\boldsymbol{X}_i$ 的条件下重复抽样，无数不显著因素对 Y_i 的均值没有影响，但是在一次抽样中，它们对 Y_i 的个值的影响是不可忽略的。如果 μ_i 仅仅是无数不显著因素对 Y_i 个值的影响，在基于随机抽样的截面数据的经典计量经济学模型中，这个“源生的”随机扰动项 μ 由大数定律保证其满足高斯假设，由中心极限定理可以证明其服从正态分布。于是，建立在高斯假设和正态分布假设基础上的统计推断具有可靠性。

2．“衍生的”随机误差项

在大部分计量经济学教科书中，在引入随机扰动项的概念时，都将它定义为“被解释变量观测值与它的期望值之间的离差”，即

$$\mu_i = y_i - E(y|\boldsymbol{X}_i) \tag{9.3.5}$$

用一个平衡式代替定义式，并且将随机扰动项与随机误差项等同。**一个“源生的”随机扰动项变成了一个“衍生的”随机误差项**。而且在解释它的具体内容时，一般都在“无数非显著因素对被解释变量的影响”之外，加上诸如“变量观测值的观测误差的影响”、“模型关系的设定误差的影响”等。

将“源生的”随机扰动变成“衍生的”随机误差，有许多理由可以为此辩解。关键在于，“源生的”随机扰动项所满足的极限法则是否适用于“衍生的”随机误差项，高斯假设和正态分布假设是否仍然成立。

3. 存在模型关系误差情况下的随机误差项

对于一个计量经济学应用模型，假定真实的数据生成过程是模型(9.3.3)，其中的随机扰动项μ_i服从经典假设。假定模型被错误地设定为

$$Y_i = m'(\boldsymbol{X}_i, \tilde{\beta}) + \nu_i \qquad i = 1, 2, \cdots, n \tag{9.3.6}$$

其中ν_i**为存在模型关系误差情况下的随机误差项**。经简单的数学变换后可得

$$\nu_i = \mu_i + m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta}) \qquad i = 1, 2, \cdots, n \tag{9.3.7}$$

显然这里ν_i的分布与$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$有密切的关系。下面分两种情况讨论。

第一种情况：$\boldsymbol{X}_i$是非随机的。这时关键是如何看待$\tilde{\beta}$，由于$\tilde{\beta}$是在模型错误设定下的参数，因此没有很好的定义。不过对每一个给定$\tilde{\beta}$，$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$是确定性变量的函数之差，因此错误模型中的误差$v_i$是一个正态随机数$\mu_i$与非随机数$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$之和，因此仍然是正态的。

第二种情况：$\boldsymbol{X}_i$是随机的。这种情况下，$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$将必然是一个随机数，而且这个随机数受到了三个因素的影响：模型的正确动力学关系$m(\cdot)$、模型被误设的动力学关系$m'(\cdot)$和随机回归元X_t的分布。注意到：

$$\nu_i - \mu_i = m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta}) \qquad i = 1, 2, \cdots, n$$

因此误差v_i是一个正态随机数的充要条件是$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$是正态的。而在上面提到的三个因素的作用下，即使在大样本下，$m(\boldsymbol{X}_i, \beta) - m'(\boldsymbol{X}_i, \tilde{\beta})$的正态性也不能为任何数学定理所保证。

三、模型关系设定的指导原则

如何才能正确设定计量经济学应用模型的变量之间的关系？必须遵循经济学理论和统计学分析相结合的原则。

一是经济学理论指导原则。经济学理论的发展，无论在宏观经济还是微观

PDG

经济领域，都有丰富的成果。这些成果大多采用变量之间的函数关系加以表达。所以，进行任何一项计量经济学应用模型研究，必须洞悉研究对象所属领域的经济学理论及数理模型。尽管如§9.2中所讨论的，这些理论揭示的是理想经济世界的简洁的规律，它们对于设定现实经济世界中变量之间的关系仍然具有重要的指导意义。例如§9.1例题中的需求理论与需求函数模型，§9.2案例中的消费理论与消费函数模型，以及本节下面案例中的生产理论与生产函数模型，它们对于我们建立实际的计量经济学需求模型、消费模型和生产模型都是有借鉴意义的。

二是统计分析指导原则。对数据进行统计分析，特别是通过变量观测值的散点图及对散点图进行的统计分析，为应用模型的关系设定提供了有效的工具，成为模型设定中不可或缺的一项工作。尽管对数据进行统计分析得到的是单个解释变量与被解释变量之间的关系，有时甚至会产生误导，但是只要正确理解统计分析的适用性和局限性，是可以避免被误导的。

四、模型关系设定的检验

正确的模型关系设定并不是一次完成的，需要经过“设定—检验—再设定—再检验”的过程。检验只能发现原来的设定是否是恰当的，如果发现是不恰当的，并不能告诉你正确的模型关系是什么。于是需要再重新设定和再进行检验。

前面已经介绍了模型关系设定偏误检验的原理和常用的检验方法，包括RESET检验。这里不再重复。

需要提倡的是模型残差的正态性检验。从上述的模型关系误设的后果中可以看到，最主要的是模型随机误差项对经典模型基本假设的违背，特别是正态性假设。所以进行模型残差的正态性检验是重要的。

五、案例——以要素替代性质描述为线索的生产函数模型的发展

下面将以生产函数模型为例，从不同生产函数模型中解释变量与被解释变量之间关系的变化中，认识计量经济学应用模型关系设定的原则，同时也借此学习生产函数模型。

在经济学中，生产理论是最重要的内容之一；同样，在计量经济学中，生产函数模型的研究与发展始终是一个重要且活跃的领域。在我国也是这样，从20世纪20年代末，美国数学家柯布(C·Cobb)和经济学家道格拉斯(P·Dauglas)提出了生产函数这一名词，并用1899—1922年的数据资料，导出了著名的C-D生产函数以来，不断有新的研究成果出现，使生产函数的研究与应用呈现长盛不衰的局面。

生产函数是描述生产过程中投入的生产要素的某种组合同它可能的最大产出量之间的依存关系的数学表达式，即

$$Y = f(A, K, L, \cdots) \tag{9.3.8}$$

其中Y为产出量，A，K，L分别为技术、资本、劳动等投入要素。这里“投入的生产要素”是生产过程中发挥作用、对产出量产生贡献的生产要素；“可能的最大产出量”指这种要素组合应该形成的产出量，而不一定是实际产出量。生产要素对产出量的作用与影响，主要是由一定的技术条件决定的，所以，从本质上讲，生产函数反映了生产过程中投入要素与产出量之间的技术关系。

1. 要素替代弹性

生产函数所描述的是投入要素与产出量之间的技术关系，要素替代弹性是该技术关系的重要表征。将要素替代弹性定义为两种要素的比例的变化率与边际替代率的变化率之比，一般用σ表示，则有

$$\sigma = \frac{\mathrm{d}(K/L)}{K/L} \Bigg/ \frac{\mathrm{d}(MP_L/MP_K)}{MP_L/MP_K} \tag{9.3.9}$$

其中，边际产量MP_K和MP_L是指其他条件不变时，某一种投入要素增加 1 个单位时导致的产出量的增加量，用于描述投入要素对产出量的影响程度。边际产量可以表示为

$$MP_K = \frac{\partial f}{\partial K}$$

$$MP_L = \frac{\partial f}{\partial \mathrm{L}}$$

$$\cdots$$

一般情况下，要素替代弹性σ为一个正数。如果用K替代L，则(9.3.9)式分子大于 0；由于L减少，其边际产量MP_L增大，而由于K增加，其边际产量MP_K减小，于是(9.3.9)式分母也大于 0。所以替代弹性σ大于 0，表明要素之间具有有限可替代性。在特殊情况下，要素之间不可以替代，此时K/L不变，则(9.3.9)式分子等于 0，所以替代弹性σ等于 0。另一种极端情况是，无论要素的数量增加或者减少，其边际产量不变，此时(9.3.9)式分母等于 0，替代弹性σ为∞，表明要素之间具有无限可替代性。

2. 线性生产函数模型

如果假设资本K与劳动L之间是无限可替代的，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = \alpha_0 + \alpha_1 K + \alpha_2 L \tag{9.3.10}$$

对于该模型，要素的边际产量$MP_K = \alpha_1, MP_L = \alpha_2$，边际产量之比$MP_K / MP_L = \alpha_1 / \alpha_2$。于是有

$$d(MP_K / MP_L) = 0$$

代入(9.3.9)式得到$\sigma = \infty$，即**要素替代弹性为**∞。从(9.3.10)式也可以直观地看出，一种要素可以被另一种要素替代直至减少为 0，产出量仍然不变。

3. 投入产出生产函数模型

另一种极端的情况是假设资本K与劳动L之间是完全不可替代的，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = \min\left(\frac{K}{a}, \frac{L}{b}\right) \tag{9.3.11}$$

称为投入产出型生产函数，其中a,b为生产 1 单位的产出量所必须投入的资本、劳动的数量。由于a,b为常数，所以产出量Y所必需的资本投入量$K = aY$，劳动投入量$L = bY$，二者之比$K/L = a/b$为常数，$d(K/L) = 0$。代入(9.3.9)式得到$\sigma = 0$，即**要素替代弹性为 0**，资本K与劳动L之间完全不可替代。

