高等计量经济学

李子奈　　叶阿忠　　编著

清华大学

2003年9月16日

序 言

多年来，我一直为清华大学经济管理学院的研究生开设《高等计量经济学》学位课程，本书就是以讲稿为基础编写的。

《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；在广泛调查研究和充分讨论的基础上，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。1998年7月，教育部高等学校经济学学科教学指导委员会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。随后，我受教育部高教司和教学指导委员会的委托，承担了编写计量经济学本科教材《计量经济学概论》的任务，已交由高等教育出版社出版。

《高等计量经济学》作为一门研究生课程，虽然大家都认为是十分重要的，但在我国高等院校中已经开设的并不多。原因之一是不少研究生在本科阶段并没有学习过初、中级计量经济学，所以许多学校为研究生开设的是中级水平的计量经济学。现在，既然《计量经济学》已经列入经济类专业本科核心课程，那么在研究生阶段开设《高等计量经济学》必将成为未来几年的一个普遍趋势。这是我编写本教材的原因所在。

1992年由清华大学出版社出版了由我编著的《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。《计量经济学—方法与应用》是作为清华大学本科生的教材，所以它自然地成为这本《高等计量经济学—方法与应用》的起点。也就是说，本书所谓的“高等”，是相对于《计量经济学—方法与应用》而言的，凡是那本书中没有涉及或者较少涉及的内容，都被作为“高等”而纳入本书的内容体系之中。

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。以中级计量经济学和相应的数学、经济学、经济统计学课程为先修课程。其中带“*”的章节是否作为教学内容，视学时和教学要求而定，一般可以不作为必须掌握的内容。

本书最大的特点是建立了高等计量经济学的完整、清晰的内容体系。全书共分六章。在第一章中，作为绪论，介绍了计量经济学及其内容体系，力图使读者明了本书涉及的内容在整个计量经济学中的位置；同时，还对属于初、中级计量经济学的经典线性计量经济学模型的理论方法进行了简要的回顾，起到承前启后的作用。接着的第二、三、四章，分别以模

型结构非经典的计量经济学问题、估计方法非经典的计量经济学问题和数据类型非经典的计量经济学问题为题，系统地介绍计量经济学理论方法在模型结构、估计方法和数据类型方面的扩展，尤其是近30年来的新发展。然后在第五和第六章中，对最重要的两个问题，即非线性模型和动态模型，作了较为详细的介绍。如此设计高等计量经济学的内容体系，在目前见到的国内外同类教科书中尚不多见，给读者以内容完整、结构清晰的感觉。教学实践也表明，这是一种成功的设计。

本书另一个显著特点是融理论方法与应用为一体，即方法与应用的结合。计量经济学分为理论计量经济学与应用计量经济学，而已有的为数不多的高级教程（主要是国外的）基本上都属于理论计量经济学，以理论方法的数学描述为主要内容。而本书在数学描述方面适当淡化，以讲清楚方法思路为目标，在方法的提出背景、应用过程中容易出现的问题的处理等方面适当加强，再辅之以简单的应用实例，试图使读者在阅读后能够正确地加以应用。

本书的第三个特点是内容的完整性。在模型结构方面，既介绍了早已发展并广泛应用、但在经典计量经济学中级教科书中较少涉及的非线性模型、变参数模型、增长曲线模型、时间序列分析模型等，也介绍了最近发展的误差修正模型、无参数模型、半参数模型等。在估计方法方面，同样既介绍了早已发展成熟、但在经典计量经济学建模中并不常用、因而在中级教科书中未作为重点的最大似然方法、贝叶斯方法等，也介绍了最近受到重视的局部回归估计、广义矩估计等方法。在数据类型方面，最近20年发展的平行数据模型、离散选择模型、受限被解释变量模型等，在书中都有较系统的介绍。

本书的第四个特点是简繁得当。作为一本教科书，无论是书的篇幅还是教学的学时，都是有限的。面对高等计量经济学如此庞杂的内容，必须有简有繁。本书除了将非线性模型和动态模型作为重点单独列为一章外，在其它的每章中，都有被认为是重要的需要详细介绍的重点。例如，在模型结构非经典的计量经济学问题中，对误差修正模型和无参数模型进行了较详细介绍；在估计方法非经典的计量经济学问题中，将广义矩估计作为重点；在数据类型非经典的计量经济学问题中，平行数据模型占了全章二分之一的篇幅。因为这些理论方法是近20年来计量经济学领域最重要的发展，也是最具应用价值的。

本书作为正式立项的“国家教委‘九五’重点教材”，在编著过程中得到教育部有关部门的大力指导与支持。

应我之邀，叶阿忠副教授参加了编著工作。书中第二章第六节、第三章第四、五、六节、第四章第一、二节和第五章的初稿是由他完成的，我只作了一些文字加工。

在我们的教学中，以TSP和GAUSS作为主要教学软件。关于TSP的应用，已经很普遍，在我编著的由高等教育出版社出版的《计量经济学概论》后面附有简要的使用说明。关于GAUSS的应用尚不普遍，所以将其简要的使用说明作为附录列在本书后面，由鲁传一编写。

在本书编著过程中，参考了《Econometric Analysis》（Third Edition, William H. Greene, Prentice Hall , 1997）、《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）、《动态经济计量学》（D.Hendry、秦朵著，上海人民出版社，1998年）等数十本中、外计量经济学教科书（书目列于参考文献中）和发表于中、外学术刊物上的众多学术论文，以及我曾经指导过的学生们的学位论文和综合练习，有些内容直接摘自这些文献，在此向有关作者表示衷心感谢。

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1999年10月

目录

第一章 绪论

计量经济学作为经济学的一个分支学科，经过70年，尤其是近30年的发展，形成了广泛的内容体系。本章将对计量经济学及其内容体系作概念性介绍，以说明本书所涉及的内容在计量经济学学科中的地位。为了使读者在阅读本书时对前续内容有一定了解，这里将以《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年）为代表，简要总结属于中级计量经济学范畴的经典计量经济学模型的理论与方法，以承前启后。

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。挪威经济学家R.Frish（弗里希）将它定义为经济理论、统计学和数学三者的结合。

英文"Econometrics"最早是由R.Frish于1926年模仿"Biometrics"（生物计量学）提出的，标志着计量经济学的诞生。但人们一般认为，1930年12月29日世界计量经济学会成立和由她创办的学术刊物《Econometrica》于1933年正式出版，标志着计量经济学作为一个独立学科正式诞生了。计量经济学从诞生之日起，就显示了极强的生命力，经过40、50年代的大发展和60年代的大扩张，已经在经济学科中占据极重要的地位。正如著名计量经济学家、诺贝尔经济学奖获得者R.Klein（克莱因）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P.Samuelson（萨谬尔森）甚至说，"第二次大战后的经济学是计量经济学的时代"。

计量经济学于1980年以来，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，本书以此为名，也在于这点。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。

计量经济学模型

模型，是对现实的描述和模拟。对现实的各种不同的描述和模拟方法，就构成了各种不同的模型，例如，语义模型（也称逻辑模型）、物理模型、几何模型、数学模型和计算机模拟模型等。语义模型是用语言来描述现实，例如，对供给不足下的生产活动，我们可以用“产出量是由资本、劳动、技术等投入要素决定的，在一般情况下，随着各种投入要素的增加，产出量也随之增加，但要素的边际产出是递减的”来描述。物理模型是用简化了的实物来描述现实，例如，一栋楼房的模型，一架飞机的模型。几何模型是用图形来描述现实，例如一个零部件的加工图。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。数学模型是用数学语言描述现实，也是一种重要的模型方法，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。

经济数学模型是用数学方法描述经济活动。根据所采用的数学方法不同、对经济活动揭示的程度不同，构成各类不同的经济数学模型。在这里，我们着重区分数理经济模型和计量经济模型。

数理经济模型揭示经济活动中各个因素之间的理论关系，用确定性的数学方程加以描述。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互相影响，为控制经济活动提供理论指导。但是，数理经济模型并没有揭示因素之间的定量关系，因为在上面的公式中，参数α、β、γ是未知的。

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

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

其中μ为随机误差项。这就是计量经济学模型的理论形式。如果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：

$$Q = 0.6479e^{0.0128t}K^{0.3608}L^{0.6756}$$

公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，例如结构分析、生产预测等。这就是计量经济模型得到高度重视和广泛应用的原因所在。

从这个例子中，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。

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：计量经济学是一门经济学科还是应用数学？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个分支学科，即它是一门经济学科。为什么？

首先，从计量经济学的定义看。1933年在《Econometrica》的创刊号社论中，R.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。"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。

其次，考察一下计量经济学在西方国家经济学科中的地位，也是有益的。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1998年，共有43位经济学家获奖，覆盖了经济学的各个分支学科。然而，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9人，居经济学各分支学科之首。1969年第一届获奖者，并不是P.萨缪尔森、J.希克斯这样的经济学大家，而是创立计量经济学的R.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J.丁伯根。绝大多数获奖者，即使其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。R.索罗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以及导出的增长方程；F.莫迪利尼亚由于他在家庭储蓄和金融市场作用方的首创性研究而获1985年奖，他曾是数学教师，担任过计量经济学会会长，在研究中广泛应用了计量经济学实

证分析方法；1993年得主R.福格尔和D.诺斯，属新制度经济学派，研究经济史的，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，也可以应用于其它领域，例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、样本数据的收集，则必须以对经济理论、对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、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它。

四、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。而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”。认为经济学是其它社会科学的基础，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”。这些认识显然过于偏激，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则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、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、讨论人们关心的经济现实问题。许多世界一流大学的经济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”（Toronto大学），“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”（Stanford大学）。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。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。在这方面，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。

五、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、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之外，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。计量经济学则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。

1.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，诸如产业结构、产品结构、消费结构、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它变量以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。经典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。乘数也可以从计量经济学模型很方便的求得。

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，没有定量描述变量之间关系的、包含变量和参数的计量经济学模型，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。

2.经济预测

计量经济学模型作为一类经济数学模型，是从用于经济预测，特别是短期预测而发展起来的。在50年代与60年代，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70年代以来，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，经典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。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，经典计量经济学模型显得无能为力。同时，还应该看到，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之中。本书介绍的许多模型技术，就是在近20年中发展起来的。所以，经济预测仍然是计量经济学模型的一个主要应用。将计量经济学模型与其它经济数学模型相结合，也是一个主要发展方向。

3.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，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，可以很方便的评价各种不同的政策对目标的影响。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，即我们希望达到的目标，通过求解模型，可以得到政策变量值。二是政策模拟。即将各种不同的政策代入模型，计算各自的目标值，然后比较其优劣，决定政策的取舍。三是最优控制方法。将计量经济学模型与最优化方法结合起来，选择使得目标最优的政策或政策组合。

4.经济理论的检验与发展

实践的观点是唯物辩证法的首先的和基本的观点，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的很好的方法。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，如果拟合很好，则这种经济理论得到了检验。这就是检验理论。二是用表现已经发生的经济活动的样本数据去拟合各种模型，拟合最好的模型所表现出来的数量关系，则是经济活动所遵循的经济规律，即理论。这就是发现和发展理论。

1.2 计量经济学的内容体系

这里的"计量经济学"是一个广义的概念，是一个学科的概念。关于计量经济学的内容体系，可以从不同的角度进行分类和说明。

一、从学科发展角度划分

从学科发展角度，可以将计量经济学划分为经典的计量经济学和广义的计量经济学。

1.经典的计量经济学

经典的计量经济学，或者称为狭义的计量经济学，也就是我们通常在中级教科书中所说的计量经济学，以经济理论为导向，以揭示经济现象中的因果关系为目的，以线性随机方程为理论形式，主要应用回归分析方法估计模型。本书将在下一节中对经典的计量经济学模型的理论、方法与应用进行全面、扼要的回顾与总结。

2.广义的计量经济学

广义的计量经济学是利用经济理论、数学以及统计学定量研究经济现象的经济计量方法的统称。它既包括几乎与经典的计量经济学同时发展的投入产出分析方法、时间序列分析方法等，也包括近30年来发展的许多新的计量经济学理论方法。在西方许多以"Econometrics"为名称的书中，往往包含如此广泛的内容。

二、从内容角度划分

从内容角度，可以将计量经济学划分为理论计量经济学和应用计量经济学。

1.理论计量经济学

理论计量经济学是以介绍、研究计量经济学的理论与方法为主要内容，侧重于理论与方法的数学证明与推导，与数理统计联系极为密切。除了介绍计量经济模型的数学理论基础、普遍应用的计量经济模型的参数估计方法与检验方法外，还研究特殊模型的估计方法与检验方法，应用了广泛的数学知识。

理论计量经济学也涉及方法的应用，但它不追求应用的结果，而追求为了适应应用对象而必须进行的理论与方法的发展。近30年来，为了适应70年代两次“石油危机”之后世界经济的剧烈变动，以及由此产生的对经典计量经济学理论方法的挑战，计量经济学在理论方法领域取得了迅速的发展，极大地丰富了理论计量经济学的内容体系。

2.应用计量经济学

应用计量经济学则以建立与应用计量经济学模型为主要内容，强调应用模型的经济学和经济统计学基础，侧重于建立与应用模型过程中实际问题的处理。

三、从程度角度划分

从涉及内容的程度角度，一般在高等学校的课程设置和教科书编写中，将计量经济学分为初级计量经济学、中级计量经济学和高级计量经济学，以及计量经济学专题等几个层次。

1.初级计量经济学

初级计量经济学一般包括计量经济学所必须的基础数理统计知识和矩阵代数知识、经典的线性计量经济学模型理论与方法（以单一方程模型为主）、单方程模型的应用等内容，在描述方法上，很少运用矩阵描述。

2.中级计量经济学

中级计量经济学以经典的线性计量经济学模型理论与方法及其应用为主要内容，包括单一方程模型和联立方程模型。在应用方面，主要讨论计量经济学模型在生产、需求、消费、投资、货币需求和宏观经济系统等传统领域的应用，注重于应用过程中实际问题的处理。在描述方法上，普遍运用矩阵描述。

3.高级计量经济学

高级计量经济学以扩展的线性模型理论与方法、非线性模型理论与方法和动态模型理论与方法，以及它们的应用为主要内容。

四、从模型类型角度划分

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

1.经典线性模型

经典线性模型是以揭示经济现象中的因果关系为目的、在数学上主要应用回归分析方法的线性模型。构成了初级、中级计量经济学的主要内容，在一个相当长的时间里代表了计量经济学的主流。

2.非经典线性模型

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

3.非线性模型

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

4.动态模型

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

5.无参数回归模型

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

五、从估计方法角度划分

模型的参数估计方法处于不断的发展之中，构成了任何一本计量经济学教科书的主要内容。从这个角度划分，可以将常用的参数估计方法分为4大类：最小二乘方法、最大似然方法、贝叶斯估计方法和广义矩方法。

1.最小二乘方法

最小二乘方法是一类依赖样本信息、从最小二乘原理出发的参数估计方法，其概念清楚、方法简单，是经典线性计量经济学模型的最主要的估计方法。例如，在经典线性计量经济学模型满足基本假设时采用的普通最小二乘法，在经典线性计量经济学模型存在异方差性时采用的加权最小二乘法，在经典线性计量经济学模型存在序列相关性时采用的广义最小二乘法，在经典线性计量经济学模型存在随机解释变量时采用的工具变量方

法，估计经典线性联立方程计量经济学模型的二阶段最小二乘法、三阶段最小二乘法，等等。在一些特殊计量经济学模型中采用的特殊估计方法，例如部分回归估计、偏回归估计、交叉估计、局部回归估计方法等，也是从最小二乘原理出发的，也可以归入最小二乘方法。

2.最大似然方法

最大似然方法是一类依赖样本信息、从最大似然原理出发的参数估计方法。由于其坚实的理论基础，使得它在理论计量经济学中占有更重要的位置。在经典线性计量经济学模型中，存在着一个与最小二乘方法对应的最大似然方法系列，例如与普通最小二乘法对应的最大似然法，与二阶段最小二乘法对应的有限信息最大似然法，与三阶段最小二乘法对应的完全信息最大似然法。但是，由于其方法的数学描述较为复杂，在经典计量经济学应用模型研究中并不常用。在非经典线性计量经济学模型中，最大似然方法成为主要的估计方法，正如在本书中将要看到的。

3.贝叶斯估计方法

贝叶斯估计方法在数理统计中具有悠久的历史和重要的地位。在计量经济学模型的参数估计中，它的最主要特点是利用了非样本信息，包括前验信息和后验信息。在一些特殊的计量经济学应用模型中，由于样本数量的不足，使得最小二乘方法和最大似然方法无法应用，这时贝叶斯估计方法是不可替代的。

4.广义矩方法

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

六、从数据类型角度划分

计量经济学模型离不开数据。不同的数据类型支持着不同类型的计量经济学模型。经典计量经济学模型的样本数据主要是连续的、服从某种分布的截面数据和时间序列数据。近20年来计量经济学模型的一个主要发展方向，就是利用其它类型的数据，以满足较为特殊的研究对象的需要。例如平行数据、离散数据、受限数据、持续数据等。

1.截面数据分析

截面数据分析指仅利用截面数据作为计量经济学模型的样本数据，对模型的参数进行估计。在用于因果分析的经典计量经济学模型中，截面数据是应用得最多的一类样本数据。

2.时序数据分析

时序数据分析指仅利用时间序列数据作为计量经济学模型的样本数据，对模型的参数进行估计。在用于因果分析的经典计量经济学模型中，时序数据也是常用一类样本数据。在时间序列分析模型中，毫无疑问，时

序数据是唯一可以应用的样本数据。

3.平行数据分析

平行数据（Panel Data）是若干个截面数据的组合，比较多的情况是截面上的样本点数目多于截面数目，所以也被称为“面板数据”。在计量经济学模型中利用平行数据是近20年来计量经济学研究中最活跃的一个领域，这一方面是由于平行数据为计量经济学模型的理论方法研究提供了一个更为丰富的环境，但更重要的是在实际应用中它能够用于研究仅用截面数据或者时序数据所无法研究的问题。例如，对企业的生产成本问题进行截面分析可以揭示企业规模对成本的影响，对企业的生产成本问题进行时序分析可以揭示技术进步对成本的影响，而要研究企业规模和技术进步对成本的综合影响，就要进行平行数据分析。

4.离散被解释变量数据计量经济学模型

离散被解释变量数据计量经济学模型（Model with Discrete dependent Variable），简称离散数据模型，实际上仅指模型的被解释变量的样本数据是离散的。例如二元选择模型和多元选择模型就属于这一类，前者被解释变量只能取是与非，即0和1两种数据，后者被解释变量可以取诸如0、1、2等多种数据。比较典型的是著名的Probit模型和Logit模型。显然，在计量经济研究的实践中，这类模型是具有重要实际价值的。

5.受限被解释变量数据计量经济学模型

受限被解释变量数据计量经济学模型（Limited Dependent Variable），简称受限数据模型，实际上仅指模型的被解释变量的样本数据是受到某种限制的。例如，以收入作为被解释变量建立的收入模型，由于受到条件的限制，只能得到人均月收入高于500元的样本观测值。如何从这样的样本出发对总体进行研究，就是这类计量经济学模型要解决的问题。

6.持续被解释变量数据计量经济学模型

持续被解释变量数据计量经济学模型(Duration Model)，简称持续数据模型，实际上仅指模型的被解释变量的样本观测值是事件持续的期间长度。例如。建立关于失业者等待时间的计量经济模型，就属于这类情况。由于人们关心的不仅是对于事件持续期间的解释，更重要的是关心在样本观测值的时点上立即结束事件的概率，例如已经等待了3个月时立即就业的概率。所以就不能仅采用经典模型中对被解释变量进行解释的研究方法。

1.3 经典线性计量经济学模型理论方法回顾

为了便于读者阅读本书，对经典线性计量经济学模型的理论方法作一简单的回顾是必要的。

一、经典线性计量经济学模型的形式

经典线性计量经济学模型分为单方程模型和联立方程模型。单方程模型用以研究单一的经济活动中各变量之间的关系，联立方程模型用以研究经济系统中各变量之间的关系。变量之间的关系用线性方程加以描述，

当然，实际经济活动中很少存在直接线性线性关系，但是在相当多的情况下可以通过简单的变换使之化为线性关系，所以经典线性模型具有普遍意义。

单方程模型的一般形式为：

$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i$i=1,2,. . . ,n (1.3.1)

其中y为被解释变量，x为解释变量，β为待估参数，k为解释变量的数目，n为样本容量。人们习惯上把常数项看成为一个虚变量的系数，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）。由(1.3.1)表示的n个随机方程的矩阵表达式为：

$$\mathbf{Y} = \mathbf{X}\mathrm{B} + \mathrm{N} \tag{1.3.1}$$

其中

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}_{n\times(k+1)}$$

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

联立方程模型的一般形式为：

$$\mathrm{B}\mathbf{Y} + \Gamma\mathbf{X} = \mathrm{N} \tag{1.3.2}$$

或

$$(\mathrm{B}\Gamma)\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} = \mathrm{N} \tag{1.3.3}$$

其中

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} \quad \mathrm{N} = \begin{pmatrix} \mathrm{N}_1 \\ \mathrm{N}_2 \\ \vdots \\ \mathrm{N}_g \end{pmatrix}$$

$\mathbf{Y}S)C$g为内生变量的数目，$\mathbf{X}$为先决变量，k为先决变量的数目，N为随机误差项。用n表示样本容量，则

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{pmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots \\ y_{g1} & y_{g2} & \cdots & y_{gn} \end{bmatrix}$$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix}$$

$$\mathrm{N} = \begin{pmatrix} \mathrm{N}_1 \\ \mathrm{N}_2 \\ \vdots \\ \mathrm{N}_g \end{pmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots \\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gn} \end{bmatrix}$$

参数矩阵为：

$$\mathrm{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1g} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g} \\ \vdots \\ \beta_{g1} & \beta_{g2} & \cdots & \beta_{gg} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \vdots \\ \gamma_{k1} & \gamma_{k2} & \cdots & \gamma_{kk} \end{bmatrix}$$

$$(\mathrm{B}\Gamma)(\Phi\Omega"$$

二、经典线性计量经济学模型的关系类型

经典计量经济学模型所揭示的变量之间的关系是因果关系。单方程模型用以揭示单一的经济活动中各变量之间的单向因果关系，联立方程模型用以揭示经济系统中各变量之间的互为因果关系。它们反映了经济变量之间在行为、技术、制度等方面的联系。虽然模型中也出现仅反映变量数据之间关系的统计方程，那是不得已而为之。也正因为此，一般在经典计量经济学模型中不包括时间序列分析模型。

三、经典线性计量经济学模型理论模型的设定

以经典单方程计量经济学模型为背景。

对所要研究的经济现象进行深入的分析，根据研究的目的，选择模型中将包含的因素，根据数据的可得性选择适当的变量来表征这些因素，并

根据经济行为理论和样本数据显示出的变量间的关系，设定描述这些变量之间关系的数学表达式，即理论模型。例如生产函数

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

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。

确定模型所包含的变量

在单方程模型中，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"，例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，例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，这些变量必须具有数据可得性。于是，我们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。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。

首先，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。

其次，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。计量经济学模型是要在样本数据，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是统计指标体系中存在的、有可靠的数据来源的。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，使得每一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、经济统计学和数学三者结合的思想。

在选择变量时，错误是容易发生的。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

农付产品出口额=－107.66＋0.13×社会商品零售总额

＋0.22×农付产品收购额

这里选择了无关的变量，因为社会商品零售总额与农付产品出口额无直接关系，更不是影响农付产品出口额的原因。再如

生产资料进口额=0.73×轻工业投资＋0.21×出口额

0.18×生产消费＋67.60×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。再如

农业总产值=0.78＋0.24×粮食产量＋0.05×农机动力

－0.21×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，往往要要经过多次反复。

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，那么它是不能为人们所接受的。这就要求理论模型的建立要在参数估计、模型检验的全过程中反复修改，以得到一种既能有较好的经济学解释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，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。

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、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¡α¡1, 0¡β¡1, $\alpha+\beta \approx 1$，0¡γ¡1并接近0, A¿0。

四、经典线性计量经济学模型的样本数据

样本数据的收集与整理，是建立计量经济学模型过程中最为费时费力的工作，也是对模型质量影响极大的一项工作。从工作程序上讲，它是在理论模型建立之后进行，但实际上经常是同时进行的，因为能否收集到合适的样本观测值是决定变量取舍的主要因素之一。

几类常用的样本数据

常用的样本数据有三类：时间序列数据、截面数据和虚变量数据。

时间序列数据是一批按照时间先后排列的统计数据。一般由统计部门提供，在建立计量经济学模型时应充分加以利用，以减少收集数据的工作量。在利用时间序列数据作样本时，要注意以下几个问题。一是所选择的样本区间内经济行为的一致性问题。例如，我们建立纺织行业生产模型，选择反映市场需求因素的变量，诸如居民收入、出口额等作为解释变量，而没有选择反映生产能力的变量，诸如资本、劳动等，原因是纺织行业属于供大于求的情况。对于这个模型，利用时间序列数据作样本时，只能选择80年代后期以来的数据，因为纺织行业供大于求的局面只出现在这个阶段，而在80年代中期以前的一个长时期里，我国纺织品是供不应求的，那时制约行业产出量的主要因素是投入要素。二是样本数据在不同样本点之间的可比性问题。经济变量的时间序列数据往往是以价值形态出现的，包含了价格因素，而同一件实物在不同年份的价格是不同的，这就造成样本数据在不同样本点之间不可比。需要对原始数据进行调整，消除其不可比因素，方可作为模型的样本数据。三是样本观测值过于集中的问题。经济变量在时间序列上的变化往往是缓慢的，例如，居民收入每年的变化幅度只有5%左右。如果在一个消费函数模型中，以居民消费作为被解释变量，以居民收入作为解释变量，以它的时间序列数据作为解释变量的样本数据，由于样本数据过于集中，所建立的模型很难反映两个变量之间的长期关系。这也是时间序列数据不适宜于对模型中反映长期变化关系的结构参数的估计的一个主要原因。四是模型随机误差项的序列相关问题。用时间序列数据作样本，容易引起模型随机误差项产生序列相关。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。主要由统计部门提供。用截面数据作为计量经济学模型的样本数据，应注意以下几个问题。一是样本与母体的一致性问题。计量经济学模型的参数估计，从数学上讲，是用从母体中随机抽取的个体样本估计母体的参数，那么要求母体与个体必须是一

致的。例如，估计煤炭企业的生产函数模型，只能用煤炭企业的数据作为样本，不能用煤炭行业的数据。那么，截面数据就很难用于一些总量模型的估计，例如，建立煤炭行业的生产函数模型，就无法得到合适的截面数据。二是模型随机误差项的异方差问题。用截面数据作样本，容易引起模型随机误差项产生异方差。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。也可以取0、1以外的数值，表示该因素的变化程度。例如，在工业生产模型中用虚变量表示气候对工业生产的影响，可以将不同年份气候的影响程度，分别用0、1、-1，甚至0.5、-0.5等。不过，这种方法应慎用，以免违背客观性。

样本数据的质量

样本数据的质量问题大体上可以概括为完整性、准确性、可比性和一致性四个方面。

完整性，即模型中包含的所有变量都必须得到相同容量的样本观测值。这既是模型参数估计的需要，也是经济现象本身应该具有的特征。但是，在实际中，"遗失数据"的现象是经常发生的，尤其在中国，经济体制和核算体系都处于转轨之中。在出现"遗失数据"时，如果样本容量足够大，样本点之间的联系并不紧密的情况下，可以将"遗失数据"所在的样本点整个地去掉；如果样本容量有限，或者样本点之间的联系紧密，去掉某个样本点会影响模型的估计质量，则要采取特定的技术将"遗失数据"补上。

准确性，有两方面含义，一是所得到的数据必须准确反映它所描述的经济因素的状态，即统计数据或调查数据本身是准确的；二是它必须是模型研究中所准确需要的，即满足模型对变量口径的要求。前一个方面是显而易见的，而后一个方面则容易被忽视。例如，在生产函数模型中，作为解释变量的资本、劳动等必须是投入到生产过程中的、对产出量起作用的那部分生产要素，以劳动为例，应该是投入到生产过程中的、对产出量起作用的那部分劳动者。于是，在收集样本数据时，就应该收集生产性职工人数，而不能以全体职工人数作为样本数据，尽管全体职工人数在统计上是很准确的，但其中有相当一部分与生产过程无关，不是模型所需要的。

可比性，也就是通常所说的数据口径问题，在计量经济学模型研究中可以说无处不在。而人们容易得到的经济统计数据，一般可比性较差，其原因在于统计范围口径的变化和价格口径的变化，必须进行处理后才能用于模型参数的估计。计量经济学方法，是从样本数据中寻找经济活动本身客观存在的规律性，如果数据是不可比的，得到的规律性就难以反映实

际。不同的研究者研究同一个经济现象，采用同样的变量和数学形式，选择的样本点也相同，但可能得到相差甚远的模型参数估计结果。为什么？原因在于样本数据的可比性。例如，采用时间序列数据作为生产函数模型的样本数据，产出量用不变价格计算的总产值，在不同年份间是可比的；资本用当年价格计算的固定资产原值，在不同年份间是不可比的。对于统计资料中直接提供的这个用当年价格计算的固定资产原值，有人直接用于模型估计，有人进行处理后再用于模型的估计，结果当然不会相同。

一致性，即母体与样本的一致性。上面在讨论用截面数据作为计量经济学模型的样本数据时已经作了介绍。违反一致性的情况经常会发生，例如，用企业的数据作为行业生产函数模型的样本数据，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31个省份的数据作为全国总量模型的样本数据，等等。

五、经典线性计量经济学模型参数的估计

模型参数的估方法，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（对联立方程模型而言）、估计方法的选择、软件的应用等内容。

两类参数估计方法可以用于经典线性计量经济学模型的估计：最下二乘法和最大似然法。例如，在模型满足基本假设时采用的普通最小二乘法，在模型存在异方差性时采用的加权最小二乘法，在模型存在序列相关性时采用的广义最小二乘法和广义差分法，在模型存在随机解释变量时采用的工具变量方法，估计联立方程模型的二阶段最小二乘法、三阶段最小二乘法，等等。在经典线性计量经济学模型估计方法中，存在着一个与最小二乘方法对应的最大似然方法系列，例如与普通最小二乘法对应的最大似然法，与二阶段最小二乘法对应的有限信息最大似然法，与三阶段最小二乘法对应的完全信息最大似然法。但是，由于其方法的数学描述较为复杂，在实际中并不常用。

六、经典线性计量经济学模型的检验

在模型的参数估计量已经得到后，可以说一个计量经济学模型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，单方程计量经济学模型必须通过四级检验，即经济意义检验、统计学检验、计量经济学检验和预测检验。

经济意义检验

主要检验模型参数估计量在经济意义上的合理性。主要方法是将模型参数的估计量与预先拟定的理论期望值进行比较，包括参数估计量的符号、大小、相互之间的关系，以判断其合理性。

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：

$$\text{煤炭产量} = -108.5427 + 0.00067 \times \text{固定资产原值} + 0.01527 \times \text{职工人数} - 0.00681 \times \text{电力消耗量} + 0.00256 \times \text{木材消耗量}$$

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。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（煤炭产量）=2.69＋1.85Ln（固定资产原值）＋0.51Ln（职工人数）

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，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（人均购买日用品支出额）=-3.69＋1.20Ln（人均收入）

－6.40Ln（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，即是它们各自的需求弹性。该二参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，二参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，不管其它方面的质量多么高，模型也是没有实际价值的。

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。通常最广泛应用的统计检验准则有拟合优度检验、变量和方程的显著性检验等，分别采用R^2, F, t作为检验统计量。有时也称为一级检验。

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。通常最主要的检验准则有随机误差项的序列相关检验和异方差性检验，解释变量的多重共线性检验等。有时也称为二级检验。

模型预测检验

预测检验主要检验模型参数估计量的稳定性以及相对样本容量变化时的灵敏度，确定所建立的模型是否可以用于样本观测值以外的范围，即模型的所谓超样本特性。具体检验方法为：(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，并将该预测值与实际观测值进行比较，并检验二者之间差距的显著性。

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单方程计量经济学模型，可以将它应用于预定的目的。

对于联立方程模型，与单方程计量经济学模型一样，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。凡是在单方程模型中必须进行的各项检验，对于联立方程模型中的结构方程，以及应用2SLS或3SLS方法过程中的简化式方程，都是适用的和需要的。然后对于模型系统，通常进行以下检验。

1.拟合效果检验

对于联立方程模型(6.4.2)，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

$$\hat{\mathrm{B}}\hat{\mathbf{Y}}+\hat{\Gamma}\mathbf{X}=0 \tag{1.3.4}$$

将样本期的先决变量观测值代入(6.4.5)，求解该方程组，即可得到内生变量的估计值$\hat{\mathbf{Y}}$。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(6.4.5)？模型系统虽然是线性系统，但并不排除其中存在非线性方程。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-Dauglass生产函数表示的生产方程。所以，对给定$\mathbf{X}$的值，求解内生变量的估计值$\hat{\mathbf{Y}}$的常用方法是迭代法。

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：

$$RMS_i=\sqrt{\sum_{t=1}^{n}e_{it}^2/n}$$

$$e_{it}=(y_{it}-\hat{y}_{it})/y_{it}$$

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。显然，当RMS_i=0，表示第i个内生变量估计值与观测值完全拟合。一般地，在g个内生变量中，RMS¡5%的变量数目占70%以上，并且每个变量的RMS不大于10%，则认为模型系统总体拟合效果较好。

2.预测性能检验

建立联立方程模型，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

$RE=(y_{i0}-\hat{y}_{i0})/y_{i0}$i=1,2,. . . ,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，$RE<5\%$的变量数目占70%以上，并且每个变量的相对误差不大于10%，则认为模型系统总体预测性能较好。

3.方程间误差传递检验

由于联立方程模型系统中变量之间互为解释变量，那么就存在误差的传递，需要对此进行检验。

一个总体结构清晰的计量经济学模型系统，应该存在一些明显的关键路径，描述主要经济行为主体的经济活动过程，这是由经济系统的特征所决定的。在关键路径上，方程之间存在明显的递推关系。例如，在一个中国宏观经济模型中，生产方程、收入方程、分配方程、投资方程、固定资产形成方程等，就构成一个关键路径。而且存在着递推关系，由固定资产决定总产值，由总产值决定国民收入，由国民收入决定财政收入，由财政收入决定投资，由投资决定固定资产。在关键路径上进行误差传递分析，可以检验总体模型的模拟优度和预测精度。

如果关键路径上的方程数目为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$=\frac{1}{T}\sum\limits_{i=1}^{T}e_i$

均方根误差$=\sqrt{\frac{1}{T}\sum\limits_{i=1}^{T}e_i^2}$

冯诺曼比$=\left(\sum\limits_{i=2}^{T}(e_i-e_{i-1})^2\Big/\sum\limits_{i=1}^{T}e_i^2\right)\frac{T}{T-1}$

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。在联立方程模型系统中，由于经济系统的动态性，决定了有一定数量的滞后内生变量。由于滞后内生变量的存在，使得模型预测误差不仅在方程之间传递，而且在不同的时间截面之间，即样本点之间传递。所以对模型进行滚动预测检验是必要的。

如果样本期为t=1,2,...,n，对于模型(4.9.2)，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$\hat{\mathbf{Y}}_1$；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$\hat{\mathbf{Y}}_2$；如此逐年滚动预测，直至得到t=n时的内生变量的预测值$\hat{\mathbf{Y}}_n$。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$\hat{\mathbf{Y}}_n'$。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。

七、经典计量经济学模型成功的三要素

从上述建立计量经济学模型的步骤中，不难看出，任何一项计量经济学研究、任何一个计量经济学模型赖以成功的要素应该有三个：理论、方法和数据。理论，即经济理论，所研究的经济现象的行为理论，是计量经济学研究的基础。方法，主要包括模型方法和计算方法，是计量经济学研究的工具与手段，是计量经济学不同于其它经济学分支学科的主要特征。数据，反映研究对象的活动水平、相互间联系以及外部环境的数据，或更广义讲是信息，是计量经济学研究的原料。这三方面缺一不可。

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是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。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。

1.4 本章思考题

1.为什么说计量经济学是经济理论、数学和经济统计学的结合？

2.为什么说计量经济学是一门经济学科？它在经济学科体系中的地位是什么？它在经济研究中的作用是什么？

3.建立经典线性计量经济学模型的主要步骤有哪些？

4.计量经济学模型有哪些主要应用领域？各自的原理是什么？

5.与经典线性计量经济学模型相比较，现代计量经济学模型理论方法分别在哪些方面产生了哪些重要的发展？

6.当你学完本书并对现代计量经济学模型理论方法有了较全面的了解后，再认真总结经典线性计量经济学模型理论方法在计量经济学理论上的价值和实际应用价值，以认识它在计量经济学内容体系中的重要性。

第二章 模型结构非经典的计量经济学问题

将§1.3中总结的计量经济学模型作为经典线性计量经济学模型，在本书的各章中将分别介绍它的扩展。本章主要介绍从模型结构形式上扩展的计量经济学模型。如第一章中所述，本书所谓“经典计量经济学模型”的结构的内涵包括：变量的选择依赖于经济理论，模型所揭示的是经济变量之间的因果关系，模型关系是明确的并且是线性的，模型中的参数是不变的，等等。凡是与这些相违背的，都称为模型结构非经典的计量经济学问题。

虽然非线性模型—经典线性计量经济学模型的最重要的扩展模型将单独列为一章，但是为了保持内容体系的完整性，仍然将较为传统的非线性计量经济学模型的理论方法列入本章，第五章将主要介绍现代非线性计量经济学模型理论方法。出于同样的理由，虽然协整理论是动态宏观计量经济模型的基础，在第六章中将专门详细介绍，但仍然将协整理论的基本原理和由此而来的误差修正模型的简单概念列入本章。

2.1 传统的非线性单方程计量经济学模型

经典的计量经济学模型理论与方法是在线性模型的基础上发展、完善起来的，因而线性计量经济学模型领域的理论与方法已经相当成熟。但是，现实经济活动并不都能抽象为线性模型，所以非线性计量经济学模型在计量经济学模型中占据重要的位置，关于它的理论与方法的研究是计量经济学理论与方法研究中的一个广泛的领域。尤其在70年代至80年代初，关于非线性模型理论与方法的研究成为一个热点。非线性模型理论与方法已经形成了一个与线性模型相对应的体系，包括从最小二乘原理出发的一整套方法和从最大或然原理出发的一整套方法，也包括随机误差项违背基本假设的非线性问题的估计方法。在本书第五章将专门介绍非线性方程计量经济学模型，在本节中，只涉及一些最基本的概念，以及最简单的传统的单方程非线性模型的估计方法，为读者进一步学习第五章内容与建立非线性模型打下一个基础。

一、非线性单方程计量经济学模型概述

1.解释变量非线性问题

在线性计量经济学模型中我们曾经提及，现实经济现象中变量之间往往呈现非线性关系，但在许多情况下，又可以通过简单的变换，使之变成线性。解释变量非线性问题就属于这种情况。例如需求函数模型中需求量与价格之间的关系为：

$$\frac{1}{Q} = \alpha + \beta\frac{1}{p} + \mu$$

通过变量置换就可以化为线性模型。经验表明，解释变量非线性问题一般都可以化为线性模型。

2.可以化为线性的包含参数非线性的问题

计量经济学模型，一旦包含参数非线性，一般情况下通过简单的变换难以化为线性问题。但是，由于非线性模型的估计远比线性模型复杂，所以还应该尽可能地将它们化为线性问题。例如著名的Cobb-Dauglass生产函数模型

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

和不变替代弹性(CES)生产函数模型

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

在假设随机误差项的对数形式服从正态分布的情况下，即引入随机误差项后可以写成：

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

和

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

尽管包含参数非线性，仍然可以首先化为线性问题。对前者两边取对数得到：

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

对后者两边取对数得到：

$$\ln Q = \ln A - \tfrac{1}{\rho}\ln(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}) + \ln \mu$$

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

$$\ln Q \approx \ln A + \delta_1 \ln K + \delta_2 \ln L - \frac{1}{2}\rho\delta_1\delta_2(\ln(\frac{K}{L}))^2 + \ln\mu$$

然后进行模型的估计。

不可以化为线性的包含参数非线性的问题

不可以化为线性的包含参数非线性的问题是下面要讨论的真正非线性模型。它的一般表达式为：

$y_i = f(X_i, \mathrm{B}) + \mu_i$i=1,2,. . . ,n (2.1.1)

其中f是非线性函数，$X_i = (x_{1i}, x_{2i}, \cdots, x_{ki})$，$\mathrm{B} = (\beta_1, \beta_2, \cdots, \beta_k)'$，n为样本容量。例如，上述生产函数模型，如果随机误差项直接服从正态分布，在引入随机误差项后模型写成：

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

和

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

就是典型的非线性模型。

对于这类模型，第二章中介绍的模型估计方法不再适用，必须发展新的方法估计模型，主要有非线性最小二乘法和非线性最大或然法。非线性最大或然法在第五章中是主要方法，这里只介绍非线性最小二乘法。

二、非线性普通最小二乘法

1.普通最小二乘原理

模型(2.1.1)中，如果随机误差项服从0均值、同方差的正态分布，且无序列相关，则可以从普通最小二乘原理出发，构造模型的估计方法。

对于只有一个参数的非线性模型，(2.1.1)写成：

$y_i = f(x_i, \beta) + \mu_i$i=1,2,. . . ,n (2.1.2)

如果参数估计值已经得到，则应使得残差平方和最小。即

$$S(\hat{\beta}) = \sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta}))^2 \tag{2.1.1}$$

最小。(6.5.1)取极小值的一阶条件为：

$$\frac{dS}{d\hat{\beta}} = -2\sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta})(\frac{-df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0$$

即

$$\sum_{i=1}^{n}(y_i - f(x_i, \hat{\beta})(\frac{df(x_i, \hat{\beta})}{d\hat{\beta}}) = 0 \tag{2.1.2}$$

现在的问题在于如何求解非线性方程(6.4.2)。

对于多参数非线性模型，用矩阵形式表示(2.1.1)为：

$$Y = f(\mathbf{X}, \mathrm{B}) + \mathrm{N} \tag{2.1.3}$$

其中各个符号的意义与线性模型相同。向量B的普通最小平方估计值$\hat{\mathrm{B}}$应该使得残差平方和

$$S(\hat{\mathrm{B}}) = (Y - f(\mathbf{X}, \hat{\mathrm{B}}))'(Y - f(\mathbf{X}, \hat{\mathrm{B}}))$$

达到最小值。即$\hat{\mathrm{B}}$应该满足下列条件：

$$\frac{\partial}{\partial \hat{\mathrm{B}}}(S(\hat{\mathrm{B}})) = -2\frac{\partial}{\partial \hat{\mathrm{B}}}(f(\mathbf{X}, \hat{\mathrm{B}})') \cdot (Y - f(\mathbf{X}, \hat{\mathrm{B}})) = 0$$

即

$$\frac{\partial}{\partial \hat{\mathrm{B}}}(f(\mathbf{X}, \hat{\mathrm{B}})') \cdot (Y - f(\mathbf{X}, \hat{\mathrm{B}})) = 0 \tag{2.1.4}$$

其中$\frac{\partial}{\partial \hat{\mathrm{B}}}(f(\mathbf{X}, \hat{\mathrm{B}})')$是一个（k×n）阶偏微分矩阵，其第(j,i)个元素为$\frac{\partial}{\partial \beta_j}(f(X_i, \hat{\mathrm{B}})')$。求解(6.4.5)的原理和方法与求解(6.4.2)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(6.4.2)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(6.4.2)，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（2.1.3)出发的。

根据经验给出参数估计值$\hat{\beta}$的初值$\hat{\beta}_{(0)}$，将(6.5.1)中的$f(x_i, \hat{\beta})$在$\hat{\beta}_{(0)}$处展开台劳级数，取一阶近似值。即有：

$$f(x_i, \hat{\beta}\,) \approx f(x_i, \hat{\beta}_{(0)}) + \frac{df(x_i, \hat{\beta}\,)}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}} (\hat{\beta} - \hat{\beta}_{(0)}) \tag{2.1.5}$$

令$z_i(\hat{\beta}\,) = \frac{df(x_i, \hat{\beta}\,)}{d\hat{\beta}}$

于是$z_i(\hat{\beta}_{(0)}) = \frac{df(x_i, \hat{\beta}\,)}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}}$

代入(6.5.1)，得到：

$$S(\hat{\beta}\,) = \sum_{i=1}^{n} (y_i - f(x_i, \hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})(\hat{\beta} - \hat{\beta}_{(0)}))^2$$

$$= \sum_{i=1}^{n} (y_i - f(x_i, \hat{\beta}_{(0)}) + z_i(\hat{\beta}_{(0)})\hat{\beta}_{(0)} - z_i(\hat{\beta}_{(0)})\hat{\beta}\,)^2$$

$$= \sum_{i=1}^{n} (\tilde{y}_i(\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})\hat{\beta}\,)^2 \tag{2.1.6}$$

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(6.4.7)中的$\tilde{y}_i(\hat{\beta}_{(0)})$和$z_i(\hat{\beta}_{(0)})$的确定的观测值。于是，将(6.5.1)取极小值变成对(6.4.7)取极小值。

如果有一个线性模型：

$$\tilde{y}_i(\hat{\beta}_{(0)}) = z_i(\hat{\beta}_{(0)})\beta + \varepsilon_i \tag{2.1.7}$$

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

$$S(\hat{\beta}_{(1)}) = \sum_{i=1}^{n} (\tilde{y}_i(\hat{\beta}_{(0)}) - z_i(\hat{\beta}_{(0)})\hat{\beta}_{(1)})^2 \tag{2.1.8}$$

最小。比较(6.4.7)与(6.4.9)后发现，满足使(6.4.9)达到最小的估计值$\hat{\beta}_{(1)}$同时也是使(6.4.7)达到最小的$\hat{\beta}$。换句话说，线性模型(6.4.8)的普通最小二乘估计值就是模型(2.1.2)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，将它记作为参数估计值$\hat{\beta}$的第一次迭代值$\hat{\beta}_{(1)}$。它是通过对线性模型(6.4.8)进行普通最小二乘估计而得到的，而线性模型(6.4.8)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(6.5.1)中的$f(x_i, \hat{\beta})$ 在$\hat{\beta}_{(1)}$处展开台劳级数，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(2.1.2)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$\hat{\beta}_{(0)}$处展开台劳级数，取一阶近似值；

第二步：计算$z_i = \frac{df(x_i,\hat{\beta})}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}}$ 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$\beta$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－拉夫森(Newton-Raphson)迭代法

牛顿－拉夫森迭代法作为高斯－牛顿迭代法的改进，当给出参数估计值$\hat{\beta}$的初值$\hat{\beta}_{(0)}$，将(6.5.1)式在$\hat{\beta}_{(0)}$处展开台劳级数，取二阶近似值。即

$$S(\hat{\beta}) \approx S(\hat{\beta}_{(0)}) + \frac{dS(\hat{\beta})}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}} (\hat{\beta} - \hat{\beta}_{(0)}) + \frac{1}{2}\frac{d^2S(\hat{\beta})}{d\hat{\beta}^2}\Big|_{\hat{\beta}_{(0)}} (\hat{\beta} - \hat{\beta}_{(0)})^2 \quad (2.1.9)$$

这里与高斯－牛顿迭代法有两点不同：一是直接对$S(\hat{\beta})$展开台劳级数，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(6.4.10)达到极小的条件是：

$$\frac{dS(\hat{\beta})}{d\hat{\beta}} = 0$$

注意，这里的$S(\hat{\beta})$已经用(6.4.10)的近似式代入，而不是(6.5.1)。再对$\frac{dS(\hat{\beta})}{d\hat{\beta}}$取一阶近似，则有：

$$\frac{dS(\hat{\beta})}{d\hat{\beta}} \approx \frac{dS(\hat{\beta})}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}} + \frac{d^2S(\hat{\beta})}{d\hat{\beta}^2}\Big|_{\hat{\beta}_{(0)}} \cdot (\hat{\beta} - \hat{\beta}_{(0)}) = 0$$

于是得到

$$\hat{\beta} = \hat{\beta}_{(0)} - \left(\frac{d^2S(\hat{\beta})}{d\hat{\beta}^2}\Big|_{\hat{\beta}_{(0)}}\right)^{-1} \cdot \frac{dS(\hat{\beta})}{d\hat{\beta}}\Big|_{\hat{\beta}_{(0)}} \quad (2.1.10)$$

由(6.4.11)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－拉夫森迭代法，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经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TSP6.5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，例如将Cobb-Dauglass生产函数模型输入为：

Q=c(1)*K c(2)*L c(3)

就可以得到参数A, α, β的估计量c(1)、c(2)和c(3)。

三、讨论：一个例子

下面分别用线性化后的普通最小二乘法和非线性普通最小二乘法进行一个实际模型的估计。模型的目的是分析农民收入的增长是由哪些因素决定的，以及各个因素的贡献，进行研究提高农民收入的措施。

例2.1.1 经过反复模拟，剔除从直观上看可能对农民收入产生影响但实际上并不显著的变量后，得到如下结论：改革开放以来，影响我国农民收入总量水平的主要因素是从事非农产业的农村劳动者人数、农副产品收购价格和农业生产的发展规模。用I表示农民纯收入总量水平、Q表示农业生产的发展规模、P表示农副产品收购价格、L表示从事非农产业的农村劳动者人数。收入采用当年价格；农业生产的发展规模以按可比价格计算的、包括种植业、林业、牧业、副业和渔业的农业总产值指数为样本数据；农副产品收购价格以价格指数为样本数据。所有样本数据列于表2.1.1中。

表2.1.1 农民收入及相关变量数据

年份	I（10亿元）	Q (1978=100)	P (1978=100)	L（100万人）
1978	62.45	100.0	100.0	31.52
1979	79.30	107.5	122.1	31.90
1980	96.50	109.0	130.8	35.02
1981	107.65	115.3	138.5	36.92
1982	120.80	128.4	141.5	38.05
1983	142.40	138.4	147.8	43.40
1984	185.85	155.4	153.7	58.88
1985	238.70	160.7	166.9	67.13
1986	285.52	166.1	177.6	75.22
1987	343.80	175.8	198.9	81.30
1988	442.60	182.6	244.6	86.11
1989	495.30	188.3	281.3	84.98
1990	524.66	202.6	274.0	86.74
1991	559.30	210.1	268.4	89.06
1992	613.66	223.5	277.5	97.65
1993	743.49	241.0	314.7	109.98
1994	979.39	261.7	440.3	119.64
1995	1271.16	290.2	527.9	127.07
1996	1567.33	317.5	550.1	130.28
1997	1721.71	333.7	525.3	135.27

分别以各种形式的模型对样本数据进行拟合，经检验、比较后，选择如下关于农民收入增长因素分析模型的数理形式：

$$I = AQ^{\alpha_1}P^{\alpha_2}L^{\alpha_3}$$

将模型线性化后用广义差分法（为了克服序列相关问题）进行估计，得到

$$
\begin{aligned}
&Ln\hat{I} = -4.7410 + 0.5224LnQ + 0.7819LnP + 0.8511LnL + 0.8254\hat{\mu}_{t-1} - 0.6555\hat{\mu}_{t-2}\\
&\qquad (-8.38) \quad (1.64) \qquad\quad (5.17) \qquad (6.06) \qquad (3.31) \qquad (-2.78)\\
&R^2 = 0.9984 \qquad D.W. = 2.30 \qquad F = 1516.6
\end{aligned}
$$

从模型中不难得到，农民收入的农业总产值弹性为$\alpha_1 = 0.5224$，农民收入的农副产品收购价格弹性为$\alpha_2 = 0.7819$，农民收入的非农产业的农村劳动者人数弹性为$\alpha_3 = 0.8511$。i表示农民收入年均增长率、q表示农业总产值年均增长率、p表示农副产品收购价格年均增长率、l表示非农产业的农村劳动者人数年均增长率，于是应该有：

$$i = \alpha + \alpha_1 q + \alpha_2 p + \alpha_3 l$$

其中α为没有包括在模型中的因素对农民收入增长的贡献。由样本数据计算出1980-1997年间，i=18.47%、q=6.80%、p=8.52%、l=8.27%；进一步计算得到，在农民收入的增长中，农业总产值增长、农副产品收购价格提高、非农产业的农村劳动者人数增加和其它因素的贡献率分别为：

$$\frac{\alpha_1 q}{i} = 19.23\% \qquad \frac{\alpha_2 p}{i} = 36.07\% \qquad \frac{\alpha_3 l}{i} = 38.11\% \qquad \frac{\alpha}{i} = 5.59\%$$

按照经典线性计量经济学模型理论方法的要求，这个例子应该是相当成功了。但是，如果用非线性普通最小二乘法直接估计模型，给定迭代初值0.01、0.5、0.8和0.9，迭代得到的参数估计量为：

$$A = 0.00153 \qquad \alpha_1 = 1.77640 \qquad \alpha_2 = 0.26547 \qquad \alpha_3 = 0.39543$$

改变迭代初值0.5、1、1和1，迭代次数增加了，但得到的参数估计量是相同的。

从例中看到，用两种方法估计的模型参数估计量相差甚远。为什么？留待读者思考与讨论。

2.2 变参数线性计量经济学模型

在经典线性计量经济学模型中，以一元线性模型为例，在模型

$y_t = \alpha + \beta\, x_t + \mu_t$t=1,2,...,n (2.2.1)

中，认为参数α, β在样本期内是常数，即认为产生样本观测值的经济结构保持不变，解释变量对被解释变量的影响保持不变。我们称之为常参数模型。但是，在实际上，真正的常参数模型只存在于假设之中，变参数

的情况是经常发生的。如果将参数α, β作为变量，(2.2.1)就是一个变参数模型。根据参数变化类型不同，变参数模型以及估计方法也不同。下面仅介绍几类较简单然而较常用的变参数模型。

一、确定性变参数模型

将(2.2.1)写成变参数模型形式：

$y_t = \alpha_t + \beta_t x_t + \mu_t$t=1,2,. . . ,n (2.2.2)

如果参数α_t, β_t是变量，但不是随机变量，而是确定性变量，那么称(2.2.2)为确定性变参数模型。经常出现的确定性变参数模型有以下几种类型。

1.参数随某一个变量呈规律性变化

如果有

$$\begin{aligned} \alpha_t &= \alpha_0 + \alpha_1 p_t \\ \beta_t &= \beta_0 + \beta_1 p_t \end{aligned} \tag{2.2.1}$$

其中参数$\alpha_0, \beta_0, \alpha_1, \beta_1$是常数。表示(2.2.2)中的参数随着某一个变量变化。在实际经济问题中，p往往是一个政策变量，表示由于政策的变化改变了解释变量对被解释变量的影响程度。例如，如果(2.2.1)是消费模型，描述消费是如何决定于收入的。从经济学意义上讲，参数β表示边际消费倾向，边际消费倾向与边际储蓄倾向之和等于1，而边际储蓄倾向与当时的利率有关，所以边际消费倾向也随利率而变化，这时p表示利率。再如，如果(2.2.1)是一个生产函数模型，是由Cobb-Dauglass生产函数经过对数化后得到的，那么，从经济学意义上讲，参数β表示某一种投入要素的产出弹性，根据一般规律，投入要素的产出弹性也是一个变数，例如它会随着不同投入要素之间的比例而变化，这时p表示不同投入要素的比例。

将(6.5.1)代入(2.2.2)得到

$$y_t = \alpha_0 + \alpha_1 p_t + \beta_0 x_t + \beta_1 p_t x_t + \mu_t \quad t = 1, 2, \ldots, n \tag{2.2.2}$$

因为p为确定性变量，与随机误差项不相关，可以用OLS方法估计(2.2.4)，得到参数估计量$\hat{\alpha}_0, \hat{\beta}_0, \hat{\alpha}_1, \hat{\beta}_1$。可以通过检验$\hat{\alpha}_1, \hat{\beta}_1$是否为0来检验变量$p$是否对$\alpha, \beta$有影响。

2.参数作间断性变化

如果有

$$\begin{aligned} \alpha_t &= \alpha_0 + \alpha_1 p_t \\ \beta_t &= \beta_0 + \beta_1 p_t \end{aligned} \qquad \text{当} \begin{cases} 1 \le t \le n_0 & p_t = 0 \\ n_0 < t \le n & p_t = 1 \end{cases} \tag{2.2.3}$$

表示(2.2.2)中的参数在n_0处发生了突发性变化。在实际经济问题中，往往表示某项政策的实施所产生的影响。例如，如果(2.2.1)是某种商品的出口模型，描述出口量是如何决定于国内总产量的。但是对于这种商品的出口政策在某一年发生了大的变化，之前是限制出口政策，之后是鼓励出口政策，于是就出现了(2.2.5)描述的情况。

关于这类变参数模型的估计，又分3种不同情况。

(1)n_0已知

如果n_0已知，则可以分段建立模型，分段估计模型。将(2.2.2)改写为：

$y_t = \alpha_0 + \beta_0 x_t + \mu_{1t}$t=1,2,. . . , n_0

$y_t = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1)x_t + \mu_{2t}$t=$n_0$+1,. . . ,n (2.2.6)

分别估计该两个方程，得到参数估计量$\hat{\alpha}_0, \hat{\beta}_0, \hat{\alpha}_1, \hat{\beta}_1$。

也可以建立一个统一的模型：

$y_t = (\alpha_0 + \alpha_1 D_t) + (\beta_0 + \beta_1 D_t)x_t + \mu_t$t=1,2,. . . ,n

即

$y_t = \alpha_0 + \alpha_1 D_t + \beta_0 x_t + \beta_1 D_t x_t + \mu_t$t=1,2,. . . ,n (2.2.7)

其中D为虚变量，其样本观测值为：

$$\begin{cases} 1 \le t \le n_0 & D = 0 \\ n_0 < t \le n & D = 1 \end{cases}$$

直接估计(2.2.7)，得到参数估计量$\hat{\alpha}_0, \hat{\beta}_0, \hat{\alpha}_1, \hat{\beta}_1$。

前一种方法是由G.C. Chow于1960年提出，被称为Chow方法；后一种方法是由Gujarati于1970年提出，被称为Gujarati方法。一些实例表明，两种方法得到的参数估计量具有很好的一致性。

陈正澄用1964-1981年台湾个人收入和储蓄额的数据，对两种方法进行了验证。

表2.2.1 台湾个人收入和储蓄额的数据单位：百万新台币

年份	收入(X)	储蓄(Y)	年份	收入(X)	储蓄(Y)
1964	8.8	0.36	1973	15.5	0.59
1965	9.4	0.21	1974	16.7	0.90
1966	10.0	0.08	1975	17.7	0.95
1967	10.6	0.20	1976	18.6	0.82
1968	11.0	0.10	1977	19.7	1.04
1969	11.9	0.12	1978	21.1	1.53
1970	12.7	0.41	1979	22.8	1.94
1971	13.5	0.50	1980	23.9	1.75
1972	14.3	0.43	1981	25.2	1.99

采用Chow方法，分别以1964-1972年和1973-1981年数据为样本，估计一元线性模型。得到

$$\hat{Y}_1 = -0.2645 + 0.0474X_1 \qquad (1964 - 1972)$$
$$\hat{Y}_2 = -1.75017 + 0.15045X_2 \qquad (1973 - 1981)$$

采用Gujarati方法，以1964–1981年数据为样本，估计一元线性模型

$$Y = \alpha_1 + \alpha_2 X + \alpha_3 D + \alpha_4 (DX) + \mu$$

其中D的样本观测值为：

$$D = \begin{cases} 1 & (1964-1972) \\ 0 & (1973-1981) \end{cases}$$

得到：

$$\hat{Y} = -1.7502 + 0.1505X + 1.4843D - 0.1034(DX)$$

容易计算得到：

$$\begin{array}{ll} \hat{Y}_1 = -0.2659 + 0.0471X_1 & (1964-1972) \\ \hat{Y}_2 = -1.7502 + 0.1505X_2 & (1973-1981) \end{array}$$

与Chow方法的估计结果几乎完全相同。

该类模型可以很容易推广到多阶段和多解释变量的情况.

(2)n_0未知，但$Var(\mu_{1t}) = Var(\mu_{2t})$

这时，一般可以选择不同的n_0，按照(1)的方法进行试估计，然后从多次试估计中选择最优者。选择的标准是使得(2.2.6)中两段方程的残差平方和之和最小。

(3)n_0未知，且$Var(\mu_{1t}) \neq Var(\mu_{2t})$

此时，将n_0看作待估参数。模型采用(2.2.6)的形式，假设

$$\mu_{1t} \sim N(0, \sigma_1^2)$$

$$\mu_{2t} \sim N(0, \sigma_2^2)$$

且不存在自相关。Goldfeld和Quandt于1973年研究并提出用最大或然法进行估计。构造关于n_0的对数或然函数为：

$$\ln L(\beta, \sigma^2 | n_0) = -\tfrac{n}{2}\ln(2\pi) - n_0 \ln\sigma_1 - (n-n_0)\ln\sigma_2$$

$$-\frac{1}{2\sigma_1^2}\sum_{t=1}^{n_0}(y_t - \alpha_0 - \beta_0 x_t)^2 - \frac{1}{2\sigma_2^2}\sum_{t=n_0+1}^{n}(y_t - (\alpha_0+\alpha_1) - (\beta_0+\beta_1)x_t)^2$$

遍取1,2,. . . ,n作为n_0的可能值，代入对数或然函数，选择使得对数或然函数最大的n_0值作为突变点的估计值。

二、随机变参数模型

对于模型(2.2.2)，如果参数α_t,β_t不仅是变量，而且是随机变量，那么称(2.2.2)为随机变参数模型。经常出现的随机变参数模型有以下几种类型。

1.参数在一常数附近随机变化

如果模型参数只在一常数附近随机变化，即

$$\alpha_t = \alpha + \varepsilon_t \quad \beta_t = \beta + \eta_t$$

其中ε_t,η_t为具有0均值的随机项。于是有：

$$y_t = \alpha + \beta x_t + \omega_t \tag{2.2.4}$$

其中

$$\omega_t = \varepsilon_t + \mu_t + \eta_t x_t$$
$$E\omega_t = 0$$
$$E(x_t\omega_t) = E(\varepsilon_t x_t + \mu_t x_t + \eta_t x_t^2) = 0$$
$$var(\omega_t) = E(\varepsilon_t + \mu_t + \eta_t\omega_t)^2 = E(\mu_t^2) + E(\varepsilon_t^2) + E(\eta_t x_t)^2$$

$$= 2\sigma^2 + x_t^2\sigma^2 = (2 + x_t^2)\sigma^2$$

显然，模型(6.4.2)具有异方差性，而且已经推导出随机误差项的方差与解释变量之间的函数关系，所以可以采用经典线性计量经济学模型中介绍的估计方法，例如加权最小二乘法等方法很方便地估计参数。

由Hildreth和Houck于1968年提出了如下的变参数模型：

$y_t = \beta_{0t} + \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \cdots + \beta_{kt}x_{kt} + \mu_t$t=1,2,. . . ,n

$\beta_{jt} = \beta_j + \varepsilon_{jt} \quad Var(\varepsilon_{jt}) = \sigma_j^2$j=0,1,2,. . . ,k

被称为Hildreth-Houck模型。模型(6.4.2)是它的特殊情况，但模型(6.4.2)的分析与估计方法完全适用于Hildreth-Houck模型。

2.参数随某一变量作规律性变化，同时受随机因素影响

在这种情况下，参数可以表示为：

$$\alpha_t = \alpha + \theta\, p_t + \varepsilon_t \quad \beta_t = \beta + \delta\, p_t + \eta_t$$

模型(2.2.2)表示为：

$y_t = \alpha + \theta\, p_t + \beta\, x_t + \delta\, p_t x_t + \mu_t + \varepsilon_t + \eta_t + \eta_t x_t$t=1,2,. . . ,n (2.2.9)

容易导出(2.2.9)是一具有异方差性的多元线性模型，也可以采用经典线性计量经济学模型中介绍的估计方法，例如加权最小二乘法等方法很方便地估计参数。

3.自适应回归模型

如果模型(2.2.2)

$y_t = \alpha_t + \beta_t x_t + \mu_t$t=1,2,. . . ,n

中的参数α_t可以表示为：

$$\begin{aligned}&\alpha_t=\alpha_{t-1}+\varepsilon_{t-1}\\&E(\varepsilon_t)=0\\&Var(\varepsilon_t)=\sigma^2\end{aligned}$$

$$\beta_t=\beta$$

则称该模型为自适应回归模型。它是由影响α_t的变量具有一阶自相关性所引起的。例如

$$\alpha_t=\alpha_0+\delta\, p_t$$

$$p_t=\rho\, p_{t-1}+\varepsilon_t$$

则

$$\alpha_t=\alpha_0+\delta\rho\, p_{t-1}+\delta\varepsilon_t$$

如果$\rho\approx 1$，即具有较高程度自相关，于是有：

$$\alpha_t=\alpha_{t-1}+\delta\varepsilon_t$$

就引出了自适应回归模型。而影响α_t的变量具有一阶自相关性是实际经济活动中常见的现象。例如，如果(2.2.2)是一个消费方程，α_t表示自发性消费（即在收入等于0时的消费水平），国家的消费政策（刺激、鼓励、一般或抑制的政策）使得自发性消费是一个随机变量，而国家的消费政策往往具有一阶自相关性，引起自发性消费也具有一阶自相关性。所以自适应回归模型在实际经济研究中是经常出现的。

将$\alpha_t=\alpha_{t-1}+\varepsilon_{t-1}$代入(2.2.2)得到：

$$y_t=\alpha_{t-1}+\beta\, x_t+\mu_t+\varepsilon_{t-1} \tag{2.2.5}$$

选择$\alpha_{n+1}=\alpha$，则有

$$\begin{aligned}&\alpha_n=\alpha-\varepsilon_n\\&\alpha_{n-1}=\alpha-\varepsilon_n-\varepsilon_{n-1}\\&\cdots\\&\alpha_1=\alpha-\varepsilon_n-\varepsilon_{n-1}-\cdots-\varepsilon_1\end{aligned}$$

于是(6.4.4)写为：

$y_t=\alpha+\beta\, x_t+\mu_t-\varepsilon_n-\varepsilon_{n-1}-\cdots-\varepsilon_{t-1}$t=2,3,...,n

相当于

$$y_t=\alpha+\beta\, x_t+\omega_t$$

$$
\begin{aligned}
&Var(\mu_t)=\sigma_\mu^2\\
&Var(\varepsilon_t)=\sigma_\varepsilon^2\\
&Cov(\omega)=\begin{bmatrix}
1+n\lambda & (n-1)\lambda & \cdots & 3\lambda & 2\lambda & \lambda\\
(n-1)\lambda & 1+(n-1)\lambda & \cdots & 3\lambda & 2\lambda & \lambda\\
\vdots & & & & & \\
2\lambda & 2\lambda & \cdots & 2\lambda & 1+2\lambda & \lambda\\
\lambda & \lambda & \cdots & \lambda & \lambda & 1+\lambda
\end{bmatrix}\sigma_\mu^2
\end{aligned}
$$

其中

$$\lambda=\sigma_\varepsilon^2/\sigma_\mu^2$$

当λ已知时，容易采用广义最小二乘法（GLS）估计模型参数；如果λ未知，可以选择不同的λ值试估计，选择拟合最好者。

自适应回归模型的另一种形式是(2.2.2)中α不变，β具有一阶自相关。即

$$
\begin{aligned}
&y_t=\alpha+\beta_t x_t+\mu_t\\
&\beta_t=\beta_{t-1}+\varepsilon_{t-1}
\end{aligned}
$$

这类模型也是具有代表性的。例如，在消费方程中β_t表示边际消费倾向，在生产方程中β_t表示某种投入要素的产出弹性，而前面提到的影响边际消费倾向的利率、影响投入要素产出弹性的投入要素的比例，都具有一阶自相关性，引起βt的一阶自相关性。

假设

$$
\begin{aligned}
&Var(\mu_t)=(1-\theta)\sigma^2\Sigma_\mu\\
&Var(\varepsilon_t)=\theta\sigma^2\Sigma_\varepsilon
\end{aligned}
$$

其中，$0<\theta<1$，为了简单但又失去一般性，可以假设$\Sigma_\mu,\Sigma_\varepsilon$为单位阵。取

$$\beta_{n+1}=\beta$$

类似于上面模型的推导和估计过程，完成该模型的估计。

变参数计量经济学模型是计量经济学模型研究领域的一个重要方向，在其理论、方法与应用方面都有广泛的内容，以上只介绍其中最简单、然而是最常见的几种模型，作为概论与入门。

2.3 增长曲线模型

所谓经典计量经济学模型，专指那些在数学上采用回归分析的方法，在经济意义上揭示因果关系的经济数学模型。在本节和下一节中，作为"扩展的"计量经济学模型，将简单介绍得到普遍应用的增长曲线模型和随机线性时间序列模型，它们的共同特点是采用回归分析的方法估计模型的参数，揭示经济变量的某种变化规律，但并不反映经济活动中的因果关系。

一、增长曲线模型概述

增长曲线模型描述经济变量随时间变化的规律性，从已经发生的经济活动中寻找这种规律性，并用于未来的经济预测。但是，时间并不是经济活动变化的原因，所以增长曲线模型不属于因果关系模型。

常用的增长曲线模型主要包括以下一些形式。

1.多项式增长曲线模型

多项式增长曲线模型的一般数学形式为：

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_k t^k \tag{2.3.1}$$

式中y_t是第t个时间单位的某个经济指标值，t是时间，$\alpha_0, \alpha_1, \cdots, \alpha_k$为模型参数。如果k=0，增长曲线为一条与时间轴平行的直线；如果k=1，增长曲线为一条截距为α_0、斜率为α_1的直线；如果k=2，增长曲线为一条抛物线。比较常见的是这3种。

(6.5.1)中的参数可以采用回归分析方法进行估计，只需进行简单的变量置换，就可以将(6.5.1)化为线性模型。

2.简单指数型增长曲线模型

简单指数型增长曲线模型表示为：

$$y_t = ab^t \tag{2.3.2}$$

显然，当a¿0、b¿1时，y随着t的增加无限制的增大；当a¿0、0¡b¡1时，y随着t的增加趋向于0。这是两类常见的经济指标。

通过两边取对数后变量置换，将原模型化成线性模型，可以采用回归分析方法估计其参数a和b。

3.修正指数型增长曲线模型

在指数型增长曲线模型中增加一个常数项，即将指数型增长曲线沿y轴平移。其数学表达为：

$$y_t = k + ab^t \tag{2.3.3}$$

其中k是y的逼近值。当a¿0、b¿1时，y随着t的减少直至$-\infty$而逼近于k；当a¿0、0¡b¡1时，y随着t的增加直至$+\infty$而趋向于k。

由于模型中增加一个常数项，使得参数估计相当困难，因为难以通过简单的变换而使模型线性化。估计方法之一是将模型变换为：

$$y_t - k = ab^t$$

给定k的值，通过两边取对数后变量置换，将模型化成线性模型，采用回归分析方法估计参数a和b。反复试算，选择使y的估计值与观测值拟合最好的k以及在该k值下估计得到的a和b的估计值，作为原模型参数k、a和b的估计结果。

4.逻辑(Logistic)增长曲线模型

逻辑增长曲线模型是一种著名的增长曲线模型，在经济预测中有广泛的应用。下面将专门于以介绍。

5.龚珀兹(Gompertz)增长曲线模型

龚珀兹增长曲线模型也是一种著名的增长曲线模型，下面也将专门介绍。

二、逻辑(Logistic)增长曲线模型

逻辑增长曲线模型，俗称"S曲线"，由Verhulst于1845年提出，当时主要目的是模拟人口的增长。其一般形式为：

$$y_t = \frac{K}{1 + e^{\varphi(t)}}$$

其中$\varphi(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_k t^k$。后来经过逐步简化，目前最常见的形式是：

$$y_t = \frac{K}{1 + ae^{-bt}} \tag{2.3.4}$$

也称狭义的逻辑增长曲线模型。

由(6.4.5)表示的增长曲线有两个重要特征。一是y随着t的增加直至$+\infty$而趋向于K，K即是y的饱和值；反过来，当$t \to -\infty$时，$y \to 0$。二是增长曲线具有一个拐点，在拐点之前，y的增长速度越来越快；在拐点之后，y的增长速度越来越慢，逐渐趋近于0。在现实经济生活中，许多指标的增长过程具有这两个特征。例如，一种新产品、新技术的普及率，一种耐用品的存量，它们的增长过程都遵循逻辑增长曲线模型。所以，逻辑增长曲线模型在经济预测中有广泛的应用，是一种重要的预测模型。

在(6.4.5)中，K、a和b是待估参数，一般的线性模型的估计方法已不再适用。人们研究并提出了多种参数估计方法，其中较多应用的如下两种。

1.逻辑增长曲线模型的线性化估计

设法将(6.4.5)式线性化，然后用线性模型的参数估计方法估计其参数。将(6.4.5)写成：

$$\frac{1}{y} = \frac{1}{K} + \frac{a}{K}e^{-bt} \tag{2.3.5}$$

如果给定K，(6.4.6)即可写成：

$$\frac{1}{y}-\frac{1}{K}=\frac{a}{K}e^{-bt}$$

$$\ln(\frac{1}{y}-\frac{1}{K})=\ln(\frac{a}{K})-bt \tag{2.3.6}$$

可以用其它变量表示为：

$$z_t=\alpha+\beta\cdot t$$

显然可以很方便地估计模型的参数。

关键在于如何给定K。首先根据模型的经济背景给出K的上下限。例如一种新产品、新技术的普及率的上限为100%，下限是已经实现的普及率，比如60%。然后分别以上、下限作为K的给定值，估计模型，计算残差平方和。接下来可以根据优选法提供的思路给定新的K值，反复试算，直到得到残差平方和最小时的K值和a、b的估计值。

2.逻辑增长曲线模型的“三和法”估计

所谓“三和法”，是增长曲线模型参数的一种代数估计方法。

当$t=1,2,\cdots,n$时，可以将样本分成3段，分别为$t=1,2,\cdots,r$、$t=r+1,r+2,\cdots,2r$、$t=2r+1,2r+2,\cdots,n$。分别计算每段中(6.4.6)式的和：

$$S_1=\sum_{t=1}^{r}\frac{1}{y_t}=\frac{r}{K}+\frac{a}{K}\frac{e^{-b}(1-e^{-rb})}{1-e^{-b}}$$

$$S_2=\sum_{t=r+1}^{2r}\frac{1}{y_t}=\frac{r}{K}+\frac{a}{K}\frac{e^{-(r+1)b}(1-e^{-rb})}{1-e^{-b}}$$

$$S_3=\sum_{t=2r+1}^{3r}\frac{1}{y_t}=\frac{r}{K}+\frac{a}{K}\frac{e^{-(2r+1)b}(1-e^{-rb})}{1-e^{-b}}$$

设$D_1=S_1-S_2,D_2=S_2-S_3$

得到

$$D_1=\frac{a}{K}\frac{e^{-b}(1-e^{-rb})^2}{1-e^{-b}}\quad D_2=\frac{a}{K}\frac{e^{-(r+1)b}(1-e^{-rb})^2}{1-e^{-b}}$$

于是$D_1/D_2=e^{rb},b=\frac{1}{r}(\ln D_1-\ln D_2)$。当得到y的样本数据后，就可以由此式计算得到b的值。

又因为：

$$D_1 - D_2 = \frac{a}{K}\frac{e^{-b}}{1-e^{-b}}(1-e^{-rb})^3$$

$$D_1^2 = \frac{a^2}{K^2}\frac{e^{-2b}}{(1-e^{-b})^2}(1-e^{-rb})^4$$

$$D_1^2/(D_1 - D_2) = \frac{a}{K} \cdot \frac{e^{-b}}{1-e^{-b}}(1-e^{-rb}) = S_1 - \frac{r}{K}$$

所以有

$$K = \frac{r}{S_1 - D_1^2/(D_1 - D_2)}$$

$$a = K\frac{D_1^2}{D_1 - D_2} \cdot \frac{e^b - 1}{1 - e^{-rb}}$$

由该二式可以计算得到K和a的值。至此，模型的所有参数估计值都已求得。

3.逻辑增长曲线模型的非线性估计

将(6.4.6)式写成：

$$z = \alpha + \beta\gamma^t$$

对该模型直接应用非线性最小二乘法，得到α, β, γ的估计值，然后计算K, a, b的估计值。在一般的软件中很容易实现这一估计过程。

4.一个逻辑增长曲线模型实例

例2.3.1 某城市人均用电量数据如表2.3.1所示。其增长曲线为一逻辑曲线，如图2.3.1。

某城市人均用电量数据表单位：KWH

年份t	人均用电量y	年份t	人均用电量y	年份t	人均用电量y
1(1971)	330	10(1980)	764	19(1989)	1468
2(1972)	359	11(1981)	848	20(1990)	1513
3(1973)	405	12(1982)	930	21(1991)	1548
4(1974)	456	13(1983)	1035	22(1992)	1601
5(1975)	484	14(1984)	1113	23(1993)	1609
6(1976)	522	15(1985)	1150	24(1994)	1641
7(1977)	554	16(1986)	1227	25(1995)	1672
8(1978)	586	17(1987)	1341	26(1996)	1695
9(1979)	652	18(1988)	1376	27(1997)	1710

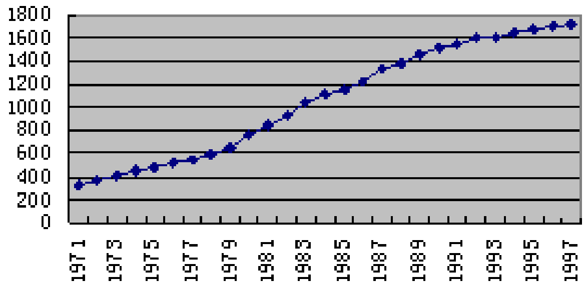


图2.3.1 某城市人均用电量增长曲线

建立逻辑增长曲线模型：

$$y_t = \frac{K}{1 + ae^{-bt}}$$

其中t的观测值取1、2、. . . 、27。利用“三和法”估计模型参数，结果为：

$$\hat{K} = 1995 \qquad \hat{a} = 6.4006 \qquad \hat{b} = 0.1447$$

利用非线性最小二乘法（TSP6.5）估计模型参数，结果为：

$$\hat{K} = 2479 \qquad \hat{a} = 7.4236 \qquad \hat{b} = 0.1054$$

读者可以用两种结果计算人均用电量的模拟值，以对估计方法进行比较。

三、龚珀兹（Gompertz）增长曲线模型

龚珀兹增长曲线由B.Gompertz于1825年提出，其数学形式为：

$$y_t = Ka^{bt} \tag{2.3.7}$$

其中K和a、b为待估参数，K为y的上限逼近值，0为y的下限逼近值。可见龚珀兹增长曲线与逻辑增长曲线相似，只是二者的拐点的位置不同。所以它也有广泛的适用性。

类似于逻辑增长曲线，龚珀兹增长曲线参数估计方法也有许多。可以对(6.4.8)模型直接应用非线性最小二乘法。也可以采用线性估计方法，将(6.4.8)变换为

$$\frac{y_t}{K} = a^{bt}$$

则有

$$\ln(\ln \frac{y_t}{K}) = \ln(\ln a) + t \ln b \tag{2.3.8}$$

如果给定K的值，(6.4.9)就是一个简单线性模型，可以方便地进行估计。关于K值的选择，仍然类似于逻辑增长曲线参数估计过程，不再重复。

同样可以用“三和法”估计龚珀兹增长曲线模型。将(6.4.8)写成：

$$\ln y_t = \ln K + bt \ln a$$

就可以进行类似于逻辑增长曲线模型采用“三和法”估计参数的估计过程，这里也不再重复。

2.4 线性时间序列分析模型

时间序列分析模型包含广泛的内容，在经济预测中得到广泛的应用。在许多高等院校中将它作为一门专门课程，有关时间序列分析的教科书也不泛其数。在本节书中，只对几类常用的、在数学上采用回归分析的方法估计参数的线性时间序列分析模型作一介绍，作为经典计量经济学模型的一类扩展，使读者建立一个基本的概念与框架，为进一步专门学习与应用时间序列分析模型建立一个基础。

一、时间序列分析模型概述

所谓时间序列，就是各种社会、经济、自然现象的数量指标按照时间次序排列起来的统计数据。所谓时间序列分析模型，就是揭示时间序列自身的变化规律和相互联系的数学表达式。

时间序列分析模型分确定性模型和随机模型两大类。

1.确定性时间序列分析模型

对于一个时间序列

$$y_1, y_2, y_3, \cdots, y_n$$

确定性模型主要有以下几种。

(1)滑动平均模型

将下列平均数

$$\hat{y}_t = \frac{y_t + y_{t-1} + \cdots + y_{t-N+1}}{N} \quad t \geq N \tag{2.4.1}$$

称为时间序列y_t的滑动平均数序列。该式表达的模型称为滑动平均模型。滑动平均模型的主要作用是消除干扰，显示序列的趋势性变化，并用于趋势预测。

(2)加权滑动平均模型

将下列平均数

$$\hat{y}_{tw} = \frac{a_0 y_t + a_1 y_{t-1} + \cdots + a_{N-1} y_{t-N+1}}{N} \quad t \geq N \tag{2.4.2}$$

称为时间序列y_t的加权滑动平均数序列。$a_0, a_1, \cdots, a_{N-1}$为加权因子，满足

$$\frac{\sum\limits_{i=0}^{N-1} a_i}{N} = 1$$

由(6.4.2)式表达的模型称为加权滑动平均模型。加权滑动平均模型的主要作用除了消除干扰，显示序列的趋势性变化外，还可以通过加权因子的选取，增加新数据的权重，使趋势预测更加准确。

(3)二次滑动平均模型

所谓二次滑动平均是对经过一次滑动平均产生的序列再进行滑动平均。即

$$\hat{\hat{y}}_t = \frac{\hat{y}_t + \hat{y}_{t-1} + \cdots + \hat{y}_{t-N+1}}{N} \quad t \geq N \tag{2.4.3}$$

由此构成的序列称为时间序列y_t的二次滑动平均数序列，该式表达的模型称为二次滑动平均模型。

(4)指数平滑模型

如果采用下式求得序列的平滑预测值

$$\hat{y}_t = \hat{y}_{t-1} + \alpha(y_{t-1} - \hat{y}_{t-1}) \tag{2.4.4}$$

则称此预测模型为指数平滑模型，其中α称为平滑常数，$0 < \alpha < 1$。(6.4.5)也可以写成：

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1} \tag{2.4.5}$$

即预测值是前期实际值与预测值的加权和。

如何选择α？一种方法是优选法。按照优选法的规则，选择不同的α，代入模型，计算预测值序列。以实际值与预测值的差的平方和最小为准则，确定α值。

(5)二次指数平滑模型

在一次指数平滑模型的基础上再进行指数平滑计算，即构成二次指数平滑模型。同样，还可以构成三次指数平滑模型。不再赘述。

2.随机时间序列分析模型

随机时间序列分析模型分为3种类型：自回归模型（Auto-regressive Model, AR）、滑动平均模型(Moving Average Model,MA) 和自回归滑动平均模型（Auto-regressive Moving Average Model, ARMA)。

(1)自回归模型

若时间序列y_t为它的前期值和随机项的线性函数，可以表示为：

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mu_t \tag{2.4.6}$$

则称该时间序列y_t为自回归序列，该模型为p阶自回归模型，记为AR(p)。参数$\varphi_1, \varphi_2, \cdots, \varphi_p$为自回归参数，是模型的待估参数。随机项$\mu_t$为服从0均值、方差为$\sigma_\mu^2$的正态分布，且互相独立的白噪声序列。而且随机项$\mu_t$与$y_{t-1}, y_{t-2}, \cdots, y_{t-p}$不相关。

引入滞后算子B，模型(6.4.7)可以表示为：

$$y_t = \varphi_1 B y_t + \varphi_2 B^2 y_t + \cdots + \varphi_p B^p y_t + \mu_t \tag{2.4.7}$$

其中

$$B y_t = y_{t-1}, B^2 y_t = y_{t-2}, \cdots, B^p y_t = y_{t-p}$$

进一步有

$$(1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p) y_t = \mu_t \tag{2.4.8}$$

令$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p)$

则模型可写成：

$$\varphi(B) y_t = \mu_t$$

(2)滑动平均模型

若时间序列y_t为它的当前与前期的误差和随机项的线性函数，可以表示为：

$$y_t = \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \tag{2.4.9}$$

则称该时间序列y_t为滑动平均序列，该模型为q阶滑动平均模型，记为MA(q)。参数$\theta_1, \theta_2, \cdots, \theta_q$为滑动平均参数，是模型的待估参数。

同样，(6.4.10)可以写成：

$$(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \mu_t = y_t \tag{2.4.10}$$

令$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$

则模型可写成：

$$y_t = \theta(B) \mu_t$$

(3)自回归滑动平均模型

若时间序列y_t为它的当前与前期的误差和随机项，以及它的前期值的线性函数，可以表示为：

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \tag{2.4.11}$$

则称该时间序列y_t为自回归滑动平均序列，该模型为(p,q)阶自回归滑动平均模型，记为ARMA(p,q)。参数$\varphi_1, \varphi_2, \cdots, \varphi_p$为自回归参数，$\theta_1, \theta_2, \cdots, \theta_q$ 为滑动平均参数，是模型的待估参数。引入引入滞后算子B，(6.4.12)可以表示为：

$$\varphi(B)y_t = \theta(B)\mu_t$$

二、随机时间序列分析模型（AR、MA、ARMA）的识别

自回归滑动平均模型（ARMA）是随机时间序列分析模型的普遍形式，自回归模型(AR)和滑动平均模型(MA)是它的特殊情况。关于这几类模型的研究，是时间序列分析的重点内容，主要包括模型的平稳性分析、模型的识别和模型的估计。

1.自相关函数和偏自相关函数

对于ARMA模型，在进行参数估计之前，需要进行模型的识别。识别的基本任务是找出ARMA(p,q)、AR(p)、MA(q)模型的具体特征，最主要的，是确定模型的阶，即ARMA(p,q)中的p和q、AR(p)中的p、MA(q)中的q。识别的方法是利用时间序列样本的自相关函数和偏自相关函数。

(1)MA(q)的自相关函数

模型(6.4.10)

$$y_t = \mu_t - \theta_1\mu_{t-1} - \theta_2\mu_{t-2} - \cdots - \theta_q\mu_{t-q}$$

的自协方差函数为：

$$r_k = \mathrm{E}(y_{t+k}y_t) = \begin{cases} \sigma_\mu^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2) & \text{当}k=0 \\ \sigma_\mu^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q) & \text{当}1\le k\le q \\ 0 & \text{当}k>q \end{cases}$$

其中利用了

$$\begin{aligned} \mathrm{E}\,\mu_t &= 0 \\ \mathrm{E}(\mu_{t+j}\mu_t) &= \begin{cases} 0 & \text{当}j\neq 0 \\ \sigma_\mu^2 & \text{当}j=0 \end{cases} \end{aligned}$$

从而根据自相关函数的定义，有自相关函数：

$$\rho_k = \frac{r_k}{r_0} = \begin{cases} 1 & \text{当}k=0 \\ (-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q)/(1+\theta_1^2+\cdots+\theta_q^2) & \text{当}1\le k\le q \quad 6.4.12 \\ 0 & \text{当}k>q \end{cases} \tag{2.4.12}$$

由此可见，当$k > q$时，y_t与y_{t+k}不相关，这种现象称为截尾，因此，当$k > q$时，$\rho_k = 0$是MA(q)的一个特征。换句话说，可以根据自相关系数是否从某一点开始一直为0来判断MA(q)模型的阶。

⑵AR(p)的自相关函数

模型(6.4.8)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \mu_t$$

的自协方差函数为：

$$\begin{aligned} r_k &= \mathrm{E}(y_{t+k}y_t) = \mathrm{E}(\varphi_1 y_{t+k-1} + \varphi_2 y_{t+k-2} + \cdots + \varphi_p y_{t+k-p} + \mu_{t+k})y_t \\ &= \varphi_1 r_{k-1} + \varphi_2 r_{k-2} + \cdots + \varphi_p r_{k-p} \end{aligned}$$

从而有自相关函数：

$$\rho_k = \frac{r_k}{r_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p} \tag{2.4.13}$$

由此可见，AR(p)序列的自相关函数是非截尾序列，称为拖尾序列。因此自相关函数拖尾，是AR(p)序列的一个特征。

由(6.4.13)，利用$\rho_k = \rho_{-k}$，得到如下方程组：

$$\begin{aligned} \rho_1 &= \varphi_1 + \varphi_2\rho_1 + \cdots + \varphi_p\rho_{p-1} \\ \rho_2 &= \varphi_1\rho_1 + \varphi_2 + \cdots + \varphi_p\rho_{p-2} \end{aligned} \tag{2.4.14}$$

$$\ldots \rho_p = \varphi_1\rho_{p-1} + \varphi_2\rho_{p-1} + \cdots + \varphi_p\rho_{p-k}$$

此方程组被称为Yule Walker方程组。若已知模型参数$\varphi_1, \varphi_2, \cdots, \varphi_p$，可求$\rho_1, \rho_2, \cdots, \rho_p$，然后递推下去，可求所有的$\rho_k, k > p$；反过来，若已知$\rho_1, \rho_2, \cdots, \rho_p$，可求模型参数$\varphi_1, \varphi_2, \cdots, \varphi_p$以及$\sigma_\mu^2$。模型参数通过求解方程组得到，$\sigma_\mu^2$通过下式计算：

$$\sigma_\mu^2 = \mathrm{E}\,\mu_t^2 = \cdots = r_0 - \sum_{i,j=1}^{p} \varphi_i\varphi_j r_{j-i}$$

⑶ARMA(p,q)的自相关函数

ARMA(p,q)的自相关函数，可以看作MA(q)的自相关函数和AR(p)的自相关函数的混合物。当p=0时，它具有截尾性质；当q=0时，它具有拖尾性质；当p、q都不为0时，它具有拖尾性质。经过推导得到，ARMA(p,q)的自协方差函数为：

$$\begin{aligned} r_k &= \mathrm{E}(y_{t+k}y_t) = \varphi_1 r_{k-1} + \varphi_2 r_{k-2} + \cdots + \varphi_p r_{k-p} \\ &\quad + r_{y\mu}(k) - \theta_1 r_{y\mu}(k-1) - \cdots \theta_q r_{y\mu}(k-q) \end{aligned}$$

其中

$$r_{y\mu}(k) = \mathrm{E}(y_t \mu_{t+k}) = \begin{cases} 0 & \text{当}k > 0 \\ \sigma_\mu^2 \psi_{-k} & \text{当}k < 0 \end{cases}$$

所以，当$k > q$时，

$$r_k = \varphi_1 r_{k-1} + \varphi_2 r_{k-2} + \cdots + \varphi_p r_{k-p}$$

ARMA(p,q)的自相关函数为：

$$\rho_k = \frac{r_k}{r_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \cdots + \varphi_p \rho_{k-p} \tag{2.4.15}$$

可见，ARMA(p,q)的自相关函数ρ_k，当$k > q$时，仅依赖于模型参数$\varphi_1, \varphi_2, \cdots, \varphi_p$以及$\rho_{k-1}, \rho_{k-2}, \cdots, \rho_{k-p}$。

⑷ARMA(p,q)偏自相关函数

所谓偏自相关函数，是ARMA(p,q)模型的另一个统计特征，它是在已知序列值$y_{t-1}, y_{t-2}, \cdots, y_{t-k-1}$的条件下，$y_t, y_{t-k}$之间的关系的度量。

下面以AR(p)为例，认识偏自相关函数的含义。假定先以AR(k-1)去拟合一个序列，然后又用AR(k)去拟合，后者比前者增加了一个滞后变量y_{t-k}。如果φ_{kj}表示后者的自回归系数，那么相应于滞后变量y_{t-k}的系数就是φ_{kk}，称为偏自相关系数。根据AR(p)的拖尾性质以及偏自相关系数的含义，可以采用方差最小为原则来求得偏自相关系数：

$$\varphi_{kj} = \begin{cases} \varphi_j & \text{当}1 \le j \le p, k = p, p+1, \cdots \\ 0 & \text{当}j > p \end{cases}$$

由此得到AR(p)的主要特征是$k > p$时，$\varphi_{kk} = 0$，即是φ_{kk}在p以后截尾。

对于ARMA(p,q)与MA(q)模型，可以证明它们的偏自相关函数是拖尾的。

2.模型的识别

⑴AR(p)模型的识别

若y_t的偏自相关函数φ_{kk}在p以后截尾，即k¿p时，$\varphi_{kk} = 0$，而且它的自相关函数ρ_k是拖尾的，则此序列是自回归AR(p)序列。

⑵MA(q)模型的识别

若随机序列的自相关函数截尾，即自q以后$\rho_k = 0, k > q$，而它的偏自相关函数是拖尾的，则此序列是AR(p)序列是滑动平均MA(q)序列。

⑶ARMA(p,q)的识别

若随机序列的自相关函数和偏自相关函数都是拖尾的，则此序列是自回归滑动平均序列。至于模型中p和q的识别，则要从低阶开始逐步试探，直到定出合适的模型为止。

利用时间序列分析软件包中的自相关、偏相关函数的计算功能，可以方便地识别一个序列是何种序列，以及AR(p)、MA(q)序列的阶数。例

如，中国国内生产总值GDP序列（以当年价格计算，1978-1987年），根据自相关、偏相关函数的计算结果，当$k=4$时，φ_{kk}=0；而且当$k>4$时，所有$\varphi_{kk}\approx 0$；同时它的自相关函数ρ_k是拖尾的。所以，该序列是自回归AR(4)序列。

三、随机时间序列分析模型（AR、MA、ARMA）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

Yule Walker方程组(6.4.14)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$\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$以及$\hat{\sigma}_\mu^2$。结果如下：

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \cdots & \hat{\rho}_{p-2} \\ \vdots & & & \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & \hat{\rho}_0 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix} \tag{2.4.16}$$

$$\hat{\sigma}_\mu^2 = \hat{r}_0 - \sum_{j=1}^{p} \hat{\varphi}_j \hat{r}_j = \hat{r}_0 - \sum_{i,j=1}^{p} \hat{\varphi}_i \hat{\varphi}_i \hat{r}_{j-i} \tag{2.4.17}$$

例如，对于上述GDP序列，有：

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \hat{\varphi}_3 \end{pmatrix} = \begin{bmatrix} 1 & 0.809 & 0.604 \\ 0.809 & 1 & 0.809 \\ 0.604 & 0.809 & 1 \end{bmatrix}^{-1} \begin{pmatrix} 0.809 \\ 0.604 \\ 0.410 \end{pmatrix} = \begin{pmatrix} 0.574 \\ 0.432 \\ 0.210 \end{pmatrix}$$

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：

$$\hat{r}_k = \begin{cases} \hat{\sigma}_\mu^2(1+\hat{\theta}_1^2+\hat{\theta}_2^2+\cdots+\hat{\theta}_q^2) & \text{当}k=0 \\ \hat{\sigma}_\mu^2(-\hat{\theta}_k+\hat{\theta}_1\hat{\theta}_{k+1}+\cdots+\hat{\theta}_{q-k}\hat{\theta}_q) & \text{当}1\le k\le q \\ 0 & \text{当}k>q \end{cases} \tag{2.4.18}$$

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(6.4.18)是一个包含$(q+1)$个待估参数$\hat{\theta}_1,\hat{\theta}_2\cdots\hat{\theta}_q,\hat{\sigma}_\mu^2$的非线性方程组，可以用直接法或迭代法求解。常用的迭代方法有线性迭代法和Newton-Raphsan迭代法。具体的求解过程不再赘述，读者可参考时间序列分析的专门教科书。

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$\sigma_\mu^2$，其估计量计算步骤及公式如下：

(1)估计$\varphi_1,\varphi_2,\cdots,\varphi_p$

计算公式为：

$$\begin{bmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{bmatrix}=\begin{bmatrix}\hat{\rho}_q & \hat{\rho}_{q-1} & \cdots & \hat{\rho}_{q-p+1}\\ \hat{\rho}_{q+1} & \hat{\rho}_q & \cdots & \hat{\rho}_{q-p}\\ \vdots & & & \\ \hat{\rho}_{q+p-1} & \hat{\rho}_{q+p-2} & \cdots & \hat{\rho}_q\end{bmatrix}^{-1}\begin{bmatrix}\hat{\rho}_{q+1}\\ \hat{\rho}_{q+2}\\ \vdots\\ \hat{\rho}_{q+p}\end{bmatrix}$$

其中$\hat{\rho}_k$是样本的标准自相关函数的估计值，由观测数据计算得到。

(2)改写模型，求$\theta_1,\theta_2,\cdots,\theta_q$以及$\sigma_\mu^2$的估计值

将模型

$$y_t=\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+\cdots+\varphi_p y_{t-p}+\mu_t-\theta_1\mu_{t-1}-\theta_2\mu_{t-2}-\cdots-\theta_q\mu_{t-q}$$

改写为：

$$y_t-\varphi_1 y_{t-1}-\varphi_2 y_{t-2}-\cdots-\varphi_p y_{t-p}=\mu_t-\theta_1\mu_{t-1}-\theta_2\mu_{t-2}-\cdots-\theta_q\mu_{t-q} \quad (2.4.19)$$

令$\tilde{y}_t=y_t-\hat{\varphi}_1 y_{t-1}-\hat{\varphi}_2 y_{t-2}-\cdots-\hat{\varphi}_p y_{t-p}$

于是(6.4.19)可以写成：

$$\tilde{y}_t=\mu_t-\theta_1\mu_{t-1}-\theta_2\mu_{t-2}-\cdots-\theta_q\mu_{t-q}$$

构成一个MA模型。按照估计MA模型参数的方法，可以得到$\theta_1,\theta_2,\cdots,\theta_q$以及$\sigma_\mu^2$的估计值。

4.AR(p)的最小二乘估计

假设模型(6.4.8)的参数估计值$\hat{\varphi}_1,\hat{\varphi}_2,\cdots,\hat{\varphi}_p$已经得到，即有

$$y_t=\hat{\varphi}_1 y_{t-1}+\hat{\varphi}_2 y_{t-2}+\cdots+\hat{\varphi}_p y_{t-p}+\hat{\mu}_t$$

残差的平方和为：

$$S(\hat{\varphi})=\sum_{t=p+1}^{n}\hat{\mu}_t^2=\sum_{t=p+1}^{n}(y_t-\hat{\varphi}_1 y_{t-1}-\hat{\varphi}_2 y_{t-2}-\cdots-\hat{\varphi}_p y_{t-p})^2 \quad (2.4.20)$$

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(6.4.20)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

$$\sum_{t=p+1}^{n}(y_t-\hat{\varphi}_1y_{t-1}-\hat{\varphi}_2y_{t-2}-\cdots-\hat{\varphi}_py_{t-p})y_{t-j}=0\quad j=1,2,\ldots,p \tag{2.4.21}$$

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(2.4.20)改写成：

$$\frac{\hat{\varphi}_1}{n}\sum_{t=p+1}^{n}y_{t-1}y_{t-j}+\frac{\hat{\varphi}_2}{n}\sum_{t=p+1}^{n}y_{t-2}y_{t-j}+\cdots+\frac{\hat{\varphi}_p}{n}\sum_{t=p+1}^{n}y_{t-p}y_{t-j}=\frac{1}{n}\sum_{t=p+1}^{n}y_ty_{t-j}$$

j=1,2,. . . ,p

由自协方差函数的定义，并用自协方差函数的估计值

$$\hat{r}_k=\frac{1}{n}\sum_{t=p+1}^{n-k}y_{t+k}y_t$$

代入，上式表示的方程组即为：

$\hat{\varphi}_1\hat{r}_{j-1}+\hat{\varphi}_2\hat{r}_{j-2}+\cdots+\hat{\varphi}_p\hat{r}_{j-p}=\hat{r}_j$j=1,2,. . . ,p

解决该方程组，得到：

$$\begin{bmatrix}\hat{\varphi}_1\\ \hat{\varphi}_2\\ \vdots\\ \hat{\varphi}_p\end{bmatrix}=\begin{bmatrix}\hat{r}_0 & \hat{r}_1 & \cdots & \hat{r}_{p-1}\\ \hat{r}_1 & \hat{r}_0 & \cdots & \hat{r}_{p-2}\\ \vdots & & & \\ \hat{r}_{p-1} & \hat{r}_{p-2} & \cdots & \hat{r}_0\end{bmatrix}^{-1}\begin{bmatrix}\hat{r}_1\\ \hat{r}_2\\ \vdots\\ \hat{r}_p\end{bmatrix}$$

即为参数的最小二乘估计。与(6.4.16)的估计值比较发现，当n足够大时，二者是相似的。σ_μ^2的估计值为：

$$\hat{\sigma}_\mu^2=\frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\mu}_t^2=\frac{S}{n-p}$$

在实际应用中，最小二乘估计很难被应用，因为存在严重的共线性。例如，对于上述GDP序列，利用OLS方法，估计结果为：

$\hat{\varphi}_1 = 2.548 \quad (t = 14.03)$
$\hat{\varphi}_2 = -2.780 \quad (t = -7.19)$
$\hat{\varphi}_3 = 1.393 \quad (t = 4.82)$
$R^2 = 0.9976$

参数估计值显然不合理，而且显然是由于存在严重的共线性所导致。

2.5 协整理论与误差修正模型

当许多传统的计量经济学模型在70年代的经济动荡面前预测失灵时，误差修正模型却显示了它的稳定性和可靠性。对其原因进行深入分析之后发现，误差修正模型的非稳定的单整变量之间存在一种长期稳定关系。C.J.Granger把这种长期稳定关系称为“协整关系”，于是，一种新的理论—协整理论诞生了。虽然协整理论诞生于误差修正模型之后，但在本节中，为了便于理解，我们首先介绍协整理论，然后引出误差修正模型。

将协整理论与误差修正模型作为经典单方程计量经济学模型的扩展的理由在于，经典的计量经济学模型是以某种经济理论或对经济行为的认识来确立模型的理论关系形式，而在这里，则是从经济变量的数据中所显示的关系出发，确定模型包含的变量和变量之间的理论关系。这是80年代以来计量经济学模型建模理论的一个重大发展。

一、单整(Integration)

1.稳定序列

如果一个时间序列x_t是稳定的，则：

⑴其均值$E(x_t)$与时间t无关；

⑵其方差$Var(x_t)$是有限的，并不随着t的推移产生系统的变化。

于是，时间序列x_t将趋于返回它的均值，以一种相对不变的振幅围绕均值波动。

如果一个时间序列x_t是非稳定的，则其均值、方差将随t而改变。例如，随机游动序列

$$x_t = x_{t-1} + \varepsilon_t \quad \varepsilon_t \sim N(0, \delta^2)$$

若$x_0 = 0$

则$x_t = \sum\limits_{i=1}^{t} \varepsilon_i \quad Var(x_t) = t\delta^2$

当$t \to \infty$时，$Var(x_t) \to \infty$，均值就无意义了，实际上序列x_t返回曾经达到过的某一点的期望时间是无穷大。

一个稳定序列一般可以用一个自回归移动平均表达式ARMA(p,q)表示：

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \xi_t + \theta\xi_{t-1} + \cdots + \theta_q \xi_{t-q}$$

2.单整

如果一个序列在成为稳定序列之前必须经过d次差分，则该序列被称为d阶单整。记为I(d)。换句话说，如果序列x_t是非稳定序列，$\Delta^d x_t$是稳定序列，则序列x_t是I(d)。其中：

$$\Delta x_t = x_t - x_{t-1}, \Delta^2 x_t = \Delta(\Delta x_t), \cdots, \Delta^d x_t = \Delta(\Delta^{d-1} x_t)$$

如果有两个序列分别为d阶单整和e阶单整，即

$$x_t \sim I(d), y_t \sim I(e), e > d$$

则二序列的线性组合是e阶单整序列，即

$$z_t = \alpha x_t + \beta y_t \sim I(\max(d, e))$$

二、单整的单位根检验

1.单整的DF检验

对于时间序列x_t，建立下列方程：

$$x_t = \rho x_{t-1} + \varepsilon_t \quad \Delta x_t = (\rho - 1) x_{t-1} + \varepsilon_t$$

如果ρ不显著为0，则序列x_t存在单位根，至少为1阶单整I(1)。问题在于如何判断ρ是否不显著为0。

构造t统计量，但这时t统计量服从由Dickey和Fuller于1976年提出的Dickey—Fuller分布，即DF分布。象在经典线性单方程模型中介绍的变量显著性检验t统计量的计算一样，估计上述方程并计算得到t统计量的值；从DF分布表中查出给定显著性水平下的临界值；如果t统计量的绝对值大于临界值的绝对值，则拒绝$\rho = 0$假设，序列x_t至少为1阶单整I(1)。这就是Dickey—Fuller检验，也称为单位根（unit root）检验。

通过了1阶单整检验后，再建立如下方程：

$$\Delta^2 x_t = (\rho - 1)\Delta x_{t-1} + \varepsilon_t$$

进行同样过程的检验，如果通过检验，则序列x_t至少为2阶单整I(2)。...直到不能通过检验为止。通过该检验，同时也就确定了序列x_t的单整的阶数。

一般讲，在经济数据中，表示流量的序列，例如以不变价格表示的消费额、收入等经常表现为1阶单整；表示存量的序列，例如以不变价格表示的资产总值、储蓄余额等经常表现为2阶单整；用当年价格表示的流量的序列，例如以当年价格表示的消费额、收入等，由于价格指数的作用，也经常表现为2阶单整；而象利率等序列，经常表现为0阶单整。了解这些，对于选择什么变量进入模型是十分重要的。

2.单整的ADF检验

在DF检验中，由于不能保证方程中的ε_t是白躁声，所以得到的$\rho-1$的估计值不是无偏的。于是Dickey和Fuller于1979、1980年对DF检验进行了扩充，形成了ADF(Augment Dickey-Fuller)检验，这是目前普遍应用的单整检验方法。

在ADF检验中，为了保证方程中的ε_t是白躁声，在方程右边加了一些滞后项。于是单位根检验的回归方程为：

$$\Delta x_t = (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \theta_i \Delta x_{t-i} + \varepsilon_t$$