4. C-D 生产函数模型

1928 年美国数学家柯布和经济学家道格拉斯提出的生产函数的数学形式为

$$Y = AK^{\alpha} L^{\beta} \tag{9.3.12}$$

根据要素的产出弹性的定义，很容易推出：

$$E_K = \frac{\partial Y}{\partial K} \cdot \frac{K}{Y} = A\alpha K^{\alpha-1} L^{\beta} \frac{Y}{K} = \alpha$$

$$E_L = \frac{\partial Y}{\partial L} \cdot \frac{L}{Y} = AK^{\alpha} \beta L^{\beta-1} \frac{Y}{L} = \beta$$

即参数α，β分别是资本与劳动的产出弹性。那么由产出弹性的经济意义，应该有

$$0 \leqslant \alpha \leqslant 1, \quad 0 \leqslant \beta \leqslant 1$$

在最初提出的 C-D 生产函数中，假定参数满足$\alpha + \beta = 1$，即生产函数的一阶齐次性，也就是假定研究对象满足规模报酬不变。1937 年，杜兰德提出了 C-D 生产函数的改进型，即取消了$\alpha + \beta = 1$的假定，允许要素的产出弹性之和大于 1 或小于 1，也即承认研究对象可以是规模报酬递增的，也可以是规模报酬递减的，取决于参数的估计结果。模型(9.3.12)中的待估参数A为效率系数，是广义技术进步水平的反映。显然，应该有A>0。由此可见，C-D 生产函数模型的参数具有明确的经济意义，这是它的一个显著特点，是它被广泛应用的一个重要原因。该生产函数可以通过简单的对数变换，化为线性模型进行估计。

现在来介绍模型(9.3.12)对要素替代弹性的假设。根据(9.3.9)式，可以得到

$$\begin{aligned}\sigma &= \frac{\mathrm{d}(K/L)}{K/L} \bigg/ \frac{\mathrm{d}(MP_L/MP_K)}{MP_L/MP_K} \\ &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\ln\left(\frac{MP_L}{MP_K}\right) \\ &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\ln\left(\frac{\beta K}{\alpha L}\right) \\ &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\left(\ln\left(\frac{\beta}{\alpha}\right) + \ln\left(\frac{K}{L}\right)\right) \\ &= 1\end{aligned}$$

这是一个重要的结论，它表明 **C-D 生产函数模型假设要素替代弹性为 1。**

显然，与上述要素之间可无限替代的线性生产函数模型和要素之间完全不可替代的投入产出生产函数模型相比较，C-D 生产函数模型假设要素替代弹性为 1，是更加逼近于生产活动的实际，这是一个很大的进步。正因为此，加之 C-D 生产函数模型的参数具有明确的经济意义，使得它一经提出，就得到广泛的应用。直到今天，它仍然是应用最广泛的一种生产函数模型。

但是，C-D 生产函数模型关于要素替代弹性为 1 的假设仍然具有缺陷。根据这一假设，不管研究对象是什么，样本区间是什么，也不管样本观测值是什么，要素替代弹性都为 1，这是与实际不符的。例如，劳动密集型的农业与资本密集型的现代工业，资本与劳动之间的替代性质是明显不同的；再例如，对于同一个研究对象，如果样本区间不同，即考察的区间不同，要素之间的替代性质也应该是不同的；即使研究对象相同、样本区间相同，对于不同的样本点，由于要素的比例不同，相互之间的替代性质也应该是不同的。所有这些，都需要人们发展新的生产函数模型。

5. 不变替代弹性生产函数模型

在 1961 年，由阿罗(Arrow)、钱纳利(Chenery)、明海斯(Mihas)和索洛(Solow)四位学者提出了两要素不变替代弹性(constant elasticity of substitution)生产函数模型，简称 CES 生产函数模型，其基本形式如下：

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}} \tag{9.3.13}$$

其中，待估参数 A 表示广义技术进步水平，应该有 $A>0$；δ_1 和 δ_2 为分配系数，$0<\delta_1<1$，$0<\delta_2<1$，并且满足 $\delta_1+\delta_2=1$；ρ 为替代参数，下面将专门讨论。(9.3.13)式假定研究对象具有不变规模报酬，因为

$$A(\delta_1(\lambda K)^{-\rho} + \delta_2(\lambda L)^{-\rho})^{-\frac{1}{\rho}} = \lambda(A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}})$$

即当资本与劳动的数量同时增长 λ 倍时，产出量也增长 λ 倍。后来，在应用中取消了这一假定，将(9.3.13)式改写为

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{m}{\rho}} \tag{9.3.14}$$

即承认研究对象可以是规模报酬递增的，也可以是规模报酬递减的，取决于参数 m 的估计结果。于是参数 m 为规模报酬参数，当 $m = 1(<1, >1)$ 时，表明研究对象是规模报酬不变(递减、递增)的。公式(9.3.14)为实际应用的 CES 生产函数模型的理论形式。

现在来看看模型(9.3.14)对要素替代弹性的假设。根据(9.3.9)式，要素替代弹性为

$$\begin{aligned}\sigma &= \frac{\mathrm{d}(K/L)}{K/L} \Big/ \frac{\mathrm{d}(MP_L/MP_K)}{MP_L/MP_K} \\ &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\ln\left(\frac{MP_L}{MP_K}\right)\end{aligned}$$

因为

$$\begin{aligned}MP_K &= \frac{\partial Y}{\partial K} \\ &= A\left(-\frac{1}{\rho}\right)\left(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}\right)^{-\frac{1}{\rho}-1} \cdot \delta_1(-\rho)K^{-\rho-1} \\ &= AK^{-1-\rho}(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}-1}\delta_1 \\ MP_L &= AL^{-1-\rho}(\delta_1 K^{-\rho} + \delta_2 L^{-\rho})^{-\frac{1}{\rho}-1}\delta_2 \\ \frac{MP_L}{MP_K} &= \frac{\delta_2}{\delta_1}\left(\frac{K}{L}\right)^{1+\rho}\end{aligned}$$

所以

$$\begin{aligned}\sigma &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\ln\left(\frac{\delta_2}{\delta_1}\left(\frac{K}{L}\right)^{1+\rho}\right) \\ &= \mathrm{d}\ln\left(\frac{K}{L}\right) \Big/ \mathrm{d}\left(\ln\left(\frac{\delta_2}{\delta_1}\right) + (1+\rho)\ln\left(\frac{K}{L}\right)\right) \\ &= \frac{1}{1+\rho}\end{aligned} \tag{9.3.15}$$

由于要素替代弹性 σ 为一个正数，所以参数 ρ 的数值范围为 $-1 < \rho < +\infty$ 。

由(9.3.15)式可以看出，一旦研究对象确定、样本观测值给定，可以得到参数 ρ 的估计值，并计算得到要素替代弹性的估计值。**对于不同的研究对象，或者同一研究对象的不同的样本区间，由于样本观测值不同，要素替代弹性是不同的。这使得 CES 生产函数比 C-D 生产函数更接近现实。**但是，在 CES 生产函数中，仍然假定要素替代弹性与样本点无关，这就是不变替代弹性生产函数

PDG

模型的“不变”的含义。而这一点，仍然是与实际不符的。对于不同的样本点，由于要素的比例不同，相互之间的替代性质也应该是不同的。所以，不变替代弹性生产函数模型还需要发展。

将 CES 生产函数模型两边取对数，将其中的 $\ln(\delta_1 K^{-\rho}+\delta_2 L^{-\rho})$ 在 $\rho=0$ 处展开泰勒级数，取 0 阶、1 阶和 2 阶项，得到

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

为一个简单线性模型，采用单方程模型的估计方法，利用参数对应关系和 $\delta_1+\delta_2=1$，可以计算得到关于参数 $A,\rho,m,\delta_1,\delta_2$ 的估计值。

在不变替代弹性生产函数模型中，如果参数 ρ 的估计值等于 0，则要素替代弹性 σ 的估计值为 1，此时 CES 生产函数退化为 C-D 生产函数。

6. 变替代弹性生产函数模型

变替代弹性(Variable Elasticity of Substitution)生产函数模型，简称 VES 生产函数模型，有许多理论和方法方面的研究成果。较著名的是瑞宛卡(Revankar)于 1971 年提出的模型。模型假定**要素替代弹性 σ 为要素比例的线性函数**，即

$$\sigma=a+b\cdot\frac{K}{L}$$

容易理解，要素比例不同，要素之间的替代性能是不同的。当 K/L 较大时，资本替代劳动就比较困难；当 K/L 较小时，资本替代劳动就比较容易。生产函数的一般形式为

$$Z=A\exp\int\frac{\mathrm{d}k}{k+c\left(\dfrac{k}{a+bk}\right)^{1/a}}\tag{9.3.16}$$

其中 $Z=Y/L,k=K/L$。

当 b=0 时，(9.3.16)式变为

$$\frac{Y}{L}=A\exp\int\frac{\mathrm{d}k}{k+c\left(\dfrac{k}{a}\right)^{1/a}}=A\exp\left(\frac{a}{1-a}\ln\frac{k^{\frac{1-a}{a}}}{1+\dfrac{c}{a^{1/a}}k^{\frac{1-a}{a}}}+\mu\right)$$