该方程称为模型1。如果包含常数项，则为模型2：

$$\Delta x_t = \alpha + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \theta_i \Delta x_{t-i} + \varepsilon_t$$

如果再加入时间趋势项，则为模型3：

$$\Delta x_t = \alpha + \beta t + (\rho - 1)x_{t-1} + \sum_{i=1}^{p} \theta_i \Delta x_{t-i} + \varepsilon_t$$

实际检验时从模型3开始，然后模型2、模型1，何时检验拒绝零假设，即原序列不存在单位根，为平稳序列，何时检验停止。注意，这里的零假设为：$H_0 : \rho = 1$。否则，就要继续检验。换句话说，只有检验到最后，才能得到原序列存在单位根的结论。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。

3.ADF检验的具体步骤

⑴估计和检验模型3

估计模型3，并得到参数的t统计量。

第一步，检验$H_0 : \rho = 1$。从ADF分布临界值表中查得给定显著性水平下的、用于模型3检验的τ_ρ的临界值。如果参数$\rho-1$的t统计量小于临界值（因为τ_ρ的临界值为负值，参数$\rho-1$的估计量一般为负，其t统计量也为负。所以t统计量小于临界值等价于其绝对值大于临界值的绝对值，与以前的概念不矛盾），则拒绝零假设，此时即可以得出序列不存在单位根的结论，不再进入下面的步骤；否则，进入下一步骤。

第二步，给定$\rho = 1$，检验$H_0 : \beta = 0$。从ADF分布临界值表中查得给定置信水平下的、用于模型3检验的τ_β的临界值。如果参数β的t统计量大于临界值（因为τ_β的临界值正值，参数β的估计量一般为正，其t统计量也为正。所以t统计量大于临界值等价于其绝对值大于临界值的绝对值），则拒绝零假设，进入下一步骤。否则，即说明模型不包含时间趋势项，应该采用模型2的形式，则要继续估计和检验模型2。

第三步，用一般的t分布检验$H_0: \rho = 1$。如果拒绝零假设，则原序列不存在单位根，为平稳序列；否则，说明原序列是不平稳的，必须对其差分后进一步检验其单位根。

⑵估计和检验模型2

估计模型2，并得到参数的t统计量。

第一步，检验$H_0: \rho = 1$。从ADF分布临界值表中查得给定显著性水平下的、用于模型2检验的τ_ρ的临界值。如果参数$\rho - 1$的t统计量小于临界值，则拒绝零假设，此时即可以得出序列不存在单位根的结论，不再进入下面的步骤；否则，进入下一步骤。

第二步，给定$\rho = 1$，检验$H_0: \alpha = 0$。从ADF分布临界值表中查得给定置信水平下的、用于模型2检验的τ_α的临界值。如果参数α的t统计量大于临界值，则拒绝零假设，进入下一步骤。否则，即说明模型不包含常数项，应该采用模型1的形式，要继续估计和检验模型1。

第三步，用一般的t分布检验$H_0: \rho = 1$。如果拒绝零假设，则原序列不存在单位根，为平稳序列；否则，说明原序列是不平稳的，必须对其差分后进一步检验其单位根。

⑶估计和检验模型1

估计模型1，并得到参数的t统计量。

检验$H_0: \rho = 1$。从ADF分布临界值表中查得给定显著性水平下的、用于模型1检验的τ_ρ的临界值。如果参数$\rho - 1$的t统计量小于临界值，则拒绝零假设，得出序列不存在单位根的结论；否则，说明原序列是不平稳的，必须对其差分后进一步检验其单位根。

三、协整（Cointegration）

1.定义及意义

如果序列$X_{1t}, X_{2t}, \cdots, X_{kt}$都是d阶单整，存在一个向量$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$，使得$\mathbf{Z}_t = \alpha \mathbf{X'}_t \sim I(d-b)$，其中，$b > 0, \mathbf{X}_t = (X_{1t}, X_{2t}, \cdots, X_{kt})'$，则认为序列$X_{1t}, X_{2t}, \cdots, X_{kt}$是(d,b)阶协整，记为$\mathbf{X}_t \sim CI(d, b), \alpha$为协整向量。

例如，居民收入时间序列Y_t为1阶单整序列，居民消费时间序列C_t也为1阶单整序列，如果二者的线性组合$\alpha_1 Y_t + \alpha_2 C_t$构成的新序列为0阶单整序列，于是认为序列$Y_t$与$C_t$是(1,1)阶协整。

由此可见，如果两个变量都是单整变量，只有当它们的单整阶相同时，才可能协整，例如上面的居民收入Y_t和居民消费C_t；如果它们的单整阶不相同时，就不可能协整，例如居民消费C_t和居民储蓄余额S_t（一般讲作为存量的居民储蓄余额S_t为2阶单整）。

三个以上的变量，如果具有不同的单整阶数，有可能经过线性组合构成低阶单整变量。例如，如果存在：

$$W_t \sim I(1), V_t \sim I(2), U_t \sim I(2)$$

并且

$$P_t = aV_t + bU_t \sim I(1)$$
$$Q_t = cW_t + eP_t \sim I(0)$$

那么认为：

$$V_t, U_t \sim CI(2,1)$$
$$W_t, P_t \sim CI(1,1)$$

从协整的定义可以看出协整的经济意义在于：两个变量，虽然它们具有各自的长期波动规律，但是如果它们是协整的，则它们之间存在着一个长期稳定的比例关系。例如居民收入Y_t和居民消费C_t，如果它们各自都是1阶单整，并且它们是(1,1)阶协整，则说明它们之间存在着一个长期稳定的比例关系，而这个比例关系就是消费倾向，也就是说消费倾向是不变的。从计量经济学模型的意义上讲，建立如下消费函数模型

$$C_t = \alpha_0 + \alpha_1 Y_t + \mu_t$$

变量选择是合理的，随机误差项一定是"白噪声"（即均值为0，方差不变的稳定随机序列），模型参数有合理的经济解释。

反过来，如果两个变量，具有各自的长期波动规律，但是它们不是协整的，则它们之间就不存在着一个长期稳定的比例关系。例如居民消费C_t和居民储蓄余额S_t，由于它们单整阶数不同，所以它们不是协整的，则说明它们之间不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，建立如下消费函数模型

$$C_t = \alpha_0 + \alpha_1 S_t + \mu_t$$

或者

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 S_t + \mu_t$$

变量选择是不合理的，随机误差项一定不是"白噪声"，模型参数没有合理的经济解释。

从这里，我们已经初步认识，检验变量之间的协整关系，在建立计量经济学模型中的重要性。而且，从变量之间是否具有协整关系出发选择模型的变量，其数据基础是牢固的，其统计性质是优良的。从协整理论出发，在建立消费函数模型时，就不会选择居民储蓄余额作为居民消费的解释变量；但是，按照传统的计量经济学建模理论，从已经认识的经济理论出发选择模型的变量，那么选择居民储蓄余额和居民收入共同作为居民消费的解释变量，不仅不感到奇怪，而且被认为是完全合理的，按照"生命周期消费理论"建立的消费函数模型正是这样的。

2.协整的检验

⑴两变量的Engle-Granger检验

为了检验两变量Y_t, X_t是否为协整，Engle和Granger于1987年提出两步检验法，也称为EG检验。

第一步，用OLS方法估计下列方程：

$$Y_t = \alpha X_t + \varepsilon_t$$

得到

$$\hat{Y}_t = \hat{\alpha} X_t$$
$$\hat{e}_t = Y_t - \hat{Y}_t$$

称为协整回归。

第二步，检验$\hat{e}_t$的单整性。如果$\hat{e}_t$为稳定序列，则认为变量Y_t, X_t为(1,1)阶协整；如果$\hat{e}_t$为1阶单整，则认为变量Y_t, X_t为(2,1)阶协整；...检验$\hat{e}_t$的单整性的方法即是上述的DF检验或者ADF检验。

⑵多变量协整关系的检验

上述Engle-Granger检验通常用于检验两变量之间的协整关系，对于多变量之间的协整关系，Johansen于1988年，以及与Juselius于1990年提出了一种用向量自回归模型进行检验的方法，通常称为Johansen检验，或JJ检验，将在本书§6.4中加以介绍。

四、误差修正模型（ECM）

误差修正模型（Error Correction Model）是一种具有特定形式的计量经济学模型，它的主要形式是由Davidson、Hendry、Srba和Yeo于1978年提出的，称为DHSY模型。为了便于理解，我们通过一个具体的模型来介绍它的结构。

对于(1,1)阶自回归分布滞后模型

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \varepsilon_t \tag{2.5.1}$$

移项后得到

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1) y_{t-1} + \beta_3 z_{t-1} + \beta_1 z_{t-1} + \varepsilon_t$$

$$= \beta_0 + \beta_1 \Delta z_t + (\beta_2 - 1)(y - \frac{\beta_1 + \beta_3}{1 - \beta_2} z)_{t-1} + \varepsilon_t \tag{2.5.2}$$

方程(6.4.2)即为误差修正模型。其中$y - \frac{\beta_1+\beta_3}{1-\beta_2} z$为误差修正项。

显然，(6.4.2)实际上是一个短期模型，反映了y_t的短期波动Δy_t是如何被决定的。如果变量y和z之间存在长期均衡关系，即存在$y = az$，例如在(6.5.1)中，若$z = \bar{z}$，那么y的均衡值与$\bar{z}$有下列均衡关系

$$\bar{y} = \frac{\beta_1 + \beta_3}{1 - \beta_2} \bar{z}$$

(6.4.2)中的误差修正项正是与它相一致的。所以它反映长期均衡对短期波动的影响；(6.4.2)中的差分项反映变量短期波动的影响。于是，被解释变量的波动被分成两部分：一部分为短期波动，一部分为长期均衡。

模型(6.4.2)可以写成：

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t + \gamma\ ecm + \varepsilon_t \tag{2.5.3}$$

其中ecm表示误差修正项。由(6.5.1)可知，一般情况下$|\beta_2| < 1$，所以有$\gamma = \beta_2 - 1 < 0$。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$\frac{\beta_1+\beta_3}{1-\beta_2}z$，$ecm$为正，$\gamma \times ecm$为负，使得$\Delta y_t$减少；若(t−1)时刻$y$小于其长期均衡解$\frac{\beta_1+\beta_3}{1-\beta_2}z$，$ecm$为负，$\gamma \times ecm$为正，使得$\Delta y_t$增大。体现了长期均衡误差对$y_t$的控制。

2.ecm与协整的关系

对于上述(1,1)阶自回归分布滞后模型，如果

$$y_t \sim I(1), z_t \sim I(1)$$

那么，(3.4.2)式左边

$$\Delta y_t \sim I(0)$$

右边的$\Delta z_t \sim I(0)$，只有y与z协整，才能保证右边也是I(0)。此时，$\frac{\beta_1+\beta_3}{1-\beta_2}$为协整系数，$y_t - \frac{\beta_1+\beta_3}{1-\beta_2}z_t$即为均衡误差。

3.从协整理论到误差修正模型

前面提到，实际上是先有误差修正模型，然后用协整理论去解释误差修正模型。那么在今天，我们就可以首先对变量进行协整分析，以发现变量之间的协整关系，即长期均衡关系，求出协整系数，并以这种关系构成误差修正项。然后建立短期模型，将误差修正项看作一个解释变量，连同其它反映短期波动的解释变量一起，建立短期模型，即误差修正模型。

五、一个误差修正模型实例：中国居民消费方程

下面结合中国居民消费模型的建立，具体说明误差修正模型的建立过程。数据选自《中国统计年鉴》，样本区间为1978-1997年。

1.初步分析

在对与居民消费相关的变量序列进行一般的分析后发现，用以表示中国居民消费水平的居民消费总额(C)取对数后随时间呈现线性变化，经过一阶差分后，其图形类似白噪声。于是可以初步判断$\ln C_t$为一个I(1)变量。同样可以初步判断国内生产总值的对数$\ln G_t$也为一个I(1)变量。

2.$\ln C_t$的单整检验

经过尝试，在用于ADF检验的模型1、模型2、模型3中选取滞后阶数p=2。对于模型3，为：

$$\Delta\ln C_t=\alpha+\beta\,t+(\rho-1)\ln C_{t-1}+\sum_{i=1}^{2}\theta_i\Delta\ln C_{t-i}+\varepsilon_t$$

估计该模型，变量$\ln C_{t-1}$的t统计量为-1.6553，大于ADF分布表中的临界值-3.60（给定显著性水平为5%，样本容量为25），不能拒绝存在单位根的零假设。于是进入模型3检验的第二步，变量t的t统计量为1.8305，小于ADF分布表中的临界值2.85（给定显著性水平为5%），得出不能拒绝零假设的结论。这样则要进行模型2的检验。

模型2为：

$$\Delta\ln C_t=\alpha+(\rho-1)\ln C_{t-1}+\sum_{i=1}^{2}\theta_i\Delta\ln C_{t-i}+\varepsilon_t$$

估计该模型，变量$\ln C_{t-1}$的t统计量为1.3577，大于ADF分布表中的临界值-3.00（给定显著性水平为5%，样本容量为25），不能拒绝存在单位根的零假设。于是进入模型2检验的第二步，常数项的t统计量为-0.7050，小于ADF分布表中的临界值2.61（给定显著性水平为5%），同样得出不能拒绝零假设的结论。这样则要进行模型1的检验。

模型1为：

$$\Delta\ln C_t=(\rho-1)\ln C_{t-1}+\sum_{i=1}^{2}\theta_i\Delta\ln C_{t-i}+\varepsilon_t$$

估计该模型，变量$\ln C_{t-1}$的t统计量为2.9528，大于ADF分布表中的临界值-1.95（给定显著性水平为5%，样本容量为25），不能拒绝存在单位根的零假设。

至此，方可以断定序列$\ln C_{t-1}$中含有单位根。一般来说，检验某一时间序列是否含有单位根，很少需要检验完模型1、2、3才能得出结论。该例是比较特殊的。

为了判断序列$\ln C_{t-1}$的单整阶数，必须检验差分之后的序列是否含有单位根。首先检验模型3：

$$\Delta^2\ln C_t=\alpha+\beta\,t+(\rho-1)\Delta\ln C_{t-1}+\sum_{i=1}^{2}\theta_i\Delta^2\ln C_{t-i}+\varepsilon_t$$

估计模型，得出变量$\Delta\ln C_{t-1}$的t统计量为-3.2770，小于ADF分布表中的临界值-3.24（给定显著性水平为10%，样本容量为25），所以在10%的显著性水平下，拒绝存在单位根的零假设。即$\Delta\ln C_{t-1}$为I(0)序列。那么序列$\ln C_{t-1}$为1阶单整。

3.$\ln G_t$的单整检验

采用同样的步骤，检验得到国内生产总值经过对数化后的序列为为1阶单整。

4.方程的初步设定

在本书第六章中将详细介绍，按照“从一般到简单”的思想，最初设定的方程是一个一般的自回归分布滞后模型。模型的右边包含被解释变量的滞后、解释变量及其滞后。初始的模型必须保证足够的阶数，以尽量包含被解释变量的有效信息，使残差成为不包含有效信息的白噪声。

对于居民消费模型，首先设定为：

$$\ln C_t = \alpha_0 + \alpha_1 \ln C_{t-1} + \alpha_2 \ln C_{t-2} + \alpha_3 \ln G_t + \alpha_4 \ln G_{t-1} + \alpha_5 \ln G_{t-2} + \varepsilon_t \tag{2.5.4}$$

5.方程的估计与简化

估计自回归分布滞后模型的常用软件是PC-Give。采用其中的最小二乘法估计(2.5.4)的参数，逐次剔除不显著的变量$\ln G_{t-2}$和$\ln G_{t-1}$，得到：

$$\ln \hat{C}_t = -0.13905 + 0.48253 \ln C_{t-1} - 0.21189 \ln C_{t-2} + 0.69223 \ln G_t$$

此时，各变量系数都显著不为0，拟合优度为0.9995，残差近似白噪声。关于对该模型的其它方面的专门检验，这里不作介绍。在PC-Give中为这些检验提供了方便的功能。如果仅从估计的角度，一般的软件也是适用的。

6.求长期均衡方程

由于$\ln C_t$和$\ln G_t$都为I(1)变量，它们之间存在长期稳定的均衡关系。由PC-Give求出的长期均衡方程为：

$$\ln C = 0.9491 \ln G - 0.1907$$

如果没有PC-Give，一般的软件也是可以采用的。

7.求出误差修正序列(ecm)

利用

$$ecm = \ln C - 0.9491 \ln G + 0.1907$$

代入$\ln C_t$和$\ln G_t$的实际观测值，求出ecm序列，作为误差修正模型中解释变量的观测值。通过对该序列的检验，可以得到它是0阶单整的结论，证明了$\ln C_t$和$\ln G_t$确实存在协整关系。

8.建立误差修正模型

对下列误差修正模型进行估计：

$$\Delta \ln C_t = \alpha \Delta \ln C_{t-1} + \beta \Delta \ln G_t + \gamma ecm + \varepsilon_t$$

估计后得到：

$$\Delta \ln \hat{C}_t = 0.1830\Delta \ln C_{t-1} + 0.7451\Delta \ln G_t - 0.5215ecm$$

各项检验均通过。

在模型中，各差分项反映了变量短期波动的影响。被解释变量的波动可以分为两部分：一部分是短期波动，一部分是长期均衡。根据模型的参数估计量，短期国内生产总值的变化将引起居民消费的相同方向的变化，如果国内生产总值变化1%，引起居民消费变化0.7451%；而上期居民消费的变化，也引起居民消费的相同方向的变化，弹性为0.183，反映消费惯性的延续。ecm项系数的大小反映了对偏离长期均衡的调整力度，从系数估计值（-0.5215）看，调整力度是比较大的。这一分析结果，就是误差修正模型的优势所在。

2.6 无参数回归模型

经典计量经济学模型首先根据经济理论和样本数据设定模型的函数关系，然后估计关系参数并检验所设定的关系。如果模型的函数关系通过检验被证明是成立的，那么回归结果可以外延，其推断有较高的精度，模型的参数一般具有明确的经济意义，可以方便于各方面的应用。但关于模型及参数的一些假定在现实中未必成立。例如C-D生产函数模型假定技术进步是中性的、技术进步独立于要素投入量的变化、要素替代弹性为1、具有一次齐次性即不变规模报酬等，这些假定在现实中很难同时成立。因而当模型及参数的假定与实际背离（也包括模型的随机干扰项的正态性假定与实际背离）时，就容易造成模型设定误差。此时，基于经典假设模型所作出的推断的表现可能很差，这就促使人们寻找别的出路，而无参数回归模型则是朝着这个方向的一种努力。无参数回归模型的特点是：回归函数的形式可以任意,没有任何约束，解释变量和被解释变量的分布也很少限制，因而有较大的适应性，但无参数回归结果外延困难。无参数技术的目的是放松回归函数形式的限制，为确定或建议回归函数的参数表达式提供有用的工具。无参数技术并不能取代参数技术，两者相结合将会得到用单一方法无法获得的结论。

一、无参数回归模型的概念

设随机变量Y是被解释变量，p维向量X是解释变量，它既可以是确定性的也可以是随机性的。在本节中，为了书写方便，并不将解释变量写成矩阵形式。读者应该将它作为矩阵来理解。称

$$m(x) = E(Y|X = x) \tag{2.6.1}$$

为Y对X的回归函数或条件回归函数，其中$x \in R^p$。无参数回归模

表 2.5.1: ADF分布临界值表

模型	统计量	样本容量	1%	2.5%	5%	10%
1	τ_ρ	25	-2.66	-2.26	-1.95	-1.60
		50	-2.62	-2.25	-1.95	-1.61
		100	-2.60	-2.24	-1.95	-1.61
		250	-2.58	-2.23	-1.95	-1.61
		500	-2.58	-2.23	-1.95	-1.61
		¿500	-2.58	-2.23	-1.95	-1.61
2	τ_ρ	25	-3.75	-3.33	-3.00	-2.62
		50	-3.58	-3.22	-2.93	-2.60
		100	-3.51	-3.17	-2.89	-2.58
		250	-3.46	-3.14	-2.88	-2.57
		500	-3.44	-3.13	-2.87	-2.57
		¿500	-3.43	-3.12	-2.86	-2.57
3	τ_ρ	25	-4.38	-3.95	-3.60	-3.24
		50	-4.15	-3.80	-3.50	-3.18
		100	-4.04	-3.73	-3.45	-3.15
		250	-3.99	-3.69	-3.43	-3.13
		500	-3.98	-3.68	-3.42	-3.13
		¿500	-3.96	-3.66	-3.41	-3.12
2	τ_α	25	3.41	2.97	2.61	2.20
		50	3.28	2.89	2.56	2.18
		100	3.22	2.86	2.54	2.17
		250	3.19	2.84	2.53	2.16
		500	3.18	2.83	2.52	2.16
		¿500	3.18	2.83	2.52	2.16
3	τ_α	25	4.05	3.59	3.20	2.77
		50	3.87	3.47	3.14	2.75
		100	3.78	3.42	3.11	2.73
		250	3.74	3.39	3.09	2.73
		500	3.72	3.38	3.08	2.72
		¿500	3.71	3.38	3.08	2.72
3	τ_β	25	3.74	3.25	2.85	2.39
		50	3.60	3.18	2.81	2.38
		100	3.53	3.14	2.79	2.38
		250	3.49	3.12	2.79	2.38
		500	3.48	3.11	2.78	2.38
		¿500	3.46	3.11	2.78	2.38

型就是要在给定样本$\{(X_i,Y_i)\}_{i=1}^n$下，得到条件回归函数$m(x)$的一个估计$\hat{m}_n(x)$。

如果联合密度函数$f(x,y)$存在，则$m(x)$可表成

$$m(x)=\int yf(x,y)dy/f(x)$$

其中$f(x)=\int f(x,y)dy$是X的边缘密度函数。此时，称(2.6.1)为随机设定模型。

如果X是确定性变量，称(2.6.1)为固定设定模型，可以表示为：

$$Y_i=m(X_i)+\varepsilon_i,\quad i=1,\cdots,n \tag{2.6.2}$$

其中$\{\varepsilon_i\}_{i=1}^n$是相互独立、均值为0、方差为$\sigma^2$的序列。

无参数回归模型的估计方法有三大类，一是权函数方法，二是最小二乘估计，三是稳健估计。

二、权函数估计

权函数方法的系统研究始于Stone(1977)，最常用的权函数方法是核估计和近邻估计。使用无参数平滑技术时应该意识到关于被估回归曲线的最终确定存在一定的主观性，因为即使是渐近最优平滑也含有相当部分的噪音，留给人们主观选择的余地。

1. 权函数估计

给定样本$\{(X_i,Y_i)\}_{i=1}^n$，权函数估计就是条件回归函数$m(x)$的估计$\hat{m}_n(x)$，可表示成下述形式：

$$\hat{m}_n(x)=\sum_{i=1}^n W_{ni}(x)Y_i \tag{2.6.3}$$

其中权函数$W_{ni}(x)=W_{ni}(x;X_1,\cdots,X_n)$，$i=1,\cdots,n$，满足$W_{ni}(x)\geq 0$，$\sum\limits_{i=1}^n W_{ni}(x)=1$。

从(6.4.2)可看出，条件回归函数的估计是Y_i的线性组合，对应$x\in R^p$所得到的被解释

变量的估计$\hat{m}_n(x)$是Y_i的加权平均，权数利用了解释变量的信息且由解释变量的数值来确定每个Y_i的权数的大小。

2. 核估计

核估计是权函数估计的一种方法，在于找到核权函数作为(6.4.2)中的权函数。核权函数是最重要的一种权函数，由此产生的估计，即核估计已有广泛的应用。

(1) Nadaraya-Watson核估计

Nadaraya(1964)及Watson (1964)提出了一种既适合解释变量是确定性变量，也适合解释变量是随机变量的核估计，其思路如下：选定概率密度$K(\cdot)$

$$\int K(u)du = 1 \tag{2.6.4}$$

为核函数及窗宽$h > 0$，定义核权函数为：

$$W_{ni}(x) = K_h(x - X_i) / \sum_{j=1}^{n} K_h(x - X_j) \tag{2.6.5}$$

其中$K_h(u) = h^{-1}K(uh^{-1})$也是一个概率密度。于是Nadaraya-Watson核估计为：

$$\hat{m}_n(x) = \sum_{i=1}^{n} K_h(x - X_i)Y_i / \sum_{j=1}^{n} K_h(x - X_j) \tag{2.6.6}$$

容易推得：

$$\min_{\theta} n^{-1} \sum_{i=1}^{n} W_{ni}(x)(Y_i - \theta)^2 = n^{-1} \sum_{i=1}^{n} W_{ni}(x)(Y_i - \hat{m}_n(x))^2 \tag{2.6.7}$$

所以，核估计等价于局部加权最小二乘估计。关于局部加权最小二乘估计，将在§3.4中介绍。

当$h \to 0$时,有

$$\hat{m}_n(X_i) \to K(0)Y_i / K(0) = Y_i$$

$$\hat{m}_n(x) \to 0, (x \neq X_i, i = 1, \cdots, n) \tag{2.6.8}$$

可见，太小的窗宽得到除了数据点外其它点的函数值都为0的函数。

当$h \to \infty$时,$K(\frac{x - X_i}{h}) \to K(0)$，所以

$$\hat{m}_n(x) \to n^{-1} \sum_{i=1}^{n} K(0)Y_i / n^{-1} \sum_{i=1}^{n} K(0) = n^{-1} \sum_{i=1}^{n} Y_i \tag{2.6.9}$$

可见，太大的窗宽得到过分光滑的曲线，接近于直线。

由上述可见，窗宽是控制核估计精度的重要参数。

(2) 密度核估计

密度估计本身不属于无参数回归内容，但随机设定模型的回归方法里经常用到它，且在需求和均衡理论中，家庭收入的密度函数估计是一个重要的应用内容。另外，在条件回归函数的导数期望估计ADE(Average derivative estimation)中也要用到密度函数的估计。

a)一元密度函数估计

可从经验分布函数导出密度核估计，经验分布函数为：

$$F_n(x) = \frac{1}{n}(X_1, \cdots, X_n\text{中小于}X\text{的个数}) \tag{2.6.10}$$

取核函数为均匀核：

$$K_0(x) = \begin{cases} \frac{1}{2} & \text{当} -1 \leq x < 1 \\ 0 & \text{其他} \end{cases} \tag{2.6.11}$$

则密度估计为

$$\hat{f}_n(x) = [F_n(x+h) - F_n(x-h)]/2h = \frac{1}{2h}\int_{x-h}^{x+h} dF_n(t) = \int_{-\infty}^{+\infty} \frac{1}{h} K_0(\frac{x-t}{h}) dF_n(t)$$

$$= \frac{1}{nh}\sum_{i=1}^{n} K_0(\frac{x-X_i}{h}) \tag{2.6.12}$$

将核函数放宽就得到一般的密度核估计

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-X_i}{h}) \tag{2.6.13}$$

要求当$n \to +\infty$时，$h \to 0, nh \to +\infty$。

下面是除了均匀核外的其它核函数的例子：

$$K_1(x) = \begin{cases} 1-|x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{2.6.14}$$

$$K_2(x) = [\pi(1+x^2)]^{-1} \tag{2.6.15}$$

$$K_3(x) = \begin{cases} (2\pi)^{-1}\left[\frac{\sin(x/2)}{x/2}\right]^2 & x \neq 0 \\ (2\pi)^{-1} & x = 0 \end{cases} \tag{2.6.16}$$

$$K_4(x) = \begin{cases} 0.75\lambda^{-3}(\lambda^2 - x^2) & x^2 \leq \lambda^2 \\ 0 & x^2 > \lambda^2 \end{cases} \tag{2.6.17}$$

其中$\lambda > 0$。

此外，还有高斯核和Epanechnikov核等。

b) 多元密度函数估计

设p维随机向量X的密度函数$f(x) = f(x_1,\cdots,x_n)$未知，$X_1,X_2,\cdots,X_n$是它的一个独立同分布的样本，则$f(x)$的核估计为：

$$\hat{f}_n(x) = \frac{1}{nh^p}\sum_{i=1}^{n} K(\frac{x-X_i}{h}) \tag{2.6.18}$$

其中$K(\cdot)$是多元核函数。对窗宽的要求为：当$n \to +\infty$时，$h \to 0$，$nh^p \to +\infty$。

此外，密度函数估计方法还有近邻(Nearest Neighbors)、正交序列(Orthogonal Series)、惩罚函数(Penalty Functions)、付里叶逆变换（Fourier Inversion）、δ序列(Delta Sequences)、小波估计、贝叶斯方法等。

(3) 其它形式的核权函数

在随机设定模型中，如果$f(x)$已知，可取核权函数（Greblicki 1974;Johnston 1979,1982;Greblicki和Krzyzak 1980; Georgiev 1984a,1984b）为：

$$W_{ni}^{(1)}(x) = K_h(x-X_i)/[nf(x)] \tag{2.6.19}$$

假设$\{X_i\}_{i=1}^n$在[0,1]上取值，不妨设$X_1 \le X_2 \le \cdots \le X_n$，Priestley和Chao(1972)和Benedetti(1977)采用了下面的核权：

$$W_{ni}^{(2)}(x) = (X_i - X_{i-1})K_h(x-X_i) \quad X_0 = 0 \tag{2.6.20}$$

这个核权的解释是$f(x)$的估计为$\hat{f}(x) = [n(X_i - X_{i-1})]^{-1}$,$x \in (X_{i-1},X_i)$。

Gasser和Müller(1979)定义了如下核权：

$$W_{ni}^{(3)}(x) = \int_{S_{i-1}}^{S_i} K_h(x-u)du \tag{2.6.21}$$

其中$X_{i-1} \le S_{i-1} \le X_i$。

（2.6.20）和(6.4.21) 主要用于固定设定模型。

回归函数估计的方差随着窗宽减少而增大，偏随着窗宽减少而减少。所以，无参数估计就是在估计的偏和方差中寻求平衡，使得均方误差达最小。

(4)导数的核估计

a)固定设定模型

回归函数的导数的核估计采用Priestley-Chao核权定义为：

$$\hat{m}_n^{(k)}(x) = h^{-(k+1)}\sum_{i=1}^{n}(X_i - X_{i-1})K^{(k)}(\frac{x-X_i}{h})Y_i \tag{2.6.22}$$

即回归函数的k阶导数用回归函数的核估计的k阶导数来估计。

b)随机设定模型

对于随机设定模型，仍可利用核权函数的导数得到条件回归函数导数的估计：

$$\hat{m}_n^{(k)}(x) = \sum_{i=1}^{n} W_{ni}^{(k)}(x) Y_i$$

但此时的权数的表达式很复杂。例如：

$$W_{ni}^{(1)}(x) = \frac{K_h^{(1)}(x - X_i)}{\hat{f}_n(x)} - \frac{K_h(x - X_i)\hat{f}_n'(x)}{(\hat{f}_n'(x))^2}$$

其中$K_h^{(1)}(u) = h^{-2} K^{(1)}(u/h)$，$\hat{f}_n'(x) = n^{-1} \sum_{i=1}^{n} K_h^{(1)}(x - X_i)$

（5）窗宽的选择

当$p = 1$，$K(\cdot)$为[-1,1]上对称、单峰的概率密度时，$\hat{m}_n(x)$是集中在x附近一个邻域的样本的加权平均，而h正好是这个邻域的宽度。当h大时，参加平均的样本就多，会提高估计精度，但可能会增大偏差。反之，h小则正好相反。所以，窗宽h是控制核估计精度的最主要的参数。

因为平方拟合误差

$$p(h) = n^{-1} \sum_{i=1}^{n} (Y_i - \hat{m}_n(X_i^{)})^2 w(X_i) \tag{2.6.23}$$

($w(x) \geq 0$为某权数)随着h增加而增加(参见(6.4.8)和（2.6.9）)。所以，使得$p(h)$达最小的h为0。因而，最佳窗宽选择的基本思想是使得平均平方误差：

$$d_A(h) = n^{-1} \sum_{i=1}^{n} (\hat{m}_n(X_i) - m(X_i))^2 w(X_i) \tag{2.6.24}$$

达到最小。我们希望通过$p(h)$调整而得到$d_A(h)$的估计量。为此，将$p(h)$展开：

$$p(h) = n^{-1} \sum_{j=1}^{n} \varepsilon_j^2 w(X_j) + d_A(h) - 2n^{-1} \sum_{j=1}^{n} \varepsilon_j (\hat{m}_n(X_j) - m(X_j)) w(X_j) \tag{2.6.25}$$

最后一项可重写为：

$$C_{1n}(h) = -2n^{-1} \sum_{j=1}^{n} \varepsilon_j \left[n^{-1} \sum_{i=1}^{n} W_{ni}(X_j) Y_i - m(X_j) \right] w(X_j) \tag{2.6.26}$$

其在给定$\{X_1,\cdots,X_n\}$下的条件期望为：

$$-2n^{-1}\sum_{j=1}^{n}[n^{-1}W_{nj}(X_j)\sigma^2(X_j)]w(X_j) \tag{2.6.27}$$

至少有三种可能的途径得到$d_A(h)$的无偏估计：

- 应用剔除技术使得6.2.24)的期望为0。
- 对$p(h)$进行修正，消除(6.2.24)的偏。
- 利用插入法，将$d_A(h)$中未知量用其无偏估计量代入。

对应于这三种不同的途径产生了选择窗宽的交错鉴定方法、惩罚函数法和插入法。

交错鉴定方法（Cross-validation method）是Stone（1977）提出的，在给定核权函数$W_{ni}(x)$下，选择带宽h的一个常用方法。其基本思路是：在每个局部观察点$x=X_i$，首先在样本中剔除该观察点(X_i,Y_i)；其次将剩下的$n-1$个观察点在$x=X_i$处进行核估计：

$$\hat{m}_{n,-i}(X_i)=\sum_{j\neq i}W_{nj}(X_i)Y_j$$

最后，通过比较平方拟合误差

$$CV(h)=n^{-1}\sum_{i=1}^{n}(Y_i-\hat{m}_{n,-i}(X_i))^2w(X_i)$$

($w(x)\geq 0$为某权数)的大小，选择使平方拟合误差达最小的带宽h。该方法的关键是在样本中剔除观察点(X_i,Y_i)。如果不这样的话，由于核权函数$W_{ni}(x)$在观察点$x=X_i$达最大值，就会使得$x=X_i$的重要程度过份夸大且其它观察点数据的重要程度降低。所以采用交错鉴定方法就避免了因没剔除观察点(X_i,Y_i)，而将有用的数据排除在外的情况。将$CV(h)$展开，类似于（2.6.26）的项为：

$$C_{2n}(h)=-2n^{-1}\sum_{j=1}^{n}\varepsilon_j\left[n^{-1}\sum_{\substack{i\neq j\\ i=1}}^{n}W_{hi}(X_j)Y_i-m(X_j)\right]w(X_j)$$

其期望为0。

交错鉴定方法和惩罚函数法有时在实践中不太可靠，尤其是在高维情况。插入法（The plug-in method），即把未知函数的估计插入到渐近公式里以选择最佳窗宽。插入法的思想最初由Woodroofe(1970)在进行密度估计时引入，后由Scoott,Tapia和Thompson(1977)提出插入法选择窗宽的迭代计算步骤，Sheather和Jones(1991)，Gasser（1991）改进了这一方法，改进的插入法有好的理论分析性质与好的实际效果，并且被认为比交错鉴定方法和惩罚函数法要好。

(5)常用核函数

核估计的核心问题就是核权函数的选择和窗宽的选择。核权函数在核估计中起光滑的作用，即消除扰动的随机因素，使所得曲线反映变量之间的实际经济关系。常用的核权函数有高斯（Gaussian）和Epanechnikov核权函数。

a) 高斯核函数(Gaussian Kernel)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)'(X_i-\mu) \tag{2.6.28}$$

当X是一元时, Σ写成$\sigma^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$。高斯核为：

$$K(u) = \frac{1}{(2\pi)^{p/2}(|\det(\Sigma)|)^{1/2}}\exp(-\frac{1}{2}u\Sigma^{-1}u') \tag{2.6.29}$$

$$K_h(x-X_i) = \frac{1}{(2\pi)^{p/2}(|\det(h^2\Sigma)|)^{1/2}}\exp(-\frac{1}{2h^2}(x-X_i)\Sigma^{-1}(x-X_i)')$$

(2.6.30)

当X是一元时,$K_h(x-X_i) = \frac{1}{(2\pi)^{1/2}(h\sigma)}\exp(-\frac{1}{2h^2}\frac{(x-X_i)^2}{\sigma^2})$。

定义高斯核权函数为：

$$W_{ni}(x) = K_h(x-X_i)/\sum_{j=1}^{n}K_h(x-X_j)$$

b) Epanechnikov 核

Epanechnikov核为：

$$K(u) = \frac{p(p+2)}{2S_p}(1-u_1^2-\cdots-u_p^2)_+ \tag{2.6.30}$$

其中$S_p = 2\pi^{p/2}/\Gamma(p/2)$。当$p=1$时,$K(u) = 0.75(1-u^2)I(|u|\le 1)$

$$K_h(x-X_i)=\frac{p(p+2)}{2hS_p}[1-(\frac{x_1-X_{i1}}{h})^2-\cdots-(\frac{x_p-X_{ip}}{h})^2)_+ \tag{2.6.31}$$

Epanechnikov核权函数为：

$$W_{ni}(x)=K_h(x-X_i)/\sum_{j=1}^{n}K_h(x-X_j)$$

多元的核函数也可由一元的核函数来构造：

$$K(u_1,\cdots,u_p)=\prod_{i=1}^{p}K_i(u_i) \tag{2.6.32}$$

其中$K_i(t), i=1,\cdots,p$为一元的核函数。

$$K_h(x-X_i)=\prod_{j=1}^{p}h_j^{-1}K_j(\frac{x_j-X_{ij}}{h_j}) \tag{2.6.33}$$

其中$h=(h_1,\cdots,h_p)$为窗宽向量。
多元核权函数为

$$W_{ni}(x)=\prod_{j=1}^{p}K_j(\frac{x_j-X_{ij}}{h_j})/\sum_{k=1}^{n}\prod_{j=1}^{p}K_j(\frac{x_j-X_{kj}}{h_j}) \tag{2.6.34}$$

窗宽甚至可以为窗宽矩阵，设$K(u)$为多元核函数，

$$K_B(u)=|B|^{-1}K(uB^{-1}) \tag{2.6.35}$$

其中B是$p\times p$正定矩阵,称之为窗宽矩阵。
多元核权函数为

$$W_{ni}(x)=K_B(x-X_i)/\sum_{j=1}^{n}K_B(x-X_j) \tag{2.6.36}$$

(6)核函数的选择

$$\hat{m}_n(x)\text{œff}L$$

$$C_Vc_Kn^{-1}h^{-1}+C_B^2d_K^2h^4 \tag{2.6.37}$$

其中C_V和C_B是与核函数K无关的量。最小化(5.8.35)得到窗宽为：

$$h_o = \left(\frac{C_V}{4C_B^2}\right)^{1/5} \left(\frac{c_K}{d_K^2}\right)^{1/5} n^{-1/5} \tag{2.6.38}$$

代入(5.8.35),可得：

$$MSE_{opt} = n^{-4/5}(C_V)^{4/5}(C_B)^{2/5}(4^{1/5} + 4^{-4/5})c_K^{4/5}d_K^{2/5} \tag{2.6.39}$$

所以，最优的核函数是使

$$c_K^2 d_K = \left(\int K^2(u)du\right)^2 \int u^2 K(u)du \tag{2.6.40}$$

达最小的核函数。

更一般地，我们考虑估计p次可微的m的k阶导数$m^{(k)}$。使得$\hat{m}^{(k)}$的均方误差达最小的核函数为使得

$$V(K)B(K) = \left[\int_{-1}^{1} (K^{(k)}(u))^2 du\right]^{p-k} \left|\int_{-1}^{1} K^{(k)}(u)u^p du\right|^{2k+1} \tag{2.6.41}$$

达最小的核函数。

因为核函数关于尺度变换仍为核函数，

$$K^{(k)}(u) \to s^{-(k+1)}K^{(k)}(u/s) \tag{2.6.42}$$

所以，我们要对核函数加以约束：有支撑[-1,1]。例如，核函数

$$K(u) = C_\alpha(1-u^2)^\alpha I(|u| \le 1) \tag{2.6.43}$$

表 2.6.1: 最小化$V(K)B(K)$的核函数

k	p	$K(u)$
0	2	$(3/4)(-u^2+1)I(\|u\| \le 1)$
0	4	$(15/32)(7u^4-10u^2+3)I(\|u\| \le 1)$
1	3	$(15/4)(u^3-u)I(\|u\| \le 1)$
1	5	$(105/32)(-9u^5+14u^3-5u)I(\|u\| \le 1)$
2	4	$(105/16)(-5u^4+6u^2-1)I(\|u\| \le 1)$
2	6	$(315/64)(77u^6-135u^4+63u^2-5)I(\|u\| \le 1)$

从表2.6.1可看出，最优核的导数并不是条件回归函数导数估计的最优核。例如，$(k,p)=(1,3)$的核并不是$(k,p)=(0,4)$的核函数。

(7) 边界核

在接近观察区间的边界，由于较少的观察点用于平均，所以任何无参数回归方法的精确度都有所下降，此时估计的偏和方差都将增大。

考虑固定设定模型，核函数有[-1,1]支撑，条件回归函数的核估计为：

$$\hat{m}_n(x)=n^{-1}\sum_{i=1}^{n}K_h(x-X_i)Y_i$$

则它的期望可表成：

$$\int_{(x-1)/h}^{x/h}K(u)m(x-uh)du+O(n^{-1}h^{-1}) \tag{2.6.44}$$

设$x=\rho h\le 1-h$,利用台劳级数展开，(5.8.42)可由下式近似

$$m(x)\int_{-1}^{\rho}K(u)du-hm'(x)\int_{-1}^{\rho}uK(u)du+(1/2)h^2m''(x)\int_{-1}^{\rho}u^2K(u)du$$

$$=m(x)\omega_K(0,\rho)-hm'(x)\omega_K(1,\rho)+(1/2)h^2m''(x)\omega_K(2,\rho) \tag{2.6.45}$$

显然，若取核（不是概率密度）为：

$$K_\rho(\cdot)=K(\cdot)/\omega_K(0,\rho) \tag{2.6.46}$$

则$\hat{m}_n(x)$随着$h\to 0,n\to\infty,nh\to\infty$为渐近无偏估计。

当然，如果$\rho\ge 1$，则$\omega_K(0,\rho)=1,\omega_K(1,\rho)=0,\omega_K(2,\rho)=d_K$。因而，当$x$远离左边界时，偏由展开式（2.6.46）的第三项给出。如果$\rho<1$,即x在边界时，偏基本上是由展开式的第二项给出，广义刀切技术（Gray和Schucany 1972）可消除这一项的偏。记$\hat{m}_{h,\rho}(x)$为核为$K_\rho(\cdot)$的核估计，称

$$\hat{m}_h^J(x)=(1-R)\hat{m}_{h,\rho}(x)+R\hat{m}_{\alpha h,\rho}(x) \tag{2.6.47}$$

为$m(x)$的刀切估计。当

$$R=-\frac{\omega_K(1,\rho)/\omega_K(0,\rho)}{\alpha\omega_K(1,\rho/\alpha)/\omega_K(0,\rho/\alpha)-\omega_K(1,\rho)\omega_K(0,\rho)} \tag{2.6.48}$$

时，$\hat{m}_h^J(x)$的估计偏展开式的主要项将被消除。事实上，刀切估计是使用了核函数

$$K_\rho^J(u) = (1-R)K(u) - (R/\alpha)K(u/\alpha) \tag{2.6.49}$$

故称$K_\rho^J(\cdot)$为边界核。Rrice(1984b)建议α选择

$$\alpha = 2 - \rho$$

3．k-近邻估计

$$k - C\text{ßﬀl}OQu\text{œ}^\circ\text{ø}.u^\circ\text{ø}.\text{”}$$

(1) k-近邻核权函数(K-Nearest Neighbor Kernel Weights)
让$k = int(hn)$,其中0¡h ≤1.定义

$$J_x = \{i : \mathbf{X}_i\text{是离}x\text{最近的}k\text{个观察值之一}T\}$$
$$d(x) = (xx')^{1/2} \tag{2.6.50}$$
$$R(x) = \max\{d(x-z) : z \in J_x\} \tag{2.6.51}$$
$$u_i(x) = d(x - X_i)/R(X_i) \tag{2.6.52}$$

$k-$ 近邻估计常用的核函数有：

- Uniform Kernel

$$K(t) = 0.5 I_{[-1,1]}(t) \tag{2.6.53}$$

- Triangular Kernel

$$K(t) = (1-|t|)I_{[-1,1]}(t) \tag{2.6.54}$$

- Quartic Kernel

$$K(t) = \frac{15}{16}(1-|t|^2)^2 I_{[-1,1]}(t) \tag{2.6.55}$$

- Tricube Kernel

$$K(t) = \frac{70}{81}(1-|t|^3)^3 I_{[-1,1]}(t) \tag{2.6.56}$$

定义KNN核权函数为：

$$W_{ni}(x)=\frac{K(u_i(x))}{\sum\limits_{j=1}^{n}K(u_j(x))} \tag{2.6.57}$$

$k-$近邻估计的核权也可直接由多元核函数K产生：

$$W_{Ri}(x)=\frac{K_R(x-X_i)}{n\hat{f}_R(x)} \tag{2.6.58}$$

其中$\hat{f}_R(x) = n^{-1}\sum\limits_{i=1}^{n}K_R(x-X_i)$ 是$f(x)$核密度估计，$K_R(u) = R^{-1}K(u/R)$,

$$R=\max\{|x-z|,z\in J_x\}"$$

(2)k-近邻估计

回归函数$m(x)$的k-近邻估计为

$$\hat{m}_n(x)=\sum_{i=1}^{n}W_{ni}(x)Y_i \tag{2.6.59}$$

对于一元随机设定模型的k-近邻估计，如果$k\to\infty$，$h\to 0$，$n\to\infty$，则均匀核(4.2.51)的k-近邻估计的偏和方差为：

$$\begin{aligned}&\mathrm{E}\,\hat{m}_n(x)-m(x)\approx \tfrac{1}{24f(x)^3}[(m''f+2m'f')(x)](k/n)^2\\&\mathrm{Var}\{\hat{m}_n(x)\}\approx \tfrac{\sigma^2(x)}{k}\end{aligned}$$

另一类近邻估计是一元对称化近邻估计，它由Yang(1981)提出，Stute(1984)进一步研究。即

$$\hat{m}_{k(h)}(x)=(nh)^{-1}\sum_{i=1}^{n}K\left(\frac{F_n(X_i)-F_n(x)}{h}\right)Y_i \tag{2.6.60}$$

其中F_n是X样本的经验分布。定义期望为

$$\bar{m}_{k(h)}(x)=h^{-1}\int m(u)K\left(\frac{F(u)-F(x)}{h}\right)du \tag{2.6.61}$$

则当$n\to\infty, h\to 0, nh^3\to\infty$时，

$$(nh)^{1/2}(\hat{m}_{k(h)}(x)-\bar{m}_{k(h)}(x))\longrightarrow N(0,c_K\sigma^2(x)) \tag{2.6.62}$$

不难证明，估计的偏为：

$$h^2 d_K \frac{(m''f - m'f')(x)}{2f^3(x)} O(h^2) \tag{2.6.63}$$

表 2.6.2: 固定设定模型核估计和k-近邻估计的逐点偏和方差

权函数	偏	方差
核权$\{W_{hi}\}$	$h^2 \frac{m''(x)}{2} d_K$	$\frac{\sigma^2(x)}{nh} c_K$
$k-$近邻权	$\left(\frac{k}{n}\right)^2 \frac{m''(x)}{8} d_K$	$\frac{2\sigma^2(x)}{k} c_K$

表 2.6.3: 随机设定模型核估计和k-近邻估计的逐点偏和方差

权函数	偏	方差
核权$\{W_{hi}\}$	$h^2 \frac{(m''f+2m'f')(x)}{2f(x)} d_K$	$\frac{\sigma^2(x)}{nhf(x)} c_K$
$k-$ 近邻权	$\left(\frac{k}{n}\right)^2 \frac{(m''f+2m'f')(x)}{8f(x)^3} d_K$	$\frac{2\sigma^2(x)}{k} c_K$

三、一个用权函数估计的无参数回归模型实例

例2.6.1 我国对外经济联系与国内通货膨胀关系的无参数估计模型

选定商品进出口总额x和外汇储备fc表示我国对外经济联系。代表通货膨胀的变量则采用居民消费价格指数。从《中国物价》得到1993年4月到1998年11月每月与上年同月相比的居民消费价格指数，再换算成每月与1992年4月相比的居民消费价格指数，我们用它作为被解释变量变量y。商品进出口总额资料来自《海关统计》外汇储备资料来自《中国金融》。数据见表2.6.4。

为了进行比较，经典线性回归模型的估计结果如下：

$$\hat{y} = \underset{(21.446)}{122.06} + \underset{(2.2246)}{0.065978}\, x + \underset{(13.373)}{0.052493}\, fc$$
$$R^2 = 0.84582 \quad F = 178.29$$

图2.6.1是经典线性回归模型的拟合图，拟合的均方误差为11.035。

无参数回归模型采取高斯核估计方法，采用交错鉴定法选择最佳窗宽为$h = 0.25$，图2.6.2是高斯核估计的拟合图，无参数回归模型拟合的均方误差为3.809。再对无参数回归模型采取k-近邻Tricube核估计，采用交错鉴定法选择最佳窗宽为$h = 0.1$，图2.6.3是Tricube核估计的拟合图，拟合的均方误差为1.8644。 表2.6.5是我国通货膨胀经典线性回归模型、高斯核和k-近邻估计的无参数回归模型的拟合值，其中Y1是经典线性回归模型的拟合值，Y2是高斯核估计的拟合值，Y3是k-近邻Tricube核估计的拟合值。

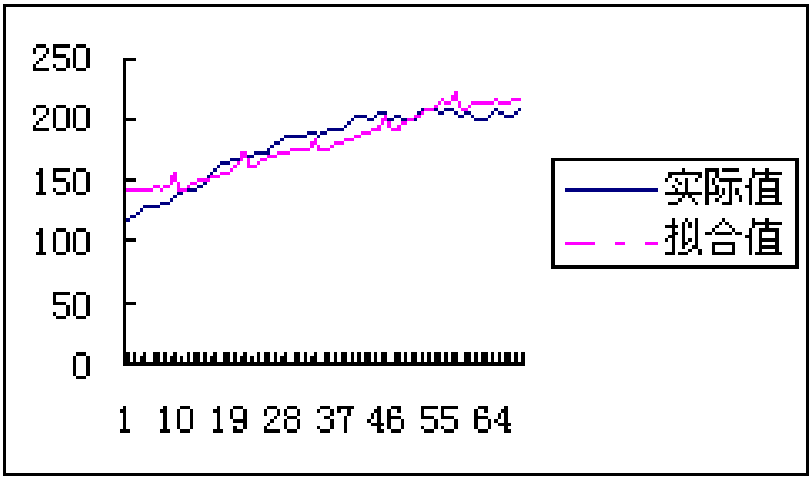


图 2.1: 线性回归拟合图

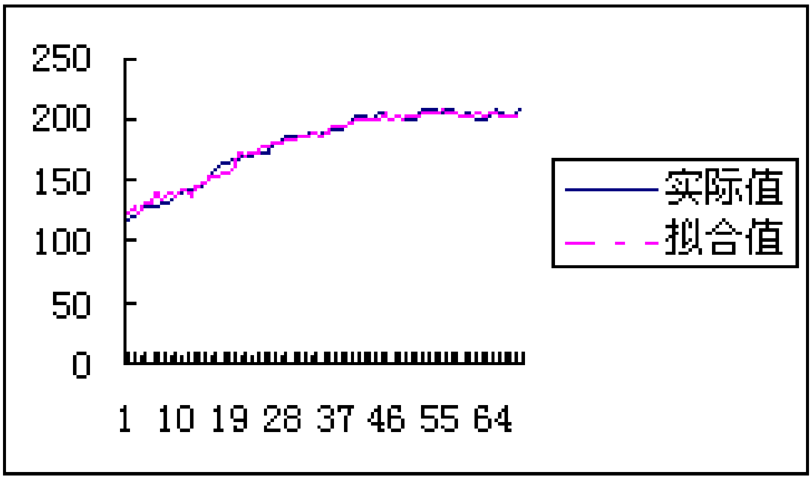


图 2.2: 高斯核估计的拟合图

表 2.6.4: 我国对外经济和通货膨胀的数据（1993.4-1998.11）

obs	P	Y	X	FC	obs	P	Y	X	FC
93.04	117	117	143.79	196.42	96.02	110.3	190.57	167.77	797.67
93.05	119.5	121.07	157.22	194.51	96.03	111.2	190.89	215.28	808.29
93.06	121.6	124.82	143.35	188.77	96.04	110.9	191.35	222.85	818.24
93.07	123.3	128.24	163.88	189.95	96.05	109.6	193.46	241.51	850.02
93.08	122.2	128.77	163.93	191.58	96.06	109.8	198.61	231.51	866.16
93.09	120.7	128.86	183.96	196.2	96.07	108.9	202.41	236.95	896.93
93.10	121.1	131	166.87	210.05	96.08	108.7	202.99	252.48	931.93
93.11	121.9	133.59	184.78	223.9	96.09	108.3	200.89	236.88	953.63
93.12	123.9	137.57	315.9	237.76	96.10	108.1	201.23	265.06	988.5
94.01	123.2	138.6	105.95	251.61	96.11	108.1	204.63	252.4	1023.06
94.02	125.9	143.5	116.95	278.49	96.12	107.9	204.66	383.58	1050.29
94.03	124.5	143.77	172.12	286.18	97.01	106.8	199.1	216.36	1086.79
94.04	123.2	144.144	183.42	301.42	97.02	106.4	202.77	175.23	1103.15
94.05	121.8	147.47	189.19	305.5	97.03	104.2	198.9	251.86	1120.55
94.06	122.7	153.16	208.13	318.63	97.04	103.4	197.85	259.96	1144.08
94.07	124.2	159.27	199.66	348.26	97.05	103	199.27	269.06	1175.75
94.08	127.1	163.67	202.44	377.9	97.06	103.1	204.77	266.13	1209.41
94.09	127.5	164.29	197.01	407.53	97.07	103.5	209.4	283.86	1259.46
94.10	127	166.36	198.26	437.16	97.08	102.7	208.47	270.9	1303.4
94.11	126.4	168.86	229.71	490.23	97.09	102.7	206.31	280.64	1340.7
94.12	124.1	170.73	364.46	516.2	97.10	102.1	205.46	306.45	1378.97
95.01	122.5	169.78	156.79	547.89	97.11	102	208.72	288.37	1388.73
95.02	120.4	172.77	160.08	573.86	97.12	101.5	207.73	382.8	1398.9
95.03	119.4	171.66	231.22	579.6	98.01	101.4	201.89	214.71	1404.86
95.04	119.7	172.54	222.69	599.22	98.02	100.9	204.95	213.82	1403.33
95.05	119.7	179.56	243.1	606.29	98.03	101.9	202.69	272.24	1406.17
95.06	118.1	180.88	249.92	626.59	98.04	100.7	199.24	276.44	1405.67
95.07	116.7	185.86	233.94	655.94	98.05	99.7	198.67	261.99	1409.05
95.08	114.1	186.74	241.07	675.19	98.06	99.4	203.54	274.55	1405.1
95.09	112.9	185.49	237.11	698	98.07	98.8	206.98	281.2	1405.99
95.1	111.9	186.15	233.14	729.98	98.08	98.7	205.76	264.35	1407.38
95.11	112.1	189.3	242.25	744.96	98.09	98.3	202.8	270.18	1411.07
95.12	111.1	189.68	357.17	735.97	98.10	98.6	202.59	263.79	1437
96.01	109.8	186.42	192.76	761.07	98.11	98.7	206	274.73	1445.88

表 2.6.5: 我国通货膨胀线性回归、高斯核和k-近邻估计的拟合值

年月	Y0	Y1	Y2	Y3	年月	Y0	Y1	Y2	Y3
93.04	117	141.86	123.51	121.62	96.02	190.57	175	188.47	189.69
93.05	121.07	142.64	128.93	124.99	96.03	190.89	178.69	193.79	191.1
93.06	124.82	141.43	123.49	122.67	96.04	191.35	179.72	194.93	192.1
93.07	128.24	142.84	131.91	126.31	96.05	193.46	182.62	195.14	195.31
93.08	128.77	142.93	131.93	126.27	96.06	198.61	182.8	196.71	197.18
93.09	128.86	144.5	140.37	128.95	96.07	202.41	184.78	197.8	200.09
93.10	131	144.1	133.35	129.93	96.08	202.99	187.64	198.83	201.94
93.11	133.59	146.01	141.88	132.03	96.09	200.89	187.75	199.05	201.91
93.12	137.57	155.38	137.66	137.63	96.10	201.23	191.44	200.54	202.39
94.01	138.6	142.26	140.66	140.54	96.11	204.63	192.42	200.84	202.52
94.02	143.5	144.4	141.57	141.74	96.12	204.66	202.5	204.44	204.09
94.03	143.77	148.44	137.13	145.46	97.01	199.1	193.38	198.96	200.36
94.04	144.14	149.98	144.67	147.69	97.02	202.77	191.53	202.71	200.58
94.05	147.47	150.58	147.71	148.49	97.03	198.9	197.5	201.32	198.71
94.06	153.16	152.52	154.55	150.77	97.04	197.85	199.27	201.94	199.21
94.07	159.27	153.52	153.78	157.53	97.05	199.27	201.53	202.91	200.53
94.08	163.67	155.25	156.27	162.3	97.06	204.77	203.1	203.23	203.21
94.09	164.29	156.45	156.29	164.13	97.07	209.4	206.9	205.34	206.96
94.10	166.36	158.09	158.39	165.33	97.08	208.47	208.35	204.4	208
94.11	168.86	162.95	173.02	168.02	97.09	206.31	210.95	205.2	206.99
94.12	170.73	173.2	170.84	170.72	97.10	205.46	214.67	206.44	206.77
95.01	169.78	161.17	171.69	171.31	97.11	208.72	213.98	205.65	208.04
95.02	172.77	162.75	172.35	171.69	97.12	207.73	220.75	207.73	207.34
95.03	171.66	167.74	178.39	175.09	98.01	201.89	209.97	203.41	203.28
95.04	172.54	168.21	178.77	175.69	98.02	204.95	209.83	203.42	203.29
95.05	179.56	169.93	179.67	176.95	98.03	202.69	213.84	203.8	202.2
95.06	180.88	171.44	180.17	180.59	98.04	199.24	214.09	204.16	202.94
95.07	185.86	171.93	182.34	185.05	98.05	198.67	213.31	203.09	202.36
95.08	186.74	173.41	182.8	185.65	98.06	203.54	213.93	204	202.47
95.09	185.49	174.34	184.36	186.41	98.07	206.98	214.42	204.64	203.19
95.10	186.15	175.76	187.11	186.99	98.08	205.76	213.38	203.24	202.46
95.11	189.3	177.15	186.53	187.35	98.09	202.8	213.96	203.6	202.96
95.12	189.68	184.26	189.7	189.24	98.10	202.59	214.9	203.05	203.93
96.01	186.42	174.73	186.8	187.65	98.11	206	216.09	203.67	204.38

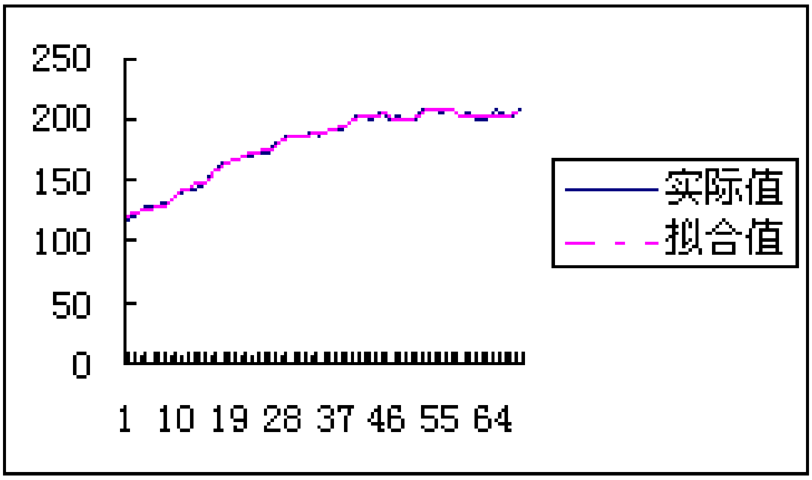


图 2.3: k-近邻Tricube核估计拟合图

四、确定性解释变量无参数回归模型的最小二乘法估计

最小二乘法是确定性解释变量无参数回归模型估计的另一种主要方法，主要有正交序列估计、样条估计和小波估计。关于小波估计，这里不作介绍。

1. 正交序列估计

一元自变量的取值区间本来可以从∞到∞，但经过一个线性变换，总可以变到[-1,1]，不妨设回归函数$m(x)$的定义域为[-1,1]，假设$\{\varphi_j\}_{j=0}^{\infty}$构成[-1,1]上的一组正交基，即

$$\int_{-1}^{1} \varphi_j(x)\varphi_k(x)dx = \delta_{jk} = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} \tag{2.6.64}$$

则$m(x)$有正交序列展开：

$$m(x) = \sum_{j=1}^{\infty} \beta_j \varphi_j(x) \tag{2.6.65}$$

显然有

$$\beta_j = \sum_{k=0}^{\infty} \beta_k \delta_{jk} = \sum_{k=0}^{\infty} \beta_k \int_{-1}^{1} \varphi_k(x)\varphi_j(x)dx = \int_{-1}^{1} m(x)\varphi_j(x)dx \tag{2.6.66}$$

于是，模型的估计为：

$$\hat{m}_n(x) = \sum_{j=1}^{N(n)} \hat{\beta}_j \varphi_j(x) \tag{2.6.67}$$

其中

$$\hat{\beta}_j = n^{-1} \sum_{i=1}^{n} \varphi_j(X_i) Y_i \tag{2.6.68}$$

$NJj)$,ˇ`”

若$\{A_i\}_{i=1}^n$是一组不交的区间集合，且满足$\sum\limits_{i=1}^n A_i = [-1,1]$和$X_i \in A_i$，$i = 1, \cdots, n$，则

$$\beta_j = \int_{-1}^{1} m(x)\varphi_j(x)dx = \sum_{i=1}^{n} \int\limits_{A_i} m(x)\varphi_j(x)dx \approx \sum_{i=1}^{n} m(X_i) \int\limits_{A_i} \varphi_j(x)dx \tag{2.6.69}$$

于是，

$$\hat{\beta}_j = \sum_{i=1}^{n} Y_i \int\limits_{A_i} \varphi_j(x)dx \tag{2.6.70}$$

常用的正交基有标准的Legendre多项式和Fourier基。
(1) Legendre多项式

$$\begin{aligned}
&P_0(x) = 1/\sqrt{2} \\
&P_1(x) = x/\sqrt{2/3} \\
&P_2(x) = \tfrac{1}{2}(3x^2 - 1)/\sqrt{2/5} \\
&P_3(x) = \tfrac{1}{2}(5x^3 - 3x)/\sqrt{2/7} \\
&P_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3)/\sqrt{2/9} \\
&P_5(x) = \tfrac{1}{8}(63x^5 - 70x^3 + 15x)/\sqrt{2/11}
\end{aligned}$$

其它高阶Legendre多项式可由下式递推地推出：

$$(m+1)P_{m+1}(x) = (2m+1)xP_m(x) - mP_{m-1}(x) \tag{2.6.71}$$

Legendre多项式序列$\{P_j(x)\}_{j=0}^{\infty}$构成[-1,1]上的一组正交基。
(2) Fourier基

$$q_1(x) = 1$$

$$\begin{aligned} q_{2k}(x) &= \sqrt{2}\cos(2\pi kx), \\ q_{2k+1}(x) &= \sqrt{2}\sin(2\pi kx) \end{aligned} \quad k = 1, 2, \cdots \tag{2.6.72}$$

则序列$\{q_j(x)\}_{j=0}^{\infty}$构成[0,1]上的一组正交基。

2. 样条估计

样条光滑最初是由Whittaker(1923)作为一种数值分析的工具提出。到六十年代Schoenberg(1964)用于数值模拟，导出光滑样条估计，但真正作为一种无参数回归分析方法研究光滑样条则始于七十年代Grace Wahba的一系列工作。现在的理论和经验都表明光滑样条是一种有效的统计数据分析的工具。

(1) 多项式样条

考虑

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \cdots, n$$

其中ε_i是i.i.d. ，均值为0，方差为σ^2。

设$t_1, \cdots, t_J$是固定节点序列，$-\infty < t_1 < \cdots < t_J < +\infty$。

$B_j(x) = (x - t_j)_+^3, (j = 1, \cdots, J)$

$$B_{J+1}(x) = 1, B_{J+2}(x) = x, B_{J+3}(x) = x^2, B_{J+4}(x) = x^3 \tag{2.6.73}$$

则立方样条函数可表示成：

$$\sum_{j=1}^{J+4} \theta_j B_j(x)$$

于是，多项式样条估计为最小化

$$\sum_{i=1}^{n} \{Y_i - \sum_{j=1}^{J+4} \theta_j B_j(x)\}^2 \tag{2.6.74}$$

得到$\hat{\theta}_j$ $(j = 1, \cdots, J+4)$,再得到

$$\hat{m}(x) = \sum_{j=1}^{J+4} \hat{\theta}_j B_j(x)\text{''} \tag{2.6.75}$$

上述的样条估计方法取决于节点个数和节点位置的选择，节点位置应选择曲线的曲率明显有较大变换的位置。

(2) 平滑样条

一种自动选择节点的方法是平滑样条。如果考虑单纯最小二乘问题，求函数m使得

$$\frac{1}{n}\sum_{j=1}^{n}[Y_j - m(X_j)]^2 \tag{2.6.76}$$

达最小，则函数m为通过所有观测点的折线的表达式。但这样的解没有应用价值，因为它的残差全为0，与现实不吻合。为了保证m有一定的光滑性，应采用如下的惩罚最小二乘法。

假定未知的回归函数$m(x)$的定义域为[0,1]，它的d阶导数是平方可积的，即

$$J(m) = \int_0^1 [m^{(d)}(x)]^2 dx < +\infty \tag{2.6.77}$$

$$m(x)\text{œȷ}Nm$$

$$m(x) = \sum_{i=0}^{d-1}\theta_i x^i + Re(x)$$

其中余项为

$$Re(x) = \frac{1}{(d-1)!}\int_0^1 m^{(d)}(t)(x-t)_+^{d-1}dt$$

如余项$Re(\cdot)$可以被控制得很小，则m的合理估计是一个$d-1$次多项式。使用Cauchy-Schwarz不等式有

$$[Re(x)]^2 \le \frac{1}{(2d-1)[(d-1)!]^2}J(m)$$

因而，存在常数$c = c(m)$¿0,使得

$$\sup_{x\in(0,1)} |Re(x)| \le cJ(m)^{1/2}$$

也就是说，作为m的光滑程度的指标$J(m)$直接关系到用多项式逼近时的余项的大小。例如对某个充分小的ρ¿0,有$J(m)\le\rho$，则这个ρ值表明在多大程度上多项式回归是否可用于m的估计。

现设$J(m)\le\rho$，回归函数$m(x)$的样条估计为使

$$S_\lambda(m) = \frac{1}{n}\sum_{j=1}^{n}[Y_j - m(X_j)]^2 + \lambda\int_0^1 [m^{(d)}(x)]^2 dx \tag{2.6.78}$$

达最小的解$\hat{m}_{n,\lambda}(x)$。这里λ称为光滑因子。这相当于在约束$J(m)\le\rho$下，使得(2.6.76) 达最小的约束最小二乘估计：

$$\frac{1}{n}\sum_{j=1}^{n}[Y_j - m(X_j)]^2 + \lambda F(m) - \rho) = \min$$

当$d=2$，Silverman(1984)指出平滑样条估计相当于变窗宽的核估计：

$$\hat{m}_{n\lambda}(x) = \sum_{i=1}^{n} W_{\lambda i}(x) Y_i$$

其中

$$W_{\lambda i}(\cdot) \approx n^{-1} f(X_i)^{-1} h(X_i)^{-1} K_S\left(\frac{\cdot - X_i}{h(X_i)}\right) \tag{2.6.79}$$

$$K_s(u) = 1/2\exp(-|u|/\sqrt{2})\sin(|u|/\sqrt{2} + \pi/4)$$

$$h(X_i) = \lambda^{1/4} n^{-1/4} f(X_i)^{-1/4}$$

为了确定(2.6.78)中的λ，定义$n\times n$矩阵$A(\lambda)$：

$$\begin{pmatrix} \hat{m}_{n,\lambda}(X_1) \\ \hat{m}_{n,\lambda}(X_2) \\ \vdots \\ \hat{m}_{n,\lambda}(X_n) \end{pmatrix} = A(\lambda) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \tag{2.6.80}$$

因为$\hat{m}_{n,\lambda}(X_i)$,$i=1,\cdots,n$是$Y_1,\cdots,Y_n$的线性函数，所以$A(\lambda)$是存在的。

Wahba提出，λ的估计可取作下式的最小化解：

$$V(\lambda) = \frac{1}{n}||(I - A(\lambda))Y||^2 / \left[\frac{1}{n}tr(I - A(\lambda))\right]^2 \tag{2.6.81}$$

并把这个估计称为广义交错鉴定(Generalized Cross-Validation,GCV)估计。

如果希望得到$\hat{m}_{n,\lambda}(x)$和$A(\lambda)$的显示表达式，可借助于贝努里（Bernoulli）多项式。这里不作具体介绍。

另外，分块多项式估计是一种重要的无参数回归估计，它较之多项式估计有更多的灵活性。事实上样条估计也是一种特殊的分块多项式估计，只是通过不同途径得到。较早的文献，如MCGtl和Catleton(1970)、Major(1973)、Stone(1985)对这种估计有深入的理论分析。但这些工作都是基于L.S.方法，因而缺乏稳健性。我国学者从稳健性出发，研究分块多项式M-估计，并已作出一些深入的理论工作（见SHI Pei de(1992)博士论文）。

五、稳健估计

1. L-平滑

考虑条件L-函数

$$l(x)=\int_0^1 J(v)F^{-1}(v|x)dv \tag{2.6.82}$$

其中$J(v)=I(\alpha\le v\le 1-\alpha)/(1-2\alpha)$)，$0<\alpha<1/2$；

$$F^{-1}(v|x)=\inf\{y:F(y|x)\ge v\}0<v<1\text{ˆ''}$$

可用核技术估计$F(\cdot|x)$如下：

$$F_n(t|x)=\frac{n^{-1}\sum_{i=1}^n K_h(x-X_i)I(Y_i\le t)}{\hat{f}_n(x)} \tag{2.6.83}$$

于是，$m(x)$的L-平滑估计为：

$$\hat{m}_n^L(x)=\int_0^1 J(v)F_n^{-1}(v|x)dv \tag{2.6.84}$$

2. M-平滑

M-平滑估计的思想是利用非二次损失函数减少异常观测值的影响。这样损失函数的一个例子（Huber 1981）是：

$$\rho(u)=\begin{cases}(1/2)u^2 & |u|\le c\\ c|u|-(1/2)c^2 & |u|>c\end{cases} \tag{2.6.85}$$

其中常数c调整稳健的程度。

M-平滑样条估计为：

$$\arg\min_g\left\{n^{-1}\sum_{i=1}^n\rho(Y_i-g(X_i))+\lambda\int[g''(x)]^2dx\right\} \tag{2.6.86}$$

其中$z=\arg\min\limits_u v(u)$为$v(u)$有唯一最小值且在z处达到。

M-平滑核估计为

$$\arg\min_\theta\left\{\sum_{i=1}^n W_{ni}(x)\rho(Y_i-\theta)\right\} \tag{2.6.87}$$

3. LOWESS(Locally Weighted Scatter plot Smoothing)

Cleveland（1979）提出了基于局部多项式拟合的稳健估计方法LOWESS。基本思想是先用局部多项式估计进行拟合，然后定义稳健

的权数并进行平滑，重复几次后就可得到稳健估计。这里局部估计方法是采用k-近邻估计。

第一步：在x的邻域进行局部多项式回归估计，得到$\{\beta_j\}_{j=0}^p$的估计，使得

$$\frac{1}{n}\sum_{i=1}^{n}W_{ni}(x)\left(Y_i-\sum_{j=0}^{p}\beta_j(X_i-x)^j\right)^2 \tag{2.6.88}$$

达最小。其中$\{W_{ni}(x)\}$是k-近邻权，最佳窗宽由交错鉴定法确定。

第二步：计算残差,$\hat{\varepsilon}_i = Y_i-\hat{\beta}_0$.其中$\hat{\beta}_0$是在$x$的邻域进行局部多项式回归的常数项$\beta_0 = m(x)$的估计量。计算$\hat{\sigma}ied\{|\hat{\varepsilon}_i|\}$并定义稳健权数$\delta_i = K(\hat{\varepsilon}_i/(6\hat{\sigma}))$，其中$K(u)=(15/16)(1-u^2)^2I(|u|\le 1)$。

第三步：重复第一步，进行局部多项式拟合，但权数用$\{\delta_iW_{ni}(x)\}$.重复s次后就可得到稳健估计。

由于稳健权数$\delta_i = K(\hat{\varepsilon}_i/(6\hat{\sigma}))$可将异常值排除在外，并且初始残差大（小）的观测值在下一次局部多项式回归中的权数就小（大）。因而，重复几次后就可将异常值不断地排除在外，并最终得到稳健的估计。

Cleveland(1979)推荐$p=1, s=3$。

2.7 本章思考题与综合练习题

一、思考题

1.理解非线性普通最小二乘法的方法原理和估计模型参数的步骤。

2.在采用非线性普通最小二乘法时，为什么给定不同的参数估计值的初值经常会得到不同的估计结果？

3.为什么许多经济问题需要建立变参数计量经济学模型？何谓确定性变参数模型？何谓随机变参数模型？

4.为什么说时间序列分析模型没有揭示经济因素之间的因果关系？

5.画出逻辑增长曲线和龚珀兹增长曲线，比较二者的异同；分别指出一些服从逻辑增长曲线和龚珀兹增长曲线的经济指标，描出它们的增长曲线。

6.理解逻辑增长曲线和龚珀兹增长曲线模型的参数估计方法。

7.熟悉各种常见的时间序列分析模型的数学形式。

8.使用时间序列分析软件对主要的宏观经济指标序列进行识别，判断它们属于哪一类时间序列分析模型。

9.理解关于稳定序列、单整和协整的概念，以及研究变量之间的协整关系对于建立计量经济学模型的意义。

10.从概念上解释：为什么以不变价格计算的消费额为1阶单整，而以当年价格计算的消费额为2阶单整？为什么消费额和收入之间可能产生协整，

而消费额和储蓄余额之间不可能产生协整？为什么选择收入作为消费的解释变量是合理的，而选择储蓄余额作为消费的解释变量是不合理的？

11.熟悉ADF检验的步骤，选择一个宏观经济指标序列完成检验的全过程。

12.理解误差修正模型的形式。

13.理解无参数回归模型的概念、权函数估计和最小二乘估计的基本原理。

14.分析书中的实例，思考为什么无参数回归模型比线性回归模型有较好的拟合效果。

二、综合练习题

1.用例2.1.1的问题和数据，分别采用线性模型和非线性模型的方法估计模型，比较结果。如果出现矛盾，试分析原因。

2.以我国进出口总额为被解释变量，国内生产总值为解释变量，建立(2.2.6)或(2.2.7)形式的确定性变参数模型；或者以我国粮食产量为被解释变量，播种面积为解释变量，建立(2.2.6)或(2.2.7)形式的确定性变参数模型；并对结果进行分析。

3.选择一个简单的实际问题，例如中国城镇居民消费，建立误差修正模型，完成单整检验、双变量协整检验以及建立模型的全过程。

第三章 估计方法非经典的计量经济学问题

在第二章中，主要介绍了经典线性计量经济学模型在模型结构形式方面的一些扩展。本章中，将主要讨论在模型估计方法方面的一些扩展，其中也包括某些虽属于经典的方法，但是在一般以经典线性计量经济学模型为主要内容的中级教科书中未作为重点，例如最大似然方法、贝叶斯估计等。

3.1 经典线性计量经济学模型的最大似然估计

在本书后续内容中，最大似然方法(Maximum Likelihood，简称ML，也称最大或然法)被广泛应用。正如在绪论中谈及的，最大似然法是不同于最小二乘法的另一种参数估计方法，是从最大似然原理出发发展起来的其它估计方法的基础。虽然在经典线性计量经济学模型中其应用没有最小二乘法普遍，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母体参数的内在机理。计量经济学理论的发展，更多地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。所以，在本节中，首先对经典线性计量经济学模型的最大似然估计作一介绍。

3.1.1 经典线性单方程计量经济学模型的最大似然估计

对于最小二乘法，当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得模型能最好地拟合样本数据。而对于最大或然法，当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大。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。通过似然函数极大化以求得总体参数估计量的方法被称为极大似然法。

1.一元线性模型的最大似然估计

对于一元线性回归模型：

$$y_i = \beta_0 + \beta_1 x_i + \mu_i \quad i = 1, 2, \dots n \tag{3.1.1}$$

$$\begin{array}{l} \mathrm{E}(\mu_i) = 0 \\ \mathrm{Var}(\mu_i) = \sigma_\mu^2 \end{array}$$

$$\mu_i \sim N(0, \sigma_\mu^2)$$

随机抽取n组样本观测值$y_i, x_i (i = 1, 2, \dots n)$，假如模型的参数估计量已经求得到，为$\hat{\beta}_0$和$\hat{\beta}_1$，那么$y_i$服从如下的正态分布：

$$y_i \sim N(\hat{\beta}_0 + \hat{\beta}_1 x_i, \sigma_\mu^2)$$

于是，y_i的概率函数为

$$P(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma_\mu^2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2} \quad i = 1, 2, \dots, n$$

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：

$$\begin{aligned} L(\hat{\beta}_0, \hat{\beta}_1, \sigma_\mu^2) &= P(y_1, y_2, \cdots, y_n) \\ &= \tfrac{1}{(2\pi)^{\frac{n}{2}} \sigma_\mu^n} e^{-\frac{1}{2\sigma_\mu^2}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2} \end{aligned} \tag{3.1.2}$$

将该似然函数极大化，即可求得到模型参数的极大似然估计量。

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

$$\begin{aligned} L^* &= \ln(L) \\ &= -n\ln(\sqrt{2\pi}\sigma_\mu) - \tfrac{1}{2\sigma_\mu^2}\Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{aligned} \tag{3.1.3}$$

对L^*求极大值，等价于对$\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2$ 求极小值。$\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2$极小值的条件为：

$$\left\{\begin{array}{l}\frac{\partial}{\partial\hat\beta_0}\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2=0\\ \frac{\partial}{\partial\hat\beta_1}\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2=0\end{array}\right.$$

解得模型的参数估计量为：

$$\left\{\begin{array}{l}\hat\beta_0=\frac{\Sigma x_i^2\Sigma y_i-\Sigma x_i\Sigma y_ix_i}{n\Sigma x_i^2-(\Sigma x_i)^2}\\ \hat\beta_1=\frac{n\Sigma y_ix_i-\Sigma y_i\Sigma x_i}{n\Sigma x_i^2-(\Sigma x_i)^2}\end{array}\right.$$

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。但是，随机误差项的方差的估计量是不同的。解似然方程

$$\frac{\partial}{\partial\sigma_\mu^2}L^*=-textstyle\frac{n}{2\sigma_\mu^2}+\tfrac{1}{2\sigma_\mu^4}\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2=0$$

即可得到σ_μ^2的最大似然估计量为：

$$\hat\sigma_\mu^2=\tfrac{1}{n}\Sigma(y_i-\hat\beta_0-\hat\beta_1x_i)^2 \tag{3.1.4}$$

至此，完成了用最大似然法估计参数的任务。

2.多元线性模型的最大似然估计

对于多元线性回归模型

$$\mathbf{Y}=\mathbf{X}\mathrm{B}+\mathrm{N} \tag{3.1.5}$$

由于

$$\mu_i\sim N(0,\sigma_\mu^2)$$

所以

$$y_i\sim N(\mathbf{X_i}\hat{\mathrm{B}},\sigma_\mu^2)$$

其中

$$\mathbf{X_i}=\left[\begin{array}{ccccc}1 & x_{1i} & x_{2i} & \cdots & x_{ki}\end{array}\right]$$

Yœ`œn组样本观测值的联合概率为：

$$
\begin{aligned}
L(\hat{\mathrm{B}}, \sigma_\mu^2) &= P(y_1, y_2, \cdots, y_n) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_\mu^n} e^{-\frac{1}{2\sigma_\mu^2}\Sigma(y_i-(\hat{\beta}_0+\hat{\beta}_1 x_{1i}+\hat{\beta}_2 x_{2i}+\cdots+\hat{\beta}_k x_{ki}))^2} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_\mu^n} e^{-\frac{1}{2\sigma_\mu^2}(\mathbf{Y}-\mathbf{X}\hat{\mathrm{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathrm{B}})}
\end{aligned}
$$

这就是变量$\mathbf{Y}$的似然函数。对数似然函数为：

$$
\begin{aligned}
L^* &= Ln(L) \\
&= -\frac{n}{2}\ln(\sqrt{2\pi}\sigma_\mu^2) - \frac{1}{2\sigma_\mu^2}(\mathbf{Y}-\mathbf{X}\hat{\mathrm{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathrm{B}})
\end{aligned} \tag{3.1.6}
$$

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

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

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

$$\hat{\mathrm{B}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y} \tag{3.1.7}$$

显然，其结果与参数的普通最小二乘估计是相同的。

这里仅介绍了模型满足基本假使情况下似然函数和对数似然函数。关于模型违背基本假设，即存在异方差性和序列相关性情况下似然函数和对数似然函数以及极大化，将在本书第五章中介绍。

3.1.2 经典线性联立方程计量经济学模型的有限信息最大似然估计

有限信息最大似然法(LIML,Limited Information Maximum Likelihood) 是一种以最大似然为准则、通过对简化式模型进行最大似然估计，以得到结构方程参数估计量的联立方程模型的单方程估计方法。由Anderson和Rubin于1949年提出，早于两阶段最小二乘法。适用于恰好识别和过度识别结构方程的估计。在该方法中，以下两个概念是重要的：

一是这里的“有限信息”指的是每次估计只考虑一个结构方程的信息，而没有考虑模型系统中其它结构方程的信息；

二是这里的“最大似然法”是针对结构方程中包含的内生变量的简化式模型的，即应用最大似然法求得的是简化式参数估计量，而不是结构式参数估计量。

对于联立方程模型

$$\mathrm{B}\mathbf{Y} + \Gamma\mathbf{X} = \mathrm{N} \tag{3.1.8}$$

的每一个结构方程，例如第1个方程，可以写成如下形式：

$$Y_1=\beta_{12}Y_2+\beta_{13}Y_3+\cdots+\beta_{1g_1}Y_{g_1}+\gamma_{11}X_1+\gamma_{12}X_2+\cdots+\gamma_{1k_1}X_{k_1}+\mathrm{N}_1 \quad (3.1.9)$$

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：

$$Y_1=(\mathbf{Y}_0,\mathbf{X}_0)\begin{pmatrix}\mathrm{B}_\mathbf{0}\\ \Gamma_\mathbf{0}\end{pmatrix}+\mathrm{N}_1 \quad (3.1.10)$$

其中

$$\mathbf{Y}_0=\begin{bmatrix}Y_2 & Y_3 & \cdots & Y_{g_1}\end{bmatrix}=\begin{bmatrix}y_{21} & y_{31} & \cdots & y_{g_11}\\ y_{22} & y_{32} & & y_{g_12}\\ \vdots & \vdots & & \vdots\\ y_{2n} & y_{3n} & & y_{g_1n}\end{bmatrix}$$

$$\mathbf{X}_0=\begin{bmatrix}X_1 & X_2 & \cdots & X_{k_1}\end{bmatrix}=\begin{bmatrix}x_{11} & x_{21} & \cdots & x_{k_11}\\ x_{12} & x_{22} & & x_{k_12}\\ \vdots & \vdots & & \vdots\\ x_{1n} & x_{2n} & & x_{k_1n}\end{bmatrix}$$

$$\mathrm{B}_0=\begin{bmatrix}\beta_{12}\\ \beta_{13}\\ \vdots\\ \beta_{1g_1}\end{bmatrix}\quad \Gamma_0=\begin{bmatrix}\gamma_{11}\\ \gamma_{12}\\ \vdots\\ \gamma_{1k_1}\end{bmatrix}\quad Y_1=\begin{bmatrix}y_{11}\\ y_{12}\\ \vdots\\ y_{1n}\end{bmatrix}\quad \mathrm{N}_1=\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \vdots\\ \mu_{1n}\end{bmatrix}$$

n为样本容量。

1.简化式模型的最大似然估计

将联立方程模型(6.4.8)中的第1个结构方程(6.4.10)改写成如下形式：

$$(\mathbf{Y}_0^1,\mathbf{X}_0)\begin{pmatrix}\mathrm{B}_0^1\\ \Gamma_\mathbf{0}\end{pmatrix}+\mathrm{N}_1=0 \quad (3.1.11)$$

其中

$$\mathbf{Y}_0^1=(Y_1,\mathbf{Y}_0)$$
$$\mathrm{B}_0^1=\begin{pmatrix}-1\\ \mathrm{B}_0\end{pmatrix}$$

该方程包含的内生变量的简化式模型为：

$$\mathbf{Y}_0^1 = \mathbf{X}\Pi_0^1 + \mathrm{E}_0^1 \tag{3.1.12}$$

或

$$\mathbf{Y}_0^1 = \mathbf{X}_0\Pi_{01}^1 + \mathbf{X}_0^*\Pi_{02}^1 + \mathrm{E}_0^1 \tag{3.1.13}$$

其中$\mathbf{X}_0^*$为结构方程中未包含的先决变量。

根据上面介绍的建立似然函数的原理，建立关于简化式模型(3.1.128)的对数似然函数：

$$LnL(\mathbf{Y}_0^1) = c + \frac{n}{2}Ln\left|\Omega_0^{-1}\right| - \frac{1}{2}tr\Omega_0^{-1}(\mathbf{Y}_0^1 - \mathbf{X}\Pi_0^1)'(\mathbf{Y}_0^1 - \mathbf{X}\Pi_0^1) \tag{3.1.14}$$

其中Ω_0是简化式模型随机误差项E_0^1的方差-协方差矩阵，n为样本容量，c为常数。

对似然函数(3.1.1410)极大化等价于广义方差

$$(\mathbf{Y}_0^1 - \mathbf{X}\Pi_0^1)'(\mathbf{Y}_0^1 - \mathbf{X}\Pi_0^1)$$

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

2.参数关系体系的应用

用B_0^1乘(6.4.13)两边：

$$\mathbf{Y}_0^1\mathrm{B}_0^1 = \mathbf{X}_0\Pi_{01}^1\mathrm{B}_0^1 + \mathbf{X}_0^*\Pi_{02}^1\mathrm{B}_0^1 + \mathrm{E}_0^1\mathrm{B}_0^1 \tag{3.1.15}$$

比较(6.4.11)与(6.4.15)，得到：

$$\begin{aligned} \Pi_{02}^1\mathrm{B}_0^1 &= 0 \\ \Pi_{01}^1\mathrm{B}_0^1 &= -\Gamma_0 \end{aligned} \tag{3.1.16}$$

即为结构式模型与简化式模型的参数关系体系。

如果结构方程是恰好识别的，即有$k - k_1 = g_1 - 1$，(6.4.16)中前一个矩阵方程表示$(g_1 - 1)$个方程组成的方程组，在已知$\hat{\Pi}_{02}^1$时可以解出唯一确定的$\hat{\mathrm{B}}_0^1$。然后再由(6.4.16)中的后一组方程，求得$-\hat{\Gamma}_0$。把计算得到的结构参数估计量称为有限信息最大似然法参数估计量。

如果结构方程是过度识别的，即有$k - k_1 > g_1 - 1$，(6.4.16)中前一个矩阵方程表示的方程组

$$\Pi_{02}^1\mathrm{B}_0^1 = 0$$

构成了似然函数(6.4.14)极大化的约束条件。进一步求解参数的LIML估计量的数学过程比较复杂，在此不作介绍。有兴趣的读者，尤其是有兴趣于理论计量经济学的学生可以参考《经济计量学》（邹至庄著，郑宗成译，中国友谊出版社，1988年3月）第258-268页。

3.1.3 经典线性联立方程计量经济学模型的完全信息最大似然估计

联立方程计量经济学模型的系统估计方法是相对于单方程估计方法而言。单方程方法每次只对一个结构方程进行估计，利用了有限信息，对于没有包含在所估计结构方程中的变量的样本数据信息，只是部分地利用了，而对于方程之间的关系信息，则完全没有利用。系统估计方法，正是针对单方程方法的问题提出来的，它同时估计全部结构方程，利用了模型系统的全部信息。因此系统估计方法的参数估计量具有良好的统计特性。也正因为此，系统估计方法也相当复杂，目前商用计量经济学软件中几乎没有包含系统估计方法的，因而在实际模型研究中也很少应用。这里只介绍完全信息最大似然法的方法原理与思路，对于了解完整的最大似然法理论方法体系是重要的。

1.联立方程模型随机误差项方差—协方差矩阵

联立方程模型

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

可以改写为

$$\mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathrm{N}} \tag{3.1.17}$$

其中

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{bmatrix} \quad Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{bmatrix}$$

$$Y_i = \mathbf{Z}_i\Delta_i + \tilde{\mathrm{N}}_i \tag{3.1.18}$$

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_0^i & \mathbf{X}_0^i \end{pmatrix}$$

$$\Delta_i = \begin{pmatrix} \mathrm{B}_0^i \\ \Gamma_0^i \end{pmatrix}$$

每个符号的含义基本同前，只是用上标i表示第i个方程。

对于(6.4.17)中的随机误差项$\tilde{\mathrm{N}}$，为了使问题适当简单，作如下假设：

(1) 对于一个结构方程的随机误差项，在不同样本点之间，具有同方差性和序列不相关性。即

$$Cov(\tilde{\mathrm{N}}_i) = \sigma_{ii}^2\mathbf{I}$$

(2)对于不同结构方程的随机误差项之间，具有且仅具有同期相关性。即

$$Cov(\tilde{N}_i, \tilde{N}_j) = \sigma_{ij}^{\mathbf{I}}$$

于是，联立方程模型系统随机误差项方差—协方差矩阵为：

$$Cov(\tilde{N}) = \begin{bmatrix} \sigma_{11}^2\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1g}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}^2\mathbf{I} & \cdots & \sigma_{2g}\mathbf{I} \\ \vdots & & & \\ \sigma_{g1}\mathbf{I} & \sigma_{g2}\mathbf{I} & & \sigma_{gg}^2\mathbf{I} \end{bmatrix}$$

$$= \Sigma \otimes \mathbf{I}$$
$$= \Omega$$

其中⊗表示“直积”，即用符号后面的矩阵去乘符号前面矩阵的每个元素。协方差矩阵Ω是由(g×g)个子矩阵组成，每个子矩阵都是一个主对角阵，且主对角线元素相同。如果放弃两条假设，每个子矩阵就不是一个主对角阵，且主对角线元素也不相同。

2.完全信息最大或然法

完全信息最大似然法(Full Information Maximum Likelihood, FIML)是一种已有实际应用的联立方程模型的系统估计方法。Rothenberg和Leenders于1964年提出一个线性化的FIML估计量。FIML是ML的直接推广，是在已经得到样本观测值的情况下，使整个联立方程模型系统的似然函数达到最大以得到所有结构参数的估计量。

对于联立方程模型的变换形式(6.4.17)

$$\mathbf{Y} = \mathbf{Z}\Delta + \tilde{N}$$

仍然维持关于随机项的假设

$$\mathrm{E}(\tilde{N}) = 0$$
$$\mathrm{Cov}(\tilde{N}) = \Sigma \otimes \mathbf{I} = \Omega$$
$$\tilde{N} \sim 正态(\mathbf{0}, \Sigma \otimes \mathbf{I})$$

$$\mathbf{Y}œq,$$

$$L(\mathbf{Y}) = \frac{1}{(2\pi)^{gn/2}\left|\Sigma \otimes \mathbf{I}\right|^{-\frac{1}{2}}} \left|\frac{\partial\,\tilde{N}}{\partial\,\mathbf{Y}}\right| e^{-\frac{1}{2}}(\mathbf{Y} - \mathbf{Z}\Delta)'(\Sigma^{-1} \otimes \mathbf{I})(\mathbf{Y} - \mathbf{Z}\Delta)$$

$$= \frac{1}{(2\pi)^{gn/2}\left|\Sigma\otimes\mathbf{I}\right|^{-\frac{1}{2}}}\left|\mathrm{B}\right|^{n} e^{-\frac{1}{2}(\mathbf{Y}-\mathbf{Z}\Delta)'(\Sigma^{-\mathbf{1}}\otimes\mathbf{I})(\mathbf{Y}-\mathbf{Z}\Delta)} \tag{3.1.19}$$

$$\mathbf{Y}œq,$$

$$\ln L(\mathbf{Y}) = -\frac{gn}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\Sigma\otimes\mathbf{I}\right| + n\ln\left|\mathrm{B}\right|$$

$$-\frac{1}{2}(\mathbf{Y}-\mathbf{Z}\Delta)'(\Sigma^{-1}\otimes\mathbf{I})(\mathbf{Y}-\mathbf{Z}\Delta) \tag{3.1.20}$$

其中
$-\frac{1}{2}\ln\left|\Sigma\otimes\mathbf{I}\right| = -\frac{1}{2}\ln\left|\Sigma\right|^{n} = -\frac{n}{2}\ln\left|\Sigma\right|$
在无约束情况下

$$-\frac{1}{2}(\mathbf{Y}-\mathbf{Z}\Delta)'(\Sigma^{-1}\otimes\mathbf{I})(\mathbf{Y}-\mathbf{Z}\Delta)$$

$$= -\frac{1}{2}\sum_{h=1}^{g}\sum_{h'=1}^{g}\sigma^{hh'}(Y_h-\mathbf{Z}_h\delta_h)'(Y_h'-\mathbf{Z}_h'\delta_h')$$

其中$\sigma^{hh'}$为协方差的逆矩阵的元素，δ_h为Δ的元素，未出现的$\sigma_{hh'}$为协方差矩阵的元素。即

$$\Sigma = (\sigma_{hh'}) \quad \Sigma^{-1} = (\sigma^{hh'}) \quad \Delta = (\delta_h)$$

于是，对数似然函数对于协方差逆矩阵的元素取极大值的一阶条件为

$$\frac{\partial\ \ln L(\mathbf{Y})}{\partial\sigma^{hh'}} = \frac{n}{2}\sigma_{hh'} - \frac{1}{2}(Y_h-\mathbf{Z}_h\delta_h)'(Y_h'-\mathbf{Z}_h'\delta_h') = 0$$

得到协方差矩阵的元素的FIML估计量为：

$$\hat{\Sigma} = (\hat{\sigma}_{hh'})$$

$$\hat{\sigma}_{hh'} = \frac{1}{n}(Y_h-\mathbf{Z}_h\delta_h)'(Y_h'-\mathbf{Z}_h'\delta_h') \quad h,h' = 1,2,\cdots,g$$

$$\hat{\Sigma}\otimes\mathbf{I} = \frac{1}{n}(\mathbf{Y}-\mathbf{Z}\hat{\Delta})(\mathbf{Y}-\mathbf{Z}\hat{\Delta})' \tag{3.1.21}$$

将(6.4.21)代入(6.4.20),得到

$$\ln L(\mathbf{Y}) = -\frac{gn}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma \otimes \mathbf{I}| + n\ln|\mathrm{B}| - \frac{n}{2}$$

于是，对于待估计参数Δ取极大值的条件为：

$$\frac{\partial\ \ln L(\mathbf{Y})}{\partial\ \Delta} = -\frac{n}{2}\frac{\partial\ \ln|\Sigma|}{\partial\ \Delta} + n\frac{\partial\ \ln|\mathrm{B}|}{\partial\ \Delta} = 0 \tag{3.1.22}$$

求解该方程系统，即可得到结构参数的FIML估计量。

但是(6.4.22)是个非线性方程系统，求解相当复杂，需要采用Newton迭代方法和阻尼Newton迭代方法等。有兴趣的读者可以参考《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年）第256-262页。在具体的应用模型中，可以寻求简化的算法。例如，如果B是三角阵、Σ是对角阵，可以推出：

$$\hat{\Delta}_h = (\mathbf{Z'}_h\mathbf{Z}_h)^{-1}\mathbf{Z'}_hY_h \quad h = 1, 2, \ldots, g$$

与OLS估计量等价。

3.2 经典线性计量经济学模型最小二乘估计的扩展

在经典线性计量经济学模型中，最小二乘方法是普遍应用的参数估计方法。本节将介绍几种在有关经典线性计量经济学模型的教科书中较介绍的从最小二乘原理出发的估计方法，包括可行的广义最小二乘估计（FGLS, Feasible Generalized Least Squares）、分部回归方法(Partitioned Regression)、偏回归方法(Partial Regression) 和交叉估计方法(Across Regression)，作为对经典的普通最小二乘法的扩展。

3.2.1 经典单方程线性计量经济学模型可行的广义最小二乘估计

1.广义最小二乘估计

广义最小二乘估计是经典线性计量经济学模型违背基本假设、且随机误差项方差—协方差矩阵未知情况下的一种参数估计方法，在一般中级计量经济学教科书中对它的基本内容都有所涉及。

对于模型

$$\mathbf{Y} = \mathbf{X}\mathrm{B} + \mathrm{N} \tag{3.2.1}$$

如果存在序列相关，同时存在异方差，即有

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

设

$$\Omega = \mathbf{DD}'$$

用$\mathbf{D^{-1}}$左乘(2.7.3)两边，得到一个新的模型：

$$\mathbf{D}^{-1}\mathbf{Y} = \mathbf{D}^{-1}\mathbf{X}\mathrm{B} + \mathbf{D}^{-1}\mathrm{N} \tag{3.2.2}$$

即

$$\mathbf{Y}^* = \mathbf{X}^*\mathrm{B} + \mathrm{N}^*$$

该模型具有同方差性和随机误差项互相独立性。因为

$$\mathrm{E}(\mathrm{N}^*\mathrm{N}^*\prime) = \mathrm{E}(\mathbf{D}^{-1}\mathrm{NN}'\mathbf{D}^{-1}\prime)$$

$$
\begin{aligned}
&= \mathbf{D}^{-1}\,\mathrm{E}(\mathrm{NN}')\mathbf{D}^{-1}\prime \\
&= \mathbf{D}^{-1}\sigma^2\mathbf{W}\mathbf{D}^{-1}\prime \\
&= \mathbf{D}^{-1}\sigma^2\mathbf{DD}'\mathbf{D}^{-1}\prime \\
&= \sigma^2\mathbf{I}
\end{aligned}
$$

于是，可以用普通最小二乘法估计模型(6.4.2)，得到参数估计量为：

$$\hat{\mathrm{B}} = (\mathbf{X}^*\prime\mathbf{X}^*)^{-1}\mathbf{X}^*\prime\mathbf{Y}^*$$

$$
\begin{aligned}
&= (\mathbf{X}'\mathbf{D}^{-1}\prime\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^{-1}\prime\mathbf{D}^{-1}\mathbf{Y} \\
&= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}
\end{aligned} \tag{3.2.3}
$$

这就是原模型(6.5.1)的广义最小二乘估计量，是无偏的、有效的估计量。

如果矩阵Ω是已知的，利用(6.4.4)可以计算参数估计值。

2.可行广义最小二乘估计

如果(6.4.4)中矩阵Ω是未知的，为了使广义最小二乘估计成为可行，就需要利用Ω的估计值$\hat{\Omega}$代替实际的Ω。于是得到

$$\hat{\mathrm{B}} = (\mathbf{X}'\hat{\Omega}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\hat{\Omega}^{-\mathbf{1}}\mathbf{Y} \tag{3.2.4}$$

该估计量称为FGLS估计量。

一般地，Ω可能包含某些参数θ，即

$$\Omega = \Omega(\theta)$$

例如，在模型存在一阶自相关时，Ω仅包含一个未知道参数ρ，即

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

如果能够得到Ω的估计值$\hat{\Omega}$为

$$\hat{\Omega} = \Omega(\hat{\theta})$$

且$\hat{\theta}$是θ的一致性估计量。那么，可以证明(6.4.5)所表示的FGLS估计量渐近等价于(6.4.4)所表示的GLS估计量。

在一些计量经济学教科书中，建议采用如下方法得到矩阵Ω的估计值$\hat{\Omega}$，即对原模型首先采用普通最小二乘法，得到随机误差项的近似估计量$\tilde{e}_1, \tilde{e}_2, \cdots, \tilde{e}_n$，以此构成矩阵$\Omega$的估计值：

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

将该估计值代入(6.4.5)得到原模型的参数估计量。请读者自己证明，该估计量与GLS估计量是否具有渐近等价性？

3.2.2 经典单方程线性计量经济学模型的分部回归估计

1.分部回归估计

对于模型(6.5.1)，将解释变量分为两部分$\mathbf{X_1}$和$\mathbf{X_2}$，对应的参数也分为两部分。有

$$\mathbf{Y} = \mathbf{X}_1\mathrm{B}_1 + \mathbf{X}_2\mathrm{B}_2 + \mathrm{N} \tag{3.2.5}$$

在满足解释变量与随机误差项不相关的情况下，可以写出关于参数估计量的方程组；

$$\begin{bmatrix} \mathbf{X'}_1\mathbf{Y} \\ \mathbf{X'}_2\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{x'}_1\mathbf{x}_1 & \mathbf{x'}_1\mathbf{x}_2 \\ \mathbf{x'}_2\mathbf{x}_1 & \mathbf{x'}_2\mathbf{x}_2 \end{bmatrix} \begin{bmatrix} \hat{\mathrm{B}}_1 \\ \hat{\mathrm{B}}_2 \end{bmatrix} \tag{3.2.6}$$

于是

$$\begin{aligned} \hat{\mathrm{B}}_1 &= (\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1\mathbf{Y} - (\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1\mathbf{X}_2\hat{\mathrm{B}}_2 \\ &= (\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1(\mathbf{Y} - \mathbf{X}_2\hat{\mathrm{B}}_2) \end{aligned} \tag{3.2.7}$$

显然，如果存在$\mathbf{X'}_1\mathbf{X}_2 = \mathbf{0}$，则有

$$\hat{\mathrm{B}}_1 = (\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1\mathbf{Y} \tag{3.2.8}$$

这就是仅以$\mathbf{X_1}$作为解释变量时的参数估计量。
同样，如果存在$\mathbf{X'}_2\mathbf{X}_1 = \mathbf{0}$，则有

$$\hat{\mathrm{B}}_2 = (\mathbf{X'}_2\mathbf{X}_2)^{-1}\mathbf{X'}_2\mathbf{Y} \tag{3.2.9}$$

根据以上推导，可以得到如下结论：如果两组解释变量是正交的，那么相应的参数估计量可以分别在仅以一部分变量为解释变量的情况下加以估计。再进一步，如果模型的所有解释变量都是互相正交的，那么可以将多元线性模型化成若干个一元模型加以分别估计。这就是所谓的“分部回归估计”。

当然，上述两组解释变量是完全正交的情况在实际中是很难发现的，或者说是不可能出现的。所以，“分部回归估计”并没有实际意义。在应用主分量法时，所选择的主分量是互相正交的，那么对多个主分量回归可以化为对每个主分量分别回归。

2.由分部回归引起的思考

在经典计量经济学模型中，读者已经很熟悉如下问题：一是当模型中某些变量被检验为不显著时，就要将这些变量从模型中剔除；二是当发现解释变量出现多重共线性时，一个有效的处理方法是从模型中去掉产生多重共线性的部分变量。这些无疑是可行的。但是，值得注意的是，由于模型原有的变量之间并不是正交的，所以当剔除部分变量后，所保留变量的参数估计值将发生变化，它所反映的不再仅仅是该变量与被解释变量之间的关系。这一点从分部回归中得到了证明。

于是，它提醒人们，不能轻易剔除“不显著”的变量，除非它的显著性低到对被解释变量几乎没有影响。更不能轻易采用剔除变量的方法消除多重共线性，应该多采用差分法或主分量法等。

3.2.3 经典单方程线性计量经济学模型的偏回归估计

将(6.4.8) 代入(6.4.7)的第2组方程

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

中，得到：

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

可导出：

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

于是得到：

$$(\mathbf{X'}_2(\mathbf{I} - \mathbf{X}_1(\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1)\mathbf{X}_2)^{-1}\mathbf{X'}_2(\mathbf{I} - \mathbf{X}_1(\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1)\mathbf{Y} = \hat{\mathrm{B}}_\mathbf{2} \quad (3.2.10)$$

令$(\mathbf{I} - \mathbf{X}_1(\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1) = \mathbf{M}$，有

$$\hat{\mathrm{B}}_\mathbf{2} = (\mathbf{X'}_2\mathbf{M}\mathbf{X}_2)^{-1}(\mathbf{X'}_2\mathbf{M}\mathbf{Y}) \quad (3.2.11)$$

利用$\mathbf{M}$为等幂矩阵的性质，$\mathbf{M} = \mathbf{M'M}$，(6.4.12)可以写成为：

$$\begin{aligned}\hat{\mathrm{B}}_\mathbf{2} &= (\mathbf{X'}_2\mathbf{M'M}\mathbf{X}_2)^{-1}(\mathbf{X'}_2\mathbf{M'M}\mathbf{Y}) \\ &= (\mathbf{X}_2^{*\prime}\mathbf{X}_2^*)^{-1}(\mathbf{X}_2^{*\prime}\mathbf{Y}^*)\end{aligned} \quad (3.2.12)$$