令 $\dfrac{1-a}{a}=\rho, A\mathrm{e}^{\mu}=A'$，则有

$$\frac{Y}{L}=A'\left(\frac{a^{1/a}+ck^{\rho}}{a^{1/a}k^{\rho}}\right)^{-\frac{1}{\rho}}=A''(a^{1/a}k^{-\rho}+c)^{-\frac{1}{\rho}}$$

$$Y=A''\left(a^{1/a}\left(\frac{K}{L}\right)^{-\rho}+c\right)^{-\frac{1}{\rho}}\cdot L=A''(a^{1/a}K^{-\rho}+cL^{-\rho})^{-\frac{1}{\rho}}\tag{9.3.17}$$

此时，VES 生产函数模型退化为(9.3.13)式所表示的 CES 生产函数模型。

当 b=0，a=1 时，(9.3.16)式变为

$$\frac{Y}{L}=A\exp\int\frac{\mathrm{d}k}{k(1+c)}=A'\exp\left(\frac{\ln k}{1+c}\right)=A'k^{\frac{1}{1+c}}$$

$$Y=A'K^{\frac{1}{1+c}}\cdot L^{-\frac{1}{1+c}}\cdot L=A'K^{\frac{1}{1+c}}\cdot L^{\frac{c}{1+c}} \tag{9.3.18}$$

此时，**VES 生产函数模型退化为(9.3.12)式所表示的 C-D 生产函数模型**。

当 a=1 时，$\sigma=1+bk$，(9.3.16)式可写成

$$Y=AK^{\frac{1}{1+c}}\left(L+\left(\frac{b}{1+c}\right)K\right)^{\frac{c}{1+c}} \tag{9.3.19}$$

即为一般常用的 VES 生产函数模型，其中 A，b，c 是待估参数。(9.3.19)式为规模报酬不变的情况，如果将规模报酬系数 m 作为一个待估参数，则 VES 生产函数模型的理论形式为

$$Y=AK^{\left(\frac{1}{1+c}\right)m}\left(L+\left(\frac{b}{1+c}\right)K\right)^{\left(\frac{c}{1+c}\right)m} \tag{9.3.20}$$

将(9.3.20)式的计量形态假设为

$$Y=AK^{\left(\frac{1}{1+c}\right)m}\left(L+\left(\frac{b}{1+c}\right)K\right)^{\left(\frac{c}{1+c}\right)m}\cdot\mu$$

其对数形式为

$$\ln Y=\ln A+\frac{m}{1+c}\ln K+\frac{cm}{1+c}\ln\left(L+\frac{b}{1+c}K\right)+\varepsilon \tag{9.3.21}$$

令

$$\ln\left(L+\frac{b}{1+c}K\right)=\ln(L+\lambda\cdot K)=Z(\lambda)$$

在 $\lambda=0$，即 $b=0$ 处展开泰勒级数：

$$Z(\lambda)=\ln L+\frac{K}{L}\cdot\lambda+0(\lambda)$$

代入(9.3.21)式得到

$$\ln Y=\ln A+\frac{m}{1+c}\ln K+\frac{cm}{1+c}\ln L+\frac{cmb}{(1+c)^2}\frac{K}{L}+\varepsilon \tag{9.3.22}$$

对(9.3.22)式进行变量置换，得到

$$Z=\alpha_0+\alpha_1X_1+\alpha_2X_2+\alpha_3X_3+\varepsilon$$

采用单方程模型的估计方法，得到 $\alpha_0,\alpha_1,\alpha_2,\alpha_3$ 的估计值，利用对应关系可以计算得到关于参数 A,c,m,b 的估计值。

7．多要素生产函数模型

如果作为产出量的解释变量的投入要素多于 2 个，可以有不同的处理方法，

PDG

关键在于对要素之间替代性质的认识。下面以三要素(资本K、劳动L和能源E)为例介绍几种多要素生产函数模型。

(1) 多要素线性生产函数模型

如果资本K、劳动L和能源E互相之间都是无限可替代的，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = \alpha_0 + \alpha_1 K + \alpha_2 L + \alpha_3 E$$

(2) 多要素投入产出生产函数模型

假设资本K、劳动L和能源E互相之间都是完全不可替代的，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = \min\left(\frac{K}{a}, \frac{L}{b}, \frac{E}{c}\right)$$

(3) 多要素 C-D 生产函数模型

假设资本K、劳动L和能源E互相之间的替代弹性都为 1，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = AK^{\alpha} L^{\beta} E^{\gamma}$$

(4) 多要素一级 CES 生产函数模型

假设资本K、劳动L和能源E相互之间的替代弹性相同，为同一个待估参数，则产出量Y与投入要素组合之间的关系可以用如下形式的模型描述：

$$Y = A(\delta_1 K^{-\rho} + \delta_2 L^{-\rho} + \delta_3 E^{-\rho})^{-\frac{m}{\rho}} \tag{9.3.23}$$

其中δ_1，δ_2和δ_3为分配系数，$0<\delta_1<1$，$0<\delta_2<1$，$0<\delta_3<1$，并且满足$\delta_1+\delta_2+\delta_3=1$。要素之间的替代弹性为

$$\sigma = \frac{1}{1+\rho}$$

(5) 多要素二级 CES 生产函数模型

假设资本K、劳动L和能源E相互之间的替代弹性不相同，例如资本与能源之间的替代弹性不同于它们与劳动之间的替代弹性，这是比较符合实际的，那么一级 CES 生产函数模型就不能描述要素之间的替代性质。许多人在探索如何既保持 CES 生产函数的性质，又能解决多要素之间不同替代弹性的问题。1967 年加藤(Sato)提出的多要素二级 CES 生产函数模型，是一个比较成功的具有实用价值的成果。以三要素为例，二级 CES 生产函数模型表达如下：

$$\begin{aligned} Y_{KE} &= (a_1 K^{-\rho_1} + a_2 E^{-\rho_1})^{-\frac{1}{\rho_1}} \\ Y &= A(b_1 Y_{KE}^{-\rho} + b_2 L^{-\rho})^{-\frac{m}{\rho}} \end{aligned} \tag{9.3.24}$$

其中Y_{KE}为第一级 CES 生产函数，在第二级 CES 生产函数中，将它作为一个组合要素。

当投入要素多于3个时，还可以根据要素之间的替代性质，构造三级CES生产函数模型，其原理与二级CES生产函数模型相同，不再赘述。

8. 超越对数生产函数模型

一个更具有一般性的变替代弹性生产函数模型是由克里斯蒂森(L.Christensen)、乔根森(D.Jorgenson)和刘(Lau)于1973年提出的超越对数生产函数模型，其形式为

$$\ln Y = \beta_0 + \beta_K \ln K + \beta_L \ln L + \beta_{KK}(\ln K)^2 + \beta_{LL}(\ln L)^2 + \beta_{KL} \ln K \cdot \ln L \qquad (9.3.25)$$

该生产函数模型的显著特点是它的易估计和包容性。它是一个简单线性模型，可以直接采用单方程线性模型的估计方法进行估计。所谓包容性，是它可以被认为是任何形式的生产函数的近似。例如，如果 $\beta_{KK} = \beta_{LL} = \beta_{KL} = 0$，则表现为C-D生产函数；如果 $\beta_{KK} = \beta_{LL} = -\frac{1}{2}\beta_{KL}$，则表现为CES生产函数。所以可以根据该生产函数的估计结果判断要素的替代性质。

9. 重要启示

以上是以要素之间的替代性质为线索发展的一系列生产函数模型。**从中可以看出，解释变量与被解释变量之间的函数关系并不是随意设定的。它既受一定的经济理论假设(例如生产函数中关于要素之间的替代性质的假设)的指导，又要受到实际经济活动中客观表现出的变量之间关系(例如实际生产活动中投入要素之间的替代能力)的检验。**

在引用已有研究成果时不能盲目，因为我们面对的研究对象是一个新的现实经济活动。正确设定模型中变量之间的关系，最重要的仍然是关于经济行为的分析，下面用一个例题加以说明。

例

在一篇以资本、劳动和各种能源为投入要素的生产函数模型研究中，研究者设计的多要素CES生产函数模型为

$$Y_t = A\left[\delta_0 K_t^{-\alpha\rho} L_t^{-(1-\alpha)\rho} + \sum_{i=1}^{k} \delta_i G_{it}^{-\rho}\right]^{-\frac{1}{\rho}}$$

其中，Y 为产出量，K，L 为资本和劳动投入量，G_{i} 为第 i 种能源投入量，其他为参数。

该模型首先将 K 和 L 之间的替代弹性设为1，将二者组合形成组合要素：

$$y_{klt} = K_t^{\alpha} L_t^{1-\alpha}$$

然后，将该组合要素 y_{kl} 与每种能源投入量 G_i 一起，建立多要素一级CES生产函数。模型假设了 y_{kl} 与 k 种能源之间，以及每种能源之间具有相同的替代弹性，为 $1/(1+\rho)$，这显然是错误的。各种能源之间，例如煤炭和石油

具有很强的替代性，而每种能源与 y_{kl} 之间的替代性显然要差得多。

应该采用多级 CES 生产函数。例如第一级包含两个函数：

$$y_{klt} = f(K_t, L_t) \qquad y_{Gt} = g(G_1, G_2, \cdots)$$

第二级为

$$Y_t = A(\delta_1 y_{klt}{}^{-\rho} + \delta_2 y_{Gt}{}^{-\rho})^{-\frac{m}{\rho}}$$