其中$\mathbf{X}_2^* = \mathbf{M}\mathbf{X}_2, \mathbf{Y}^* = \mathbf{M}\mathbf{Y}$。因为

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

所以，它实际上是当$\mathbf{Y}$仅以$\mathbf{X_1}$为解释变量所建立的模型的残差。

暂时放下(6.4.12)和(6.4.13)，看另外一个问题：如果以$\mathbf{X_2}$中的每个变量分别作为被解释变量，以$\mathbf{X_1}$中的变量作为它们的解释变量逐个进行回归估计，然后再写成矩阵的形式，有

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

残差为：

$$\begin{aligned}\mathbf{X_2} - \hat{\mathbf{X}}_\mathbf{2} &= \mathbf{X_2} - \mathbf{X}_1((\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1\mathbf{X}_\mathbf{2}) \\ &= (\mathbf{I} - \mathbf{X}_1(\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1)\mathbf{X_2} \\ &= \mathbf{M}\mathbf{X_2} \\ &= \mathbf{X_2^*}\end{aligned} \quad (3.2.13)$$

可见，$\mathbf{X}_2^*$ 实际上是以$\mathbf{X}_2$中的每个变量分别对$\mathbf{X}_1$ 中的变量进行回归所建立的模型的残差。

于是，得到了一个十分重要的结论：由(6.4.13)所表示的原模型(6.5.1)中一部分解释变量$\mathbf{X}_2$的参数的估计量，实际上等于以$\mathbf{Y}^*$为被解释变量、$\mathbf{X}_2^*$为解释变量所建立的模型的参数估计量，而$\mathbf{Y}^*$是当原模型的被解释变量$\mathbf{Y}$ 仅以另一部分解释变量$\mathbf{X}_1$单独作为解释变量所建立的模型的残差，$\mathbf{X}_2^*$是以$\mathbf{X}_2$中的每个变量分别对$\mathbf{X}_1$进行回归所建立的模型的残差。

类似地，对于原模型中一部分解释变量$\mathbf{X}_1$的参数的估计量，也可以得到同样的结论。当然，如果已经得到$\hat{\mathrm{B}}_2$，也可以通过(6.4.8)求解$\hat{\mathrm{B}}_1$。

正是因为上述过程中反复出现了"偏"的含义，所以用(6.4.13)所揭示的方法估计原模型参数估计量被称为"偏回归估计"。有时甚至也将多元线性回归模型中的系数统称为"偏回归系数"。

偏回归估计除了在理论上的意义外，当计算技术还不发达的时候，采用这种方法可以减少估计模型的计算工作量。例如，将模型的4个解释变量分成2组，每组包含2个变量，就可以将一个4元回归问题变成4个二元回归问题进行估计计算。

例3.2.1 以我国国内生产总值(GDP)为被解释变量，居民消费总额(COM)、政府消费额(GOV)、投资总额(INV)为解释变量建立线性模型，用以分析各项最终需求与用生产法计算的国内生产总值之间的统计关系。样本数据见表3.2.1。

采用OLS法估计模型，得到：

$$\begin{aligned}&GDP1_t = \underset{(-4.6313)}{-521.7277} + \underset{(20.6161)}{1.2040}\,COM_t + \underset{(8.0895)}{2.4328}\,GOV_t + \underset{(4.0007)}{0.3691}\,INV_t\\&R^2 = 0.9999 \qquad F = 53321 \qquad D.W. = 2.26\end{aligned} \tag{3.2.14}$$

其中GDP1表示GDP的估计值。

让GDP对常数项和INV回归，得到如下回归方程：

$$\begin{aligned}&GDP2_t = \underset{(2.0821)}{974.0084} + \underset{(65.2919)}{2.4817}\,INV_t\\&R^2 = 0.9958 \qquad F = 4263 \qquad D.W. = 1.09\end{aligned} \tag{3.2.15}$$

其中GDP2表示GDP在该情况下的估计值。令

$$\mathrm{E}_t = GDP_t - GDP2_t \qquad t = 1978, 1979, \cdots, 1997$$

让COM对常数项和INV回归，得到如下回归方程：

$$\begin{aligned}&COM1_t = \underset{(2.0460)}{664.5436} + \underset{(44.5615)}{1.1761}\,INV_t\\&R^2 = 0.9910 \qquad F = 1986 \qquad D.W. = 0.82\end{aligned} \tag{3.2.16}$$

其中COM1表示COM在该情况下的估计值。令

表 3.2.1: 例3.2.1 有关数据表单位：亿元

年份	GDP	COM	GOV	INV
1978	3624	1759	480	1378
1979	4038	2005	614	1474
1980	4518	2317	659	1590
1981	4862	2604	705	1581
1982	5293	2868	770	1760
1983	5935	3183	838	2005
1984	7171	3675	1020	2469
1985	8964	4589	1184	3386
1986	10202	5175	1367	3846
1987	11962	5961	1490	4322
1988	14928	7633	1727	5495
1989	16909	8524	2033	6095
1990	18548	9113	2252	6444
1991	21618	10316	2830	7517
1992	26638	12460	3492	9636
1993	34634	15682	4499	14998
1994	46759	21230	5986	19260
1995	58478	27839	6691	23877
1996	67885	33188	7852	26867
1997	74772	36118	8650	28564

$$E1_t = COM_t - COM1_t \qquad t = 1978, 1979, \cdots, 1997$$

让GOV对常数项和INV回归，得到如下回归方程：

$$\begin{aligned} GOV1_t &= \underset{(4.5337)}{285.9544} + \underset{(55.8796)}{0.2864}\, INV_t \\ R^2 &= 0.9943 \qquad F = 3122 \qquad D.W. = 1.46 \end{aligned} \tag{3.2.17}$$

其中GOV1表示GOV在该情况下的估计值。令让

$$E2_t = GOV_t - GOV1_t \qquad t = 1978, 1979, \cdots, 1997$$

让E对E1和E2回归，得到如下回归方程：

$$\begin{aligned} \hat{E}_t &= \underset{(21.8667)}{1.2040}\, E1_t + \underset{(8.5802)}{2.4328}\, E2_t \\ R^2 &= 0.9762 \qquad F = 738 \qquad D.W. = 2.26 \end{aligned} \tag{3.2.18}$$

比较(6.4.19)和(6.4.15)中的回归系数，E1与COM的回归系数完全相同，E2与GOV的回归系数完全相同。验证了偏回归估计方法。

但是，请读者注意，对方程(6.4.16)、(6.4.17)、(6.4.18)并没有进行检验，从显示的统计量看，至少存在一阶自相关。如果消除它们的自相关性，结果将发生变化。当采用广义差分法消除了(6.4.16)、(6.4.17)、(6.4.18) 的自相关性后，得到的E对E1和E2回归结果为：

$$\hat{E}_t = \underset{(11.8426)}{1.4123}\, E1_t + \underset{(5.5814)}{2.4031}\, E2_t + \underset{(1.4725)}{0.3788}\, AR(2)$$
$$R^2 = 0.9571 \qquad F = 167 \qquad D.W. = 1.64$$

其中系数与(6.4.19)略有不同。

所以，如果采用偏回归估计方法，对中间方程的不必进行检验。

3.2.4 经典单方程线性计量经济学模型的交叉回归估计

交叉估计也是对模型参数进行部分回归的一种估计方法，但是与上述部分回归估计不同的是，它将模型的参数按照其性质分类，然后分别用不同的样本观测值，包括被解释变量的样本观测值，估计各类参数。那么自然地，它只是相对于某类应用模型而言。下面将通过需求函数模型来介绍。

1.问题的提出

在需求函数模型中，解释变量一般为收入和价格，这两类变量对商品需求量的影响是不同的。按照第二章介绍的协整理论，商品需求量和收入为流量指标，一般情况下为一阶单整，它们之间可能存在协整关系，反映了二者之间的长期关系；而价格水平一般是0阶单整，它对商品需求量具有短期影响。从直观上也可以看出，收入对商品需求量具有长期影响，价格对商品需求量只具有短期影响。它们的参数分别属于长期弹性和短期弹性，具有不同的性质。而一般说来，时间序列数据适合于短期弹性的估计，截面数据适合于长期弹性的估计。所以用同一组样本数据同时估计需求函数模型的所有参数，在理论上是存在问题的。于是就提出了合并时间序列数据和截面数据的估计方法，即交叉估计方法。即用截面数据为样本估计模型中的一部分反映长期影响的参数，然后再用时间序列数据为样本估计模型中的另一部分反映短期影响的参数，分两阶段完成模型的估计。

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

观测值数据变化较小，不宜于揭示收入对需求量的长期影响。反过来，价格的时间序列数据适宜于揭示价格对需求量的短期影响。

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

2.估计方法

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

$$\ln q = \alpha_0 + \alpha_1 \ln I + \alpha_2 \ln p + \mu \tag{3.2.19}$$

其中q为商品需求量，I为收入，P为商品的价格。现有第T年的截面数据$q_j, I_j (j = 1, 2, \cdots, m)$，即将消费者按照收入分成m组。在这个截面上，认为价格是常数。于是(6.4.20)变为：

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

采用经典线性单方程模型的估计方法估计得到$\hat{\alpha}_1$。

当以时间序列数据为样本时，将模型写成：

$$\ln q_t = \alpha_0 + \alpha_1 \ln I_t + \alpha_2 \ln p_t + \mu_t \quad t = 1, 2, \cdots, T$$

此时认为$\hat{\alpha}_1$已知，令$y_t = \ln q_t - \hat{\alpha}_1 \ln I_t$，有

$$y_t = \alpha_0 + \alpha_2 \ln p_t + \mu_t \quad t = 1, 2, \cdots, T$$

采用经典线性单方程模型的估计方法估计得到$\hat{\alpha}_0, \hat{\alpha}_2$。连同前面的$\hat{\alpha}_1$，模型的全部参数得到估计。

关于交叉估计量的性质，以及交叉估计的问题，这里不再详细讨论。例如，选择第T年的截面数据，可以得到一个$\hat{\alpha}_1$，选择第T−1年的截面数据，也可以得到一个$\hat{\alpha}_1$，...于是带来了任意性。再如，在截面分析时，样本观测值取自分组人均数据，而在时序分析时，样本观测值一般取总体平均数据，这里存在一致性问题。所以，交叉估计作为一种实用方法，尚缺少理论计量经济学的支持。

3.3 经典线性计量经济学模型的贝叶斯估计

贝叶斯(Bayes)统计是由T.R.Bayes于19世纪创立的数理统计的一个重要分支，20世纪50年代，以H.Robbins为代表提出了在计量经济学模型估计中将经验贝叶斯方法与经典方法相结合，引起了广泛的重视，得到了

广泛的应用。贝叶斯估计对经典计量经济学模型估计方法的扩展在于，它不仅利用样本信息，同时利用非样本信息。贝叶斯方法的数学描述较为复杂，内容也很广泛，所以在一般中级水平的计量经济学教科书中不包括这部分内容。即使在本节中，也仅主要介绍其基本概念，并尽可能用简单的语言介绍其用于计量经济学模型估计的过程，只涉及一些简单的内容，使读者对这种方法有一个概念性了解，为深入学习与应用建立一个基础。有兴趣的读者可以进一步阅读有关贝叶斯分析的专门教科书，例如《统计决策论及贝叶斯分析》（James O.Berger著，贾乃光译，中国统计出版社，1998年）等。

3.3.1 概念

1.贝叶斯方法的基本思路

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

贝叶斯方法与经典估计方法的主要不同之处是：

(1)关于参数的解释不同

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

⑵所利用的信息不同

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

⑶对随机误差项的要求不同

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

⑷选择参数估计量的准则不同

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

2.贝叶斯定理

贝叶斯定理是贝叶斯估计方法的理论基础。贝叶斯定理表达如下：

$$g(\theta\ |Y) = \frac{f(Y|\theta)g(\theta)}{f(Y)} \tag{3.3.1}$$

其中θ为待估参数；Y为样本观测值信息，即样本信息；$g(\theta)$是待估参数θ的先验分布密度函数；$g(\theta\ \ |Y)$ 为θ的后验分布密度函数；$f(Y)$和$f(Y|\theta)$是Y的密度函数。因为对θ而言，$f(Y)$)可以认为是常数（样本观测值独立于待估参数），$f(Y|\theta)$在形式上又同θ的或然函数$L(\theta\ |Y)$一致（在最大或然法中已经了解），于是(6.5.1)可以改写为：

$$g(\theta\ |Y) \propto L(\theta\ |Y) \cdot g(\theta) \tag{3.3.2}$$

即后验信息正比于样本信息与先验信息的乘积。(6.4.2)表明，可以通过样本信息对先验信息的修正来得到更准确的后验信息。得到后验分布的密度函数后，就可以此为基础进行参数的点估计、区间估计与假设检验。

3.损失函数

常用的损失函数有线性函数和二次函数。不同的损失函数，得到的参数估计值是不同的。在下面的估计过程中再作进一步的说明。

3.3.2 单方程计量经济学模型贝叶斯估计的过程

采用贝叶斯方法估计单方程计量经济学模型的过程可概括如下：

1.确定模型的形式，指出待估参数B

2.给出待估参数B的先验分布

通常待估参数B是一个多维向量，即需要给出多参数的联合信息先验。如果对参数一无所知，可以认为B是均匀分布。实际上常用的也是均匀分布和共轭分布（尤以自然共轭分布为多）。

3.利用样本信息，修正先验分布

利用贝叶斯定理，导出B的后验密度函数（密度函数简写为p.d.f.）。到此为止，实际上已经得到B的所有信息。

4.利用B的后验密度函数，进一步推断出B的点估计值，或进行区间估计与假设检验

一般利用损失函数推断出B的点估计值，利用最高后验密度区间进行B的区间估计，利用最高后验密度区间或者利用后验优势比进行假设检验。

5.预测

根据求得的参数估计值进行预测。这并不是贝叶斯估计过程所必须，但对于检验估计结果是重要的。

3.3.3 正态线性单方程计量经济学模型的贝叶斯估计

下面以正态线性单方程计量经济学模型为例分析贝叶斯估计方法。选择正态线性单方程计量经济学模型，主要因为：(1)多元线性单方程计量经

济学模型具有普遍性意义；(2)随计误差项是大量随机扰动之总和，根据中心极限定理，可以认为它是渐近正态分布；(3)计算简单，使用方便，并能完整地体现贝叶斯估计方法的主要内容。

正态线性单方程计量经济学模型又分为随机误差项方差已知和方差未知两种情况。作为方法的演示分析，我们只讨论方差已知的情况。

1.有先验信息的后验分布

对于正态线性单方程计量经济学模型

$$Y = \mathbf{X}\mathrm{B} + \mu$$

其中$\mu \sim N(0, \sigma^2\mathbf{I})$，其它变量及参数的含义同前。

为方便起见，在下面的讨论中将只涉及正态分布的“核”，即其指数部分，而忽略其常数部分，不影响讨论结果。

选择B的先验分布为自然共轭分布，即B的密度函数和它的或然函数，以及二者结合后产生的函数服从同一分布。

Bœg,ΔkΦˇ˘

$$g(\mathrm{B}) \propto e^{-\frac{1}{2}}(\mathrm{B} - \bar{\mathrm{B}})'\bar{\Sigma}_{\mathrm{B}}^{-1}(\mathrm{B} - \bar{\mathrm{B}}) \tag{3.3.3}$$

其中$\bar{\mathrm{B}}$为待估参数先验均值，$\bar{\Sigma}_{\mathrm{B}}$为待估参数先验协方差矩阵。

Bœ,$L(\mathrm{B}\,|Y\,)$Øuœ=

$$L(\mathrm{B}\,|Y\,) \propto e^{-\frac{1}{2\sigma^2}}(Y - \mathbf{X}\mathrm{B})'(Y - X\mathrm{B}) \tag{3.3.4}$$

利用贝叶斯定理，得到B的后验密度函数为：

$$\begin{aligned} g(\mathrm{B}\,|Y\,) &\propto g(\mathrm{B}) \cdot L(\mathrm{B}\,|Y\,) \\ &\propto \exp\left\{-\tfrac{1}{2\sigma^2}\left[(A^{\frac{1}{2}}\bar{\mathrm{B}} - A^{\frac{1}{2}}\mathrm{B})'(A^{\frac{1}{2}}\bar{\mathrm{B}} - A^{\frac{1}{2}}\mathrm{B}) + (Y - \mathbf{X}\mathrm{B})'(Y - \mathbf{X}\mathrm{B})\right]\right\} \end{aligned} \tag{3.3.5}$$

其中$A = \sigma^2\bar{\Sigma}_{\mathrm{B}}^{-1}$

令$W = \begin{pmatrix} A^{\frac{1}{2}}\bar{\mathrm{B}} \\ Y \end{pmatrix}_{(k+n)\times !} \qquad G = \begin{pmatrix} A^{\frac{1}{2}} \\ \mathbf{X} \end{pmatrix}_{(k+n)\times n}$

于是有：

$$g(\mathrm{B}\,|Y\,) \propto \exp\left\{-\frac{1}{2\sigma^2}(W - G\mathrm{B})'(W - G\mathrm{B})\right\} \tag{3.3.6}$$

用$\bar{\bar{\mathrm{B}}}$表示待估参数后验均值，$\bar{\bar{\Sigma}}_{\mathrm{B}}$表示待估参数后验协方差矩阵。并且应用下列结论：

$$(W-G\mathrm{B})'(W-G\mathrm{B})=(\mathrm{B}-\bar{\bar{\mathrm{B}}})'G'G(\mathrm{B}-\bar{\bar{\mathrm{B}}})+(W-G\bar{\bar{\mathrm{B}}})'(W-G\bar{\bar{\mathrm{B}}}) \tag{3.3.7}$$

其中

$$\begin{aligned}\bar{\bar{\mathrm{B}}} &= (G'G)^{-1}G'W=(A+\mathbf{X}'\mathbf{X})^{-1}(A\bar{\mathrm{B}}+\mathbf{X}'\mathbf{X}\beta)\\ \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y\end{aligned}$$

βBœ OLS估计值。将(3.3.7)代入(6.4.6)，代入时舍去(3.3.7)的右边第二项，因为其不含B，可以作为常数项处理，只考虑核。得到

$$\begin{aligned}g(\mathrm{B}\,|Y\,) &\propto \exp\left\{-\tfrac{1}{2\sigma^2}(\mathrm{B}-\bar{\bar{\mathrm{B}}})'G'G(\mathrm{B}-\bar{\bar{\mathrm{B}}})\right\}\\ &\propto \exp\left\{-\tfrac{1}{2\sigma^2}(\mathrm{B}-\bar{\bar{\mathrm{B}}})'(A+\mathbf{X}'\mathbf{X})(\mathrm{B}-\bar{\bar{\mathrm{B}}})\right\}\\ &\propto \exp\left\{-\tfrac{1}{2}(\mathrm{B}-\bar{\bar{\mathrm{B}}})'(\bar{\Sigma}_{\mathrm{B}}^{-1}+\mathbf{X}'\mathbf{X}/\sigma^2)(\mathrm{B}-\bar{\bar{\mathrm{B}}})\right\}\\ &\propto \exp\left\{-\tfrac{1}{2}(\mathrm{B}-\bar{\bar{\mathrm{B}}})'(\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1})(\mathrm{B}-\bar{\bar{\mathrm{B}}})\right\}\end{aligned} \tag{3.3.8}$$

式中

$$\begin{aligned}\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1} &= \bar{\Sigma}_{\mathrm{B}}^{-1}+\mathbf{X}'\mathbf{X}/\sigma^2\\ \bar{\bar{\mathrm{B}}} &= (A+\mathbf{X}'\mathbf{X})^{-1}(A\bar{\mathrm{B}}+\mathbf{X}'\mathbf{X}\beta)\\ &= (\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1}+\mathbf{X}'\mathbf{X}/\sigma^2)^{-1}(\bar{\Sigma}_{\mathrm{B}}^{-1}\bar{\mathrm{B}}+(\mathbf{X}'\mathbf{X}/\sigma^2)\beta\\ &= \bar{\bar{\Sigma}}_{\mathrm{B}}^{-1}(\bar{\Sigma}_{\mathrm{B}}^{-1}\bar{\mathrm{B}}+(\mathbf{X}'\mathbf{X}/\sigma^2)\beta)\end{aligned} \tag{3.3.9}$$

于是，(6.4.7)正好是均值为$\bar{\bar{\mathrm{B}}}$、方差为$\bar{\bar{\Sigma}}_{\mathrm{B}}$的多元正态分布的核，即B的后验密度函数为：

$$(\mathrm{B}\,|Y\,)\sim N(\bar{\bar{\mathrm{B}}},\bar{\bar{\Sigma}}_{\mathrm{B}}) \tag{3.3.10}$$

将协方差矩阵的逆$\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1}$定义为精确度矩阵，那么可以看出：(1)验精确度矩阵$\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1}$是先验精确度矩阵$\bar{\Sigma}_{\mathrm{B}}^{-1}$ 与样本信息精确度矩阵$\mathbf{X}'\mathbf{X}/\sigma^2$之和，故后验精确度总是高于先验精确度；⑵后验均值$\bar{\bar{\mathrm{B}}}$是先验均值$\bar{\mathrm{B}}$ 与样本信息OLS估计值β的加权平均和，权数为各自的精确度。

2.无先验信息的后验分布

在对待估参数一无所知的情况下，仍然可以用贝叶斯方法求得待估参数的后验分布。这时可以认为待估参数的所有元素服从$(-\infty,+\infty)$上的均匀分布，且互不相关。于是，B的先验密度函数为：

$$g(\mathrm{B})=g(\beta_1)\cdot g(\beta_2)\cdots g(\beta_k)\propto c$$

B的或然函数$L(\mathrm{B}\,|Y\,)$仍然与(3.5.4)相同，则后验密度函数与或然函数形式相同：

$$\begin{aligned} g(\mathrm{B}\,|Y\,) &\propto g(\mathrm{B})\cdot L(\mathrm{B}\,|Y\,)\propto L(\mathrm{B}\,|Y\,) \\ &\propto \exp\left\{-\tfrac{1}{2\sigma^2}(Y-\mathbf{X}\mathrm{B})'(Y-\mathbf{X}\mathrm{B})\right\} \\ &\propto \exp\left\{-\tfrac{1}{2\sigma^2}\left[(\mathrm{B}-\beta)'\mathbf{X}'\mathbf{X}(\mathrm{B}-\beta)+(Y-\mathbf{X}b)'(Y-\mathbf{X}\beta)\right]\right\} \\ &\propto \exp\left\{-\tfrac{1}{2\sigma^2}(\mathrm{B}-\beta)'\mathbf{X}'\mathbf{X}(\mathrm{B}-\beta)\right\} \end{aligned} \tag{3.3.11}$$

得到的B的后验密度函数仍然是正态的，均值为β，协方差矩阵为$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$，即

$$(\mathrm{B}\,|Y\,)\sim N(\beta,\sigma^2(\mathbf{X}'\mathbf{X})^{-1}) \tag{3.3.12}$$

从形式上看，无信息先验得到的后验分布均值与样本信息的OLS估计相同，但二者有不同的含义。贝叶斯结论(6.4.11)中的B是随机的，而均值β在样本确定后是固定的；样本信息的结论中，B作为期望值，而β是随机变量，即有$\beta\sim N(\mathrm{B},\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$。

事实上，也可以直接由(6.4.7)、(6.4.8)、(6.4.9)得到无信息先验下的后验密度，只要把无信息先验作为有信息先验的一种特殊情况，即无信息先验的精确度为0。即

令$\bar{\Sigma}_{\mathrm{B}}^{-1}=0$，代入(6.4.8)、(6.4.9)可得到：

$$\bar{\bar{\Sigma}}_{\mathrm{B}}^{-1}=\mathbf{X}'\mathbf{X}/\sigma^2$$

从而$\bar{\bar{\Sigma}}_{\mathrm{B}}=\sigma^2\mathbf{X}'\mathbf{X}$

$$\bar{\bar{\mathrm{B}}}=(\mathbf{X}'\mathbf{X}/\sigma^2)^{-1}\cdot(\mathbf{X}'\mathbf{X}/\sigma^2)\cdot\beta=\beta$$

所以

$$(\mathrm{B}\,|Y\,)\sim N(\beta,\sigma^2(\mathbf{X}'\mathbf{X})^{-1})$$

与(6.4.11)相同。

3.点估计

在得到贝叶斯估计的后验密度函数后，即可以此为出发点，进行点估计。其思路是利用损失函数并使平均损失最小。为此需要确定一个损失函数。

先讨论一般情况。假设$\hat{\mathrm{B}}$为B的估计量，损失函数为$Lo(\mathrm{B},\hat{\mathrm{B}})$，表示参数为B时采用$\hat{\mathrm{B}}$为估计量所造成的损失，故总存在$Lo(\mathrm{B},\hat{\mathrm{B}})\geq 0$。于是满足$Min(Lo(\mathrm{B},\hat{\mathrm{B}}))$的$\hat{\mathrm{B}}$即是需要的估计量。由于损失函数依赖样本，故取加权平均损失，权数为后验密度函数$g(\mathrm{B}\,|Y\,)$，表示为：

$$\mathrm{E}(Lo(\mathrm{B},\hat{\mathrm{B}})\,|Y\,)=\int Lo(\mathrm{B},\hat{\mathrm{B}})\cdot g(\mathrm{B}\,|Y\,)d\mathrm{B} \tag{3.3.13}$$

能使(6.4.12)所表示的加权平均后验损失最小的$\hat{\mathrm{B}}$的值，即为B的贝叶斯点估计值。

下面以二次损失函数为例说明点估计过程。

损失函数的二次型为：

$$Lo=(\mathrm{B}-\hat{\mathrm{B}})'\mathbf{M}(\mathrm{B}-\hat{\mathrm{B}}) \tag{3.3.14}$$

$\mathbf{M}$˘Ω$\hat{\mathrm{B}}$BœﬄO”　$\mathrm{E}(Lo)\Theta$

$$\begin{aligned}
\mathrm{E}(Lo) &= \mathrm{E}((\mathrm{B}-\hat{\mathrm{B}})'\mathbf{M}(\mathrm{B}-\hat{\mathrm{B}})) \\
&= \mathrm{E}\left\{(\mathrm{B}-\mathrm{E}(\mathrm{B}))-(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))\right\}'\mathbf{M}\left\{(\mathrm{B}-\mathrm{E}(\mathrm{B}))-(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))\right\} \\
&= \mathrm{E}\left\{(\mathrm{B}-\mathrm{E}(\mathrm{B}))'\mathbf{M}(\mathrm{B}-\mathrm{E}(\mathrm{B}))+(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))'\mathbf{M}(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))\right\}
\end{aligned} \tag{3.3.15}$$

上述推导过程中出现的交叉项为0，即：

$$\begin{aligned}
\mathrm{E}\left\{(\mathrm{B}-\mathrm{E}(\mathrm{B}))'\mathbf{M}(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))\right\} &= \mathrm{E}(\mathrm{B}-\mathrm{E}(\mathrm{B}))'\mathbf{M}(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B})) \\
&= 0\cdot\mathbf{M}\cdot(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))=0
\end{aligned}$$

考察(3.3.16)式，第1项不含$\hat{\mathrm{B}}$，又因为$\mathbf{M}$为一个正定矩阵，故第2项

$$(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))'\mathbf{M}(\hat{\mathrm{B}}-\mathrm{E}(\mathrm{B}))\geq 0 \tag{3.3.16}$$

所以使(3.3.16)式最小的$\hat{\mathrm{B}}$应使(6.4.14)中等号成立，即为：

$$\hat{\mathrm{B}}=\mathrm{E}(\mathrm{B})$$

即二次损失函数的点估计值为后验均值。

4.区间估计

类似于点估计，可以根据B的后验密度函数进行区间估计。这里需要引入最高后验密度区间的概念：区间内每点的后验密度函数值大于区间外任何一点的后验密度函数值，这样的区间称为最高后验密度区间（HPD区间）。

对于单参数（k=1）模型的区间估计，可参照下式：

$$\int_a^b g(\beta\ |Y)d\beta=1-\alpha \tag{3.3.17}$$

其中$(1-\alpha)$为置信水平。这种区间估计与经典样本信息理论中的区间估计是一致的。特别在正态分布情况下，只要稍作变换，查正态分布表便可很容易地得到结论。

对于多参数（$k > 1$）模型的区间估计，可根据B的后验密度函数，求出B的每一元素的边际后验密度函数，再按照单参数情况求出每一参数的最高后验密度区间。

需要说明的是，参数的最高后验密度区间在形式上与经典样本信息理论中的置信区间是一致的，但解释并不相同。在贝叶斯估计中参数的最高后验密度区间的含义是参数以$(1-\alpha)$的概率位于该区间内；经典样本信息理论中的置信区间的含义是该区间以$(1-\alpha)$的概率包含参数。含义上的区别源于是否把待估参数作为随机变量。

5.假设检验

可以用上述的最高后验密度区间进行假设检验，但常用的方法是利用后验优势比检验。

3.3.4 一个贝叶斯估计的实例

由于贝叶斯估计利用了样本信息和先验信息，所以它比仅利用样本信息的参数估计量更有效。在样本容量比较小的情况下，仅利用样本信息得到的参数估计量是很不可靠的，如果能够采用贝叶斯估计，则将提高参数估计量的可靠性。问题在于如何得到先验信息。也正是由于这一点，使得贝叶斯估计虽然具有较大的理论意义，但是实际应用却受到了限制。下面以建立我国国防支出模型为例，对贝叶斯估计的过程进行实际演算。

经过理论分析、数据散点图分析和变量显著性检验，得到我国国防支出（用DE表示）主要取决于当年国内生产总值，而且呈线性关系。于是模型为：

$$DE_t = \beta_0 + \beta_1 GDP_t + \mu_t \qquad t = 1, 2, \cdots, T$$

1.利用样本信息估计模型

选取1978-1997年的数据为样本，采用经典模型的估计方法，得到：

$$\hat{\beta}_0 = 126.18 \qquad \hat{\beta}_1 = 0.008886 \qquad \hat{\sigma}_\mu^2 = 453$$

2.求先验均值与方差

这里采用一种特别的方法获得先验信息。选取1952-1977年的数据为样本观测值，估计模型，将估计结果作为先验信息。得到参数的先验均值为：

$$\bar{\beta}_0 = 10.981 \qquad \bar{\beta}_1 = 0.047678$$

参数的先验协方差矩阵为：

$$\bar{\Sigma}_{\mathrm{B}} = \begin{bmatrix} 76.615 & -0.0371 \\ -0.0371 & 2.1204e-005 \end{bmatrix}$$

3.利用样本信息修正先验分布

利用(6.4.8)和(6.4.9)计算得到：

$$\bar{\bar{\beta}}_0 = 87.0354 \qquad \bar{\bar{\beta}}_1 = 0.00979$$

$$\bar{\bar{\Sigma}}_{\mathrm{B}} = \begin{bmatrix} 10.0093 & -0.000262 \\ -0.000262 & 2.98683e-008 \end{bmatrix}$$

4.求参数的点估计值

根据文中的证明，上述后验均值就是参数的点估计值。于是得到采用贝叶斯估计的模型为：

$$D\hat{E}_t = 87.0354 + 0.00979GDP_t \tag{3.3.18}$$

显然与仅仅依赖于1978-1997年的样本信息估计的模型

$$D\hat{E}_t = 126.18 + 0.008886GDP_t \tag{3.3.19}$$

有明显不同。

5.预测检验

利用模型(6.4.16)预测1998年的国防支出，得到$D\hat{E}_{1998} = 865.7$，在95%的置信水平下预测值的置信区间为（780.4,951.0）。1998年实际国防支出为934.7。

为了进行比较，利用仅仅依赖于1978-1997年样本信息估计的模型(6.4.17)对1998年进行预测，得到$D\hat{E}_{1998} = 831.7$，其预测精度明显低于模型(6.4.16)；即使利用1952-1997年的所有数据为样本估计模型，并对1998年进行预测，得到的预测值为860.4，其预测精度也低于模型(6.4.16)。

3.4 局部多项式回归估计

局部多项式回归估计主要用于无参数回归模型的估计，在参数回归模型的估计中应用很少。在§2.6中介绍的无参数回归模型的估计中，正交序列估计和样条估计都需要估计出一系列的参数。为了减少参数估计的个数，可以采用局部线性回归估计，即对于给定的x,认为$m(\cdot)$在x附近的局部邻域近似于线性的，对x附近的那部分数据应用线性回归的技术，而该局部邻域的大小由窗宽来控制。这样，局部线性回归估计只需估计两个参数，而不是一系列参数。本节中主要以一元模型为例介绍局部线性回归估计。对于参数模型，局部回归的意义在于，人们可以将原始模型设定为无参数模型，然后用该方法进行估计，得到在某些方面较原始参数模型为优的结果。

3.4.1 一元局部多项式回归

对于一元无参数回归模型

$$Y = m(X) + u \tag{3.4.1}$$

Y为被解释变量，是随机变量；X为解释变量，它可以是随机的，也可以是非随机；u为随机误差项。X与Y的关系表达式$Y = m(X)$未知。

1. 局部多项式回归

假定$m(x)$在$x = x_0$处$p+1$阶导数存在，假设已有样本$(Y_1, X_1), (Y_2, X_2), \cdots, (Y_n, X_n)$，要估计$m(x_0), m'(x_0), m''(x_0), \cdots, m^{(p)}(x_0)$。为此，先将$m(x)$在$x = x_0$处进行台劳展开：

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p \tag{3.4.2}$$

$$Y_i = m(x_0) + m'(x_0)(X_i - x_0) + \cdots + \frac{m^{(p)}(x_0)}{p!}(X_i - x_0)^p + \varepsilon_i, \quad i = 1, \cdots, n \tag{3.4.3}$$

该多项式可用加权最小二乘法进行局部拟合。即最小化

$$\sum_{i=1}^{n} \{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2 K_h(X_i - x_0) I(\frac{|X_i - x_0|}{h} \le 1) \tag{3.4.4}$$

其中$K_h(\cdot) = K(\cdot/h)/h$，$h$为控制局部邻域大小的窗宽，$K(\cdot)$为核函数。$I(\cdot)$为显示性函数，当括号内的不等式成立时，取值为1，否则取值为0。

若$K(\cdot)$为[-1,1]上的核函数，则(3.4.4)简化为：

$$\sum_{i=1}^{n} \{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2 K_h(X_i - x_0) \tag{3.4.5}$$

(3.4.3)写成矩阵的形式为：

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{3.4.6}$$

其中

$$Y=\begin{pmatrix} Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix},X=\begin{pmatrix} 1 & X_1-x_0 & \cdots & (X_1-x_0)^p\\ 1 & X_2-x_0 & \cdots & (X_2-x_0)^p\\ \vdots & \vdots & \vdots & \vdots\\ 1 & X_n-x_0 & \cdots & (X_n-x_0)^p\end{pmatrix},\beta=\begin{pmatrix} \beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{pmatrix}\varepsilon=\begin{pmatrix} \varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{pmatrix}$$

β为参数向量，ε 为模型误差向量。

β 的加权最小二乘估计为极小化：

$$\min_{\beta}(Y-X\beta)'W(Y-X\beta) \quad 3.4.7$$

于是

$$\hat{\beta}=(X'WX)^{-1}X'Wy \quad 3.4.8$$

其中$W=diag\{K_h(X_i-x_0)\}$。$m^{(v)}(x_0)$的估计为$\hat{m}_v(x_0)=v!\hat{\beta}_v$，$v=0,1,\cdots,p$。

以上就是局部回归估计的基本思路。不难发现，无参数回归的Nadaraya-Watson核估计是$p=0$情况的局部回归估计。

2. 局部回归方法的优点

局部多项式拟合从理论和实践上都是吸引人的。第一，传统回归方法将经济变量局部上的变化差异掩盖了，因而无法反映经济现象的结构变化。传统回归方法处理有结构变化的经济变量之间关系时常用的手法是引入虚拟变量，但虚拟变量只能反映两个时期或若干已知时期经济现象的结构变化。而局部回归的结果能够动态地反映经济现象的结构变化。第二，以往根据经济理论建立的线性或非线性计量模型，隐含着若干假设条件。例如，C-D生产函数，隐含着技术中性、规模报酬不变等假设。而这些隐含条件对于具体问题不一定符合。局部回归假定变量之间的关系未知，因而没有隐含任何假设条件，所以更加符合实际。第三，其它普遍使用的核估计，如Nadaraya-Watson估计导致不必要的偏差，而Gasser-Müller估计在处理解释变量为随机性变量时有较大的方差（参见表3.4.1）。第四，局部回归方法既适合于解释变量为确定性变量的固定设定模型，也适合于解释变量为随机性变量的随机设定模型。第五，局部回归方法适合于随机设定模型解释变量分布均匀情形，也适合于分布不均匀的情形。第六，局部多项式拟合不必进行边界修正，它在边界的偏差自动与内部的偏差有相同的阶。在处理多维情况时，这是一个重要的优点。因为多维情况的边界相当大，其它估计方法在进行边界修正非常困难。第七，局部多项式回归估计在所有线性估计中，在极小极大效率(Minimax efficiency)意义上接近于最优，它的有效性为100%。

3. 偏和方差

表3.4.1为Nadaraya-Watson核估计、Gasser-Müller核估计和局部线性回归估计的逐点渐近偏和方差的比较。由表可见，Nadaraya-Watson核估计的方差与局部线性回归估计的相同，但偏却多了一项；Gasser-Müller核估计虽然偏与局部线性回归估计的相同，但方差却大了许多。因而，局部线性回归估计优于Nadaraya-Watson核估计和Gasser-Müller核估计。

表 3.4.1: $m(x)$的核回归估计的逐点渐近偏和方差（Fan 1992）

方法	偏	方差
Nadaraya-Watson	$(m''(x)+\frac{2m'(x)f'(x)}{f(x)})b_n$	V_n
Gasser-Müller	$m''(x)b_n$	$1.5V_n$
局部线性回归	$m''(x)b_n$	V_n

其中$b_n=\frac{h^2}{2}\int_{-\infty}^{\infty}u^2K(u)du$，$V_n=\frac{\sigma^2(x)}{f(x)nh}\int_{-\infty}^{\infty}K^2(u)du$。
记

$$\begin{array}{ll}
\mu_j=\int u^jK(u)du, & v_j=\int u^jK^2(u)du \\
S=(\mu_{j+l})_{0\le j,l\le p}, & c_p=(\mu_{p+1,}\cdots,\mu_{2p+1})' \\
\tilde{S}=(\mu_{j+l+1})_{0\le j,l\le p}, & \tilde{c}_p=(\mu_{p+2,}\cdots,\mu_{2p+2})' \\
S^*=(v_{j+l})_{0\le j,l\le p}, & S^{-1}=(S^{jl})_{0\le j,l\le p} \\
e_{\nu+1}=(0,\cdots,0,1,0,\cdots,0)', & 1\text{在第}\nu+1\text{位置上。}
\end{array}$$

$K(t)$的同等核定义为：

$$K^*_{v,p}(t)=(\sum_{l=0}^{p}S^{vl}t^l)K(t) \tag{3.4.7}$$

表3.4.2 核函数$K(t)$的同等核$K^*_{v,p}(t)$

v	p	核函数$K(t)$的同等核$K^*_{v,p}(t)$
0	1	$K(t)$
0	3	$(\mu_4-\mu_2^2)^{-1}(\mu_4-\mu_2t^2)K(t)$
1	2	$\mu_2^{-1}tK(t)$
2	3	$(\mu_4-\mu_2^2)^{-1}(t^2-\mu_2)K(t)$

假设$f(x_0)>0$，且$f(\cdot),m^{(p+1)}(\cdot)$和$\sigma^2(\cdot)$在$x_0$的邻域连续。则当$h\to 0,nh\to\infty$时，$\hat{m}_v(x_0)=v!\hat{\beta}_v$的渐近条件方差为：

$$\begin{aligned}
\mathrm{Var}\{\hat{m}_v(x_0)|X\} &= e'_{v+1}S^{-1}S^*S^{-1}e_{v+1}\tfrac{v!^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}+op(\tfrac{1}{nh^{1+2v}}) \\
&\approx \int K^*_{v,p}2(t)dt\tfrac{v!^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}+op(\tfrac{1}{nh^{1+2v}})
\end{aligned}$$

当$p-v$奇数时，渐近条件偏为：

$$\begin{aligned} Bias\{\hat{m}_v(x_0)|X\} &= e'_{v+1}S^{-1}c_p\tfrac{v!}{(p+1)!}m^{(p+1)}(x_0)h^{p+1-v}+op(h^{p+1-v}) \\ &\approx \int t^{p+1}K^*_{v,p}(t)dt\tfrac{v!}{(p+1)!}m^{(p+1)}(x_0)h^{p+1-v}+op(h^{p+1-v}) \end{aligned}$$

当$p-v$偶数时,若$f'(\cdot)$–$^{(p+2)}(\cdot)$在x_0的邻域连续,则渐近条件偏为：

$$\begin{aligned} Bias\{\hat{m}_v(x_0)|X\} &= e'_{v+1}S^{-1}\tilde{c}_p\tfrac{v!}{(p+2)!} \\ &\quad \times\{m^{(p+2)}(x_0)+(p+2)\{m^{(p+1)}(x_0)\tfrac{f'(x_0)}{f(x_0)}\}h^{p+1-v}+op(h^{p+2-v}) \end{aligned}$$

可见，当$p-v$奇数时，渐近条件偏有一个不包含$f'(x_0)$的较简单表达式；而当$p-v$偶数时，渐近条件偏却另有一个包含$f'(x_0)$的项。这也说明了为什么表3.4.1中Nadaraya-Watson核估计$(v=0,p=0)$的偏比局部线性回归估计$(v=0,p=1)$的偏多一项。

4.最佳窗宽的选择

窗宽参数h在局部回归中起相当重要的作用，太大的窗宽将使与$x=x_0$距离较远的观察点也参与局部回归，也就造成局部回归估计的偏差较大。太小的窗宽将使与$x=x_0$距离较近的观察点没能参与局部回归，也就造成局部回归估计的随机偏差较大。因而寻求合适的窗宽将是局部回归的最重要任务之一。

(1) 理论的最佳窗宽

设$\hat{\beta}_j,j=0,1,\cdots,p$是$p$阶局部多项式回归的参数估计，则$\hat{m}_v(x_0)=v!\hat{\beta}_v$是$m^{(v)}(x_0)$的估计,$v=0,1,\cdots,p$。

理论上,估计$m^{(v)}(x_0)$的局部窗宽的最优选择是使得条件均方误差(MSE)：

$$\begin{aligned} MSE(x_0) &= E[\{\hat{m}_v(x_0)-m^{(v)}(x_0)\}^2|X] \\ &= [Bias\{\hat{m}_v(x_0)|X\}]^2+Var\{\hat{m}_v(x_0)|X\} \\ &\approx \int t^{p+1}K^*_{v,p}(t)dt\tfrac{v!}{(p+1)!}m^{(p+1)}(x_0)h^{p+1-v} \end{aligned} \quad (3.4.8)$$
$$+\int K^*_{v,p}2(t)dt\tfrac{v!^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}$$

达到最小,于是得到最优的局部窗宽为：

$$h_{opt}(x_0)=C_{v,p}(K)\left[\frac{\sigma^2(x_0)}{\{m^{(p+1)}(x_0)\}^2f(x_0)}\right]^{1/(2p+3)}n^{-1/(2p+3)} \quad (3.4.9)$$

其中$C_{v,p}(K)=\left[\frac{(p+1)!^2(2v+1)\int K^{*2}_{v,p}(t)dt}{2(p+1-v)\{\int t^{p+1}K^*_{v,p}(t)dt\}^2}\right]^{1/(2p+3)}$，$\sigma^2(x_0)$是给定$Xt_0$时，$Y$的条件方差$Var(Y|X=x_0)$，$f(\cdot)$是$X$的概率密度函数。

v	p	高斯核	均匀核	Epanechnikov	Biweight	Triweight
0	1	0.773	1.351	1.719	2.036	2.312
0	3	1.161	2.813	3.243	3.633	3.987
1	2	0.884	1.963	2.275	2.586	2.869
2	3	1.006	2.604	2.893	3.208	3.503

表3.4.3 $C_{v,p}(K)$的数值

全局最优窗宽应是使得条件加权MISE（Mean Integrated Squared Error）

$$\int([Bias\{\hat{m}_v(x)|X\}]^2+Var\{\hat{m}_v(x)|X\})w(x)dx \tag{3.4.10}$$

达到最小，其中$w\geq 0$是某权函数。于是，得到最优全局窗宽

$$h_{opt}=C_{v,p}(K)\left[\frac{\int\sigma^2(x)w(x)/f(x)dx}{\int\{m^{(p+1)}(x)\}^2w(x)dx}\right]^{1/(2p+3)}n^{-1/(2p+3)} \tag{3.4.11}$$

（2）最佳窗宽的估计

IJœ˜ˆ–Θ–”

在每个局部观察点$x=X_i$，首先在样本中剔除该观察点(X_i,Y_i)；其次将剩下的$n-1$个观察点在$x=X_i$处进行核权局部回归。局部回归的一个明显的好处就是常数项$\beta_{0,-i}$的估计$\hat{\beta}_{0,-i}$刚好就是Y_i的拟合值。最后，通过比较平均拟合误差

$$RMSE=\sqrt{\frac{\sum\limits_{i=1}^{n}(Y_i-\hat{\beta}_{0,-i})^2}{n}}$$

的大小，选择使平均拟合误差达最小的窗宽h。

也可以采用插入法来选择最佳窗宽，即把未知函数的估计插入到渐近公式里以选择最佳平滑参数。

5. 最佳核函数的选择

核函数为一个对称的概率密度函数。核权函数在局部回归中起光滑的作用，即消除扰动的随机因素，使所得曲线反映变量之间的实际经济关系。在$x=x_0$处进行局部回归之前，对于不同的观察点X将赋予不同的权数，即不同的观察点在$x=x_0$处局部回归时的重要程度不同，或影响程度不同。一般核函数的选择原则是：当观察点X与x_0 “距离”近时，所赋予的权数就较大，相反就较小.核权函数可理解为每个观察点X与x_0的“距离”.“距离”的定义不同就产生不同的核权函数.