其中组合要素 y_{kl} 可以采用 C-D 生产函数或者 CES 生产函数的形式， y_G 也可以采用 C-D 生产函数或者 CES 生产函数的形式。

§9.4 计量经济学应用模型变量性质设定

前面讨论的计量经济学应用模型总体回归模型设定，目的是确定哪些变量应该作为解释变量引入模型。本节讨论的“变量设定”，是关于变量性质的设定，即它们对被解释变量具有直接影响还是间接影响，它们是内生变量还是外生变量，它们是随机变量还是确定性变量。`这些自然是模型设定的重要内容，是在进行模型估计之前必须明确的。

一、问题的提出

计量经济学模型所描述和分析的是经济主体之间的行为关系，或者称为动力学关系，是通过描述和分析表征主体行为状态的变量之间的关系来实现的。于是，变量和变量之间的关系就构成了计量经济学模型。例如，目前应用最为普遍的经典单方程计量经济学模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i \qquad \mu_i \sim N(0, \sigma^2)$$
$$i = 1, 2 \cdots, n \tag{9.4.1}$$

即是由变量 **X** 和变量 **Y** 之间的关系构成的。在一个计量经济学模型或者模型系统中，为了研究的需要，我们需要在总体回归模型设定时将一些变量(例如变量 **X**)设定为内生或者外生的、随机或者确定的，然后依此选择模型估计方法和进行模型检验。

但是，内生或者外生、随机或者确定，并不是变量本身所固有的绝对的属性，而是相对于模型的研究对象、模型系统，甚至模型中的参数而言的。经济变量，如果一定要给出它们的固有属性，只能说它们都是内生的和随机的，因为它们都是在社会经济系统中，在互相影响和作用的过程中生成的。所以，关于变量内生或者外生、随机或者确定的设定，成为计量经济学应用模型研究中一个十分重要又十分困难的问题。同一个经济变量，相对于不同的研究对象，

相对于模型不同的应用目的，甚至相对于不同的结构参数，可能有不同的设定。这就提出了计量经济学模型变量设定内生与外生的相对性、随机与确定的相对性问题。

一个变量对于另外一个变量(例如变量 $\boldsymbol{X}$ 对于变量 $\boldsymbol{Y}$)具有直接影响还是间接影响，主要依据经济行为分析加以判断，从理论上讲，它是客观的、绝对的。计量经济学模型，说到底，就是因果分析模型，引入模型的变量，应该是对研究对象产生直接影响且互相独立的变量。否则，不仅会得到错误的结论，而且也会引发变量的内生性和随机性。但是，实际上，经济系统中变量之间经常是相互影响的，所谓的直接影响和间接影响仍然具有相对性。

例 9.4.1

在一篇经验实证研究我国收入不平等对经济增长的影响的论文中，作者通过严谨的理论分析认为，收入不平等将影响固定资本投资和人力资本投资，而固定资本投资和人力资本投资直接影响经济增长。由于在增长的初期，扩大收入差距可以增加固定资本投资，而在增长达到一定水平后，缩小收入差距有利于增加人力资本投资，因此收入差距与经济增长水平之间呈现倒 U 型关系。为了检验这种倒 U 型关系，作者建立了以我国 GDP 为被解释变量的计量经济学模型，解释变量包括收入差距(SRCJ)和收入差距的二次方，以及固定资本投资(GDZB)、人力资本投资(RLZB)和其他控制变量(X)：

$$\mathrm{GDP}_t = f(\mathrm{GDZB}_t, \mathrm{RLZB}_t, \mathrm{SRCJ}_t, \mathrm{SRCJ}_t^2, X_t) + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.4.2}$$

于是问题出现了：将对 GDP 具有直接影响的固定资本投资、人力资本投资等变量和对 GDP 具有间接影响的收入差距同时引入模型。带来的后果是多方面的。首先，固定资本投资、人力资本投资将不再是外生的、确定性变量；其次，固定资本投资、人力资本投资、收入差距和收入差距的二次方的参数不再是它们各自对 GDP 的影响的客观反映，因为它们之间不具有相互独立性。即使模型显示收入差距的二次方项是显著的，并且参数为负，也不能给出收入差距与经济增长之间是否存在倒 U 型关系任何有意义的检验。

类似的应用研究实例还可以举出许多。例如，在一篇关于我国生产性公共支出的经济增长效应实证分析的论文中，通过理论分析指出，“政府通过征收资本所得税、劳动所得税和消费税为公共支出融资”，而在建立的回归方程中，在经济增长的诸多解释变量中同时出现了反映税收政策的直接税比例和反映政府公共支出的各项支出规模，那么就存在政府各项支出规模是否受到直接税比例

的影响的问题，即它们是否具有内生性，以及直接税比例是否对经济增长具有直接影响的问题。再例如，在一篇分析我国制造业企业进入和退出行为的影响因素的论文中，作者建立了包括进入方程和退出方程的联立方程模型系统，作为模型系统的外生变量包括国有企业比重、产业利润率、资本密度、企业市场规模、产业集中度、产品差异化、研发支出密度等。毫无疑问，这些变量都会对企业进入和退出行为产生影响，但是它们是否都是独立的外生变量，是否都对企业进入和退出行为产生直接影响？这些问题显然是存在的，但论文没有给予足够的讨论。

如果变量性质设定错误，那么建立在变量设定基础上的统计推断无论多么严密，也是非有效的。本节将分别对变量的内生与外生性、随机与确定性，以及直接影响与间接影响之间的相对性进行系统的讨论，提出设定的原则和检验方法。在讨论中，主要针对经典单方程计量经济学模型，即讨论模型(9.4.1)中变量 $\boldsymbol{X}$ 的内生与外生性、随机与确定性，以及它对 Y 的直接影响与间接影响问题。其原则对于 § 9.1 中讨论的各种类型的计量经济学模型都是适用的。

二、变量之间的直接影响与间接影响

1. 直接影响与间接影响

变量的直接影响与间接影响具有相对性，是针对模型的被解释变量而言的。判断的依据是经济行为分析。

在例 9.4.1 中，通过经济行为分析，已经明确固定资本投资和人力资本投资直接影响经济增长。而收入差距对经济增长只是间接影响，所以不应该将它们同时作为经济增长的解释变量。下面是一个看似相同实际上存在本质不同的例子。

例 9.4.2

关于缩小收入差距以促进消费的实证研究。经过居民消费的行为分析，发现居民的平均消费水平(或总消费水平)(JMXF)不仅取决于平均收入水平(或总收入水平)(JMSR)，还与收入差距(SRCJ)有关。为了检验收入差距是否对居民消费有显著影响，以居民的平均消费水平(或总消费水平)为被解释变量，以居民平均收入水平(或总收入水平)和居民收入差距以及其他相关控制变量(X)为解释变量，建立居民消费模型：

$$\mathrm{JMXF}_t = f(\mathrm{JMSR}_t, \mathrm{SRCJ}_t, X_t) + \mu_t \qquad t = 1, 2, \cdots, T \qquad (9.4.3)$$

模型的解释变量中虽然同时包含了居民收入和收入差距，但是它们都对居民消费产生直接影响，而且居民收入差距并不影响居民平均收入水平(或总收入水平)，它们之间是互相独立的。所以这是一个正确的模型。

对于单方程计量经济学模型或者联立方程模型系统中的每一个结构方程，

其解释变量无论是内生或者外生，随机或者确定，都必须是对被解释变量产生直接影响的变量，其相应的结构参数则描述和揭示了它们之间直接的数量关系。这样的模型设定才可能是正确的。

2. 如何判断直接影响或间接影响

如何判断一个变量对另一个变量的影响是直接影响还是间接影响，必须依据经济系统的动力学关系分析，即在经济理论的指导下分析研究对象的实际经济行为，从行为理论上理清变量之间的关系。作为模型的解释变量，应该遵循直接性原则、集合性原则、层次性原则和独立性原则。

所谓**直接性原则**，其含义是显而易见的，即该变量对被解释变量的影响在经济行为机制上是直接的，不需要经过其他任何中间变量。所谓**集合性原则**，即当一个变量由若干成分变量集合而成，尽可能采用集合变量；如果选择其中一个成分变量来代表时，该成分变量应该是最具代表性的。例如，制度作为一种投入要素，和资本、劳动、技术等投入要素一样，将对经济增长产生直接影响。但是，制度变迁水平作为一个综合变量，是由投资的分散决策程度、价格的市场化程度、对外开放程度等多个变量综合而成，应该尽可能将制度变迁水平引入模型。所谓**层次性原则**，即在分层次的变量体系中，应该将第一层次变量引入模型，尤其不能将第二层次变量与相关的第一层次变量同时引入模型。最重要的是**独立性原则**，如果违背了直接性原则、集合性原则或者层次性原则，带来的直接后果是解释变量之间不再具有相互独立性。

可以采用适当的统计检验方法对经过行为分析确定的解释变量进行必要性检验，即因果关系检验(例如在时间序列数据中常用的格兰杰因果关系检验)。但是必须注意，它们只能够检验所分析确认的解释变量与被解释变量之间的因果关系是否满足必要性条件，而不能区分直接影响和间接影响，因为间接影响也是原因。对于通过了因果关系检验的解释变量，再进行统计上的独立性检验，它通常是有效的，至少可以避免直接影响和间接影响变量同时引入模型，或者将不同层次的变量同时引入模型。在目前的应用研究中，解释变量之间的独立性检验已经被广泛采用，但是一般只检验它们之间是否存在线性相关，这是不够的，还需要检验是否存在非线性相关。例如上述的收入不平等与固定资本投资和人力资本投资的关系，就是非线性的。具体的检验方法可以参考本书 § 4.3。

三、变量的内生性与外生性

1. 变量内生性与外生性的相对性

变量的内生性与外生性具有相对性，是相对于模型或者模型系统而言的。

计量经济学应用研究中应用最多的模型类型仍然是单方程计量经济学模型，包括经典和非经典模型。对于单方程模型，因为模型系统中只有一个方程，一般假定模型的解释变量X是外生的，即它只影响模型，而不受模型的影响。但是在实际的应用研究中，X是否可以被设定为外生变量，必须从相对意义上进行判断。因为凡是经济变量，都是状态变量，都是由系统的数据生成过程生成的，因此它们在本质上都是内生的。

同一个经济变量，相对于不同的研究对象，可能有不同的设定。例如，如果我们研究居民个体的商品需求行为，以商品需求量Q为被解释变量，以收入I、商品价格P_1和其他商品价格P_2为解释变量，建立的居民商品需求模型：

$$q_i = f(I_i, p_{1i}, p_{2i}) + \mu_i \qquad i = 1, 2, \cdots, n \tag{9.4.4}$$

在模型(9.4.4)中，商品价格P_1可以被设定为外生变量，因为对于居民个体来讲，他的需求量并不足以影响价格，他只能接受由市场总供给和总需求所决定的价格，并且在该价格下决定他的商品需求量。但是，如果我们研究的是商品总需求量，以商品总需求量Q为被解释变量，以总收入I、商品价格P_1和其他商品价格P_2为解释变量，建立的居民商品需求模型：

$$q_t = f(I_t, p_{1t}, p_{2t}) + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.4.5}$$

在模型(9.4.5)中，商品价格P_1就不能设定为外生，因为它必须受到商品总需求量的影响。

再例如，如果我们研究居民个体的消费行为，以消费额C为被解释变量，以收入I以及其他对个体消费产生直接影响的变量为解释变量，建立的居民个体消费模型：

$$C_i = f(I_i, \cdots) + \mu_i \qquad i = 1, 2, \cdots, n \tag{9.4.6}$$

在模型(9.4.6)中，收入I可以被设定为外生变量，因为对于居民个体来讲，他的收入水平影响消费水平，而消费水平并不反过来影响收入水平。但是，如果我们研究的是社会总消费，以社会总消费额C为被解释变量，以社会总收入I以及其他对总消费产生直接影响的变量为解释变量，建立的社会总消费模型：

$$C_t = f(I_t, \cdots) + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.4.7}$$

在模型(9.4.7)中，收入I就不能设定为外生变量，因为收入不仅影响消费，也受消费的影响。

*2．弱外生性、强外生性和超外生性

外生性设定问题自从计量经济学诞生以来就进入了理论研究者的视野，考尔斯委员会(Cowles Committee)在20世纪50年代初就定义了外生性，即对于模型(9.4.1)式中的X_i和μ_i，若对任何s都有X_i和μ_{i+s}随机独立，则称X_i是外生

的。考尔斯委员会的这个外生性的定义一直被计量经济学界所沿用。

1983 年，恩格尔(Engle)、亨德里(Hendry)和理查德(Richard)发表了专门讨论外生性的论文，计量经济学界对内生性和外生性的认识产生了突破性的改变。恩格尔等将外生性进行分类，分为弱外生性(weak exogeneity)、强外生性(strong exogeneity)和超外生性(super exogeneity)。恩格尔的这种分类方法是相对于研究目的来定义的：**弱外生性是对模型中关注的参数进行估计和检验所必需的；强外生性则为模型预测目的而定义；超外生性则是模型用于政策评价所必需的。**

根据新的外生性定义，甚至在同一个模型系统中，同一个经济变量相对于不同的结构参数，就可能有不同的设定。例如，在一个粮食供求模型系统

$$\begin{cases} p_t = \alpha_0 + \alpha_1 q_t + \mu_t, & \mu_t \sim N(0,\sigma_\mu^2) \\ q_t = \beta_0 + \beta_1 p_{t-1} + \varepsilon_t, & \varepsilon_t \sim N(0,\sigma_\varepsilon^2) \end{cases} \qquad E(\mu_t \varepsilon_t) = 0, \qquad \forall t,s \qquad (9.4.8)$$

中，如果我们关注的参数是价格的需求弹性，即价格与需求量之间的关系，可以用$1/\alpha_1$近似表示，那么需求量q_t是弱外生性变量，具有外生性。但是如果我们关注的参数是描述该供求系统的稳定性的参数，即需要将供给方程代入需求方程，得到

$$p_t = \alpha_0 + \alpha_1 \beta_0 + \alpha_1 \beta_1 p_{t-1} + v_t$$

表示为$p_t = \gamma + \rho p_{t-1} + v_t$。由经济理论知道，参数$\rho$反映系统的稳定性，只有当$|\rho| < 1$时，模型系统收敛于一个均衡价格。显然此时的关注参数不仅包含α_1，而且包含β_1，所以q_t不再是弱外生性变量。

相对于模型的不同的应用目的，对解释变量的外生性的要求是不同的。对于结构分析模型，目的是分析各个解释变量与被解释变量之间的关系，那么只要求解释变量具备弱外生性。对于预测模型，解释变量必须不受滞后被解释变量的影响，才能给定解释变量的未来值，进而根据模型得到被解释变量的未来预测值。这就要求解释变量在弱外生性的基础上具备强外生性。而对于政策分析模型，作为政策变量的解释变量必须受到滞后被解释变量的影响，因为任何政策都是适时制定的。这就要求解释变量在弱外生性的基础上具备超外生性。

3．实际应用模型中的重点

在实际应用模型的研究中，主要需要考虑两种原因引起的解释变量的内生性。一是解释变量影响被解释变量，反过来也受被解释变量的影响。例如上述的商品总需求模型中的价格和社会总消费模型中的收入。二是解释变量本身受到其他解释变量的影响。例如在上述的模型(9.4.2)中，作为解释变量的固定资本投资、人力资本投资本身受到另一个解释变量收入差距的影响，因此它们具有内生性。这两种原因引起的解释变量的内生性的直接后果，造成了解释变量和模型的随机扰动项相关，违背了考尔斯委员会关于外生性的定义，也违背了

模型的高斯-马尔可夫假设。

*4．弱外生性检验

外生性检验是计量经济学理论方法体系中重要的组成部分，也是实际建立计量经济学应用模型时必须进行的一项工作。既然外生性按照研究需要分为三类：弱外生性、强外生性和超外生性，则检验外生性必须分为不同的类型。这里仅介绍弱外生性检验。

下面结合一个双变量的数据生成过程，进一步探讨外生性的定义、分类乃至检验：

$$\begin{aligned} y_t &= \beta x_t + \varepsilon_{1t} \\ x_t &= \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \varepsilon_{2t} \end{aligned} \tag{9.4.9}$$

其中第 1 个方程描述了 y_t 的条件分布，第 2 个方程描述了 x_t 的边缘分布，随机扰动项服从正态分布且不存在时序上的相关，只存在同期相关，即

$$\begin{matrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{matrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}\right) \tag{9.4.10}$$

对于模型(9.4.9)，若 x 对 β 是弱外生的，则要求 $\sigma_{12}=0$。弱外生性的主要优点在于人们可以忽视边缘分布，然而对有效的弱外生性检验来说，要求同时对边缘分布和条件分布建立模型。恩格尔(1984)已经提出了一种对弱外生性进行检验的一般 LM 检验。该一般程序要求检验如下联合假设，即 y_t 不会出现在条件变量的边缘方程中，而且扰动项协方差矩阵的一个适当的子矩阵为 0。在模型(9.4.9)中，已假设 y_t 不在边缘分布方程中出现，且只有一个边缘方程，因此这个适当的子矩阵中只有一个元素。于是原假设就是 $H_0:\sigma_{12}=0$。在这个极为简单的例子中，LM 检验也变得比较简单。它以(9.4.9)式的上下两式所给出的残差为基础。在上述原假设下，(9.4.9)式的上式就是条件方程，独立于下式的边缘方程式。因而，在这个原假设下，每个方程都可以用普通最小二乘法得到有效估计。将(9.4.9)式的上下两式所得到的残差分别记为 e_y 和 e_x。为书写模型简便起见，没有给出截距，但在估计这些方程时，除非有一个很好的先验理由，否则都应该包含一个常数项。

拉格朗日乘数检验统计量的构造如下：把 e_y 对一个常数、x 和 e_x 进行回归，在原假设下，从这个回归所得到的 nR^2 渐近地服从 $\chi^2(1)$ 分布。然后进行显著性或置信区间检验，若 nR^2 超过了预选的临界值则拒绝原假设，也就是认定 x_t 对参数 β 来说不是弱外生的。在这个双变量的情形中，另一种形式的检验是把 e_x 对一个常数、滞后的 x、滞后的 y 和 e_y 进行回归。在有限样本中，R^2 在这两个回归中将不相同，但它们是渐近等价的。