对于不同的v和p，通过(6.4.4)和(6.4.6)分别使MSE和$MISE$达最小而选择窗宽，而MSE和$MISE$都通过

$$T_{v,p}(K) = |\int t^{p+1}K^*_{v,p}(t)dt|^{2v+1}\{\int K^{*2}_{v,p}(t)dt\}^{p+1-v} \tag{3.4.12}$$

而依赖于核函数K。注意到对于任何对称核函数K，当$p-v$偶数时，$T_{v,p}(K)$=0。不妨设$p-v$为奇数。

于是，求使得MSE和$MISE$达最小的最优的核函数转换为求使得(6.4.7)达最小的核函数。因而得到极其简单的结论：在内点，使得MSE和$MISE$达最小的最优的核函数为Epanechnikov核函数$K(z) = 0.75(1-z^2)_+$；在左边界点-1，最优核函数为$K(z) = (1-z)I_{[0,1]}(z)$；在右边界点1，最优核函数为$K(z) = (1+z)I_{[-1,0]}(z)$。

有时，当估计的偏可忽略时，最佳核函数为使估计的方差达最小的核函数，也就是使$\int K^{*2}_{v,p}(t)dt$达最小的核函数。无论是在内点，还是边界点，均匀核$K(z) = 0.5I\{|z| \leq 1\}$是估计的方差达最小的最优核函数。

6. 最佳多项式阶数的选择

由于局部回归估计的偏差主要由窗宽控制，因而局部多项式阶数p的选择没有窗宽参数h的选择重要。对于给定的窗宽h，较大的p将减少局部回归估计的偏差，但导致估计的较大的方差。

假设我们用p阶多项式进行局部回归拟合得到函数$m_v(x_0)$的估计$\hat{m}_v(x_0)$,则估计量的方差为：

$$Var\{\hat{m}_v(x_0)|X\} = \int K^{*2}_{v,p}(t)dt\frac{v!^2\sigma^2(x_0)}{f(x_0)nh^{1+2v}}\{1+op(1)\} \tag{3.4.13}$$

下面考虑回归函数的估计(v=0)，则回归函数的估计方差为：

$$V_p\frac{\sigma^2(x_0)}{f(x_0)nh}\{1+op(1)\} \tag{3.4.14}$$

其中$V_p = \int K^{*2}_{0,p}(t)dt$。

表3.4.4 常用核函数的V_p/V_0

回归函数$m^{(v)}(x_0)$估计的方差随p增加而增大的一般规律是：当$p-v$从奇数增加到偶数时，方差增加；当$p-v$从偶数增加到奇数时，方差不增。所以，应选用$p-v$为奇数$2q+1$，$p=v+2q+1$阶多项式进行估计，它比$p=v+2q$阶多项式多一个参数，而回归函数估计量的方差却相同，但由于多一个参数就可以大大地减少在边界区域和数据密集区域的拟合造成的偏差。

因为窗宽被用于控制模型的复杂性，所以当目的是估计$m^{(v)}(x_0)$时，一般推荐使用$p=v+1$或偶尔$p=v+3$。

p	高斯	均匀	Epanechnikov	Biweight	Trriweight
1	1	1	1	1	1
2	1.6876	2.2500	2.0833	1.9703	1.9059
3	1.6876	2.2500	2.0833	1.9703	1.9059
4	2.2152	3.5156	3.1550	2.8997	2.7499
5	2.2152	3.5156	3.1550	2.8997	2.7499
6	2.6762	4.7852	4.2222	3.8133	3.5689
7	2.6762	4.7852	4.2222	3.8133	3.5689
8	3.1224	6.0562	5.2872	4.7193	4.3753
9	3.1224	6.0562	5.2872	4.7193	4.3753
10	3.5704	7.3281	6.3509	5.6210	5.1744

7. 数据类型的适应性

偏和方差表达式是在随机设定模型下给出的，但它们对于固定设定模型仍然有效。因而，局部多项式回归既适应于随机设定模型，也适应于固定设定模型。对比之下，Gasser-Müller核估计不适应于随机设定模型，其方差是局部线性回归估计方差的1.5倍。

当$p-v$偶数时，渐近条件偏多了一项。例如，$v=0, p=0$时Nadaraya-Watson核估计的渐近偏多了一项$m'(x_0)f'(x_0)/f(x_0)$。这样，估计的偏依赖于$m''(x_0)$和$m'(x_0)f'(x_0)/f(x_0)$的相互作用：当$m''(x_0)$固定时，对于解释变量高度密集且不均匀非对称的数据，其$|f'(x_0)/f(x_0)|$较大，因而估计的偏也较大。这样Nadaraya-Watson核估计不适合于解释变量分布不均匀且高度密集的数据。类似地，当$p-v$偶数时，$m^{(v)}(x_0)$的局部多项式估计不适合于解释变量分布不均匀且高度密集的数据。

当$p-v$奇数时，由于$m^{(v)}(x_0)$的局部多项式估计的偏与解释变量分布密度无关，所以，它既适合于解释变量均匀分布的情形，也适合于解释变量非均匀分布的情形。

8. 边界的自动修正

在实际应用中，数据总是有界的。在估计$m^{(v)}(x_0)$时，当x_0接近边界时，局部邻域$x_0 \pm h$落在设计区域之外。这样，一些在内点成立的核矩条件在边界的x_0局部邻域不再成立，造成大多数平滑技术在边界有较大的偏。在§2.6中，为了纠正边界偏，介绍了边界核方法。Fan和Gijbels(1992)在理论上证明了局部多项式回归拟合能自动地进行有效的边界修正，这是其它平滑技术所无法比拟的。也不象多数其它平滑方法，局部多项式拟合方法进行边界修正不需要知道支撑端点的位置。

假定解释变量的设计密度有有界支撑[0,1]，左边界点表示成$x=ch$，右边界点表示成$x=1-ch$。记

$$\mu_{j,c} = \int_{-c}^{\infty} u^j K(u)du, \quad S_c = (\mu_{j+l,c})_{0\le j,l\le p} S_c^{-1} = (S_c^{jl})_{0\le j,l\le p}$$

$$K(t)œ>.Ø$$

$$K^*_{v,p,c}(t) = (\sum_{l=0}^{p} S_c^{vl} t^l) K(t)$$

边界同等核满足Gasser,Müller和Mammitzsch(1985)的边界矩条件，因而局部多项式回归拟合具有自动边界修正性质。

9. 最佳线性平滑

让

$$C_{p+1} = \{m : |m(z) - \sum_{j=0}^{p} \frac{m^{(j)}(x_0)}{j!}(z - x_0)^j| \leq C \frac{|z - x_0|^{p+1}}{(p+1)!}\} \tag{3.4.15}$$

定义线性极小极大风险为：

$$R_{v,L}(n, C_{p+1}) = \inf_{\hat{S}_v \quad linear} \sup_{m \in C_{p+1}} E\left[\{\hat{S}_v - m^{(v)}(x_0)\}^2 | X\right] \tag{3.4.16}$$

这里仅考虑$p = 1, v = 0$的特殊情况。对于线性估计$\hat{m}_L$，定义它的极小极大效率为：

$$\left(\frac{R_{0,L}(n, C_2)}{\sup_{m \in C_2} E[\{\hat{m}_L(x_0) - m(x_0)\}^2 | X]} \right)^{5/4} \tag{3.4.17}$$

于是有结论：核函数为Epanechnikov核，窗宽为：

$$h = \left\{ \frac{15\sigma^2(x_0)}{f(x_0)C^2 n} \right\}^{1/5} \tag{3.4.18}$$

的局部线性回归估计$\hat{m}^*(x_0)$达到极小极大风险，因而它的极小极大效率为100%。

表3.4.5 核回归平滑的极大极小效率（%）

核	局部线性拟合	Gasser-Muller	Nadaraya-Watson
Epanechnikov	100	66.67	0
高斯	95.12	63.41	0
均匀	92.95	61.97	0

于是，80%的极小极大效率的线性估计仅仅利用了数据的80%的信息，而局部线性回归估计利用的数据信息可达到100%。因而用100个数据

得到的估计，若采用局部线性回归只需80个数据得到的估计将与它一样的好。注意，Nadaraya-Watson估计依赖于导数$m'(x_0)$，因而除非$f'(x_0) = 0$，不然在C_2上的极大风险为无穷大。所以，它的极小极大风险为0。

上述极小极大性质很容易推广到回归函数及其导数的局部多项式估计的一般情况，局部多项式估计将达到最高的线性极小极大效率。在边界，局部线性估计可达到线性极小极大效率为94.64%。这样，局部线性回归估计几乎是最佳的边界修正方法。详细的内容请读者参考Fan(1996)。。

3.4.2 多元局部多项式回归

1. 多元局部回归方法

Yæ)CXp)Cœ$'LY = m(X)$¨"8–œ$gkY = m(X)$3?1Taylor展开；其次，应用核权函数给每个观察点加权；最后，在局部用加权最小二乘法进行局部回归估计。

(1) Taylor展开

假定已有一个样本$Y_i, X_i = (X_{i1}, X_{i2}, \cdots, X_{id}), i = 1, 2, \cdots, n$。将未知的函数关系$m(x)$在$x = X_i$处进行Taylor展开(忽略高阶项，只写出二阶多项式)：

$$m(x) \approx m(X_i) + (x - X_i)m'(X_i) + \frac{1}{2}(x - X_i)m''(X_i)(x - X_i)' \quad (3.4.19)$$

其中$m(X_i)$是$m(x)$在$x = X_i$处的观察值;$m'(X_i) = \frac{\partial m(X_i)}{\partial x}$是$m(x)$在$x = X_i$处的斜率向量或响应系数向量;$m''(X_i) = \frac{\partial^2 m(X_i)}{\partial x \partial x'}$是$m(x)$在$x = X_i$处的二阶偏导数方阵。因为$m(x)$未知，所以这些量也未知。局部回归提供了对它们进行估计的一种方法。

(2) 核权局部回归(Kernal Weighted Local Regression)

选定概率密度$K(\cdot)$

$$\int K(u)du = 1 \quad (3.4.20)$$

为核函数及窗宽$h > 0$，则$K_h(u) = h^{-1}K(uh^{-1})$也是一个概率密度。对应观察点$x_0$，将回归模型(1)写成：

$$Y_j = \alpha + (X_j - x_0)\beta + \frac{1}{2}(X_j - x_0)\gamma(X_j - x_0)' + \varepsilon_j j = 1, \cdots, n \quad 3.4.23$$

)

假设$\varepsilon_j \sim iid(0, \sigma^2)$，核权局部回归就是：根据加权最小二乘原理，求$\alpha, \beta, \gamma$的估计$\hat{\alpha}, \hat{\beta}, \hat{\gamma}$，使得（3.4.47）：

$$\sum_{i=1}^{n}\{Y_i-\alpha(X_i-x_0)\beta-frac12(X_i-x_0)\gamma(X_i-x_0)'\}^2$$
$$\times K_h(X_i-x_0)\prod_{j=1}^{p}I(\frac{|X_{ij}-x_{0j}|}{h}\leq 1)$$

达最小。于是$\hat{m}(x_0)=\hat{\alpha}$，$\hat{m}'(x_0)=\hat{\beta}$。

权数也可用k-近邻的权数。让$K=int(hN)$,其中0¡h ≤1，令$d(x)=(xx')^{1/2}$，对$d(X_j-x_0),j=1,2,\cdots,n$进行排序，得到$d_0$。它是$x_0$的邻域只有K个观察点的最大距离。设$u(x)=d(x)/d_0$，则$k$-近邻的权数为：

$$K_h(x)=K(u(x))\quad 3.4.24$$

$$K_h(X_j-x_0)=K(u(X_j-x_0))\qquad j=1,2,\cdots n$$

其中$K(\cdot)$为[-1,1]上的核函数。

3.4.3 一个局部回归的实际例

例3.4.1 ：对我国对外经济联系与国内通货膨胀的关系进行局部回归估计。例题的目的是估计$m'(x)$，进而分别得到商品进出口总额x_1和外汇储备x_2的相对变化对通货膨胀y的相对变化的弹性系数，以反映不同时期对外经济的政策变化对国内通货膨胀的影响。

因为目的是估计$m'(x)$，故用二阶多项式进行局部回归，采用高斯核函数，利用交错鉴定法确定最佳窗宽为0.6，此时局部回归平均拟合误差为2.8051。下面的图分别为商品进出口总额的响应系数β_1图、外汇储备的响应系数β_2图、商品进出口总额的弹性系数η_1图和外汇储备的弹性系数η_2图。表3.4.6为通货膨胀局部回归的响应系数和弹性系数。商品进出口总额的弹性系数为$\eta_1=\frac{\partial m(x)}{\partial x_1}\cdot\frac{x_1}{y}$，外汇储备的弹性系数为$\eta_2=\frac{\partial m(x)}{\partial x_2}\cdot\frac{x_2}{y}$。

3.5 半参数回归模型及其参数估计

§2.6中介绍的无参数回归模型虽然有许多的优点，但是从实际应用来说，有它的局限性。例如影响y的因素(即解释变量)可分为两个部分，即x和t，根据经验或历史资料可以认为因素x是主要的，而且y同x是具有明确的关系的；而t则是某种干扰因素(或者看作协变量)，它同y的关系是完全未知的，但是又没有理由将其归入误差项。此时如果用非参数回归模型加以处理，则会失去太多的信息，不能揭示y同x的关系；若仅采用参数回归模型，一般拟合效果较差。于是自然地提出了两者的混合模型。

半参数回归模型的方法融合了参数回归模型方法和较近发展起来的无参数回归模型方法，但并非这两类方法的简单叠加。可以想象，在不少实

表 3.4.2: 通货膨胀局部回归的响应系数和弹性系数

年月	β_1	η_1	β_2	η_2	年月	β_1	η_1	β_2	η_2
93.04	-0.805	-0.989	0.32	0.538	96.02	-0.526	-0.463	0.213	0.893
93.05	-0.269	-0.349	0.296	0.475	96.03	0.138	0.155	0.164	0.695
93.06	-0.801	-0.92	0.321	0.486	96.04	0.223	0.26	0.153	0.657
93.07	0.041	0.052	0.282	0.418	96.05	0.313	0.391	0.133	0.585
93.08	0.038	0.049	0.282	0.42	96.06	0.212	0.247	0.146	0.639
93.09	0.733	1.047	0.247	0.377	96.07	0.194	0.227	0.143	0.634
93.1	0.119	0.152	0.275	0.441	96.08	0.204	0.254	0.131	0.602
93.11	0.726	1.004	0.242	0.406	96.09	0.082	0.097	0.15	0.712
93.12	-0.036	-0.083	0.151	0.262	96.1	0.165	0.217	0.125	0.618
94.01	-1.269	-0.97	0.336	0.61	96.11	0.07	0.086	0.137	0.685
94.02	-1.288	-1.049	0.333	0.646	96.12	0.27	0.507	0.173	0.891
94.03	0.145	0.174	0.258	0.515	97.01	-0.304	-0.331	0.178	0.974
94.04	0.559	0.711	0.234	0.491	97.02	-0.336	-0.291	0.179	0.974
94.05	0.727	0.933	0.224	0.466	97.03	-0.11	-0.139	0.143	0.806
94.06	1.006	1.367	0.204	0.425	97.04	-0.077	-0.101	0.131	0.758
94.07	0.894	1.121	0.204	0.447	97.05	-0.011	-0.015	0.115	0.679
94.08	0.9	1.113	0.197	0.455	97.06	-0.127	-0.165	0.121	0.718
94.09	0.727	0.872	0.201	0.5	97.07	0.148	0.201	0.083	0.499
94.1	0.688	0.821	0.197	0.519	97.08	-0.244	-0.317	0.115	0.724
94.11	0.977	1.33	0.151	0.441	97.09	-0.108	-0.147	0.096	0.624
94.12	0.127	0.272	0.19	0.575	97.1	0.359	0.535	0.061	0.41
95.01	-0.808	-0.746	0.26	0.839	97.11	0.014	0.019	0.08	0.537
95.02	-0.742	-0.687	0.251	0.836	97.12	0.014	0.026	0.14	0.947
95.03	0.873	1.176	0.139	0.471	98.01	-0.53	-0.564	0.159	1.112
95.04	0.764	0.986	0.149	0.519	98.02	-0.529	-0.55	0.16	1.097
95.05	0.755	1.023	0.132	0.448	98.03	-0.4	-0.537	0.117	0.814
95.06	0.575	0.794	0.135	0.47	98.04	-0.334	-0.463	0.111	0.783
95.07	0.726	0.914	0.133	0.471	98.05	-0.5	-0.657	0.127	0.904
95.08	0.684	0.883	0.127	0.46	98.06	-0.365	-0.492	0.114	0.787
95.09	0.638	0.816	0.129	0.489	98.07	-0.234	-0.318	0.101	0.692
95.1	0.542	0.679	0.134	0.527	98.08	-0.482	-0.619	0.125	0.86
95.11	0.534	0.683	0.126	0.498	98.09	-0.433	-0.575	0.119	0.834
95.12	0.15	0.283	0.187	0.726	98.1	-0.519	-0.676	0.125	0.887
96.01	-0.182	-0.189	0.194	0.795	98.11	-0.421	-0.562	0.115	0.808

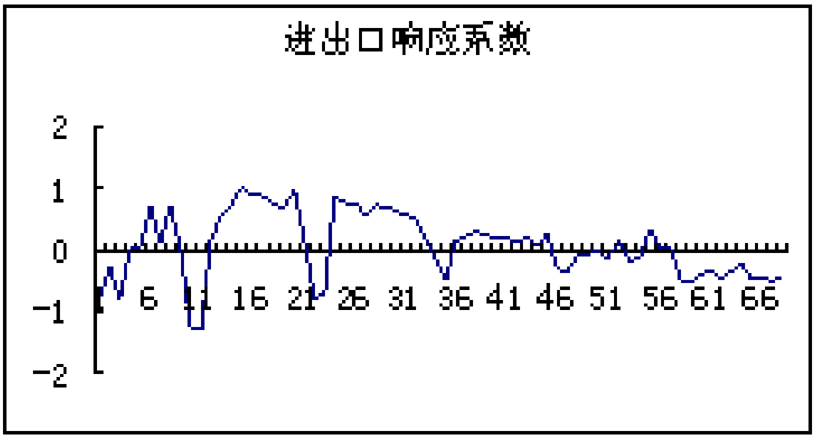


图 3.1: 商品进出口总额的响应系数

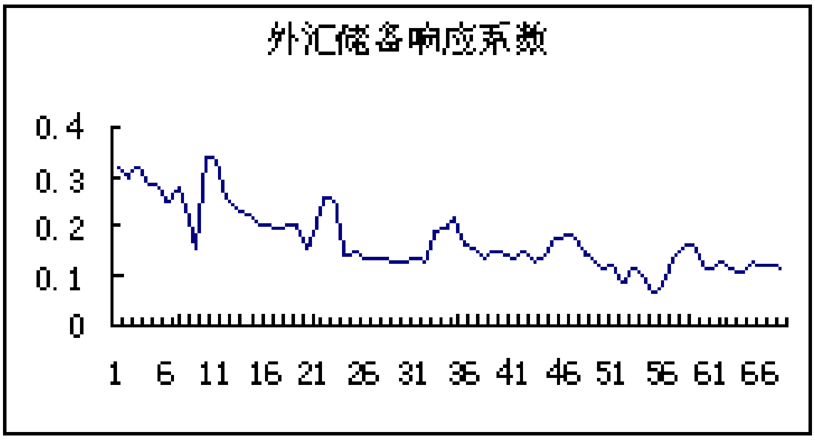


图 3.2: 外汇储备的响应系数

际问题中，它可能是一个更接近真实、更能充分利用数据中所提供的信息的方法。

半参数回归模型形式是简单的，本节将主要讨论它的估计问题。所以没有将它归入第二章作为一种特殊的模型结构，而是将它归入第三章作为一种特殊的估计方法。

3.5.1 偏残差估计

Stone于1977年提出如下回归模型：

$$\mathbf{y} = \mathbf{X}\beta + g(\mathbf{Z}) + \varepsilon \tag{3.5.1}$$

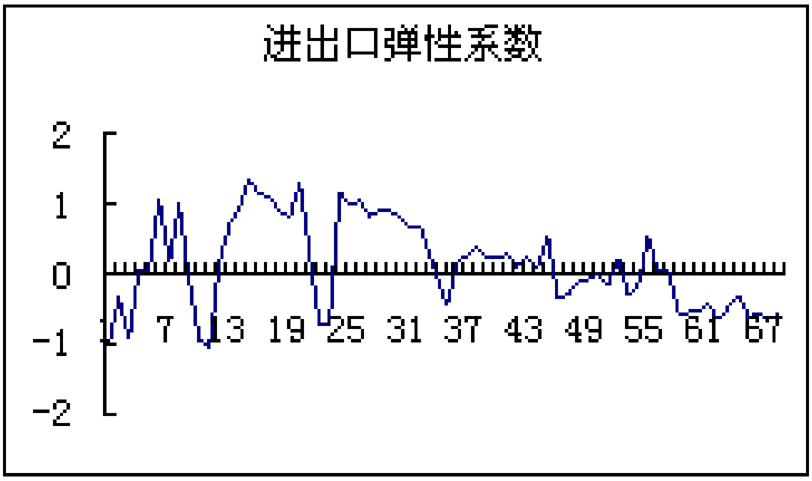


图 3.3: 商品进出口总额的弹性系数

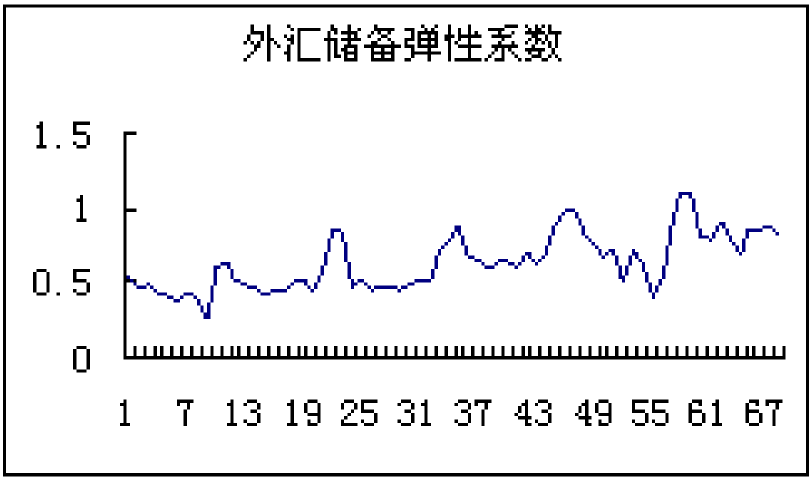


图 3.4: 外汇储备的弹性系数

它有线性主部$\mathbf{X}\beta$,把握被解释变量的大势走向，适于外延预测；还有非参数部分$g(\mathbf{Z})$，可以对被解释变量作局部调整，使模型更好地拟合样本观测值。由于(6.5.1)结合了参数模型和非参数模型，所以称为半参数回归模型。由于这种模型既含参数分量，又含非参数分量，可以概括和描述众多实际问题，因而引起广泛的重视。

(6.5.1)中$\mathbf{X} = (X_1, \cdots, X_p)$和$\mathbf{Z}$为取值于$R^p$和[0,1]上的随机或非随机向量和变量，$\beta$为p维未知参数向量, $g(\mathbf{Z})$为定义于[0,1]上的未知函数，ε是随机误差,$\mathrm{E}(\varepsilon) = 0$, $\mathrm{E}(\varepsilon^2) = \sigma^2$。如果认为$\mathbf{X}$和$\mathbf{Z}$是随机的，则需假定它们与$\varepsilon$独立。模型的任务是要从观测数据$y_i$、$X_i$和$Z_i$,$i = 1, \cdots, n$出发，估计未知函数$g(z)$、未知参数$\beta$和分布参数$\sigma^2$。

偏残差估计方法是由Denby(1984,1986)提出的，分两步进行估计。

第一步，先设β已知，估计$\hat{g}(\mathbf{Z})$。基于

$$u_i = y_i - X_i\beta = g(Z_i) + \varepsilon_i \tag{3.5.2}$$

用权函数法作出未知函数$g(\mathbf{Z})$的估计:

$$\hat{g}(\mathbf{Z}) = \sum_{i=1}^{n} W_{ni}(\mathbf{Z})u_i = \sum_{i=1}^{n} W_{ni}(\mathbf{Z})Y_i - \sum_{i=1}^{n} W_{ni}(\mathbf{Z})X_i\beta \tag{3.5.3}$$

权函数$W_{n,i}(\mathbf{Z})$可取作$W_{n,i}(\mathbf{Z}) = K(\frac{\mathbf{Z}-Z_i}{h})/\sum\limits_{j=1}^{n} K(\frac{\mathbf{Z}-Z_j}{h})$。$K$是核函数，可取概率密度核，例如Bartlett-Epanechnikov核: $K(z) = \frac{3}{4}(1 - z^2)_+, h > 0$为带宽，可根据经验选取。记$\hat{g}_1(\mathbf{Z}) = \sum\limits_{i=1}^{n} W_{ni}(\mathbf{Z})Y_i$，$\hat{g}_2(\mathbf{Z}) = \sum\limits_{i=1}^{n} W_{ni}(\mathbf{Z})X_i$，则

$$y_i - X_i\beta = \hat{g}_1(Z_i) - \hat{g}_2(Z_i)\beta + e_i$$

第二步，在已知$\hat{g}(\mathbf{Z})$的基础上估计$\hat{\beta}$。基于

$$y_i - \hat{g}_1(Z_i) = (X_i - \hat{g}_2(Z_i))\beta + e_i \tag{3.5.4}$$

得到β的最小二乘估计$\hat{\beta}$。记$\tilde{Y}_i = y_i - \hat{g}_1(Z_i)$，$\tilde{X}_i = X_i - \hat{g}_2(Z_i)$，则

$$\hat{\beta} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} \tag{3.5.5}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\beta})'(\tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\beta}) \tag{3.5.6}$$

其中$\tilde{\mathbf{Y}}=\begin{pmatrix}\tilde{Y}_1\\ \tilde{Y}_2\\ \vdots\\ \tilde{Y}_n\end{pmatrix}$，$\tilde{\mathbf{X}}=\begin{pmatrix}\tilde{X}_1\\ \tilde{X}_2\\ \vdots\\ \tilde{X}_n\end{pmatrix}$。

由于在两步估计中都分别以"偏"残差作为被解释变量观测值，所以称之为偏残差估计。

常见的权函数还有最近邻、多项式和三角级数等，而核方法在非参数估计中占有特别重要的地位，且在实际中最常用。需要指出的是$\hat{\beta}$,$\hat{\sigma}^2$,g^*的大样本性质本质上不依赖于权函数的选择。有关文献在权函数分别取多项式、最近邻和核函数时证明了$\hat{\beta}$的渐近正态性。$\hat{\sigma}^2$的渐近正态性首次由对权函数分别取最近邻和核函数的情形所证明。

如何给定$\hat{\beta}$以进行第一步的估计？一个可行的方法是直接估计原模型的参数部分$\mathbf{y}=\mathbf{X}\beta+\mu$，得到$\hat{\beta}$的预设值。

3.5.2 光滑样条估计

假设$0\le Z_1<Z_2<\cdots<Z_n\le 1$,$\beta$和$g(\mathbf{Z})$的光滑样条估计为使

$$\frac{1}{n}\sum_{i=1}^{n}[y_i-X_i\beta-g(Z_i)]^2+\lambda\int_0^1[g^{(m)}(u)]^2du \tag{3.5.7}$$

达到最小的估计。以$Z_1,Z_2,\cdots,Z_n$为节点的光滑样条估计为

$$\hat{\beta}_\lambda=(\mathbf{X}'(I-A(\lambda))\mathbf{X})^{-1}\mathbf{X}'(I-A(\lambda))\mathbf{y} \tag{3.5.8}$$

$$\hat{g}_\lambda(\mathbf{Z})=a^\tau(\mathbf{Z})(\mathbf{y}-\mathbf{X}\hat{\beta}_\lambda) \tag{3.5.9}$$

$$\begin{aligned}A(\lambda)&=\sum M^{-1}(I-T(T'M^{-1}T)^{-1}T'M^{-1})+T(T'M^{-1}T)^{-1}T'M^{-1}\\&=(a(Z_1),\cdots,a(Z_n))'\end{aligned} \tag{3.5.10}$$

选择λ，使得

$$GCV(\lambda)=\frac{1}{n}\sum_{i=1}^{n}[y_i-X_i\beta-g(Z_i)]^2/[\frac{1}{n}tr(I-A(\lambda))]^2 \tag{3.5.11}$$

达到最小。

3.5.3 两阶段最小二乘估计

1. 两阶段最小二乘估计

记$a = E(g(Z_i))$,$E(g^2(Z_i)) < \infty$，令$u_i = g(Z_i) - a + \varepsilon_i$，则

$$y_i = a + X_i\beta + u_i \tag{3.5.12}$$

使用最小二乘法得到β的估计β^*，称β^*为β的一次估计。实际上是提出了一种"偏残差估计"中首先给定β的估计量的方法。然后基于残差$y_i - X_i\beta^*$，在模型

$$\breve{Y}_i = y_i - X_i\beta^* = g(Z_i) + e_i$$

中估计g，例如记g^*为g的核权估计。最后，将模型中的g换为g^*，再次使用最小二乘法（偏残差估计中的第二步），得到β的二次估计$\hat{\beta}$。

$$g^*(\mathbf{Z}) = \sum_{i=1}^{n} W_{ni}(\mathbf{Z})\tilde{Y}_i = \sum_{i=1}^{n} W_{ni}(\mathbf{Z})(y_i - X_i\beta^*) \tag{3.5.13}$$

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta - g^*(Z_i))^2 \tag{3.5.14}$$

2. 两阶段局部回归估计

假设$a = E(g(Z_i))$,$E(g^2(Z_i)) < \infty$；令$u_i = g(Z_i) - a + \varepsilon_i$，则

$$y_i = a + X_i\beta + u_i$$

使用最小二乘法得到β的一次估计β^*。然后，在模型

$$\tilde{Y}_i = y_i - X_i\beta^* = g(Z_i) + e_i$$

中，采用核权局部回归方法估计g。即最小化：

$$\sum_{j=1}^{n} (\tilde{Y}_j - \alpha_0 - \alpha_1(Z_j - \mathbf{Z}))^2 K_h(Z_j - \mathbf{Z}) \tag{3.5.15}$$

其中$K_h(\cdot) = K(\cdot/h)/h$，$h$为控制局部邻域大小的带宽，$K(\cdot)$为核函数。

常用的核函数有高斯(Gaussian)、k-近邻(KNN)等。在给定核函数$K(\cdot)$下，采用交错鉴定方法选择窗宽h。使得(6.4.16)达到最小的α_0就是$g(Z)$的估计,记为$g^*(Z)$。

最后，在模型

$$\breve{Y}_i = y_i - g^*(Z_i) = X_i\beta + \varepsilon_i$$

中,使用最小二乘法,得到β的二次估计$\hat{\beta}$.

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta - g^*(Z_i))^2$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - X_i\hat{\beta} - g^*(Z_i))^2$$

其中"arg"为argument的缩写，表示"自变量"。

3. 一个例子

试图研究商品进出口对国内通货膨胀的影响。采用居民消费价格指数作为代表通货膨胀的变量，用它作为被解释变量y；从《中国物价》得到1993年4月到1998年11月每月与上年同月相比的居民消费价格指数，再换算成每月与1992年4月相比的居民消费价格指数，作为被解释变量y的样本观测值；选择商品进出口总额x、进出口差额z作为解释变量，样本观测值来自《海关统计》。

首先采用简单线性模型描述它们之间的关系，即

$$y = \alpha_0 + \alpha_1 x + \alpha_2 z + \varepsilon$$

采用普通最小二乘估计模型，得到的结果如下:

$$\hat{y} = \underset{(14.686)}{114.87} + \underset{(6.7547)}{0.22988}\, x + \underset{(7.8253)}{0.7536}\, z$$

$$R^2 = 0.70217 \quad \hat{\sigma} = 15.337$$

其中参数估计值的下方为其t统计量。从拟合结果看出，简单线性模型不能较好地拟合样本观测值。

然后采用半参数回归模型，商品进出口总额x与被解释变量y之间的关系仍用线性参数模型描述，进出口差额z对价格指数的影响采用无参数形式描述。即：

$$y = \alpha + \beta\, x + g(z) + \varepsilon$$

采用两阶段局部回归法估计模型。β的一次估计为0.31095；选用k近邻(KNN)核函数，用交错鉴定方法（CVM)选择最佳带宽为0.65，$g(z_i)$的估计$g^*(z_i)$的散点图如图3.5.1。β的二次估计为0.26441,$\hat{\sigma} = 14.249$。

由于14.249¡15.337，所以半参数模型的平均拟合误差小于线性回归模型的平均拟合误差。

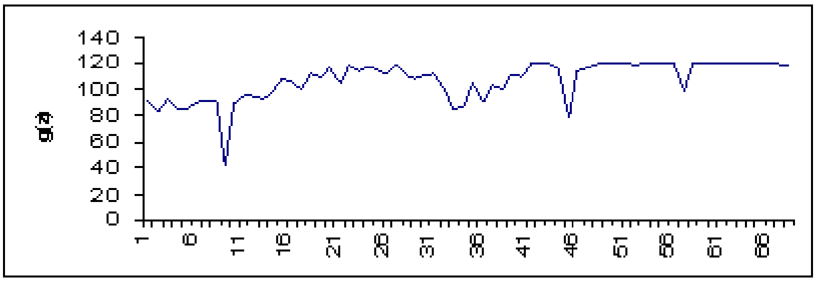


图 3.5: $g(z_i)$的估计$g^*(z_i)$的散点图

3.6 计量经济学模型的广义矩估计

广义矩估计方法（GMM, Generalized Method of Moments）是基于模型实际参数满足的一些矩条件而形成的一种参数估计方法，是矩估计方法的一般化。如果模型的设定是正确的，则总能找到该模型实际参数满足的若干矩条件而采用GMM。由于传统的计量经济学模型估计方法，例如普通最小二乘法、工具变量法、极大似然法等，都有它们的局限性，其参数估计量必须在模型满足某些假设时才具有良好的性质，诸如只有当模型的随机误差项服从正态分布或某一已知分布，极大似然法参数估计量才是可靠的估计量；而GMM允许随机误差项存在异方差和序列相关，所得到的参数估计量比其它参数估计方法更合乎实际；而且可以证明，GMM包容了许多常用的估计方法，普通最小二乘法、工具变量法、极大似然法都是它的特例。所以，GMM具有其优越性而得到应用。

3.6.1 广义矩估计的概念

1.参数的矩估计

参数的矩估计就是用样本矩去估计总体矩。

例如，$y_1, y_2, \cdots, y_n$是从正态分布总体$N(\mu, \sigma^2)$中抽取的一组样本观测值，那么可以从样本观测值计算样本一阶（原点）矩和二阶（原点）矩，然后去估计总体一阶矩和总体二阶矩，再进一步计算总体参数（期望和方差）的估计量。即

$$X^{(1)} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad X^{(2)} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

分别为样本的一阶矩和二阶矩，于是总体一阶矩和总体二阶矩的估计量为：

$$\hat{M}^{(1)} = E(Y) = X^{(1)} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{M}^{(2)} = E(Y^2) = X^{(2)} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

因为$E(Y) = \mu,\ E(Y^2) = \sigma^2 - \mu^2$，于是得到总体参数（期望和方差）的估计量为：

$$\hat{\mu} = \hat{M}^{(1)} = E(Y) = X^{(1)}$$

$$\hat{\sigma}^2 = \hat{M}^{(2)} + (\hat{M}^{(1)})^2 = X^{(2)} + (X^{(1)})^2$$

这里我们用$X^{(r)}$表示样本的r阶矩，用$M^{(r)}$表示总体的r阶矩。

再如，设某总体的分布密度函数为

$$f(t) = \frac{a}{\sqrt{2\pi}\Phi(t)t^2}\exp\left(-\frac{1}{2}\left(\frac{a}{t} - b\right)^2\right) \qquad t > 0 \tag{3.6.1}$$

其中参数为a和b，$\Phi(x)$为服从$N(0,1)$的分布函数。在进行参数估计时，可以使用负指数矩估计。因为对于该总体存在如下关系，即矩条件：

$$Et^{-1} = \frac{b}{a} + \frac{1}{a\sqrt{2\pi}\Phi(b)}\exp\left(-\frac{1}{2}b^2\right) \tag{3.6.2}$$

$$Et^{-2} = \frac{b^2+1}{a} + \frac{b}{a^2\sqrt{2\pi}\Phi(b)}\exp\left(-\frac{1}{2}b^2\right) \tag{3.6.3}$$

所以有关于参数估计量的方程组：

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_i} = \frac{\hat{b}}{\hat{a}} + \frac{1}{\hat{a}\sqrt{2\pi}\Phi(b)}\exp\left(-\frac{1}{2}\hat{b}^2\right) \tag{3.6.4}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_i^2} = \frac{\hat{b}^2+1}{\hat{a}} + \frac{\hat{b}}{\hat{a}^2\sqrt{2\pi}\Phi(b)}\exp\left(-\frac{1}{2}\hat{b}^2\right) \tag{3.6.5}$$

由样本计算方程组左端，求解方程组即可得到参数估计量。这样的负指数矩估计不仅便于计算，而且恰好等于参数的最大似然估计。

2.参数的广义矩估计

上面两个例子中都是选择两个样本矩估计总体的两个参数。为什么不选更多一些样本矩呢？如果选择的矩估计方程个数多于待估参数个数，那么该怎样确定参数估计值呢？于是，广义矩估计方法应运而生。

设样本r个矩为$X^{e)}, i = 1,\cdots,r$，对应母体r个矩为$M^{e)}(\beta), i = 1,\cdots,r$。$M^{e)}(\beta)$为待估总体参数$\beta$（向量）的函数，且$r$大于待估总体参数的个数。则最小二乘法参数估计量实际上是使得欧氏距离函数

$$Q(\beta) = \sum_{i=1}^{r}(X^{(i)} - M^{(i)}(\beta))^2 \tag{3.6.6}$$

达到最小的参数估计量$\hat{\beta}$。

但是不同的矩起的作用不同。我们希望某些矩的作用大些，这就想到加权最小二乘法，想到广义最小二乘法，从函数空间距离角度，就是要应用Mahalanobis距离。写成向量形式，记$X = (X^{(1)},\cdots,X^{(r)})'$,$M = (M^{(1)},\cdots,M^{(r)})'$，则马氏距离定义为：

$$Q(\beta) = (X-M)'S^{-1}(X-M) \tag{3.6.7}$$

其中S是关于$(X-M)$的协方差阵。参数β的GMM估计就是使得$Q(\beta)$达最小的$\hat{\beta}$。

3.6.2 计量经济学模型的广义矩估计

我们知道，如果模型的设定是正确，则存在一些为0的条件矩。所以广义矩估计的基本思想是利用矩条件估计模型参数，而且可以用于检验模型设定。

假设理论上确立了以下模型关系式：

$$y_i = h(X_i,\beta) + \varepsilon_i \quad i = 1,\cdots n \tag{3.6.8}$$

其中β是要估计的$k\times 1$参数向量。（注意：可能有$Cov[\varepsilon_i, h(X_i,\beta)] \neq 0$，或甚至$Cov[\varepsilon_i, X_j] \neq 0$，对所有$i$和$j$。）并且假设

$$E[\varepsilon] = 0$$

$$E[\varepsilon\varepsilon'] = \Omega \tag{3.6.9}$$

其中Ω是半正定的。这些设定表示，允许存在随机解释变量、异方差、序列相关等违背基本假设的情况。

1.估计方法的原理

(1)方法的引出

如果(6.4.9)中解释变量$x_j(j = 1,2,\cdots,k)$与随机误差项不相关，且随机误差项不存在异方差和序列相关，那么存在：

$$\sum_{i=1}^{n} x_{ji}\varepsilon_i = 0 \qquad j = 1,2,\cdots,k$$

即

$$\sum_{i=1}^{n} x_{ji}(y_i - h(X_i, \beta)) = 0 \qquad j = 1, 2, \cdots, k \tag{3.6.10}$$

这就是一组矩条件，由样本矩(6.4.11)条件估计模型参数β，就是一种矩估计。当然，读者已经清楚，(6.4.11)实际上是普通最小二乘估计的正规方程组。

因为当模型设定正确时，存在一些为0的条件矩。那么不妨假定由这些为0的矩条件可找到一个含$J(J > k)$个变量的$J \times 1$向量Z_i，使得Z_i与ε_i无关：

$$Cov[Z_i, \varepsilon_i] = 0 \tag{3.6.11}$$

也称此条件为矩条件，可以把Z看作一组工具变量。

定义

$$e(y_i, X_i; \beta) = y_i - h(X_i, \beta) \quad i = 1, 2, \cdots, n$$

$$m(\beta) = \frac{1}{n}\sum_{i} Z_i e(y_i, X_i; \beta) = \frac{1}{n} Z' e(y, X; \beta) \tag{3.6.12}$$

称$m(\beta)$为对应(6.4.12)的样本矩。$m(\beta)$为J×1阶向量，如果取J=k，则为k×1阶向量。即

$$m(\beta) = \begin{pmatrix} m_1(\beta) \\ m_2(\beta) \\ \vdots \\ m_k(\beta) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum\limits_i z_{1i}e_i \\ \frac{1}{n}\sum\limits_i z_{2i}e_i \\ \vdots \\ \frac{1}{n}\sum\limits_i z_{ki}e_i \end{pmatrix}$$

显然，如果$Z_j(j = 1, 2, \cdots, k)$与ε不相关，则$m(\beta) = 0$，求解该方程组可以得到参数β的估计量。这就是读者熟悉的工具变量法。方程组$m(\beta) = 0$的解，就是$m(\beta)'m(\beta)$极小化时的$\hat{\beta}$。

⑵GMM估计方法的定义

一般地，定义GMM估计方法就是极小化

$$q = m(\beta)'W^{-1}m(\beta) \tag{3.6.13}$$

其中权矩阵W为某正定矩阵。GMM估计量就是使(6.4.14)极小化时的参数估计量$\hat{\beta}$，即

$$\hat{\beta} = \arg\min\left(m(\beta)'W^{-1}m(\beta)\right) \tag{3.6.14}$$

(3)权矩阵的选择

关于权矩阵W的选择，是GMM估计方法的一个核心问题。Hansen's(1982)提出最佳的权矩阵为：

$$\begin{aligned} W = Asy.Var[m(\beta)] &= \frac{1}{n^2}\sum_i\sum_j Cov[Z_i\varepsilon_i, Z_j\varepsilon_j] \\ &= \frac{1}{n^2}\sum_i\sum_j \omega_{ij}Z_iZ_j' = \frac{1}{n^2}Z'\Omega Z \end{aligned} \tag{3.6.15}$$

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

$$\hat{W} = \frac{1}{n}S_0 \tag{3.6.16}$$

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

$$\hat{W} = \frac{1}{n}S = \frac{1}{n}(S_0 + \sum_{l=1}^{L} w(l)(S_l + S_l')) \tag{3.6.17}$$

其中

$$w(l) = 1 - \frac{l}{L+1} \quad S_l = \frac{1}{n}\sum_{i=l+1}^{n} \tilde{e}_i\tilde{e}_{i-l}z_iz_{i-l}' \quad l = 0, 1, \cdots, L,$$

$$\tilde{e}_i = e(y_i, X_i, \tilde{\beta}) = y_i - h(X_i, \tilde{\beta}) \quad i = 1, 2, \cdots, n"$$

Lœ`OKæffΓuLœSffl′Θ˘ΛO”$\tilde{\beta}$(6.4.14)中令$W = I$所得到的一个非有效但一致的估计量，或是用其它方法得到的一致估计量。

(4)其它

GMM估计量$\hat{\beta}$的渐近协方差矩阵是：

$$\Sigma = (D'W^{-1}D)^{-1} \tag{3.6.18}$$

其中$D = \frac{\partial m}{\partial \beta}$。由Greene(1997)知：

$$\hat{\beta} \xrightarrow{a} N(\beta, \Sigma)$$

注意：GMM估计是一个大样本估计，它的令人满意的性质仅在大样本情况下才有。GMM估计在大样本情况下是渐近有效的，在小样本情况下是无效的。所以，只有在大样本情况下，才使用GMM方法进行参数估计。

2.估计方法的步骤

根据上述原理，可以把(6.4.9)的GMM估计步骤归纳如下：

⑴采用OLS估计(6.4.9)，求得$\tilde{\beta}$。目的在于求得权矩阵的估计量。

⑵计算权矩阵的估计量。如果采用(6.4.18)的权矩阵估计量，则要首先选择L的值。当模型不存在序列相关时，取$L=1$；当模型存在序列相关时，可以采用广义差分法判断L的取值。权矩阵为J×J阶矩阵。

⑶将权矩阵的估计量代入(6.4.15)得到参数的GMM估计量。

3.GMM和正交性条件

假设由经济理论或先验信息得到关于总体的正交性条件，通常具有形式

$$E[g(Y,X,\beta)]=0$$

其中$g(\cdot)$是关于数据(Y,X)和参数β的$R\times 1$连续函数向量$(R\geq K)$。构造对应于总体正交性条件的样本矩：

$$m(Y,X,\beta)=\frac{1}{n}\sum_{i=1}^{n}g(Y_i,X_i,\beta)$$

GMM估计方法就是极小化

$$m(Y,X,\beta)'W^{-1}m(Y,X,\beta)$$

其中权矩阵W的最佳选择为$var[m(\cdot)]$的一致估计。如果选择了最佳的W，则上式的极小化值在条件矩成立的情况下，渐近地服从自由度为$R-K$的χ^2分布。

正交性条件是特别重要的。再次考虑简单线性回归模型：

$$Y=X\beta+u$$

在经典线性计量经济学模型的理论中，对模型有几条严格的假设条件：模型几乎包括与被解释变量相关的所有变量，随机误差项同方差且正态分布等。不幸的是，这些条件在实际中是很难符合。例如往往存在异方差性，而造成随机误差项异方差的一个经常的原因是由于模型遗漏了某些影响被解释变量的相关变量。于是产生问题：多少个解释变量才能获得一个可靠的参数估计？这个问题并不容易回答。但GMM方法使得该问题在大样本下得到答案，即满足矩条件$E(X'u)=0$。

3.6.3 OLS和ML估计是GMM估计的特例

1.OLS是GMM的特例

考虑标准的线性回归模型：

$$y_i = X_i'\beta + u_i \quad i = 1, \cdots, n \tag{3.6.19}$$

其中X_i是$(k \times 1)$阶解释变量向量。OLS估计的重要假设之一是随机误差项与解释变量不相关，即$E(X_i u_i) = 0$。换言之，隐含的假定是真实值β_0满足下列矩条件：

$$E[X_i(y_i - X_i'\beta_0)] = 0 \tag{3.6.20}$$

令

$$m(\hat{\beta}) = \frac{1}{n}\sum_i X_i(y_i - X_i'\hat{\beta}) = 0 \tag{3.6.21}$$

于是,得到OLS参数估计量为：

$$\hat{\beta}_{OLS} = (XX')^{-1}XY. \tag{3.6.22}$$

因为GMM估计量仅在随机误差项u_i与解释变量不相关的条件下得到，所以参数的GMM估计$\hat{\beta}_{GMM}$与OLS估计$\hat{\beta}_{OLS}$相同。

下面分情况讨论参数估计量的方差。

(1) u_i同方差且序列不相关

假设u_i是同方差且序列不相关，GMM估计$\hat{\beta}_{GMM}$的协方差阵为：

$$\Sigma = \hat{\sigma}^2(\sum_{i=1}^{n} X_i X_i')^{-1} \tag{3.6.23}$$

其中$\hat{\sigma}^2 = (1/n)\sum\limits_{i=1}^{n}\hat{u}_i^2$，$\hat{u}_i = y_i - X_i'\hat{\beta}_{OLS}$为OLS的残差。这与OLS估计的协方差阵相同，所以OLS是GMM的特例。

（2）u_t存在异方差或序列相关

当u_t存在异方差或序列相关时，只要(6.4.21))满足，OLS估计则为一致估计。此时，OLS估计$\hat{\beta}$的标准误差的计算公式要根据异方差或自相关调整为GMM估计的标准误差。

若随机误差项存在异方差且不存在自相关, W用White提出的估计量(6.4.16)进行估计；若随机干扰项存在自相关，W用Newey和West提出的估计量(6.4.17)进行估计。此时，GMM估计$\hat{\beta}$的方差－协方差阵为：

$$E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] \cong (1/n)\left\{(1/n)\sum_{i=1}^{n} X_i X_i'\hat{W}^{-1}(1/n)\sum_{i=1}^{n} X_i X_i'\right\}^{-1}$$

$$= n\left[\sum_{i=1}^{n} X_i X_i'\right]^{-1} \hat{W} \left[\sum_{i=1}^{n} X_i X_i'\right]^{-1} \tag{3.6.24}$$

2.ML是GMM的特例

最大似然估计也有一个GMM解释。回忆最大似然估计也就是最大化对数似然函数，令对数似然函数的导数为零：

$$m(Y, X; \beta) = \frac{\partial \ln L}{\partial \beta} = 0 \tag{3.6.25}$$

该条件为一个矩条件。

$$\min_{\theta}(m(Y, X; \beta)' \cdot H^{-1} \cdot m(Y, X; \beta)) \tag{3.6.26}$$

其中权数矩阵H是矩条件的方差，即$H = -E(\partial^2 \ln L/\partial\beta\,\partial\beta')$。一阶条件为：

$$\frac{\partial^2 \ln L}{\partial\beta\,\partial\beta'} H^{-1} \frac{\partial \ln L}{\partial \beta} = 0 \tag{3.6.27}$$

由此可看出：ML是GMM估计的特例。

GMM公式建议ML估计的方差－协方差阵可用

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \cong (1/n)\{\hat{D}\hat{H}^{-1}\hat{D}'\}^{-1} \tag{3.6.28}$$

来近似。其中

$$\hat{D}' = (1/n)\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}\Big|_{\beta=\hat{\beta}} \tag{3.6.29}$$

当随机干扰项不存在序列相关时，矩阵H的估计为：

$$\hat{H} = (1/n)\sum_{t=1}^{n}\left[\frac{\partial \log f(Y_t, X_t; \beta)}{\partial \beta}\right]\left[\frac{\partial \log f(Y_t, X_t; \beta)}{\partial \beta}\right]'\Big|_{\beta=\hat{\beta}} \tag{3.6.30}$$

然而，为什么许多研究者喜欢采用GMM而不采用ML呢？第一，有时ML估计计算困难，而GMM容易计算；尽管GMM估计的渐进有效性不如ML，但它是一致性估计。第二，尽管确定似然函数的数据分布函数未知，但实际参数满足的矩条件却已知。

此外，条件矩约束不仅可用于估计，还可以用于模型设定的检验。条件矩检验的思路如下：如果模型设定暗含某些矩约束，则在数据取自的总体中也将成立。若模型设定是正确的，则样本数据应该模拟这些暗含的关系。

3.IV是GMM的特例

考虑线性模型

$$Y = \alpha + X_1\beta_1 + \cdots + X_k\beta_k + \varepsilon = \tilde{X}\beta + \varepsilon \tag{3.6.31}$$

其中$\tilde{X} = \begin{bmatrix} 1, & X \end{bmatrix} = \begin{bmatrix} 1, & X_1, & \cdots, & X_k \end{bmatrix}_{n\times(k+1)}$,$\beta = \begin{bmatrix} \alpha, & \beta_1, & \cdots, & \beta_k \end{bmatrix}$

假设$E(X_i'\varepsilon) \neq 0$，此时参数的OLS估计不是一致估计。此时，需用工具变量法。寻找工具变量Z，它与X相关但与ε不相关。例如，Y是公司雇佣人数，X_1是合同工资。我们希望估计劳动需求曲线，但雇佣人数和工资都是需求和供给变化的产物，因而$E(X_1'\varepsilon) \neq 0$。因为合同工资是在事先协议达成的，所有它的一个可能的工具变量为通货膨胀。因为通货膨胀在合同签订时，工会和公司都不能预见，且它改变了实际工资。当通货膨胀升高时，将降低实际工资，劳动需求量将减少。

假设找到工具变量Z的个数大于或等于解释变量X的个数，$\tilde{Z} = \begin{bmatrix} 1 & Z \end{bmatrix}_{n\times(1+L)}$，该问题的正交性条件为$E(Z'\varepsilon) = 0$，所以一个好的估计应是使得对应样本矩为零的估计，即

$$\frac{1}{n}Z'(Y - X\hat{\beta}) = 0 \tag{3.6.32}$$

当方程个数等于参数个数时，明显工具变量估计法等价于GMM估计方法。当方程个数大于参数个数时，即使模型设定正确，总体矩为零，但对应的样本矩却未必为零。参数的GMM估计为

$$\hat{\beta} = Arg\min\{\frac{1}{n}[Z'(Y - X\beta)]' \cdot \hat{W}^{-1} \cdot \frac{1}{n}[Z'(Y - X\beta)]\} \tag{3.6.33}$$

其中$\hat{W}$是$var[(1/n)(Z'\varepsilon)]$的一致估计。

4.2SLS是GMM的特例

当误差项是同方差且序列不相关时，GMM估计为2SLS估计。

因为(5.8.33)中大括号部分的一阶极值条件为：

$$(X'Z)\hat{W}^{-1}(Z'Y - Z'X\hat{\beta}) = 0$$

$$(X'Z)\hat{W}^{-1}Z'Y - (X'Z)\hat{W}^{-1}Z'X\hat{\beta} = 0$$

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

而此时$\hat{W} = \frac{\hat{\sigma}^2}{n^2}Z'Z$，所以

$$\hat{\beta} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z)(Z'Z)^{-1}Z'Y \tag{3.6.34}$$

而这正是2SLS估计。

如果以Z作为X的解释变量进行第一阶段估计，有

$$
\begin{aligned}
&X = Z\pi + \mu \\
&\hat{\pi} = (Z'Z)^{-1}Z'X \\
&\hat{X} = Z\hat{\pi} = Z(Z'Z)^{-1}Z'X
\end{aligned}
$$

对于$Y = X\beta + \varepsilon$进行第二阶段估计，等价于用$\hat{X}$作为X的工具变量，于是得到

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

与(5.8.34)相同。所以2SLS是工具变量估计方法的特殊情形，而工具变量估计是GMM估计的特殊情形。

3.6.4 假设检验

1.用于矩条件检验的Wald、LM和LR检验

为了检验所构造的矩条件是否显著成立，需要进行假设检验。要检验的假设是

$$H_0 : R(\beta) = 0 \tag{3.6.35}$$

其中β是要估计的$K \times 1$参数向量，$R(\beta)$是$J \times 1$向量。$R(\beta)$的分量是β的函数，可能是线性的，也可能是非线性的。

令c_1是β无约束的ML估计，c_0表示有约束的ML估计，即当施加零假设的约束时得到的估计，则利用ML估计构造三个渐近等价的检验统计量：似然比、Wald统计量和拉格朗日乘子统计量，用于假设检验。似然比为：

$$LR = -2[\ln L(c_0) - \ln L(c_1)] \sim \chi^2[J] \tag{3.6.36}$$

其中$L(\beta)$为似然函数。似然比统计量的计算需要约束和无约束ML估计量。

Wald统计量是：

$$Wald = [R(c_1)]'\{Est.Asy.Var[R(c_1)]\}^{-1}[R(c_1)] \sim \chi^2[J] \tag{3.6.37}$$

Wald统计量利用距离度量了无约束限制估计量不满足约束限制的程度。其中渐近协方差矩阵的估计量是：

$$Est.Asy.Var[R(c_1)] = A_1\{Est.Asy.Var[c_1]\}A_1' \tag{3.6.38}$$

其中

$$A_1 = \frac{\partial R(\beta)}{\partial \beta'}|_{\beta=c_1}$$

Wald统计量的计算仅需要无约束ML估计量。
拉格朗日乘子统计量是：

$$LM = g_0'\{Est.Asy.Var[g_0]\}^{-1}g_0 \sim \chi^2[J] \tag{3.6.39}$$

其中

$$g_0 = \frac{\partial \ln L}{\partial \beta}|_{\beta=c_0} \tag{3.6.40}$$

$$Est.Asy.Var[g_0] = \hat{G}'\hat{G} \tag{3.6.41}$$

$$\hat{G} = [\ \hat{g}_1,\ \ \hat{g}_2,\ \ \cdots,\ \ \hat{g}_n\ ]'$$

$$\hat{g}_i = \frac{\partial \ln f(Y_i, X_i, \beta)}{\partial \beta}|_{\beta=c_0}, i = 1, \cdots n$$

LM统计量的计算需要有约束ML估计量。
Newey和West(1987)为GMM估计设计了上述检验统计量的对等物。
LR统计量的对等物

$$LR_{GMM} = q_0 - q_1 \tag{3.6.42}$$

其中q_1是$q = m(\beta)'W^{-1}m(\beta)$在无约束下的最小值，$q_0$是$q$在有约束下的最小值。有必要在有约束和无约束两个估计量中都使用同样的矩阵W。由于无约束估计量在H_0和H_1下都是一致的，因此，采用β的一致无约束估计量计算W。

利用GMM结果而不是ML估计，同样可计算Wald统计量：

$$Wald = [R(\beta_{GMM})]'\{Est.Asy.Var[R(\beta_{GMM})]\}^{-1}[R(\beta_{GMM})] \sim \chi^2[J]$$

(3.6.43)
和(5.8.37)计算的Wald统计量完全相同。
LM统计量的对等物是：

$$LM_{GMM} = [m(c_0)'W^{-1}A(c_0)]\{Est.Asy.Var[c_0]\}^{-1}[m(c_0)'W^{-1}A(c_0)]'$$

(3.6.44)
该LM统计量等价于检验零假设：

$$\alpha = \frac{\partial q}{\partial \beta} = 2G(\beta)'W^{-1}\bar{m}(\beta) \tag{3.6.43}$$

等于0的Wald统计量，即

$$LM_{GMM} = \alpha'(Var[\alpha])^{-1}\alpha\left.\right|_{\beta=c_0} = \bar{m}(\beta)'W^{-1}G[G'W^{-1}G]^{-1}G'W^{-1}\bar{m}(\beta)\left.\right|_{\beta=c_0}$$

(3.6.46)
2. 过度识别限制的检验
在GMM中，允许J¿k，称之为过度识别限制。过度识别限制是否有效，需要进行假设检验。对于线性模型，检验假设：

$$\mathrm{E}(Z'u) = 0 \tag{3.6.44}$$

其中工具变量Z的个数大于待估参数个数。检验分三步进行。

第一步：首先，得到参数β的一个一致估计$\tilde{\beta}$。因为只要权矩阵是正定的，GMM估计为一致估计，$W = Z'Z)^{-1}$得到的2SLS估计为一致估计。当问题是非线性时，取W为单位阵，也得到参数的一致估计。

第二步：计算残差$e = Y - X\tilde{\beta}$。只要随机误差项是独立的，最佳权矩阵$WW(1/n^2)Z'\Omega Z]^{-1}$可用White估计量进行估计。即

$$W_n = \left(\frac{1}{n^2}\sum\nolimits_i Z_i Z_i' e_i^2\right)^{-1}$$

于是可得到参数β的GMM估计$\hat{\beta}_{GMM}$。

第三步：计算检验统计量

$$Test_{GMM} = \left[Z'(Y - X\hat{\beta}_{GMM})\right]'\left(\sum_i Z_i Z_i' e_i^2\right)^{-1}\left[Z'(Y - X\hat{\beta}_{GMM})\right] \tag{3.6.45}$$

在原假设下，它渐近服从于$\chi^2(L-k)$。当随机误差项同分布且序列不相关时，检验统计量具有特别简单的形式：

$$Test_{GMM} = nR^2 \tag{3.6.46}$$

其中R^2是下面回归方程的判定系数：

$$\hat{r} = Z\pi + \varepsilon \tag{3.6.47}$$

其中$\hat{r}=Y-X\hat{\beta}_{GMM}$.