另外还有一种形式的检验，它仅涉及一组残差的计算，就是用 y 对一个常

数、x 和 e_x 的回归来取代前面用来构造拉格朗日乘数检验统计量的第一个回归，然后再检验 e_x 的系数是否显著地异于 0。这个系数的 t 检验渐近等价于基于 nR^2 所作的检验。如果同时估计这两个回归，那么无论是用 e_x 还是 y 作为回归值，都将发现所得的 e_x 系数及所估计的标准都是一样的。类似地根据对称性，还可作一个只需计算 e_y 的回归。

四、变量的随机性和确定性

1. 变量随机性和确定性的相对性

变量的随机性和确定性与变量的内生性和外生性是既有联系又有区别的两个问题。内生解释变量一定具有随机性。例如上述的模型(9.4.2)中，既然固定资本投资和人力资本投资是由收入差距与其他因素决定的，那么在固定资本投资和人力资本投资模型中是作为被解释变量出现的，当然是随机变量。**但是随机变量并不都是内生的。**如果我们将模型(9.4.2)加以修正，去掉对 GDP 只产生间接影响的收入差距，使之成为一个典型的总量生产函数模型：

$$\mathrm{GDP}_t = f(\mathrm{GDZB}_t, \mathrm{RLZB}_t, X_t) + \mu_t \qquad t = 1, 2, \cdots, T \tag{9.4.11}$$

虽然固定资本投资和人力资本投资仍然是由其他模型决定的随机变量，但是相对于我们研究的模型(9.4.11)，则可以认为是外生的，因为它们只影响模型(9.4.11)，而不受模型(9.4.11)的影响。更多的情况出现在以时间序列数据为样本的模型中，滞后被解释变量经常作为模型的解释变量，例如资本的滞后作为资本形成模型的解释变量，消费的滞后作为消费函数模型的解释变量。这些滞后被解释变量显然是随机变量，但是相对于这些模型，它们往往被作为外生变量(也称为前定变量)。

更广义地理解，经济变量都具有随机性。但是在计量经济学模型中，一部分变量需要被设定为确定的，或者说不考虑它们的随机性，这就提出了变量随机性和确定性的相对性问题。可以将模型的解释变量分为具有连续概率分布的经济变量、具有离散概率分布的经济变量和非经济变量。因为模型的被解释变量和随机扰动项具有连续概率分布，所以解释变量中的具有离散概率分布的经济变量和非经济变量可以被设定为确定性变量。而解释变量中的具有连续概率分布的经济变量，如果它们是模型的内生变量，毫无疑问应该被设定为随机性变量；如果它们相对于模型是外生的，在模型估计和推断过程中，可以不考虑它们的随机性。但是，对于解释变量中的滞后被解释变量，虽然相对于模型是外生的，是否可以被当作确定性变量看待，即在模型估计和推断过程中，是否必须考虑它们的随机性，需要专门讨论。

2. 随机性和确定性的设定

解释变量的随机性和确定性的设定，关键在于是否与模型随机扰动项相

关。解释变量是否与模型随机扰动项相关，决定了模型应该采用什么估计方法以及参数估计的性质，决定了模型的统计推断是否有效。由滞后被解释变量作为模型解释变量，是应用研究中出现最多的一类解释变量随机性问题。这里专门就此进行讨论。

对于解释变量中的滞后被解释变量性质的设定，依据就是判断它们是否与模型随机扰动项相关。例如，§9.2 中介绍的按照合理预期理论建立的消费函数模型，首先认为消费 C_t 是由对收入的预期 Y_t^e 所决定的，即

$$C_t = \beta_0 + \beta_1 Y_t^e + \mu_t$$

在预期收入 Y_t^e 与实际收入 Y_t 之间存如下关系：

$$Y_t^e = (1-\lambda)Y_t + \lambda Y_{t-1}^e$$

那么容易推导出合理预期消费函数模型为

$$\begin{aligned} C_t &= \beta_0 + \beta_1(1-\lambda)Y_t + \beta_1\lambda Y_{t-1}^e + \mu_t \\ &= \beta_0 + \beta_1(1-\lambda)Y_t + \lambda(C_{t-1} - \beta_0 - \mu_{t-1}) + \mu_t \\ &= \beta_0(1-\lambda) + \beta_1(1-\lambda)Y_t + \lambda C_{t-1} + \mu_t - \lambda\mu_{t-1} \end{aligned}$$

显然在该模型中，作为解释变量的 C_{t-1} 与模型的随机扰动项 $(\mu_t - \lambda\mu_{t-1})$ 高度相关(因为 C_{t-1} 与 μ_{t-1} 高度相关)。于是在该模型中，绝对不能将滞后被解释变量 C_{t-1} 设定为确定性变量。

但是，在模型形式完全相同的相对收入假设消费函数模型中，却能够得到不同的判断。根据消费的相对收入假设，消费不仅由收入决定，而且受曾经达到过的最高消费水平的影响，即消费具有不可逆性。在时间序列中，一般以前一期的消费 C_{t-1} 与表示曾经达到过的最高消费水平，那么消费函数模型设定为

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_t$$

在该模型中，如果随机扰动项不存在序列相关，那么 C_{t-1} 与 μ_t 不相关，作为解释变量的 C_{t-1} 尽管具有随机性，但是仍然可以采用确定性变量的模型估计方法。如果随机扰动项存在序列相关，即 μ_t 与 μ_{t-1} 相关，那么与 μ_{t-1} 高度相关的 C_{t-1} 自然与 μ_t 相关，就不能将滞后被解释变量 C_{t-1} 设定为确定性变量。

由此可见，当模型解释变量中存在滞后被解释变量时，关于它的性质的设定以及模型估计方法的选择，有效的检验方法是进行模型随机扰动项的序列相关性检验。而关于随机扰动项的序列相关性检验，可以参考本书 §4.2。

本章练习题

1．分析例 9.1.1 中的问题，回答：为什么按照(1)、(2)、(3)的方法建立的农户借贷因素分析模型都是不正确的？

2．分析例 9.1.2 中的问题，回答：如果建立某类商品的单方程需求函数模型，该模型在什么情况下是可以应用的？

3．分析例 9.1.3 中的问题，回答：如果建立我国工业资本配置效率模型的目的是为了进行国际比较，那么应该建立什么类型的模型？如何采集样本数据？

4．比较例 9.1.4 和 § 9.3 中的案例，说明：需求函数模型和生产函数模型在模型设定理论方面的区别是什么？

5．某人以我国人均食品需求量 Q 为被解释变量，以食品价格指数 P 为解释变量，以1978—2007 年的数据为样本，建立了如下的食品需求模型：

$$\ln Q_t = \alpha + \beta \ln P_t + \mu_t \qquad t = 1978,1979,\cdots,2007$$

由于我国的人均食品需求量是逐年上升的，食品价格指数也是逐年上升的，所以估计该模型得到的 $\hat{\beta}$ 为正。于是得到结论：需求法则不适合于我国。试以该问题为例，分别从经济学、逻辑学和统计学三方面理论出发，说明建立计量经济学总体回归模型必须遵循“从一般到简单”的原则。

6．回答：为什么不能直接依据先验的经济学理论进行总体回归模型设定？经济学理论在总体回归模型设定中具有什么作用？

7．回答：为什么不能直接依据数据之间的关系进行总体回归模型设定？数据关系在总体回归模型设定中具有什么作用？

8．以我国城镇居民总消费为被解释变量，建立我国城镇居民消费函数模型，试完成总体回归模型的设定。

9．回答：源生的随机扰动项和衍生的随机误差项之间的区别和联系是什么？模型函数关系误设的主要后果是什么？

10．在例 9.2.1 中，如果将资本投入 ZB 分为固定资本 $ZB1$ 和流动资本 $ZB2$，将劳动投入 LD 分为一般劳动力 $LD1$ 和专业技术人员 $LD2$，其他变量不变。试设定我国总量生产函数的总体回归模型。