直观上看很清楚，如果$E(Z'u)=0$，则工具变量应该与残差项正交。于是从回归得到的R^2将很小，此时我们将接收过度识别限制有效的假设。

3.条件矩检验

条件矩约束不仅可用于估计，还可以用于模型设定的检验。条件矩检验的思路是：如果模型设定暗含某些矩约束，则在总体中也将成立。若模型设定是正确的，则样本数据应该模拟这些暗含的关系。

例如，在古典回归模型中，同方差性假设暗含如下矩条件：

$$E\{X_i[(Y_i-X_i\beta)^2-\sigma^2]\}=E[X_i(u_i^2-\sigma^2)]=0 \tag{3.6.48}$$

若同方差假设是正确的，则样本矩

$$r=\frac{1}{n}\sum_{i=1}^{n}X_i(e_i^2-s^2) \tag{3.6.49}$$

将接近于零，其中e_i是OLS残差。

在该项检验中需要解决的理论问题是：(1)用公式表述确实对应于假设检验的合适的矩条件；(2)设计适当的样本对等物；(3)设计样本矩估计量接近于零的一个适当的度量。

假设对应于要检验的假设的合适的矩条件为：

$$E[m(Y_i,X_i,Z_i,\beta)]=0 \tag{3.6.50}$$

对应的样本矩为：

$$r=\frac{1}{n}\sum_{i=1}^{n}m_i(Y_i,X_i,Z_i,\hat{\beta})=\frac{1}{n}\sum_{i=1}^{n}\hat{m}_i \tag{3.6.51}$$

根据中心极限定理，有

$$\sqrt{n}r\xrightarrow{d}N[0,\Sigma] \tag{3.6.52}$$

其中Σ为下面要估计的协方差阵。推导Σ估计的表达式非常复杂，但当参数为最大似然估计时，就有非常简单的估计形式。由于最大似然估计为GMM估计的特例。假定用以计算上面各矩的参数向量是通过求解下面方程而得到的：

$$\sum_{i=1}^{n}d(Y_i,X_i,Z_i,\hat{\beta})=\sum_{i=1}^{n}\hat{d}_i=0 \tag{3.6.53}$$

对于线性回归模型，在假设随机误差项为正态分布下，ML等价于OLS。所以，(4.2.53)将是正规方程组：

$$X'e = \sum_{i=1}^{n} X_i'(Y_i - X_i\beta) = 0 \tag{3.6.54}$$

令$D = (d_1 \quad d_2 \quad \cdots \quad d_n)'$,$M = \begin{pmatrix} m_1 & m_2 & \cdots & m_n \end{pmatrix}'$。Pagan和Vella(1989)证明了,$\Sigma$可用下式估计

$$S = \frac{1}{n}[M'M - M'D(D'D)^{-1}D'M] \tag{3.6.55}$$

此时，可得到检验(4.2.50)的Wald统计量为：

$$Wald = nr'S^{-1}r \xrightarrow{d} \chi^2(J) \tag{3.6.56}$$

3.7 本章思考题和综合练习题

一、思考题

1.最大似然法的基本原理是什么？为什么说在计量经济学理论中它比最小二乘法有更重要的意义？

2.简述经典线性联立方程模型的有限信息最大似然估计的步骤，并说明为什么称之为“有限信息估计方法”。

3.当你学完本书后，再对最大似然法的应用进行总结，以理解它在现代计量经济学中的地位。

4.理解可行的广义最小二乘估计、分部回归估计、偏回归估计和交叉估计的概念。

5.根据分部回归估计的原理，解释在实际计量经济学模型估计中，当剔除被检验为不显著的变量后，被保留的变量的参数估计量发生了变化这一现象。

6.计量经济学模型贝叶斯估计的基本原理和工作步骤。

7.局部多项式回归估计的基本思路及优点。

8.局部多项式回归估计中窗宽的选择、核函数的选择、多项式阶数的选择对估计结果各有什么影响？

9.半参数回归模型偏残差估计的原理与步骤。

10.为什么当计量经济学模型设定正确时总存在一些等于0的条件矩？

11.GMM估计中如何求得权矩阵的估计量？

12.为什么说OLS、ML、IV、2SLS估计都是GMM估计的特例？

二、综合练习题

1.建立一个经典线性单方程计量经济学模型，分别采用最大似然法和最小二乘法进行估计。通过该练习学会使用最大似然法应用软件。在目前比较通用的低版本TSP中没有最大似然法，但其它计量经济学软件包一般都包含最大似然法。只要任意选择一种软件就能够满足本书的教学要求。

2.建立一个经典线性单方程计量经济学模型，该模型具有较大的样本容量。采用以下方法分别估计模型：

(1)利用全部样本进行OLS估计；

(2)利用一部分样本进行OLS估计；

(3)利用(2)中未利用的另外一部分样本进行OLS估计，以此构造先验分布信息，包括待估参数的先验均值和先验协方差矩阵；然后假定只有(2)中利用的那一部分样本，对模型进行贝叶斯估计。

比较(1)、(2)、(3)的估计结果，分析贝叶斯估计的优点。

3.用GMM方法估计第2题的模型。如果可能，可选择不同的工具变量，并对估计结果进行比较分析。

第四章 数据类型非经典的计量经济学问题

在经典线性计量经济学模型中，所利用的数据（样本观测值）具有两个特征。一是在一个模型中，或者只利用时间序列数据，或者只利用截面数据；二是作为被解释变量的样本观测值必须是连续的，且与随机误差项同分布。而实际上，仅利用时间序列数据或者只利用截面数据，经常不能满足经济分析的需要，而需要同时利用时间序列数据和截面数据。另外，作为被解释变量的样本观测值有时是不连续的，例如离散数据；或者由于受到条件所限，不能在整个分布域内抽取观测值，等等。凡此种种，都称之为数据类型非经典的计量经济学问题，是本章讨论的对象。这些计量经济学问题，是70年代以来计量经济学领域的重要发展，具有很好的实际价值。本章中，平行数据问题和离散被解释变量数据问题将作为重点，其它问题着重介绍思路。

4.1 平行数据计量经济学模型（一）—一般模型

所谓"平行数据"(Panel Data)，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，具有很好的应用价值。本节将介绍平行数据模型的一般问题，以及变截距模型，关于一些特殊模型，将在下节介绍。

4.2 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，即在不同的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。

平行数据模型简介

单方程平行数据模型的一般形式为：

$$y_{it} = \alpha_i + x_{it}\beta_i + u_{it}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, T \tag{4.2.1}$$

其中x_{it}为$1 \times K$向量，β_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x_{it}和β_i应该写成矩阵$\mathbf{X_{it}}$和$\mathrm{B_i}$，而且有

$$\mathbf{X_{it}} = (x_{1it}, x_{2it}, \cdots, x_{Kit}) \qquad \mathrm{B_i} = (\beta_{1i}, \beta_{2i}, \cdots, \beta_{Ki})'$$

但是为了简化书写，在本节中采用(6.5.1)的表示。误差项u_{it}均值为零，方差为σ_u^2。

模型(6.5.1)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、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，称为变截距模型，在横截面上个体影响不同，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，又分为固定影响和随机影响两种情况。对于情形3，称为变系数模型，除了存在个体影响外, 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。

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，分析方法是完全系统的

4.2.1 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，计量经济学的经验研究大大地丰富了。但平行数据包括两维的数据(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。

如果可获得的数据是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，多数平行数据是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$P(y|\beta)$(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(6.5.1)，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

$$y_{it} = \alpha + x_{it}\beta + u_{it} \quad i = 1, \cdots, n, \quad t = 1, \cdots, T$$

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(如图4.1.1)。当模型(6.5.1)中不同横截面的参数相同$\beta_i = \beta$时，也有同样的问题(如图4.1.2)。

于是，研究平行数据的第一步是检验刻划被解释变量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种情况的哪一种，以确定模型的形式。广泛使用的检验是协方差分析检验。主要检验两个假设：

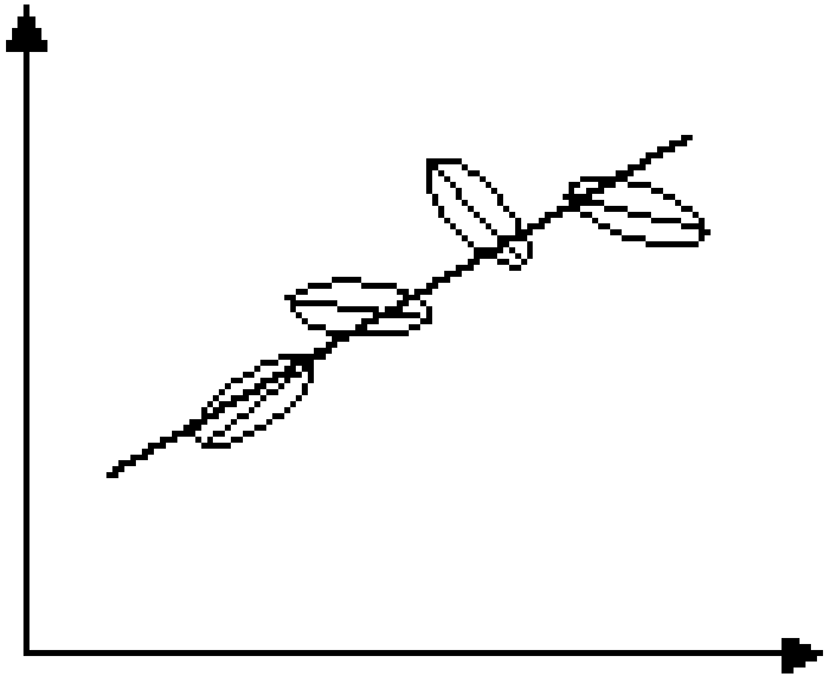
假设1：斜率在不同的横截面样本点上和时间上都相同，但截距不相同。

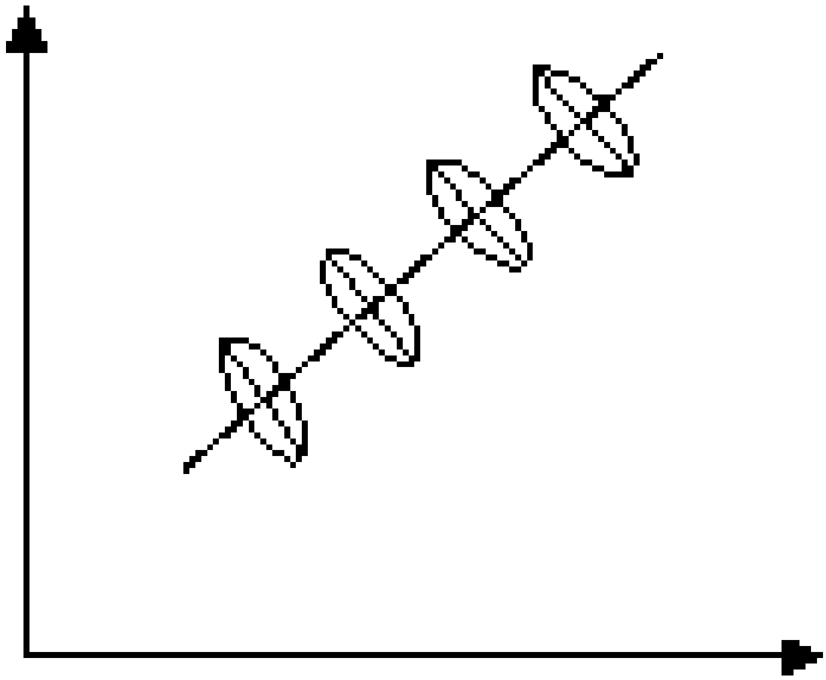
$$H_1 \quad y_{it} = \alpha_i + x_{it}\beta + u_{it} \tag{4.2.2}$$

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。

$$H_2 \quad y_{it} = \alpha + x_{it}\beta + u_{it} \quad 4.1.3$$

显然，如果接收了假设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，判断是否斜率都相等。如果假设1被拒绝，就应该采用模型(6.5.1)。





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

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

$$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it} \tag{4.2.3}$$

模型(6.5.1)参数的最小二乘估计为：

$$\hat{\beta}_i = W_{xx,i}^{-1} W_{xy,i},$$

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i \hat{\beta}_i \tag{4.2.4}$$

称为群内估计，其中

$$W_{xx,i} = \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(x_{it} - \bar{x}_i)$$

$$W_{xy,i} = \sum_{t=1}^{T} (x_{it} - \bar{x}_i)'(y_{it} - \bar{y}_i)$$

$$W_{yy,i} = \sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2 \tag{4.2.5}$$

第i群的残差平方和是$RSS_i = W_{yy,i} - W'_{xy,i} W_{xx,i}^{-1} W_{xy,i}$，模型(6.5.1)的残差平方和为：

$$S_1 = \sum_{i=1}^{n} RSS_i \tag{4.2.6}$$

模型(6.4.2)参数的最小二乘估计为

$$\hat{\beta}_w = W_{xx}^{-1} W_{xy}$$

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i \hat{\beta}_w \tag{4.2.7}$$

其中，$W_{xx} = \sum_{i=1}^{n} W_{xx,i}$,$W_{xy} = \sum_{i=1}^{n} W_{xy,i}$ (4.1.9)

让$W_{yy} = \sum_{i=1}^{n} W_{yy,i}$ (4.1.10)

模型(6.4.2)的残差平方和为

$$S_2 = W_{yy} - W'_{xy} W_{xx}^{-1} W_{xy} \tag{4.2.8}$$

模型(4.1.3)参数的最小二乘估计为

$$\hat{\beta} = T_{xx}^{-1} T_{xy}$$

$$\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta} \tag{4.2.9}$$

其中

$$T_{xx} = \sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x})'(x_{it} - \bar{x})$$

$$T_{xy} = \sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x})'(y_{it} - \bar{y})$$

$$T_{yy} = \sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \bar{y})^2 \tag{4.2.10}$$

$$\bar{y} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it} \quad , \bar{x} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} \tag{4.2.11}$$

模型(4.1.3)的残差平方和为

$$S_3 = T_{yy} - T'_{xy} T_{xx}^{-1} T_{xy} \tag{4.2.12}$$

由此可以得到下列结论：

$$S_1/\sigma_u^2 \sim \chi^2[n(T-K-1)]$$

在H_2下，$S_3/\sigma_u^2 \sim \chi^2[nT-(K+1)]$和$(S_3-S_1)/\sigma_u^2 \sim \chi^2[(n-1)(K+1)]$；

$$(S_3 - S_1)/\sigma_u^2 S_1/\sigma_u^2 "$$

所以，得到检验H_2的F统计量：

$$F_2 = \frac{(S_3 - S_1)/[(n-1)(K+1)]}{S_1/[nT - n(K+1)]} \sim F[(n-1)(K+1), n(T-K-1)] \tag{4.2.13}$$

同时得到下列结论：

在H_1下，$S_2/\sigma_u^2 \sim \chi^2[n(T-1)-K]$和$(S_2-S_1)/\sigma_u^2 \sim \chi^2[(n-1)K]$；

$$(S_2 - S_1)/\sigma_u^2 S_1/\sigma_u^2\text{''}$$

所以，得到检验H_1的F统计量：

$$F_1 = \frac{(S_2 - S_1)/[(n-1)K]}{S_1/[nT - n(K+1)]} \sim F[(n-1)K, n(T-K-1)] \tag{4.2.14}$$

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。

4.2.2 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：

$$y_{it} = \alpha_i + x_{it}\beta + u_{it} \quad i = 1, \cdots, n \quad t = 1, \cdots, T \tag{4.2.15}$$

其中x_{it}为$1 \times K$向量，β为$K \times 1$向量，α_i为个体影响，为模型中被忽略的反映个体差异变量的影响; u_{it}为随机误差项，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x_{it}不相关。

1.固定影响模型：LSDV模型及其参数估计

假定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。令y_i和X_i是第i个单位的T个观测值向量和矩阵，并令u_i是随机干扰项$T \times 1$向量,(6.4.2)可写成：

$$y_i = e\alpha_i + X_i\beta + u_i \quad i = 1, \cdots, n \tag{4.2.16}$$

其中$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T\times 1} \quad e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}_{K\times 1}$

$$u_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix} \quad X_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \cdots & x_{Ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{Ki2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{KiT} \end{pmatrix}_{T\times K}$$

(6.4.17)也可写成：

$$y = [d_1, d_2, \cdots, d_n, X]\begin{bmatrix} \alpha \\ \beta \end{bmatrix} + u \tag{4.2.17}$$

其中

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{nT\times 1} \qquad [d_1, d_2, \cdots, d_n] = \begin{pmatrix} e & 0 & \cdots & 0 \\ 0 & e & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & e \end{pmatrix}_{nT\times n}$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}_{nT\times K} \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_{n\times 1} \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}_{nT\times 1}$$

其中d_i是代表第i个单位的虚拟变量。

令$D = [d_1, d_2, \cdots, d_n]$, 则(6.4.18)等价于：

$$y = D\alpha + X\beta + u \tag{4.2.18}$$

该模型通常被称为最小二乘虚拟变量(LSDV)模型，有时也称之为协方差分析模型(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,参数可由普通最小二乘进行估计。当n很大，甚至成千上万，OLS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$Q = I_T - \frac{1}{T}ee'$，因为$I_T e = \frac{1}{T}ee'e$，所以$Qe = 0$。则由(6.4.17)有

$$Qy_i = Qe\alpha_i + QX_i\beta + Qu_i = QX_i\beta + Qu_i \tag{4.2.19}$$

于是

$$X_i'Qy_i = X_i'QX_i\beta + X_i'Qu_i$$

$$\sum_i X_i'Qy_i = (\sum_i X_i'QX_i)\beta + \sum_i X_i'Qu_i$$

$$\hat{\beta}_{CV} = \left[\sum_{i=1}^{n} X_i'QX_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'Qy_i\right] \tag{4.2.20}$$

由于模型(6.4.19)也叫协方差分析模型，所以参数β的最小二乘虚变量(LSDV)估计也叫做协方差估计。β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。它的协方差阵为：

$$Var(\hat{\beta}_{CV}) = \sigma_u^2 \left[\sum_{i=1}^{n} X_i' Q X_i \right]^{-1} \tag{4.2.21}$$

截距的估计为：

$$\hat{\alpha}_i = \bar{y}_i - \bar{X}_i \hat{\beta}_{CV} \tag{4.2.22}$$

$$Var(\hat{\alpha}_i) = \frac{\sigma_u^2}{T} + \bar{X}_i Var(\hat{\beta}_{CV}) \bar{X}_i' \tag{4.2.23}$$

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σ_u^2的估计量为：

$$s^2 = \sum\nolimits_{i=1}^{n} \sum\nolimits_{t=1}^{T} (y_{it} - \hat{\alpha}_i - x_{it}\hat{\beta}_{CV})^2 / (nT - n - K) \tag{4.2.24}$$

可以利用F检验来检验$\alpha_i = \alpha_j$的假设，在该假设下

$$F = \frac{(R_u^2 - R_p^2)/(n-1)}{(1 - R_u^2)/(nT - n - K)} \tag{4.2.25}$$

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。

2. 固定影响模型的异方差性

在固定影响模型中，在横截面单位上存在异方差，则参数的LSDV估计仍是无偏和一致估计。但参数估计的协方差阵要进行调整。

令$Z = [D, X]$，则参数LSDV估计协方差阵的White一致估计为：

$$Est.Var[\hat{\alpha}, \hat{\beta}_{CV}] = (Z'Z)^{-1} Z' E^2 Z (Z'Z)^{-1} \tag{4.2.26}$$

其中E为最小二乘(LSDV)残差的对角阵。

3. 不齐平行数据的固定影响模型

对于平行数据，常会出现有些横截面个体的数据较多，而另一些横截面个体的数据较少的情况。我们称这种平行数据为不齐平行数据。上述的估计过程只需作稍微修正就可用于这种数据模型的参数估计。第一，设第i横截面个体的数据个数为T_i，在计算$s^2, Var(\hat{\beta}_{CV}), Var(\hat{\alpha}_i)$和$F$统计量时，将总的数据个数$nT$换为$\sum_{i=1}^{n} T_i$。第二，第$i$横截面变量的均值应是$T_i$个数据的平均，变量总平均为：

$$\bar{\bar{x}} = \frac{\sum\limits_{i=1}^{n} \sum\limits_{t=1}^{T_i} x_{it}}{\sum\limits_{i=1}^{n} T_i} = \sum_{i=1}^{n} w_i \bar{x}_i \quad 4.1.29$$

其中$w_i = T_i/(\sum_{i=1}^n T_i)$。于是

$$\hat{\beta}_{CV} = \left[\sum_{i=1}^n X_i' Q_i X_i\right]^{-1} \left[\sum_{i=1}^n X_i' Q_i y_i\right] \tag{4.2.27}$$

其中$Q_i = I_{T_i} - \frac{1}{T_i} ee'$。

4.2.3 随机影响变截距模型

1.随机影响变截距模型的FGLS估计方法

当横截面的单位是总体的所有单位时，固定影响模型是一个合理的模型。如果横截面单位是随机地抽自一个大的总体，该模型仅适用于抽到的横截面单位，而不是样本之外的其它单位。在这种情况下，把总体中个体的差异认为服从随机分布可能更合适。于是(6.4.2)写成：

$$y_{it} = \mu + x_{it}\beta + \alpha_i + u_{it} \tag{4.2.28}$$

其中α_i为模型中被忽略的反映个体差异变量的影响，假定它与随机干扰项u_{it}一样是随机变量。

假定α_i与x_{it}不相关。进一步假定：

$$E[u_{it}] = E[\alpha_i] = 0, E[u_{it}^2] = \sigma_u^2, E[\alpha_i^2] = \sigma_\alpha^2$$

$E[u_{it}\alpha_j] = 0 \quad , E[u_{it}u_{js}] = 0(t \neq s,\ i \neq j),$
$E[\alpha_i\alpha_j] = 0(i \neq j)$ (4.1.32)
令$w_{it} = u_{it} + \alpha_i$，则
$E[w_{it}^2] = \sigma_u^2 + \sigma_\alpha^2, E[w_{it}w_{is}] = \sigma_\alpha^2(t \neq s),$

$$\Omega = E[w_i w_i'] = \sigma_u^2 I_T + \sigma_\alpha^2 ee'$$

$$V = (Ew_i w_j')_{nT\times nT} = I \otimes \Omega \tag{4.2.29}$$

所以，给定x_{it} 下，y_{it}的方差为$\sigma_y^2 = \sigma_u^2 + \sigma_\alpha^2$。称$\sigma_u^2$和$\sigma_\alpha^2$为方差成分，有时也称模型(5.8.28)为方差成分(或误差成分)模型。

由假定可知w_{it}与x_{it}不相关,于是OLS将得到参数的无偏和一致估计，但为什么要采用可行的广义最小二乘法(FGLS)进行参数估计？原因有两方面：第一，OLS虽得到参数的一致估计，但标准误差被低估。第二，OLS估计不如可行的广义最小二乘估计有效。

(1)Ω已知时的GLS估计

将(5.8.28)重写为：

$$y_i = \tilde{X}_i\delta + w_i \quad i = 1, \cdots, n \tag{4.2.30}$$

其中$\tilde{X}_i=(e,X_i)$,$\delta'=(\mu,\beta')$,$w_i'=(w_{i1},\cdots,w_{in})$。容易证明：

$$\Omega^{-1}=\frac{1}{\sigma_u^2}(I_T-\theta^2ee') \quad \Omega^{-1/2}=\frac{1}{\sigma_u}(I_T-\frac{\theta}{T}ee') \tag{4.2.31}$$

其中

$$\theta=1-\frac{\sigma_u}{(T\sigma_\alpha^2+\sigma_u^2)^{1/2}} \tag{4.2.32}$$

对方程(5.8.30)两边都乘以$\Omega^{-1/2}$，得到

$$\Omega^{-1/2}y_i=\Omega^{-1/2}\tilde{X}_i\delta+\Omega^{-1/2}w_i \tag{4.2.33}$$

再进行OLS，就得到参数δ的广义最小二乘估计为：

$$\hat{\delta}=\left[\sum_{i=1}^n \tilde{X}_i'\Omega^{-1}\tilde{X}_i\right]^{-1}\left[\sum_{i=1}^n \tilde{X}_i'\Omega^{-1}y_i\right] \tag{4.2.34}$$

它的协方差阵为：

$$Var(\hat{\delta})=\sigma_u^2\left[\sum_{i=1}^n \tilde{X}_i'\Omega^{-1}\tilde{X}_i\right]^{-1} \tag{4.2.35}$$

(2)Ω未知时的FGLS估计

可行的广义最小二乘估计就是在求出Ω中未知量σ_α^2 和σ_u^2的无偏估计后，进行广义最小二乘估计。对方程(5.8.28)两边在时间上求平均得到：

$$\bar{y}_i=\mu+\bar{X}_i\beta+\alpha_i+\bar{u}_i \tag{4.2.36}$$

(5.8.28)减去(5.8.36)，得到：

$$y_{it}-\bar{y}_i=[x_{it}-\bar{X}_i]\beta+[u_{it}-\bar{u}_i] \tag{4.2.37}$$

Greene(1997)推荐σ_α^2和σ_u^2的无偏估计为：

$$\hat{\sigma}_u^2=\frac{\sum_i\sum_t[(y_{it}-\bar{y}_i)-(x_{it}-\bar{X}_i)\tilde{\beta}]^2}{(nT-n-K)} \tag{4.2.38}$$

$$\hat{\sigma}_\alpha^2=\frac{\sum_{i=1}^n(\bar{y}_i-\tilde{\mu}-\bar{X}_i\tilde{\beta})^2}{n-(K+1)}-\frac{\hat{\sigma}_u^2}{T} \tag{4.2.39}$$

其中$\tilde{\mu}$和$\tilde{\beta}$是μ和β的一致估计。如

$$(\tilde{\mu},\tilde{\beta})' = \left[\sum_{i=1}^{n}(\tilde{X}_i'ee'\tilde{X}_i)\right]^{-1}\left[\sum_{i=1}^{n}(\tilde{X}_i'ee'y_i)\right]$$

(5.8.38)和(5.8.39)估计量是无偏的，但是(5.8.39)在有限样本中可能是负数。有关文献（Wallace和Hussain1969，Maddala1971，Fuller和Battese1974，Amemiya1971）推荐了它的另一个估计量，这里不作介绍。

有了σ_α^2和σ_u^2的无偏估计，就可得到未知数θ和未知矩阵$\Omega^{-1/2}$的估计：

$$\hat{\theta} = 1 - \frac{\hat{\sigma}_u}{(T\hat{\sigma}_\alpha^2+\hat{\sigma}_u^2)^{1/2}}$$

$$\hat{\Omega}^{-1/2} = \frac{1}{\hat{\sigma}_u}(I_T - \frac{\hat{\theta}}{T}ee') \tag{4.2.40}$$

对方程(5.8.30)两边乘以$\Omega^{-1/2}$的估计阵，得到

$$\hat{\Omega}^{-1/2}y_i = \hat{\Omega}^{-1/2}e\mu + \hat{\Omega}^{-1/2}X_i\beta + \hat{\Omega}^{-1/2}w_i \tag{4.2.41}$$

对(5.8.41)进行OLS估计，就得到参数的FGLS估计。

然而，β的任何一致估计量都可以在(5.8.38)和(5.8.39)中使用，这就令人怀疑模型的适应性。所以，最好对随机影响的设定进行检验。

Breush和Pagan(1980)已经基于OLS残差为随机影响模型构造了一种LM检验方法，对于

$$H_0 : \sigma_\alpha^2 = 0$$

$$H_1 : \sigma_\alpha^2 \neq 0 \tag{4.2.42}$$

检验统计量是：

$$LM = \frac{nT}{2(T-1)}\left[\frac{\sum\limits_i(\sum\limits_t\varepsilon_{it})^2}{\sum\limits_i\sum\limits_t\varepsilon_{it}^2} - 1\right]^2 = \frac{nT}{2(T-1)}\left[\frac{\varepsilon'DD'\varepsilon}{\varepsilon'\varepsilon} - 1\right]^2 \tag{4.2.43}$$

其中ε为OLS残差向量。在原假设下，LM服从自由度为1的χ^2分布。

2.Mundlak随机影响模型

Mundlak(1978)批评随机影响模型(5.8.28)忽视了随机影响α_i和解释变量x_{it}的相关性，认为有理由相信在许多情况下，α_i和x_{it}确实相关。例如，估计企业的生产函数，每个企业的产出y_{it}，可能受到不可观察的管理能力α_i的影响。管理越有效率的企业生产更多的产品且使用更多的投入，

管理越无效率的企业生产更少的产品且使用更少的投入。在这种情况下，α_i和x_{it}不是独立的，呈现正相关。忽视这种相关将导致有偏估计。

Mundlak(1978)建议$E(\alpha_i|X_i)$用线性函数近似：

$$\alpha_i = \bar{X}_i a + \omega_i, \omega_i \sim N(0, \sigma_\omega^2) \tag{4.2.44}$$

模型(5.8.28)和(5.8.44)参数的GLS估计为：

$$\hat{\mu}^*_{GLS} = \bar{y} - \bar{X}\hat{\beta}_b$$

$$\hat{\beta}^*_{GLS} = \hat{\beta}_{CV}$$

$$\hat{a}^*_{GLS} = \hat{\beta}_b - \hat{\beta}_{CV} \tag{4.2.45}$$

其中

$$\hat{\beta}_b = \left[\sum_{i=1}^{n}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'\right]^{-1}\left[\sum_{i=1}^{n}(\bar{X}_i - \bar{X})(\bar{y}_i - \bar{y})\right] \tag{4.2.46}$$

为群间估计。

3.随机影响模型的ML估计

当α_i和u_{it}是随机变量且服从正态分布时，对数似然函数为：

$$\begin{aligned}\log L = & -\frac{nT}{2}\log 2\pi - \frac{n}{2}\log|V| \\ & -\frac{1}{2}\sum_{i=1}^{n}(y_i - e\mu - X_i\beta)'V^{-1}(y_i - e\mu - X_i\beta) \\ = & -\tfrac{nT}{2}\log 2\pi - \tfrac{n(T-1)}{2}\log\sigma_u^2 - \tfrac{n}{2}\log(\sigma_u^2 + T\sigma_\alpha^2) \\ & -\frac{1}{2\sigma_u^2}\sum_{i=1}^{n}(y_i - e\mu - X_i\beta)'Q(y_i - e\mu - X_i\beta) \\ & -\frac{T}{2(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{n}(\bar{y}_i - \mu - \bar{X}_i\beta)^2\end{aligned} \tag{4.2.47}$$

其中$|V| = \sigma_u^{2(T-1)}(\sigma_u^2 + T\sigma_\alpha^2)$。对数似然函数的偏导数为：

$$\frac{\partial \log L}{\partial \mu} = \frac{T}{(\sigma_u^2 + T\sigma_\alpha^2)}\sum_{i=1}^{n}(\bar{y}_i - \mu - \bar{X}_i\beta) \tag{4.2.48}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma_u^2}\left[\sum_{i=1}^{n}(y_i - e\mu - X_i\beta)'QX_i - \frac{T\sigma_u^2}{(\sigma_u^2+T\sigma_\alpha^2)}\sum_{i=1}^{n}(\bar{y}_i - \mu - \bar{X}_i\beta)\bar{X}_i\right] \tag{4.2.49}$$

$$\frac{\partial \log L}{\partial \sigma_u^2} = -\frac{n(T-1)}{2\sigma_u^2} - \frac{n}{2(\sigma_u^2+T\sigma_\alpha^2)} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{n}(y_i - e\mu - X_i\beta)'Q(y_i - e\mu - X_i\beta) + \frac{T}{2(\sigma_u^2+T\sigma_\alpha^2)^2}\sum_{i=1}^{n}(\bar{y}_i - \mu - \bar{X}_i\beta)^2 \tag{4.2.50}$$

$$\frac{\partial \log L}{\partial \sigma_\alpha^2} = -\frac{nT}{2(\sigma_u^2+T\sigma_\alpha^2)} + \frac{T^2}{2(\sigma_u^2+T\sigma_\alpha^2)^2}\sum_{i=1}^{n}(\bar{y}_i - \mu - \bar{X}_i\beta)^2 \tag{4.2.51}$$

令上述对数似然函数的偏导数为0，就得到参数的最大似然估计。

4.含个体属性变量的模型

平行数据模型可向许多不同的方向进行推广而不改变分析的方法。例如，在解释变量中可以包括描述个体属性（象性别，种类，社会背景等在时间上不改变的变量）的变量向量z_i。

考虑

$$y_i = e\mu + Z_i\gamma + X_i\beta + e\alpha_i + u_i \tag{4.2.52}$$

其中$Z_i = ez_i$。如果假定α_i是固定常数，则模型(4.2.52)存在多重共线性，因为z_i实际上也是固定常数。因而，γ, μ和α_i不可估计。然而，β仍可以用协方差方法进行估计（只要$\sum_{i=1}^{n} X_i'QX_i$可逆）。在方程(4.2.52)乘以(协方差)变换矩阵Q（参见(6.4.20)），消除$Z_i, e\mu$和$e\alpha_i$后，有

$$Qy_i = QX_i\beta + Qu_i \tag{4.2.53}$$

应用OLS，就得到β的CV估计。

当假定α_i随机且与X_i和Z_i不相关时，同样的方法得到β的CV估计。为了估计γ,将方程(4.2.52)在每个横截面样本点上对时间求平均，得到

$$\bar{y}_i - \bar{X}_i\beta = \mu + z_i\gamma + \alpha_i + \bar{u}_i \tag{4.2.54}$$

将$\alpha_i + \bar{u}_i$视为误差项，最小化$\sum_{i=1}^{n}(\alpha_i + \bar{u}_i)^2$，得到：

$$\hat{\gamma}=\left[\sum_{i=1}^{n}(z_i-\bar{z})(z_i-\bar{z})'\right]^{-1}\left\{\sum_{i=1}^{n}(z_i-\bar{z})[(\bar{y}_i-\bar{y})-(\bar{X}_i-\bar{x})\beta]\right\}$$

$$\hat{\mu}=\bar{y}-\bar{x}\beta-\bar{z}\hat{\gamma} \tag{4.2.55}$$

其中$\bar{z}=\frac{1}{n}\sum\limits_{i=1}^{n}z_i\bar{x}=\frac{1}{n}\sum\limits_{i=1}^{n}\bar{X}_i\bar{y}=\frac{1}{n}\sum\limits_{i=1}^{n}\bar{y}_i$。

将β的CV估计代入到(4.2.55)就得到γ和μ的估计。当n趋于无穷大时，上述估计是一致估计。当n固定，T趋于无穷大时，β的估计仍为一致估计，但γ的估计不再是一致估计。

但是当α_i随机且与X_i和Z_i不相关时，CV估计不是BLUE。参数的BLUE估计是GLS估计：

$$\begin{bmatrix}\hat{\mu}\\ \hat{\gamma}\\ \hat{\beta}\end{bmatrix}=\begin{bmatrix}nT\theta & nT\theta\bar{z} & nT\theta\bar{x}\\ nT\theta\bar{z}' & T\theta\sum_{i=1}^{n}z_i'z_i & T\theta\sum_{i=1}^{n}z_i'\bar{x}_i\\ nT\theta\bar{x}' & T\theta\sum_{i=1}^{n}\bar{x}_i'z_i & \sum_{i=1}^{n}X_i'QX_i+T\theta\sum_{i=1}^{n}\bar{x}_i'\bar{x}_i\end{bmatrix}^{-1}$$

$$\cdot\begin{bmatrix}nT\theta\bar{y}\\ T\theta\sum_{i=1}^{n}z_i'\bar{y}_i\\ \sum_{i=1}^{n}X_i'Qy_i+T\theta\sum_{i=1}^{n}\bar{x}_i'\bar{y}_i\end{bmatrix} \tag{4.2.56}$$

若θ未知，可以用它的一致估计代替。当T固定，GLS比CV更有效。当T趋于无穷大，β的GLS估计趋于CV估计，详细请看Lee(1978)。

5.随机影响模型中异方差和序列相关问题

异方差

假设在不同横截面上随机影响α_i的方差不同，为$\sigma_{\alpha i}^2$。只需对同方差时参数的估计稍作修正，就可适应于异方差情形。首先，将未知参数θ改为：

$$\theta_i=1-\frac{\sigma_u}{(T\sigma_{\alpha i}^2+\sigma_u^2)^{1/2}} \tag{4.2.57}$$

此时，$Var[w_{it}]=\sigma_u^2+\sigma_{\alpha i}^2$。(5.8.38)的估计$\sigma_u^2$仍提供了$\sigma_u^2$的一致估计。$\sigma_u^2+\sigma_{\alpha i}^2$的一致估计为：

$$\frac{\sum\limits_{t=1}^{T}(e_{it}-\bar{e}_i)^2}{T-1} \tag{4.2.58}$$

其中e_{it}为OLS估计的残差。于是得到$\sigma_{\alpha i}^2$的估计为：

$$\hat{\sigma}_{\alpha i}^2=\hat{\sigma}_u^2-\frac{\sum_{t=1}^{T}(e_{it}-\bar{e}_i)^2}{T-1} \tag{4.2.59}$$

(2)序列相关

变截距模型的基本假设是随机误差项u_{it}关于个体随机影响α_i是条件序列不相关。但有时包含在u_{it}的不可观测变量的影响在时间上系统地变化，例如遗漏变量存在序列相关所产生的影响。这些变量不能用常数或在时间上独立同分布的干扰项表示。为此，有必要放松u_{it}序列无关的假定。

考虑模型：

$$y_{it}=\mu+x_{it}\beta+\alpha_i+u_{it} \tag{4.2.60}$$

$$u_{it}=\rho u_{i,t-1}+\varepsilon_{it} \tag{4.2.61}$$

其中ε_{it}是均值为0、方差为σ_ε^2独立同分布的变量。但如果知道ρ，可将模型转换为：

$$y_{it}-\rho y_{i,t-1}=\alpha(1-\rho)+(x_{it}-\rho x_{i,t-1})\beta+\alpha_i(1-\rho)+\varepsilon_{it} \tag{4.2.62}$$

可以经过如下计算过程得到β的渐近有效估计。

第一步：从(5.8.28)减均值消去个体影响α_i，即

$$y_{it}-\bar{y}_i=(x_{it}-\bar{X}_i)\beta+(u_{it}-\bar{u}_i) \tag{4.2.63}$$

第二步：利用(4.2.63)最小二乘残差去估计序列相关系数ρ，或利用Durbin(1960)方法，将$(y_{it}-\bar{y}_i)$对$(y_{i,t-1}-\bar{y}_{i,-1})$和$(x_{i,t-1}-\bar{X}_{i,-1})$进行回归，$\rho$的估计为$(y_{i,t-1}-\bar{y}_{i,-1})$的系数，其中$\bar{y}_{i,-1}=\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1}$，$\bar{X}_{i,-1}=\frac{1}{T}\sum_{t=1}^{T}x_{i,t-1}$。(为简单起见，假设$y_{i0}$和$x_{i0}$可观察.)

第三步：估计σ_ε^2和σ_α^2

$$\begin{aligned}\hat{\sigma}_\varepsilon^2=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{&(y_{it}-\bar{y}_i)-(1-\hat{\rho})\hat{\mu}-\hat{\rho}(y_{i,t-1}-\bar{y}_{i,-1})\\&-[(x_{it}-\bar{X}_i)-(x_{i,t-1}-\bar{X}_{i,-1}\hat{\rho})]\hat{\beta}\}^2\end{aligned} \tag{4.2.64}$$

$$\hat{\sigma}_\alpha^2=\frac{1}{(1-\hat{\rho})^2}\cdot\frac{1}{n}\sum_{i=1}^{n}[\bar{y}_i-\hat{\mu}(1-\hat{\rho})-\hat{\rho}\bar{y}_{i,-1}-(\bar{X}_i-\bar{X}_{i,-1}\hat{\rho})\hat{\beta}]^2-\frac{1}{T}\hat{\sigma}_\varepsilon^2$$

(4.1.72)

第四步：利用ρ、σ_ε^2和σ_α^2的估计计算(4.2.62) 中$\alpha_i(1-\rho)+\varepsilon_{it}$的协方差阵，用GLS方法估计(4.2.62)的参数。

6.不齐平行数据的随机影响模型

在随机影响模型中，不齐平行数据增加了一些估计上的困难。第一，在(5.8.29)中，V不再是$I\otimes\Omega$,这是由于对角分块阵的大小不同。$V^{-1/2}$的第i个对角分块是：

$$\Omega_i^{-1/2}=\frac{1}{\sigma_u}(I_{T_i}-\frac{\theta_i}{T_i}ee'),$$

$$\theta_i=1-\frac{\sigma_u}{(T_i\sigma_\alpha^2+\sigma_u^2)^{1/2}}$$

可见造成横截面间异方差的来源是横截面各个体的时间序列的个数差异。这样，对于GLS或需估计方差成分的FGLS，都需要在变量替换(5.8.33)或(5.8.41)中将θ换成θ_i，将Ω换成Ω_i。

第二个问题是LSDV估计仍给出σ_u^2的一致估计，但不能给出σ_α^2的一致估计。因

$$Var\left[\alpha_i+\frac{\sum_{t=1}^{T_i}u_{it}}{T_i}\right]=\sigma_\alpha^2+\frac{\sigma_u^2}{T_i}$$

随着横截面的不同而不同。若$Q_n=\frac{1}{n}\sum_{i=1}^n\frac{1}{T_i}$的极限存在，则$\sigma_\alpha^2$的一致估计为：

$$\hat\sigma_\alpha^2=\frac{\sum_{i=1}^n(\bar y_i-\tilde\mu-\bar X_i\tilde\beta)^2}{n-(K+1)}--\hat\sigma_u^2Q_n$$

接下去就可进行FGLS估计。

4.2.4 变截距模型实例

为了使读者能够了解建立平行数据变截距模型的主要步骤，下面选择两个简单的实例。

1.一个固定影响变截距模型—我国城镇居民储蓄模型

例4.1.1 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（其它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、上海等12个地区1992-1996年共84组数据，并消除了价格因素。利用上述检验方法，计算得到：

$$S_1 = 1385549 \qquad S_2 = 1744353 \qquad S_3 = 4519840$$

$$F_1 = 0.85 \qquad F_2 = 3.70$$

查F分布表，给定10%的显著性水平，得到临界值：

$$F(22,36) = 1.85 \quad F(11,36) = 2.07$$

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，为一变截距模型。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(6.4.18)的形式。对模型进行估计，本例采用Eviews软件，当然采用其它软件也是一样的。估计结果为：

$$\hat{\beta} = 0.5219$$

$$\begin{array}{lll}
\hat{\alpha}_1 = -577.3510 & \hat{\alpha}_2 = -1029.709 & \hat{\alpha}_3 = -775.1573 \\
\hat{\alpha}_4 = -720.0573 & \hat{\alpha}_5 = -877.6264 & \hat{\alpha}_6 = -1004.523 \\
\hat{\alpha}_7 = -989.0625 & \hat{\alpha}_8 = -1628.885 & \hat{\alpha}_9 = -687.3832 \\
\hat{\alpha}_{10} = -1144.251 & \hat{\alpha}_{11} = -789.0931 & \hat{\alpha}_{12} = -1191.264
\end{array}$$

(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的不同储蓄水平。

2.一个随机影响变截距模型—我国城镇居民消费模型

例4.1.2 建立我国城镇居民消费函数模型。

(1) 模型设定

根据经典模型中关于消费函数模型的讨论，以城镇居民家庭人均年生活费支出为被解释变量，以城镇居民家庭人均年可支配收入和前期人均年生活费支出为被解释变量。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、山西、内蒙古、辽宁、上海、福建、江西、河南、湖南、广东、贵州、四川、陕西、新疆等14个地区1994-1998年共70组数据，并消除了价格因素。利用上述检验方法，计算得到：

$$S_1 = 110795.4 \qquad S_2 = 178591.7 \qquad S_3 = 376378.5$$

$$F_1 = 0.658980 \qquad F_2 = 1.720966$$

查F分布表，给定10%的显著性水平，得到临界值：

$$F(39, 28) = 1.57 \quad F(26, 28) = 1.63$$

由于$F_2 > 1.57$，所以拒绝H_2；由于$F_1 < 1.63$，所以接受H_1。因此模型应该采用第二种形式，为一变截距模型。具体形式为：

$$C_{it} = \alpha_i + \beta_1 I_{it} + \beta_2 C_{i\,t-1} + \mu_{it} \qquad i = 1, 2, \cdots, 14 \qquad t = 1994, \cdots 1998$$

其中C, I分别表示人均年生活费支出和人均年可支配收入。

(2) 固定或者随机影响模型的确定

在本节中并没有介绍固定或者随机影响模型的确定问题，更多地是将它作为一种假设。William H. Greene在《Econometrics Analysis》（Third Edition，1997）中介绍了一种检验方法，称为Hausman检验。该检验构造统计量W：

$$W = [b - \hat{\beta}\,]'\hat{\Sigma}^{-1}[b - \hat{\beta}\,]$$

其中b是LSDV模型的估计结果，$\hat{\beta}$是假定模型为随机影响模型时采用FGLS估计的结果，$\hat{\Sigma}$为LSDV模型或者随机影响模型经过估计后得到的协方差距阵。该统计量服从自由度为k的χ^2分布。一些计量经济学应用软件包提供了计算该统计量的功能，例如SAS软件，并且给出检验结果，即采用随机影响模型好于固定影响模型的概率。

对本例进行Hausman检验，采用SAS软件，结果表明采用随机影响模型好于固定影响模型的概率是96.35%。所以决定采用随机影响模型：

$$C_{it} = \alpha + \beta_1 I_{it} + \beta_2 C_{i\,t-1} + \varepsilon_i + \mu_{it} \qquad i = 1, 2, \cdots, 14 \qquad t = 1994, \cdots 1998$$

⑶模型的估计

采用FGLS方法估计该随机影响变截距模型。在SAS软件中默认的是Ω未知时的FGLS方法。估计结果为：

$$\hat{C}_{it} = 260.277 + 0.5372 I_{it} + 0.2488 C_{i\,t-1}$$
$$i = 1, 2, \cdots, 14 \qquad t = 1994, \cdots 1998$$

所有检验全部通过。

⑷讨论

从结果来看，读者也许会问，它与截距与系数都不变的模型有什么区别？如果采用OLS方法估计模型：

$$C_{it} = \alpha + \beta_1 I_{it} + \beta_2 C_{i\,t-1} + \mu_{it} \qquad i = 1, 2, \cdots, 14 \qquad t = 1994, \cdots 1998$$

得到的参数估计量依次为18.119、0.5154和0.3760，显然不同于前者，而且不能通过全部检验。

从该例可见，在进行模型估计与检验时，选择合适的软件是十分重要的。尽管不同的软件都可以完成模型的估计与检验，但是在某些特定的软件中具有某类模型的特定功能，将带来很大的方便。

4.3 平行数据计量经济学模型（二）—扩展模型

上节讨论了平行数据一般线性模型中变截距模型的理论方法。本节将对具有应用价值的两种扩展模型进行简要的介绍。

4.3.1 变系数模型

在§4.1讨论的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：

$$y_{it} = X_{it}\beta_i + u_{it}, \quad i = 1, \cdots, n; t = 1, \cdots, T \tag{4.3.1}$$

其中X_{it}和β_i是解释变量和参数向量。也可写成

$$y_i = X_i\beta_i + u_i \tag{4.3.2}$$

其中

$$y_i=\begin{pmatrix} y_{i1}\\ y_{i2}\\ \vdots\\ y_{iT}\end{pmatrix}_{T\times 1}\qquad X_i=\begin{pmatrix} x_{1i1} & x_{2i1} & \cdots & x_{Ki1}\\ x_{1i2} & x_{2i2} & \cdots & x_{Ki2}\\ \vdots & \vdots & \vdots & \vdots\\ x_{1iT} & x_{2iT} & \cdots & x_{KiT}\end{pmatrix}_{T\times K}$$

$$\beta_i=\begin{pmatrix} \beta_{i1}\\ \beta_{i2}\\ \vdots\\ \beta_{iK}\end{pmatrix}\quad u_i=\begin{pmatrix} u_{i1}\\ u_{i2}\\ \vdots\\ u_{iT}\end{pmatrix}$$

1. 固定影响模型

当将β_i视为固定的不同的常数时，可写成：

$$y=X\beta+u \tag{4.3.3}$$

其中

$$y=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}_{nT\times 1}\qquad X=\begin{pmatrix} X_1 & 0 & \cdots & 0\\ 0 & X_2 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & X_n\end{pmatrix}_{nT\times nK}$$

$$\beta=\begin{pmatrix} \beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{pmatrix}_{nK\times 1}\qquad u=\begin{pmatrix} u_1\\ u_2\\ \vdots\\ u_n\end{pmatrix}_{nT\times 1}$$

显然，如果随机干扰项在不同横截面个体之间不相关，即$Eu_iu_j'=0,i\neq j$且$Eu_iu_i'=\sigma_i^2I$，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，采用经典单方程模型的估计方法分别估计其参数。即使采用GLS估计同时得到$\beta=(\beta_1',\cdots,\beta_n')'$的GLS估计量，也是与在每个横截面个体上$\beta_i$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，$Eu_iu_j'\neq 0$，则$\beta=(\beta_1',\cdots,\beta_n')'$的GLS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$\Omega_{ij}=Eu_iu_j'$

$$V=\begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1n}\\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ \Omega_{n1} & \Omega_{n2} & \cdots & \Omega_{nn}\end{bmatrix}_{nT\times nT} \tag{4.3.4}$$

参数的GLS估计为：

$$\hat{\beta}_{GLS} = X'V^{-1}X)^{-1}X'V^{-1}y \tag{4.3.5}$$

如何得到协方差矩阵的估计量？一种可行的方法是：首先采用采用经典单方程模型的估计方法分别估计每个横截面个体上β_i，计算残差估计值，以此构造协方差矩阵的估计量，类似于经典单方程模型的GLS那样。

2. 随机影响模型

令$\beta_i = \beta + \alpha_i$，Swamy(1970)假定

$$E\alpha_i = 0 \quad E\alpha_i\alpha_j' = \begin{cases} \Delta & i = j \\ 0 & i \neq j \end{cases}$$

$$Ex_{it}\alpha_j' = 0 \quad Eu_iu_j' = \begin{cases} \sigma_i^2 I_T & i = j \\ 0 & i \neq j \end{cases} \tag{4.3.6}$$

(6.4.2)也可写成：

$$y = X\beta + \tilde{X}\alpha + u \tag{4.3.7}$$

其中

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{nT\times 1} \quad X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}_{nT\times K} \quad \tilde{X} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix}_{nT\times nK}$$

$$u = (u_1', \cdots, u_n')'\alpha = (\alpha_1', \cdots, \alpha_n')'EZ6\tilde{X}\alpha + u\text{œffiffΩffiΩ}1i\text{ffi}$$

$$\Phi_i = X_i\Delta X_i' + \sigma_i^2 I_T \tag{4.3.8}$$

在Swamy假设下，当$(1/nT)X'X$收敛于非零常数阵时，y对X的简单回归将得到参数β的无偏一致估计。但该估计不是有效估计，且计算估计量的协方差阵的普通最小二乘计算公式不正确，常常导致无效的统计推断。

βœZ5ffl$O$GLS估计：

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{n} X_i'\Phi_i^{-1}X_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'\Phi_i^{-1}y_i\right] = \sum_{i=1}^{n} W_i\hat{\beta}_i \tag{4.3.9}$$

其中

$$W_i = \left\{ \sum_{i=1}^{n} [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1} \right\}^{-1} [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1}$$

和

$$\hat{\beta}_i = (X_i' X_i)^{-1} X_i' y_i$$

(6.4.10)的最后表达式说明GLS估计是每一个横截面个体上最小二乘估计的矩阵加权平均。权与它们的协方差成比例。GLS估计的协方差矩阵为：

$$Var(\hat{\beta}_{GLS}) = \left[\sum_{i=1}^{n} X_i' \Phi_i^{-1} X_i \right]^{-1} = \left\{ \sum_{i=1}^{n} [\Delta + \sigma_i^2 (X_i' X_i)^{-1}]^{-1} \right\}^{-1} \tag{4.3.10}$$

Swamy建议使用最小二乘估计$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$和它们的残差$\hat{u}_i = y_i - X_i\hat{\beta}_i$得到$\sigma_i^2$和$\Delta$的无偏估计：

$$\hat{\sigma}_i^2 = \frac{\hat{u}_i' \hat{u}_i}{T-K} = \frac{1}{T-K} y_i' [I - X_i (X_i' X_i)^{-1} X_i'] y_i \tag{4.3.11}$$

$$\hat{\Delta} = \frac{1}{n-1} \sum_{i=1}^{n} (\hat{\beta}_i - n^{-1} \sum_{j=1}^{n} \hat{\beta}_j)(\hat{\beta}_i - n^{-1} \sum_{j=1}^{n} \hat{\beta}_j)' - \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i^2 (X_i' X_i)^{-1} \tag{4.3.12}$$

$$labeleq : gd040212 \tag{4.3.13}$$

4.3.2 动态模型

这里的动态模型，即指包含滞后被解释变量作为解释变量的模型。当采用平行数据作为样本观测值时，变截距模型写为：

$$y_{it} = \gamma y_{i,t-1} + x_{it}\beta + \alpha_i + u_{it} \quad i = 1, \cdots, n; \quad t = 1, \cdots, T \tag{4.3.14}$$