11．分析例 9.4.1 中的问题，试设定一个正确的模型。

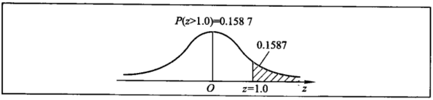
12．说明变量的内生性和随机性之间的区别和联系。

13．结合模型(9.4.8)理解：相对于不同的结构参数，变量的内生性和外生性具有相对性。

PDG

附录　统计分布表

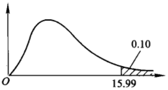
一、标准正态分布表



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500 0	0.496 0	0.492 0	0.488 0	0.484 0	0.480 1	0.476 1	0.472 1	0.468 1	0.464 1
0.1	0.460 2	0.456 2	0.452 2	0.448 3	0.444 3	0.440 4	0.436 4	0.432 5	0.428 6	0.424 7
0.2	0.420 7	0.416 8	0.412 9	0.409 0	0.405 2	0.401 3	0.397 4	0.393 6	0.389 7	0.385 9
0.3	0.382 1	0.378 3	0.374 5	0.370 7	0.366 9	0.363 2	0.359 4	0.355 7	0.352 0	0.348 3
0.4	0.344 6	0.340 9	0.337 2	0.333 6	0.330 0	0.326 4	0.322 8	0.319 2	0.315 6	0.312 1
0.5	0.308 5	0.305 0	0.301 5	0.298 1	0.294 6	0.291 2	0.287 7	0.284 3	0.281 0	0.277 6
0.6	0.274 3	0.270 9	0.267 6	0.264 3	0.261 1	0.257 8	0.254 6	0.251 4	0.248 3	0.245 1
0.7	0.242 0	0.200 9	0.235 8	0.232 7	0.229 6	0.226 6	0.223 6	0.220 6	0.217 7	0.214 8
0.8	0.211 9	0.209 0	0.206 1	0.203 3	0.200 5	0.197 7	0.194 9	0.192 2	0.189 4	0.186 7
0.9	0.181 4	0.181 4	0.178 8	0.176 2	0.173 6	0.171 1	0.168 5	0.166 0	0.163 5	0.161 1
1.0	0.158 7	0.156 2	0.153 9	0.151 5	0.149 2	0.146 9	0.144 6	0.142 3	0.140 1	0.137 9
1.1	0.135 7	0.133 5	0.131 4	0.129 2	0.127 1	0.125 1	0.123 0	0.121 0	0.119 0	0.117 0
1.2	0.115 1	0.113 1	0.111 2	0.109 3	0.107 5	0.105 6	0.103 8	0.102 0	0.100 3	0.098 5
1.3	0.096 8	0.095 1	0.093 4	0.091 8	0.090 1	0.088 5	0.086 9	0.085 3	0.083 8	0.082 3
1.4	0.080 8	0.079 3	0.077 8	0.076 4	0.074 9	0.073 5	0.072 1	0.070 8	0.069 4	0.068 1
1.5	0.066 8	0.065 5	0.064 3	0063 0	0.061 8	0.060 6	0.059 4	0.058 2	0.057 1	0.055 9
1.6	0.054 8	0.053 7	0.052 6	0.051 6	0.050 5	0.049 5	0.048 5	0.047 5	0.046 5	0.045 5
1.7	0.046 6	0.043 6	0.042 7	0.041 8	0.040 9	0.040 1	0.039 2	0.038 4	0.037 5	0.036 7
1.8	0.035 9	0.035 1	0.034 4	0.036 6	0.032 9	0.032 2	0.031 4	0.030 7	0.030 1	0.029 4
1.9	0.028 7	0.028 1	0.027 4	0.026 8	0.026 2	0.025 6	0.025 0	0.024 4	0.023 9	0.023 3
2.0	0.022 8	0.022 2	0.021 7	0.021 2	0.020 7	0.020 2	0.019 7	0.019 2	0.018 8	0.018 3
2.1	0.017 9	0.017 4	0.017 0	0.016 6	0.016 2	0.015 8	0.015 4	0.015 0	0.014 6	0.014 3
2.2	0.013 9	0.013 6	0.013 2	0.012 9	0.012 5	0.012 2	0.011 9	0.011 6	0.011 3	0.011 0
2.3	0.010 7	0.010 4	0.010 2	0.009 9	0.009 6	0.009 4	0.009 1	0.008 9	0.008 7	0.008 4
2.4	0.008 2	0.008 0	0.007 8	0.007 5	0.007 3	0.007 1	0.006 9	0.006 8	0.006 6	0.006 4
2.5	0.006 2	0.006 0	0.005 9	0.005 7	0.005 5	0.005 4	0.005 2	0.005 1	0.004 9	0.004 8
2.6	0.004 7	0.004 5	0.004 4	0.004 3	0.004 1	0.004 0	0.003 9	0.003 8	0.003 7	0.003 6
2.7	0.003 5	0.003 4	0.003 3	0.003 2	0.003 1	0.003 0	0.002 9	0.002 8	0.002 7	0.002 6
2.8	0.002 6	0.002 5	0.002 4	0.002 3	0.002 3	0.002 2	0.002 1	0.002 1	0.002 0	0.001 9
2.9	0.001 9	0.001 8	0.001 8	0.001 7	0.001 6	0.001 6	0.001 5	0.001 5	0.001 4	0.001 4
3.0	0.001 3	0.001 3	0.001 3	0.001 2	0.001 2	0.001 1	0.001 1	0.001 1	0.001 0	0.001 0

二、χ^2 分布表

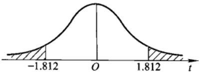
例：对于自由度 v=10，$P(\chi^2>15.99)=0.10$



α v	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.0^3157	0.0^3982	0.0^2393	0.015 8	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88
2	0.020 1	0.050 6	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60
3	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3
8	1.646	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0
9	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6
10	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2
11	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8
12	3.57	4.40	5.23	6.30	8.44	11.34	11.85	18.55	21.0	23.3	26.2	28.3
13	4.11	5.01	5.89	7.04	9.30	12.34	15.98	10.81	22.4	24.7	27.7	29.8
14	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3
15	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8
16	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3
17	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7
18	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2
19	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6
20	8.26	9.59	10.85	12.44	15.455	19.34	23.8	28.4	31.4	34.2	37.6	40.0
21	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

三、t 分 布 表

例：自由度 v=10，$P(t>1.812)=0.05$，$P(t<-1.812)=0.05$



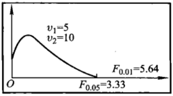
α v	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2,831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.397	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

PDG

四、*F* 分布表

例：自由度 $v_1=5$，$v_2=10$，$P(F>3.33)=0.05$，$P(F>5.64)=0.01$。

注：表中的数字是 1%的显著性水平，上面的为 5%的显著性水平。



	v_1 / v_2	分子自由度											
		1	2	3	4	5	6	7	8	9	10	11	12
分母自由度	1	161 4 052	200 4 999	216 5 403	225 5 625	230 5 764	234 5 859	237 5 928	239 5 982	241 6 022	242 6 056	243 6 082	244 6 106
	2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.34	19.37 99.36	19.38 99.38	19.39 99.40	19.40 99.41	19.41 99.42
	3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05
	4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37
	5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.27	4.78 20.15	4.74 10.05	4.70 9.96	4.68 9.89
	6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.25	4.15 8.10	4.10 7.98	4.06 7.87	4.13 7.79	4.00 7.72
	7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47
	8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67
	9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11
	10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.73	2.91 4.71
	11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40
	12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16
	13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96
	14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.56 3.86	2.53 3.80
	15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80	2.51 3.73	2.48 3.67

续表

	v_1	分子自由度										
	v_2	1	2	3	4	5	6	7	8	9	10	11	12
分母自由度	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
		8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38
		8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
		8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51	3.44	3.37
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31
		8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28
		8.10	5.85	4.94	4.43	4.10	3.87	3.71	3.56	3.45	3.37	3.30	3.23
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25
		8.02	5.78	4.87	4.37	4.04	3.81	3.65	3.51	3.40	3.31	3.24	3.17
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23
		7.94	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20
		7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18
		7.82	55.61	4.72	4.22	3.90	3.67	3.50	3.36	3.25	3.17	3.09	3.03
	25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16
		7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.21	3.13	3.05	2.99
	26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
		7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.17	3.09	3.02	2.96
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13
		7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.14	3.06	2.98	2.93
	28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12
		7.64	5.45	4.57	4.07	3.76	3.53	3.36	3.23	3.11	3.03	2.95	2.90
	29	4.18	3.33	3.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10
		7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.08	3.00	2.92	2.87
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09
		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.06	2.98	2.90	2.84
	32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07
		7.50	5.34	4.46	3.97	3.66	3.42	3.25	3.12	3.01	2.94	2.86	2.80
	34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.50
		7.44	5.29	4.42	3.93	6.61	3.38	3.21	3.08	2.97	2.89	2.82	2.76
	36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03
		7.39	5.25	4.38	3.89	3.58	3.35	3.18	3.04	2.94	2.86	2.78	2.72
	38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02
		7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.91	2.82	2.75	2.69

续表

	v_1 / v_2	分子自由度											
		1	2	3	4	5	6	7	8	9	10	11	12
分母自由度	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00
		7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.88	2.80	2.73	2.66
	42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99
		7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64
	44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98
		7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62
	46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97
		7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60
	48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96
		7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58
	50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95
		7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56
	55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93
		7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53
	60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
	65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90
		7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47
	70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89
		7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45
	80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88
		6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41
	100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85
		6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36
	125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83
		6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33
	150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82
		6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30
	200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80
		2.39	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28
	400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78
		6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23
	1 000	3.85	3.00	1.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76
		6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20
	∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75
		6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18

续表

v_2 \ v_1		分子自由度											
		14	16	20	24	30	40	50	75	100	200	500	∞
分母自由度	1	245 6 142	246 6 169	284 6 208	249 6 234	250 6 258	251 6 286	252 6 302	253 6 323	253 6 334	254 6 352	254 6 361	254 6 366
	2	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
	3	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.53 26.14	8.53 26.12
	4	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
	5	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
	6	3.96 7.60	3.92 7.52	3.87 7.33	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
	7	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
	8	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
	9	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
	10	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
	11	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
	12	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
	13	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.15	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	1.24 3.21	2.22 3.18	2.21 3.16
	14	2.48 3.70	2.44 3.62	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	2.16 3.06	2.14 3.02	2.13 3.00
	15	2.43 3.56	2.39 3.48	2.33 3.36	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.10 2.92	2.08 2.89	2.07 2.87
	16	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.18	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.89	2.07 2.86	2.04 2.80	2.02 2.77	2.01 2.75
	17	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65

续表

	v_2 \ v_1	分子自由度											
		14	16	20	24	30	40	50	75	100	200	500	∞
分母自由度	18	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92
		3.27	3.19	3.07	3.00	2.91	2.83	2.78	2.71	2.68	2.62	2.59	2.57
	19	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88
		3.19	3.12	3.00	2.92	2.84	2.76	2.70	2.63	2.60	2.54	2.51	2.49
	20	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
		3.13	3.05	2.94	2.86	2.77	2.69	2.63	2.56	2.53	2.47	2.44	2.42
	21	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81
		3.07	2.99	2.88	2.80	2.72	2.63	2.58	2.51	2.47	2.42	2.38	2.36
	22	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78
		3.02	2.94	2.83	2.75	2.67	2.58	2.53	2.46	2.42	2.37	2.33	2.31
	23	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
		2.97	2.89	2.78	2.79	2.62	2.53	2.48	2.41	2.37	2.32	2.28	2.26
	24	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73
		2.93	2.85	2.74	2.66	2.58	2.49	2.44	2.36	2.33	2.27	2.23	2.21
	25	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71
		2.89	2.81	2.70	2.62	2.54	2.45	2.40	2.32	2.29	2.23	2.19	2.17
	26	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69
		2.86	2.77	2.66	2.58	2.50	2.41	2.36	2.28	2.25	2.19	2.15	2.13
	27	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
		2.83	2.74	2.63	2.55	2.47	2.38	2.33	2.25	2.21	2.16	2.12	2.10
	28	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
		2.80	2.71	2.60	2.52	2.44	2.35	2.30	2.22	2.18	2.13	2.09	2.06
	29	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
		2.77	2.68	2.57	2.49	2.41	2.32	2.27	2.19	2.15	2.10	2.06	2.03
	30	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
		2.74	2.66	2.55	2.47	2.38	2.29	2.24	2.16	2.13	2.07	2.03	2.01
	32	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
		2.70	2.62	2.51	2.42	2.34	2.25	2.20	2.12	2.08	2.02	1.98	1.96
	34	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
		2.66	2.58	2.47	2.38	2.30	2.21	2.15	2.08	2.04	1.98	1.94	1.91
	36	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
		2.62	3.54	2.43	2.35	2.26	2.17	2.12	2.04	2.00	1.94	1.90	1.87
	38	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
		2.59	2.51	2.40	2.32	2.22	2.14	2.08	2.00	1.97	1.90	1.86	1.84

续表

v_1 / v_2		分子自由度											
		14	16	20	24	30	40	50	75	100	200	500	∞
分母自由度	40	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
		2.56	2.49	2.37	2.29	2.20	2.11	2.05	1.97	1.94	1.88	1.84	1.81
	42	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
		2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
	44	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
		2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75
	46	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
		2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
	48	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
		2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70
	50	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
		2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
	55	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
		2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
	60	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
		2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60
	65	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
		2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56
	70	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
		2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
	80	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
		2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49
	100	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
		2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
	125	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
		2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
	150	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
		2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
	200	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
		2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
	400	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
		2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
	1 000	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
		2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11
	∞	1.67	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
		2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00

五、D.W. 检验上下界表

5%的上下界

n	k=2		k=3		k=4		k=5		k=6	
	d_{L}	d_{U}	d_{L}	d_{U}	d_{L}	d_{U}	d_{L}	d_{U}	d_{L}	d_{U}
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

1%的上下界 续表

n	k=2		k=3		k=4		k=5		k=6	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

注：n 是观察值的数目；k 是解释变量的数目，包括常数项。

六、协整检验临界值表

N	模型形式	*p*	φ_∞	φ	φ
1	无常数项，无趋势项	0.01	–2.565 8	–1.960	–10.04
		0.05	–1.939 3	–3.098	0.00
		0.10	–1.615 6	–0.181	0.00
1	常数项，无趋势项	0.01	–3.433 6	–5.999	–29.25
		0.05	–2.862 1	–2.738	–8.36
		0.10	–2.567 1	–1.438	–4.48
1	常数项，趋势项	0.01	–3.963 8	–8.353	–47.44
		0.05	–3.412 6	–4.039	–17.83
		0.10	–3.127 9	–2.418	–7.58
2	常数项，无趋势项	0.01	–3.900 1	–10.534	–30.03
		0.05	–3.337 7	–5.967	–8.98
		0.10	–3.046 2	–4.069	–6.73
2	常数项，趋势项	0.01	–4.326 6	–15.531	–34.03
		0.05	–3.780 9	–9.421	–15.06
		0.10	–3.495 9	–7.203	–4.01
3	常数项,无趋势项	0.01	–4.298 1	–13.790	–46.37
		0.05	–3.742 9	–8.352	–13.41
		0.10	–3.451 8	–6.241	–2.79
3	常数项，趋势项	0.01	–4.667 6	–18.492	–49.35
		0.05	–4.119 3	–12.024	–13.13
		0.10	–3.834 4	–9.188	–4.85
4	常数项，无趋势项	0.01	–4.649 3	–17.188	–59.20
		0.05	–4.100 0	–10.745	–21.57
		0.10	–3.811 0	–8.317	–5.19
4	常数项，趋势项	0.01	–4.969 5	–22.504	–50.22
		0.05	–4.429 4	–14.501	–19.54
		0.10	–4.147 4	–11.165	–9.88
5	常数项，无趋势项	0.01	–4.958 7	–22.140	–37.29
		0.05	–4.418 5	–13.641	–21.16
		0.10	–4.132 7	–10.638	–5.48
5	常数项，趋势项	0.01	–5.249 7	–26.606	–49.56
		0.05	–4.715 4	–17.432	–16.50
		0.10	–4.434 5	–13.654	–5.77
6	常数项，无趋势项	0.01	–5.240 0	–26.278	–41.65
		0.05	–4.704 8	–17.120	–11.17
		0.10	–4.424 2	–13.347	0.00
6	常数项，趋势项	0.01	–5.512 7	–30.735	–52.50
		0.05	–4.976 7	–20.883	–9.05
		0.10	–4.699 9	–16.445	0.00

注：① 临界值计算公式为 $C_p=\varphi_\infty+\frac{\varphi_1}{T}+\frac{\varphi_2}{T^2}$，其中 *T* 表示样本容量。② *N* 表示协整回归式中所含变量个数，*p* 表示显著性水平。③ *N*=1 时，协整检验即转化为单变量平稳性的 ADF 检验。

参 考 文 献

1. 李子奈．计量经济学．北京：高等教育出版社，2000.
2. 李子奈．计量经济学——方法与应用．北京：清华大学出版社，1992.
3. 李子奈，叶阿忠．高等计量经济学．北京：清华大学出版社，2000.
4. Damodar N. Gujarrati. Basic Econometrics. 4th edition. McGraw-Hill Company, 2001.
5. Jeffrey M. Wooldridge. Introductory Econometrics: A Modern Approach. 2nd edition. Thomson, South-Western, 2003.
6. James H. Stock, Mark W. Watson. Introduction to Econometrics. Pearson Education, Inc., 2003.
7. Jack Johnston, John Dinardo. Econometric Methods. 4th edition. McGraw-Hill Company, Inc.,1997.
8. 古扎拉蒂．计量经济学．3版．林少宫译．北京：中国人民大学出版社，1999.
9. William H.Greene. Econometric Analysis. 5th edition. Prentice-Hall Inc., 2003.
10. 张保法．经济计量学．4版．北京：经济科学出版社，2000.
11. 赵国庆．计量经济学．北京：中国人民大学出版社，2001.
12. Michael D. Intriligator. Econometric Models, Techniques, and Applications. 2nd edition. Prentice-Hall Inc.,1997.
13. R. S. Pindyck, D. L.Rubinfeld. Econometric Models and Econometric Forecasts. 4th edition. McGraw-Hill Company, 1990.
14. G.S.Maddala. Introduction to Econometrics. 3rd edition. John Wiley & Sons, 2001.
15. Robert D. Mason, Douglas A. Lind. Statistical Techniques in Business and Economics. 9th edition. McGraw-Hill Company, 1996.
16. 张寿，于清文．计量经济学．上海：上海交通大学出版社，1984.
17. 唐国兴．计量经济学——理论、方法和模型．上海：复旦大学出版社，1991.
18. 陈正澄．计量经济学．台北：台湾三民书局，1980.
19. 吴承业，龚德恩．应用经济计量学教程．北京：中国铁道出版社，1996.
20. R.L.Thomas. Introductory Econometrics: Theory and Applications. Longman Inc.,1985.
21. 张晓峒．计量经济分析．北京：经济科学出版社，2000.
22. G.C.Chow．经济计量学．郑宗成，等译．北京：中国友谊出版公司，1988.
23. L.Klein．经济计量学教科书．谢嘉译．北京：商务印书馆，1983.
24. G.G.Judge，等．经济计量学理论与实践引论．周逸江，等译．北京：中国统计出版社，1993.

PDG