其中$Eu_{it} = 0$

$$Eu_{it}u_{js} = \begin{cases} \sigma_u^2 & i = j(= s \\ 0 & \end{cases} \tag{4.3.15}$$

1. 固定影响模型

首先考虑不包含外生解释变量的情况：

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + u_{it} \quad |\gamma| < 1, \quad i = 1, \cdots, n; t = 1, \cdots, T \tag{4.3.16}$$

如果u_{it}是正态分布且y_{i0}是给定的常数，则动态固定影响模型参数的ML估计在T较小时是有偏估计，详细推导请见Anderson和Hsiao(1982)。利用工具变量法可得到在T固定、$n \to \infty$时参数的一致估计。

取差分消除α_i后，有

$$(y_{it} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (u_{it} - u_{i,t-1}) \quad i = 1, \cdots, n; t = 1, \cdots, T \tag{4.3.17}$$

因为$y_{i,t-2}$和$(y_{i,t-2} - y_{i,t-3})$与$(y_{i,t-1} - y_{i,t-2})$相关，但与$(u_{it} - u_{i,t-1})$不相关，所以它们是有效的工具变量。以$(y_{i,t-2} - y_{i,t-3})$作为$(y_{i,t-1} - y_{i,t-2})$的工具变量，得到：

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{n} \sum_{t=3}^{T} (y_{it} - y_{i,t-1})(y_{i,t-2} - y_{i,t-3})}{\sum_{i=1}^{n} \sum_{t=3}^{T} (y_{i,t-1} - y_{i,t-2})(y_{i,t-2} - y_{i,t-3})} \tag{4.3.18}$$

以$y_{i,t-2}$作为$(y_{i,t-1} - y_{i,t-2})$的工具变量，得到：

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{n} \sum_{t=2}^{T} (y_{it} - y_{i,t-1}) y_{i,t-2}}{\sum_{i=1}^{n} \sum_{t=2}^{T} (y_{i,t-1} - y_{i,t-2}) y_{i,t-2}} \tag{4.3.19}$$

在$n \to \infty$或$T \to \infty$时，都是一致估计。于是进一步得到：

$$\hat{\alpha}_i = \bar{y}_i - \hat{\gamma} \bar{y}_{i,-1}, \quad i = 1, \cdots, n$$

在包含外生解释变量的情况下，即模型(6.4.13)，类似地，首先采用工具变量方法估计差分方程模型，得到γ和β的估计量，然后求得α_i的估计量。

2. 随机影响模型

如果模型(6.4.13)中α_i为随机变量，可以将模型写成：

$$y_{it} = \gamma y_{i,t-1} + z_i \rho + x_{it} \beta + v_{it} \quad i = 1, \cdots, n; t = 1, \cdots, T \tag{4.3.20}$$

其中$|\gamma| < 1 \quad v_{it} = \alpha_i + u_{it} \quad E\alpha_i = Eu_{it} = 0$

$$E\alpha_i z_i = 0 \quad E\alpha_i x_{it} = 0 \quad E\alpha_i u_{jt} = 0$$

$$E\alpha_i \alpha_j = \begin{cases} \sigma_\alpha^2 & i = j \\ 0 & \end{cases}$$

$$Eu_{it}u_{js} = \begin{cases} \sigma_u^2 & i=j, t=s \\ 0 & \end{cases}$$

其中z_i是诸如性别、民族等随时间不变的$1\times K_1$阶属性外生变量向量，x_{it}是随时间变化的$1\times K_2$阶外生变量向量，且让它的第一个元素为1，以代表截距，γ是1×1阶，ρ和β分别是$K_1\times 1$阶和$K_2\times 1$阶的参数向量。下面讨论不同的情况：

情况1：y_{i0}固定。一个横截面个体单位可能开始于某一任意位置y_{i0}，并渐近地移向稳定的位置$(\alpha_i+z_i\rho)/(1-\gamma)+\sum_{j=0}x_{i,t-j}\beta\gamma^j$。但如果何时开始抽样的决定是任意的且与$y_{i0}$的值无关，则将$y_{i0}$视为固定是值得怀疑的，因为$E\alpha_i y_{i0}=0$意味着个体影响$\alpha_i$在第0期对模型不产生影响，但影响第一期以及之后的观察值。

情况2：y_{i0}随机。可假定初始值是随机的，均值为μ_{y0}，方差为σ_{y0}^2。即

$$y_{i0}=\mu_{y0}+\varepsilon_i \quad 4.2.21$$

这个假定的合理性在于人们将y_{it}视为状态，而且并不关心怎样到达初始状态，只需要知道它的分布有有限均值和方差。

情况2a：y_{i0}独立于α_i，即$Cov(\varepsilon_i,\alpha_i)=0$。在这种情况下，初始赋值的影响逐渐随时间消失。模型有点象情况1，初始值与影响α_i是独立的，只不过现在的初始值不是固定的而是来自均值为μ_{y0}、方差为σ_{y0}^2总体的随机变量。

情况2b：y_{i0}与α_i相关。记它们的协方差为$\phi\sigma_{y0}^2$，则随着时间的推移，初始赋值$\varepsilon_i=y_{i0}-\mu_{y0}$通过它与$\alpha_i$的相关性影响$y_{it}$的未来值，并最终达到：

$$\phi\varepsilon_i/(1-\gamma)=\lim_{t\to\infty}E[y_{it}-z_i\rho/(1-\gamma)-\sum_{j=0}^{t-1}x_{i,t-j}\beta\gamma^j|\varepsilon_i]$$

(1)最大似然估计

关于初始条件的不同假定蕴含着不同形式的似然函数。在假定α_i和u_{it}正态分布下，情况1的似然函数为：

$$L_1=(2\pi)^{-nT/2}|V|^{-n/2}\exp\Big\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-y_{i,-1}\gamma$$
$$-Z_i\rho-X_i\beta)'V^{-1}(y_i-y_{i,-1}\gamma-Z_i\rho-X_i\beta\Big\} \quad (4.3.21)$$

其中$y_i=(y_{i1},\cdots,y_{iT})'$　$y_{i,-1}=(y_{i0},\cdots,y_{i,T-1})'$　$Z_i=ez_i$

$$e=(1,\cdots,1)' \quad X_i=(x_{i1},\cdots,x_{iT})' \quad V=\sigma_u^2 I_T+\sigma_\alpha^2 ee'$$

情况2a的似然函数为：

$$L_{2a}=L_1(2\pi)^{-n/2}(\sigma_{y0}^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_{y0}^2}\sum_{i=1}^{n}(y_{i0}-\mu_{y0})^2\right\} \tag{4.3.22}$$

情况2b的似然函数为：

$$L_{2b}=(2\pi)^{-\frac{nT}{2}}(\sigma_u^2)^{-\frac{n(T-1)}{2}}(\sigma_u^2+Ta)^{-\frac{n}{2}}$$
$$\exp\left\{-\tfrac{1}{2\sigma_u^2}\sum_{i=1}^{n}\sum_{t=1}^{T}[y_{it}-\gamma y_{it-1}-z_i\rho-x_{it}\beta-\phi(y_{i0}-\mu_{y0})]^2\right.$$

$$\left.+\frac{a}{2\sigma_u^2(\sigma_u^2+Ta)}\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}[y_{it}-\gamma y_{i,t-1}-z_i\rho-x_{it}\beta-\phi(y_{i0}-\mu_{y0})]\quad\right\}^2\quad\right\}$$

$$\times(2\pi)^{-\frac{n}{2}}(\sigma_{y0}^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_{y0}^2}\sum_{i=1}^{n}(y_{i0}-\mu_{y0})^2\right\}$$

(4.2.24)

其中$a=\sigma_\alpha^2-\phi^2\sigma_{y0}^2$。

使得上述似然函数达到最大化，就得到相应情况下参数的最大似然估计。表4.2.1列出了各种情况下参数估计量的性质。

表 4.3.1: 动态随机影响模型参数MLE估计的一致性质

	参数	n固定且$T\to\infty$	T固定且$n\to\infty$
情况1：y_{i0}固定	γ,β,σ_u^2	一致估计	一致估计
	ρ,σ_α^2	非一致估计	一致估计
情况2：y_{i0}随机			
2a：y_{i0}与α_i独立	γ,β,σ_u^2	一致估计	一致估计
	$\mu_{y0},\rho,\sigma_\alpha^2,\sigma_{y0}^2$	非一致估计	一致估计
2b：y_{i0}与α_i相关	γ,β,σ_u^2	一致估计	一致估计
	$\mu_{y0},\rho,\sigma_\alpha^2,\sigma_{y0}^2,\phi$	非一致估计	一致估计

(2)工具变量估计

当横截面个体单位较多、时期长度较短时，由于不同初始条件下似然函数不同，所以初始条件的错误选择将导致得到的估计与正确的估计不是渐近等价的，也可能不是一致估计。有时，关于初始条件的选择是否正确，并没有什么判断依据。此时，当然希望得到与初始条件无关的参数的

一致估计，同时也为ML迭代过程提供参数的初始值。工具变量法是一种适用的方法。下面是参数一致估计的计算步骤：

第一步：对方程(6.4.19)进行差分，有

$$y_{it} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + (x_{it} - x_{i,t-1})\beta + u_{it} - u_{i,t-1} \tag{4.3.23}$$

可利用$y_{i,t-2}$和$(y_{i,t-2} - y_{i,t-3})$作为$(y_{i,t-1} - y_{i,t-2})$的工具变量，并利用工具变量法得到γ和β的估计$\tilde{\gamma}$和$\tilde{\beta}$。

第二步：将估计出的$\tilde{\gamma}$和$\tilde{\beta}$代入对(6.4.19)在时间上求平均的方程

$$\bar{y}_i - \gamma\bar{y}_{i,-1} - \bar{X}_i\beta = z_i\rho + \alpha_i + \bar{u}_i \tag{4.3.24}$$

其中$\bar{y}_i = \sum_{t=1}^{T} y_{it}/T$，$\bar{y}_{i,-1} = \sum_{t=1}^{T} y_{i,t-1}/T$，$\bar{X}_i = \sum_{t=1}^{T} x_{it}/T$，$\bar{u}_i = \sum_{t=1}^{T} u_{it}/T$。对(6.2.22)用OLS得到$\rho$的估计$\tilde{\rho}$。

第三步：估计σ_u^2和σ_α^2，得到

$$\hat{\sigma}_u^2 = \frac{\sum\limits_{i=1}^{n}\sum\limits_{t=2}^{T}\left[(y_{it} - y_{i,t-1}) - \tilde{\gamma}(y_{i,t-1} - y_{i,t-2}) - \tilde{\beta}(x_{it} - x_{i,t-1})\right]^2}{2n(T-1)} \tag{4.3.25}$$

$$\hat{\sigma}_\alpha^2 = \frac{\sum\limits_{i=1}^{n}(\bar{y}_i - \tilde{\gamma}\bar{y}_{i,-1} - \tilde{z}_i\tilde{\rho} - \bar{X}_i\tilde{\beta})^2}{n} - \frac{1}{T}\hat{\sigma}_u^2 \tag{4.3.26}$$

上述估计与初始值无关。当n或T趋于无穷大时，γ, β和σ_u^2是一致估计。而ρ和σ_α^2的估计只有当n趋于无穷大时，才是一致估计。

4.3.3 关于平行数据模型的讨论

在建立平行数据模型并进行参数估计时，掌握以下几个要点是很有益处的。第一，在研究经济问题时，采用平行数据比单纯采用横截面数据或时间序列数据的优势在那里？第二，平行数据的局限是什么？分析平行数据的计量经济学模型方法的局限是什么？第三，使用平行数据时，怎样能增加参数估计的有效性？

为了尽可能地减少估计的偏差，区分解释变量与误差项的两种相关性是有帮助的。第一种相关是方程内包含的外生变量与应该包含但没有包含的变量的相关。第二种相关是动态模型中滞后被解释变量与误差项的相关。了解相关的不同来源将为参数的一致估计提供重要信息。

如果遗漏的变量的影响对于给定的横截面个体单位在不同时间上保持不变，或在给定的时段对所有横截面个体单位保持不变，则遗漏变量造成的偏差可由下列三种方法之一得以消除：(1)差分样本观察值以消除在横截

面或时间上不变的属性变量的影响。(2)利用虚变量来代表横截面或时间上不变的变量的影响。(3)在给定可观察的外生变量下，给定不可观察变量影响的条件分布。

对于线性回归模型，上述三种方法都可用于消除由于遗漏变量造成的偏差。如果误差随个体和时间而变化且独立同分布，虚变量(固定影响)方法和给定可观察的变量，确定影响和条件分布的随机影响模型都得到斜率系数的协方差估计。在其它假定下，尽管斜率系数的协方差估计可能不是有效的，但却是无偏和一致估计。这样，固定影响方法在经验研究中非常重要。

对于下列两种偏误来源应该加以鉴别，一种是由于忽视了与滞后被解释变量相关的时间误差；另一种是由于对初始观察值不正确的设定。残差与滞后被解释变量的相关并不受时间序列观察长度T的影响，而初始值问题仅在没有特别的信息去设定初始观察值且T较小时才产生。当T较大时，在似然函数中初始观察值的权数可忽略不计，此时才可忽视初始观察值的问题。当T较小且模型和数据都要求将初始观察值视为随机时，正确消除由初始观察值和残差的相关造成的偏误的方法依赖于误差项序列相关的类型。

时间序列数据的自由度缺少和严重的多重共线性常常使确定每个解释变量的个体影响的希望受到挫折。这个问题是由于样本提供的信息不能满足模型的需要而造成。在这种情况下，必须或是增加样本信息或是减少模型要求的信息(如对参数作先验的约束)。而平行数据由于能提供更多的自由度和个体属性方面的信息，所以能缩小模型的信息要求和数据提供的信息之间的距离。

因为平行数据通常包含大量的观察值个数，因而似乎参数估计有效性的问题不如一致性的问题重要，但并非在任何情况下都如此。在变系数模型中，假定模型设定正确，尽管忽视随机系数假定的OLS估计应该是一致估计，但在实践中它们的结果不可信。相反考虑到横截面随机属性的GLS却是有效估计。

尽管平行数据包括横截面和时间两维的数据，但大多数情况是一维的观察值少量(通常是时间序列这一维)，而另一维的观察值很多(通常是横截面这一维)。如果行为方程包含在观察值较多的一维(通常是横截面这一维)变化的变量，则有效地利用观察值较多这一维观察值之间的变化关系以使得另一维提供的较少的信息能被最有利地用于估计公共的参数。

与有效使用数据密切相关的问题是固定影响还是随机影响的判断问题。当推断限于横截面单位时，应考虑固定影响模型。当推断是关于总体的时候，应考虑随机影响模型。

应该说明的是，为了得到未知参数的一致估计，往往需要增加与估计参数相关维的样本数据。这样，区别平行数据n或T或两者是否趋于无穷大是重要的。基于这些信息，人们可判断从平行数据得到的参数估计，哪些是一致估计，哪些不是一致估计。

4.4 离散被解释变量数据计量经济学模型（一）—二元选择模型

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、支持和强烈支持5种态度，可以分别用0、1、2、3和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（Models with Discrete Dependent Variables），或者称为离散选择模型(DCM, Discrete Choice Model)。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)。本节主要介绍二元选择模型。

离散选择模型起源于Fechner于1860年进行的动物条件二元反射研究，1962年，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。70、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80年代初期。

4.4.1 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。选择利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，就需要建立计量经济模型。

例如，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模型。

再如，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模型。

由此可见，二元选择问题在我们的经济生活中是大量存在的。

4.4.2 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：

$$\mathbf{Y} = \mathbf{X}\mathrm{B} + \mathrm{N} \tag{4.4.1}$$

其中$\mathbf{Y}$为观测值为1和0的决策被解释变量，$\mathbf{X}$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(6.5.1)中，对于

$$y_i = \mathbf{X_i}\mathrm{B} + \mu_\mathbf{i}$$

因为$E(\mu_i) = 0$，所以$E(y_i) = \mathbf{X_i}\mathrm{B}$。令

$$p_i = P(y_i = 1) \qquad 1 - p_i = P(y_i = 0)$$

于是

$$E(y_i) = 1 \cdot P(y_i = 1) + 0 \cdot P(y_i = 0) = p_i$$

所以有

$$E(y_i) = P(y_i = 1) = \mathbf{X_i}\mathrm{B}$$

对于该式右端的$\mathbf{X_i}\mathrm{B}$，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，对于随机误差项，有

$$\mu_i = \begin{cases} 1 - \mathbf{X}_i\mathrm{B} & y_i = 1, \text{其概率为}\mathbf{X}_i\mathrm{B} \\ -\mathbf{X}_i\mathrm{B} & y_i = 0, \text{其概率为}1 - \mathbf{X}_i\mathrm{B} \end{cases}$$

显然，具有这种概率结构的随机误差项具有异方差性。由于存在这两方面的问题，所以模型(6.5.1)不能作为实际研究二元选择问题的模型。

2.效用模型

为了使得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：

$$U_i^1 = \mathbf{X}_i \mathrm{B}^{\mathbf{1}} + \varepsilon_i^1 \tag{4.4.2}$$

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：

$$U_i^0 = \mathbf{X}_i \mathrm{B}^0 + \varepsilon_i^0 \tag{4.4.3}$$

请注意，在模型(6.4.2)和(6.4.4)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，即对应于观测值为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，他当然要选择公共交通工具；相反，如果不可观测的$U_i^1 \leq U_i^0$，即对应于观测值为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，他当然要选择私人交通工具。

将(6.4.2)与(6.4.4)相减：

$$U_i^1 - U_i^0 = \mathbf{X}_i(\mathrm{B}^{\mathbf{1}} - \mathrm{B}^{\mathbf{0}}) + (\varepsilon_i^1 - \varepsilon_i^0)$$

记为

$$y_i^* = \mathbf{X_i}\mathrm{B}^+\mu_i^* \tag{4.4.4}$$

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$y_i^*, \mathbf{X_i}, \mathrm{B}$·$\mu_i^*$分别为模型的被解释变量、解释变量、待估计参数和随机误差项。

再来看个体选择$y_i = 1$的概率。显然应该有：

$$P(y_i = 1) = P(y_i^* > 0) = P(\mu_i^* > -\mathbf{X_i}\mathrm{B}) \tag{4.4.5}$$

3.最大似然估计

欲使得模型(6.4.5)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（logistic）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(6.4.6)可以改写为：

$$
\begin{aligned}
P(y_i=1) &= P(y_i^*>0)=P(\mu_i^*>-\mathbf{X_i}\mathrm{B}) \\
&= 1-P(\mu_i^*\le -\mathbf{X_i}\mathrm{B}) \\
&= 1-F(-\mathbf{X_i}\mathrm{B})=F(\mathbf{X_i}\mathrm{B})
\end{aligned} \tag{4.4.6}
$$

至此，可以得到模型(6.4.5)的似然函数

$$
P(y_1,y_2,\cdots,y_n)=\prod_{y_i=0}(1-F(\mathbf{X_i}\mathrm{B}))\prod_{y_i=1}F(\mathbf{X_i}\mathrm{B}) \tag{4.4.7}
$$

即

$$
L=\prod_{i=1}^{n}(F(\mathbf{X_i}\mathrm{B}))^{\mathbf{y_i}}(1-F(\mathbf{X_i}\mathrm{B}))^{\mathbf{1-y_i}} \tag{4.4.8}
$$

对数似然函数为

$$
\ln L=\sum_{i=1}^{n}(y_i\ln F(\mathbf{X_i}\mathrm{B})+(1-y_i)\ln(1-F(\mathbf{X_i}\mathrm{B}))) \tag{4.4.9}
$$

对数似然函数最大化的一阶条件为

$$
\frac{\partial\ \ln L}{\partial\ \mathrm{B}}=\sum_{i=1}^{n}\left[\frac{y_if_i}{F_i}+(1-y_i)\frac{-f_i}{(1-F_i)}\right]\mathbf{X_i}=\mathbf{0} \tag{4.4.10}
$$

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(6.4.11) 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。

4.4.3 二元Probit离散选择模型及其参数估计

Probit模型（在一些教科书中将其译成“概率单位模型”）是将标准正态分布作为(6.4.5) 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冒险。标准正态分布的概率分布函数是

$$
F(t)=\int_{-\infty}^{t}(2\pi)^{-1/2}\exp\left(-x^2/2\right)dx \tag{4.4.11}
$$

概率密度函数是

$$
f(x)=(2\pi)^{-1/2}\exp(-x^2/2) \tag{4.4.12}
$$

1.重复观测值不可以得到情况下二元Probit离散选择模型的参数估计

在重复观测值不可以得到情况下，(6.4.11)写成：

$$\begin{aligned}\frac{\partial \ln L}{\partial \mathrm{B}} &= \sum_{y_i=0} \frac{-f_i}{1-F_i}\mathbf{X_i} + \sum_{y_i=1} \frac{f_i}{F_i}\mathbf{X}_i \\ &= \sum_{i=1}^{n} \left(\frac{q_i f(q_i \mathbf{X}_i \mathrm{B})}{F(q_i \mathbf{X}_i \mathrm{B})} \right) \mathbf{X_i} \\ &= \sum_{i=1}^{n} \lambda_i \mathbf{X_i} \\ &= \mathbf{0}\end{aligned} \tag{4.4.13}$$

其中$q_i = 2y_i - 1$

(6.4.14)是关于β的非线性函数，不能直接求解，需采用§3.1 中提及的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，也将其看成为多个不同的决策者。

例4.3.1 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4.3.1。

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Probit二元离散选择模型参数估计软件（例如TSP中的Probit估计），估计结果为：

$$\hat{\alpha} = -4.7539 \qquad \hat{\beta} = 0.00307 \qquad \ln L = -6.0962$$

计算得到模型(6.4.5)的$\hat{y}_i^*$列于表4.3.1中。可见，虽然输人的是y的观测值，但是作为估计对象的不是原始模型，而是模型(6.4.5)。

2.重复观测值可以得到情况下二元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

$$p_i = P_i + e_i = F(\mathbf{X_i}\mathrm{B}) + e_i \tag{4.4.14}$$

其中

表 4.4.1: 例4.3.1 样本观测值及模拟值

x_i	y_i	$\hat{y}_i{}^*$	$\hat{y}_i{}^*$	x_i	y_i	$\hat{y}_i{}^*$	$\hat{y}_i{}^*$
x_i	y_i	(probit)	(logit)	x_i	y_i	(probit)	(logit)
100	0	-4.4472	-7.6029	1600	0	0.1533	0.2622
200	0	-4.1405	-7.0786	1700	1	0.46	0.7865
300	0	-3.8338	-6.5543	1800	0	0.7667	1.3108
400	0	-3.5271	-6.0299	1900	1	1.0734	1.8352
500	0	-3.2204	-5.5056	2000	1	1.3801	2.3595
600	0	-2.9137	-4.9812	2100	1	1.6868	2.8839
700	0	-2.607	-4.4569	2200	1	1.9935	3.4082
800	0	-2.3003	-3.9326	2300	1	2.3002	3.9325
900	0	-1.9936	-3.4082	2400	1	2.6069	4.4569
1000	0	-1.6869	-2.8838	2500	1	2.9136	4.9812
1100	0	-1.3802	-2.3595	2600	1	3.2203	5.5056
1200	0	-1.0735	-1.8352	2700	1	3.527	6.0299
1300	1	-0.7668	-1.3109	2800	1	3.8337	6.5542
1400	0	-0.4601	-0.7865	2900	1	4.1404	7.0786
1500	1	-0.1534	-0.2622	3000	1	4.4471	7.6029

$$
\begin{aligned}
&E(e_i) = 0 \\
&Var(e_i) = p_i(1 - p_i)/n_i
\end{aligned}
$$

对于标准正态分布的概率分布函数(6.4.12)，定义“观测到的”概率单位为

$$
v_i = F^{-1}(p_i) = F^{-1}(P_i + e_i) \tag{4.4.15}
$$

其中$F^{-1}(\bullet)$是标准正态分布的概率分布函数的反函数。用台劳级数展开(6.4.16)，只保留一阶项，则有

$$
F^{-1}(P_i + e_i) = F^{-1}(P_i) + \frac{e_i}{f(F^{-1}(P_i))} \tag{4.4.16}
$$

于是(6.4.16)可以改写为

$$
v_i = F^{-1}(P_i) + u_i
$$

其中

$$
\begin{aligned}
&E(u_i) = 0 \\
&Var(u_i) = \tfrac{P_i(1-P_i)}{n_i(f(F^{-1}(P_i)))^2}
\end{aligned}
$$

因为

$$F^{-1}(P_i) = \mathbf{X_i}\mathrm{B}$$

有

$$\begin{aligned} v_i &= \mathbf{X_i}\mathrm{B} + u_i \\ \mathbf{V} &= \mathbf{X}\mathrm{B} + \mathbf{U} \end{aligned} \tag{4.4.17}$$

采用广义最小二乘法估计(6.4.18)，得到

$$\hat{\mathrm{B}} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{V} \tag{4.4.18}$$

其中Ω为$\mathbf{U}$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

$$\hat{\mathrm{B}} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{V} \tag{4.4.19}$$

而$\hat{\Omega}$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。

(6.4.20)中$\mathbf{V}$的观测值通过求解标准正态分布的概率分布函数

$$p_i = \int_{-\infty}^{v_i} (2\pi)^{-1/2} \exp\left(-t^2/2\right) dt$$

的反函数得到，而其中的p_i是实际观测得到的。为了使得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次数（例如10次左右）的观测。

4.4.4 二元Logit离散选择模型及其参数估计

Logit模型（在一些教科书中将其译成“对数成败比例模型”）是将逻辑（logistic）分布作为(6.4.6) 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

$$F(t) = \frac{1}{1+e^{-t}} \tag{4.4.20}$$

概率密度函数是

$$f(t) = \frac{e^{-t}}{(1+e^{-t})^2} \tag{4.4.21}$$

(6.4.21)可以改写成：

$$F(t)=\frac{e^t}{1+e^t}=\Lambda(t) \tag{4.4.22}$$

这里$\Lambda(\bullet)$是通常用来表示逻辑分布的概率分布的符号。(6.4.22)可以改写成：

$$f(t)=\frac{e^t}{(1+e^t)^2}=\Lambda(t)(1-\Lambda(t)) \tag{4.4.23}$$

1.重复观测值不可以得到情况下二元logit离散选择模型的参数估计

在重复观测值不可以得到情况下，将(6.2.22) 和(6.2.23)代入(6.4.11)，得到：

$$\begin{aligned}\frac{\partial\ \ln L}{\partial\ \mathrm{B}} &= \sum_{i=1}^{n}\left[\frac{y_i f_i}{F_i}+(1-y_i)\frac{-f_i}{(1-F_i)}\right]\mathbf{X_i} \\ &= \sum_{i=1}^{n}(y_i-\Lambda(\mathbf{X_i}\mathrm{B})\ )\mathbf{X_i}=\mathbf{0}\end{aligned} \tag{4.4.24}$$

(6.2.24)是关于β的非线性函数，不能直接求解，需采用§3.1中提及的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4.3.1，利用最简单的logit二元离散选择模型参数估计软件（例如TSP中的logit估计），估计结果为：

$$\hat{\alpha}=-8.1273 \qquad \hat{\beta}=0.00524 \qquad \ln L=-6.2599$$

计算得到模型(6.4.5)的$\hat{y}_i^*$列于表4.3.1中。同样可见，虽然输人的是y的观测值，但是作为估计对象的不是原始模型，而是模型(6.4.5)。

2.重复观测值可以得到情况下二元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(6.4.21)可以得到：

$$\frac{F(t)}{1-F(t)}=e^t \tag{4.4.25}$$

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

$$p_i=P_i+e_i=F(\mathbf{X_i}\mathrm{B})+e_i \tag{4.4.26}$$

其中

$$E(e_i) = 0$$
$$Var(e_i) = p_i(1-p_i)/n_i$$

用样本重复观测得到的p_i构成“成败比例” $p_i/(1-p_i)$，取对数并进行台劳展开，有

$$\ln(\frac{p_i}{1-p_i}) \approx \ln(\frac{P_i}{1-P_i}) + \frac{e_i}{P_i(1-P_i)} \tag{4.4.27}$$

在(6.2.25)中，用P_i代替$F(t)$，再用$\mathbf{X}_i\mathrm{B}$代入t，然后代入(5.8.27)，得到

$$\ln(\frac{p_i}{1-p_i}) \approx \ln(e^{\mathbf{X}_i\mathrm{B}}) + u_i = \mathbf{X}_i\mathrm{B} + u_i \tag{4.4.28}$$

令$v_i = \ln(\frac{p_i}{1-p_i})$
则有

$$\begin{aligned} v_i &= \mathbf{X_i}\mathrm{B} + u_i \\ \mathbf{V} &= \mathbf{X}\mathrm{B} + \mathbf{U} \end{aligned} \tag{4.4.29}$$

采用广义最小二乘法估计(5.8.29)，得到

$$\hat{\mathrm{B}} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{V} \tag{4.4.30}$$

其中$\hat{\Omega}$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。$\mathbf{V}$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。

4.4.5 二元离散选择模型的变量显著性检验

对于二元离散选择模型，经典模型中采用的变量显著性t检验仍然是有效的。在进行参数估计的同时，计算得到各个解释变量的t统计量，可以决定省略哪些变量。现在的问题是，如果省略的变量与保留的变量不是正交的，那么对参数估计量将产生影响，需要进一步检验这种省略是否恰当。

零假设为：$H_0 : \mathbf{Y}^* = \mathbf{X}_1\mathrm{B}_1 + \mu^*$

备择假设为：$H_1 : \mathbf{Y}^* = \mathbf{X}_1\mathrm{B}_1 + \mathbf{X}_2\mathrm{B}_2 + \mu^*$

其中$\mathbf{X}_1$是保留的变量向量，$\mathbf{X}_2$是省略的变量向量。用于检验的统计量为Wald统计量、LR统计量（最大似然比）和LM统计量（拉格朗日乘子），具体计算方法如下：

$$W = \hat{\mathrm{B}}_2' V_2^{-1} \hat{\mathrm{B}}_2 \tag{4.4.31}$$

其中$V_2 = AsyVar(\hat{\mathrm{B}}_2)$。

$$LR = -2(\ln \hat{L}_0 - Ln\hat{L}_1) \tag{4.4.32}$$

其中$\hat{L}_0$、$\hat{L}_1$分别为H_0情形和H_1情形下的似然函数值的估计量。

$$LM = g_0' V_0^{-1} g_0 \tag{4.4.33}$$

其中g_0是H_1情形下的对数似然函数对参数估计量的一阶导数向量，用H_0情形下的最大似然参数估计量代入计算；V_0是H_1情形下参数最大似然估计量的方差矩阵估计量。

上述3个统计量服从χ^2分布，自由度为$\mathbf{X}_2$中变量数目。给定显著性水平α，查χ^2分布临界值，与计算得到的统计量实际值进行比较，如果统计量实际值大于临界值，拒绝H_0，接受H_1。

直观上也可以看出上述统计量的检验功能。例如对于LR统计量，如果$\mathbf{X}_2$中的变量省略后对参数估计量没有影响，那么H_0和H_1情况下的对数最大似然函数值应该相差不大，此时LR统计量的值很小，自然会小于临界值，不拒绝H_0。再例如对于Wald统计量，如果忽略$V_2 = AsyVar(\hat{\mathrm{B}}_2)$中不同参数估计量之间的协方差，它是一个对角阵，元素为参数估计量的方差。于是$W = \hat{\mathrm{B}}_2' V_2^{-1} \hat{\mathrm{B}}_2$ 实际上相当于H_1情形下$\mathbf{X}_2$中的变量对应的显著性t检验统计量的平方和。如果$\mathbf{X}_2$中的变量确实不显著，所有的t检验统计量的值都比较小，所以Wald统计量的值也会很小，因小于临界值而不拒绝H_0。

4.5 离散被解释变量数据计量经济学模型（二）—多元选择模型

除了二元选择模型之外，70、80年代以来，还发展了许多其它类型的离散被解释变量计量经济学模型，例如多元选择模型（Model or Multiple Choices）、离散计数数据模型(Model for Count Data)、联立方程二元选择模型（Multivariate Binary Choice Model）等。本节将对多元选择模型作简单的概念性介绍，选自《Econometric Analysis》（Third Edition, William H. Greene, Prentice-Hall Inc. 1997）一书，对多元选择模型和其它类型的离散被解释变量计量经济学模型有兴趣的读者可以直接查阅该书和其它专门著作，例如《Econometric Analysis of Count Data》（Rainer Winkelmann, Springer,1997）等。

4.5.1 经济生活中的多元离散选择问题

在实际经济生活中，经常遇到多元离散选择问题。

一类问题是将选择对象按照某个准则排队，由决策者从中选择。例如，一个人购买某类商品，将可供选择的商品按照质量排队，

以0、1、2、3等序数表示。影响消费者选择的因素有两类，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，就需要建立多元选择计量经济模型。这类问题的例子还有职业选择、保险等级的选择等等。

另一类问题是决策者对同一个选择对象的偏好程度。例如，仍然是购买问题。同一种商品，不同的消费者对它的偏好是不同的，诸如十分喜欢、一般喜欢、无所谓、一般厌恶、十分厌恶，以0、1、2、3、4表示。影响消费者偏好的因素有两类，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、对该商品的需求程度等。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，同样可以建立多元选择计量经济模型。再如，求职者对某种职业的选择问题。也可以描述为一个多元选择问题。还可以扩展到其它领域，一个典型的例子就是投票问题。在我们进行的任何投票中，在选票上一般都有3栏：同意、弃权、反对，这就是一个3元选择问题。如果进行投票前的调查，甚至可以在调查表上列出5栏：强烈支持、一般支持、弃权、一般反对、强烈反对，这就是一个5元选择问题。

以上两类问题都属于排序选择问题。还有一类问题是同一个决策者面临多项选择，而且多项选择之间没有排序。例如一个人如何去上班，自己开车、乘出租车、乘公共车、骑自行车，也是一个多元选择问题。这类问题更具有普遍性，我们将它称为一般的多元选择问题。

接下来首先讨论描述一般的多元选择问题的模型方法。关于排序选择问题，即排序的多元选择模型(Model for Ordered Data)，将在（三）中介绍。

4.5.2 一般多元离散选择Logit模型

在多元离散选择模型中，因为Probit模型需要对多元正态分布的整体进行评价，所以它的应用受到限制。而逻辑分布更适合于效用最大化时的分布选择，所以应用最多的多元离散选择模型是Logit模型，这里也仅介绍这类模型。

1.一般多元选择Logit模型的思路

在二元选择模型中，通过构造选择的效用模型，将选择问题转化为效用比较问题，克服了直接构造选择结果模型（即以选择结果为被解释变量）所带来的障碍。在一般多元选择模型中，仍然沿着这一思路，通过构造选择的效用模型，以效用的最大化来表示对某一方案的选择，达到估计模型总体参数的目的。

类似于二元选择模型中的(4.3.2)，如果决策者i在（J+1）项可供选择方案中选择了第j项，那么其效用模型为：

$$U_{ij} = \mathbf{X_{ij}}\mathrm{B} + \varepsilon_{\mathbf{ij}} \tag{4.5.1}$$

并且应该有：

$$P(U_{ij} > U_{ik}) \qquad k = 0, 1, 2, \cdots, J \quad k \neq j$$

McFadden于1973年指出，如果（J+1）个随机误差项互不相关，并且服从Weibull分布，即

$$F(\varepsilon_{ij}) = e^{-\varepsilon_{ij}}$$

于是有

$$P(y_i = j) = \frac{e^{\mathbf{X_{ij}}\mathrm{B}}}{\sum\limits_{j=0}^{J} e^{\mathbf{X_{ij}}\mathrm{B}}} \tag{4.5.2}$$

这就导出了后面的条件Logit模型（Conditional Logit Model）和多元名义Logit模型(Multinomial Logit Model)。

效用模型的解释变量中包括所有影响选择的因素，既包括决策者所具有的属性，也包括备选方案所具有的属性。备选方案所具有的属性是随着方案的变化而变化的，例如在不同的交通方式选择中，费用、耗费时间、安全性等因素是随着不同的交通方式变化的；决策者所具有的属性中一部分是随着方案的变化而变化的，例如决策者对不同交通方式的偏好程度等，而一部分是不随着方案的变化而变化的，例如决策者的年龄、收入等。用$\mathbf{Z_{ij}}$表示随着方案的变化而变化的那部分解释变量，$\mathbf{W_i}$表示不随着方案的变化而变化的那部分解释变量，于是(6.4.2)变为：

$$P(y_i = j) = \frac{e^{\mathbf{Z_{ij}}\Gamma + \mathbf{W_i}\mathrm{A}}}{\sum\limits_{j=0}^{J} e^{\mathbf{Z_{ij}}\Gamma + \mathbf{W_i}\mathrm{A}}} = \frac{e^{\mathbf{Z_{ij}}\Gamma} e^{\mathbf{W_i}\mathrm{A}}}{\sum\limits_{j=0}^{J} e^{\mathbf{Z_{ij}}\Gamma} e^{\mathbf{W_i}\mathrm{A}}} \tag{4.5.3}$$

式中Γ, A分别表示$\mathbf{Z_{ij}}$和$\mathbf{W_i}$的系数向量，用γ和α的大写字母表示。

显然，$\mathbf{W_i}$对于选择第j个方案的概率不产生影响，可以从模型中去掉。因为(6.4.4)可以写成：

$$P(y_i = j) = \frac{e^{\mathbf{Z_{ij}}\Gamma} e^{\mathbf{W_i}\mathrm{A}}}{e^{\mathbf{W_i}\mathrm{A}} \sum\limits_{j=0}^{J} e^{\mathbf{Z_{ij}}\Gamma}} = \frac{e^{\mathbf{Z_{ij}}\Gamma}}{\sum\limits_{j=0}^{J} e^{\mathbf{Z_{ij}}\Gamma}} \tag{4.5.4}$$

比较(6.4.5)和(6.4.2)，从形式上没有区别，只是包含的解释变量数量不同。为了书写方便，在后面的模型中，我们仍然采用(6.4.2)的形式。

实用的一般多元Logit选择模型又分3种情况。一是研究选择某种方案的概率与决策者的特征变量之间的关系；二是研究选择某种方案的概率与决策者的特征变量以及方案的特征变量之间的关系；三是考虑到不同方案之间的相关性的情况。下面将分别介绍特定的模型及其估计。

2.多元名义Logit离散选择模型及其参数估计

研究一类特定的问题。例如职业选择问题，有（J+1）种职业，以受教育程度、工作经历、性别、种族作为从事各种职业的个人的特征变量，利用大量的样本，试图研究每种职业与这些特征变量的关系。那么具有特征$\mathbf{X_i}$的个体选择第j种职业的概率为：

$$P(y_i=j)=\frac{e^{\mathbf{X_i}\mathrm{B}_j}}{\sum\limits_{j=0}^{J}e^{\mathbf{X_i}\mathrm{B}_j}} \tag{4.5.5}$$

这里矩阵的下标与(6.4.2)不同，因为在特定的研究问题中，$\mathbf{X}$中未包含备选方案所具有的属性的变量，而参数向量B对不同的选择方案（即不同的方程）是不同的。

为了研究方便，进行标准化处理，令$\mathrm{B}_0=\mathbf{0}$。于是有

$$P(y_i=j)=\frac{e^{\mathbf{X_i}\mathrm{B}_j}}{1+\sum\limits_{k=1}^{J}e^{\mathbf{X_i}\mathrm{B}_k}} \quad j=1,2,\cdots,J$$

$$P(y_i=0)=\frac{1}{1+\sum\limits_{k=1}^{J}e^{\mathbf{X_i}\mathrm{B}_k}} \tag{4.5.6}$$

可以计算得到对数成败比为：

$$\ln(\frac{P_{ij}}{P_{i0}})=\mathbf{X_i}\mathrm{B}_\mathbf{j}$$

$$\ln(\frac{P_{ij}}{P_{ik}})=\mathbf{X_i}(\mathrm{B}_\mathbf{j}-\mathrm{B}_\mathbf{k}) \tag{4.5.7}$$

由于假设了原模型(6.5.1)中（J+1）个随机误差项互不相关，所以成败比率P_j/P_k 与其它备选方案无关。实际上这一假设从经济行为上讲并不合理，本节后面将讨论放弃这一假设的情况。

多元名义Logit离散选择模型的参数估计并不复杂。对于第i个决策者，如果选择了第j个备选方案，令$d_{ij}=1$；如果没有选择第j个备选方案，令$d_{ij}=0$。同时，对于第i个决策者，在（J+1）个备选方案中，只能选择其中之一，即只能存在1个$d_{ij}=1$。于是，可以写出$y_{ij}(i=1,2,\cdots,n;j=0,1,2,\cdots,J)$的联合概率函数，由联合概率函数导出似然函数，进而得到对数似然函数为：

$$\ln L = \sum_{i=1}^{n}\sum_{j=0}^{J} d_{ij} \ln P(y_i = j) \tag{4.5.8}$$

其微分形式为：

$$\frac{\partial \ \ln L}{\partial \ \mathrm{B_j}} = \sum_{i} (d_{ij} - P_{ij})\mathbf{X_i} \qquad j = 1, 2, \cdots, J$$

$$\frac{\partial^2 \ln L}{\partial \mathrm{B_j} \partial \mathrm{B_l'}} = -\sum_{i=1}^{n} P_{ij}(\mathbf{1}(j = l) - P_{il})\mathbf{X}_i\mathbf{X'}_i \tag{4.5.9}$$

其中

$$\mathbf{l}(j = l) = \begin{cases} 1 & \text{如果} j = l \\ 0 & \text{如果} j \neq l \end{cases}$$

(6.4.10)=0即为对数似然函数最大化的一阶条件，利用Newton迭代方法可以迅速地得到方程组的解，得到模型的参数估计量。

但是，该模型参数估计量的经济意义很难直接解释。将(6.4.7)式对$\mathbf{X_i}$微分，得到：

$$\Delta_{\mathbf{j}} = \frac{\partial \ P_j}{\partial \ \mathbf{X_i}} = P_i(\mathrm{B_j} - \sum_{k=0}^{l} P_k \mathrm{B}_k) = P_j(\mathrm{B}_j - \bar{\mathrm{B}}) \tag{4.5.10}$$

从中可以看出，模型参数具有对概率的边际贡献率的经济意义。

3.多元条件Logit离散选择模型及其参数估计

从(6.4.6)可见，在上面的模型中，只考虑了选择某种方案的概率与决策者的特征变量之间的关系。如果选择某种方案的概率不仅与决策者的特征变量有关，而且也与方案的特征变量有关，那么(6.4.6)应该改写为：

$$P(y_i = j) = \frac{e^{\mathbf{X_{ij}}\mathrm{B}_j}}{\sum\limits_{j=1}^{J} e^{\mathbf{X_{ij}}\mathrm{B}_j}} \tag{4.5.11}$$

称为条件Logit模型（Conditional Logit Model）.

类似于名义Logit 模型的参数估计方法，模型的对数似然函数为：

$$\ln L = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \ln P(y_i = j) \tag{4.5.12}$$

其微分形式为：

$$\frac{\partial \ \ln L}{\partial \, \mathrm{B}} = \sum_{i=1}^{n} \sum_{j=1}^{J} d_{ij}(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)$$

$$\frac{\partial^2 \ln L}{\partial \, \mathrm{B} \partial \, \mathrm{B}'} = -\sum_{i=1}^{n} \sum_{j=1}^{J} P_{ij}(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)' \tag{4.5.13}$$

其中

$$\mathbf{l}(j = l) = \begin{cases} 1 & \text{如果} j = l \\ 0 & \text{如果} j \neq l \end{cases}$$

$$\bar{\mathbf{X}}_i = \sum_{j=1}^{J} P_{ij} \mathbf{X}_{ij}$$

(6.4.14)=0即为对数似然函数最大化的一阶条件，利用Newton迭代方法可以迅速地得到方程组的解，得到模型的参数估计量。

4.Nested Logit模型

如果在多元离散选择模型(6.5.1)中，放弃（J+1）个随机误差项互不相关的假设，即不同的选择方案之间具有相关性，而且必须考虑这种相关性，那么一种可行的思路是将（J+1）个选择方案分为l组，在每组内部的选择方案之间不具有相关性，而组间则具有相关性。也就是将条件Logit模型中隐含的齐次方差性条件放松，允许方差在组间可以不同，但在组内仍然是同方差的。这样的模型被称为Nested Logit模型。

用$\mathbf{X_{j|l}}$表示在第l组内对选择第j种方案产生影响的变量，$\mathbf{Z_l}$表示对选择第l组产生影响的变量。于是

$$P(j,l) = P_{jl} = \frac{e^{\mathbf{X_{j|l}}\mathrm{B} + \mathbf{Z_l}\Gamma}}{\sum\limits_{l=1}^{L} \sum\limits_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B} + \mathbf{Z_l}\Gamma}} \tag{4.5.14}$$

即

$$P_{jl} = P_{j|l} P_l = \left(\frac{e^{\mathbf{X_{j|l}}\mathrm{B}}}{\sum\limits_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}} \right) \left(\frac{e^{\mathbf{Z_l}\Gamma}}{\sum\limits_{l=1}^{L} e^{\mathbf{Z_l}\Gamma}} \right) \frac{\left(\sum\limits_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}} \right) \left(\sum\limits_{l=1}^{L} e^{\mathbf{Z_l}\Gamma} \right)}{\left(\sum\limits_{l=1}^{L} \sum\limits_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B} + \mathbf{Z_l}\Gamma} \right)} \tag{4.5.15}$$

(6.4.15) 与(6.4.16)没有任何区别，只是(6.4.16)用条件概率表示。定义

$$I_l = \ln \sum_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}$$

为第l组的“内值”（Inclusive Value）。化简(6.4.16)得到：

$$P_{j|l} = \frac{e^{\mathbf{X_{j|l}}\mathrm{B}}}{\sum_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}} \quad P_l = \frac{e^{\mathbf{Z_l}\Gamma + \mathbf{I_l}\mathrm{T}_l}}{\sum_{l=1}^{L} e^{\mathbf{Z_l}\Gamma + \mathbf{I_l}\mathrm{T}_l}} \frac{e^{\mathbf{X_{j|l}}\mathrm{B}}}{\sum_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}} \frac{e^{\mathbf{X_{j|l}}\mathrm{B}}}{\sum_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}} \frac{e^{\mathbf{X_{j|l}}\mathrm{B}}}{\sum_{j=1}^{J_l} e^{\mathbf{X_{j|l}}\mathrm{B}}} \tag{4.5.16}$$

这里新的参数T_l必须等于1才能产生原模型。所以可以利用T_l=1的约束重新描述条件Logit模型，然后再放松约束得到Nested Logit模型。

Nested Logit模型具有灵活的结构，在消费者选择中被广泛应用，而且被扩展到3层以上的结构。当然随着Nested Logit模型层次结构的扩展，其复杂程度也呈几何级数增长。

Nested Logit模型的参数估计方法有两类。一类是两阶段最大似然法，是一种有限信息估计方法。其具体步骤是：

⑴在组内，作为一个简单的条件Logit模型，估计参数B；

⑵计算每组的“内值”；

⑶将每组看成是一种选择方案，再进行简单的条件Logit模型的估计，得到参数Γ和T的估计量。此时用到的贡献变量是$\mathbf{Z_l}$和$\mathbf{I_l}$。

另一类方法是完全信息最大似然法。将对数似然函数写为：

$$\ln L = \sum_{i=1}^{n} \ln \left(P_{j|l} \times P_l\right)_i \tag{4.5.17}$$

完全信息最大似然法估计比两阶段最大似然法估计更有效。

4.5.3 排序多元离散选择模型

排序多元离散选择问题普遍存在于经济生活中。

对于排序多元离散选择问题，上面讨论的一般多元离散选择模型就不适合了。因为在(6.5.1)中，如果决策者i选择了j，那么就意味着：

$$U_{ij} > U_{ik} \qquad k = 0, 1, 2, \cdots, J \quad k \neq j$$

但是在排序多元离散选择中，即使决策者i选择了j，也不意味着上述效用比较关系一定成立。

直接将选择结果作为被解释变量，以0、1、2、3等作为样本观测值，采用普通回归分析的方法建立模型，当这些离散数据能够真实反映不同选

择方案的效用差异时，也是可行的。但是实际上，这些离散数据只表示不同选择方案之间的顺序，并不反映不同选择方案的效用的真实差异。

排序Probit和Logit模型是用与二元选择模型类似的方法建立的。以

$$y^* = \mathbf{X}\mathrm{B} + \mu$$

开始，这里的y^*是无法观测的。人们观测到的是：

$$\begin{cases} y = 0 & \text{如果} \quad y^* \leqslant 0 \\ \quad = 1 & \text{如果} \quad 0 < y^* \leqslant u_1 \\ \quad = 2 & \text{如果} \quad u_1 < y^* \leqslant u_2 \\ \quad \vdots & \quad \vdots \\ \quad = j & \text{如果} \quad u_{j-1} \leqslant y^* \end{cases}$$

假定μ服从正态分布，并且标准化为服从期望为0、方差为1的正态分布。那么可以得到如下的概率：

$$\begin{aligned} &P(y=0) = \Phi(-\mathbf{X}\mathrm{B}) \\ &P(y=1) = \Phi(u_1 - \mathbf{X}\mathrm{B}) - \Phi(-\mathbf{X}\mathrm{B}) \\ &P(y=2) = \Phi(u_2 - \mathbf{X}\mathrm{B}) - \Phi(u_1 - \mathbf{X}\mathrm{B}) \\ &\vdots \\ &P(y=J) = 1 - \Phi(u_{J-1} - \mathbf{X}\mathrm{B}) \end{aligned}$$

其中符号Φ表示正态分布的概率函数，即

$$\Phi(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} (\mathbf{2}\pi)^{-\mathbf{1/2}} \exp(-t^2/2)dt$$

为了保证所有的概率都是正的，必须有：

$$0 < u_1 < u_2 < \cdots < u_{J-1}$$

显然，这是§4.3中二元Probit选择模型的推广，其对数似然函数很容易得到，采用最大似然法可以估计模型的参数。

4.6 受限被解释变量数据计量经济学模型

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，得到的观测值并不完全反映被解释变量的实际状态。通过这样的样本观测值估计母体的参数，就是受限被解释变量数据计量经济学模型要解决的问题。

4.6.1 经济生活中的受限被解释变量问题

经济生活中的受限被解释变量问题很多，这里主要讨论常见的两类。

一类是“截断”（truncation）问题，即“掐头”或者“去尾”。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0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。从这样的样本数据出发，如果采用经典的方法估计母体的参数，显然是不合适的。

一类是“归并”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”(censoring)问题。例如，以居民对某一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，正象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”。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模型，显然也是不合适的。

4.6.2 “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，抽取每一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：

$$f(\xi\,|\xi > a) = \frac{f(\xi)}{P(\xi > a)} \tag{4.6.1}$$

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，那么取得每一个样本观测值的概率为：

$$f(\xi\,|\xi > c) = \frac{f(\xi)}{P(\xi > c)} = \frac{1/(b-a)}{\int\limits_c^b \frac{1}{b-a}d\xi} = \frac{1}{b-c}$$

请读者注意，原来均匀分布随机变量的概率密度函数是$1/(b-a)$，而"截断随机变量"的概率密度函数是$1/(b-c)$，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点。

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，那么取得每一个样本观测值的概率为：

$$\begin{aligned} f(\xi\,|\xi > a) &= \tfrac{f(\xi)}{P(\xi>a)} \\ &= \tfrac{(2\pi\,\sigma^2)^{-1/2}e^{-(\xi-\mu)^2/(2\sigma^2)}}{1-\Phi(\alpha)} \\ &= \tfrac{\frac{1}{\sigma}\phi(\frac{\xi-\mu}{\sigma})}{1-\Phi(\alpha)} \end{aligned} \tag{4.6.2}$$

其中，$\alpha = (a-\mu)/\sigma$，$\phi(\cdot)$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，

$$P(\xi > a) = 1 - \Phi(\frac{a-\mu}{\sigma}) = 1 - \Phi(\cdot)$$

2.截断被解释变量数据计量经济学模型的最大似然估计

已知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

$$y_i = \mathrm{B}'\mathbf{X_i} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \tag{4.6.3}$$

有

$$y_i\,|\mathbf{X_i} \sim N(\mathrm{B}'\mathbf{X_i}, \sigma^2)$$

如果y_i只能在大于a的范围内取得观测值，从(6.4.2)可以得到y_i的概率密度函数为：

$$f(y_i) = \frac{\frac{1}{\sigma}\phi((y_i - \mathrm{B}'\mathbf{X_i})/\sigma)}{1 - \Phi((a - \mathrm{B}'\mathbf{X_i})/\sigma)}$$

于是(6.4.4)的对数似然函数为：

$$\begin{aligned}\ln L = -\tfrac{n}{2}(\ln(2\pi) + \ln\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathrm{B}'\mathbf{X_i})^2 \\ - \sum_{i=1}^{n}\ln\left(1 - \Phi\left(\tfrac{a-\mathrm{B}'\mathbf{X_i}}{\sigma}\right)\right)\end{aligned} \tag{4.6.4}$$

该对数似然函数的极大化条件为：

$$\frac{\partial \ln L}{\partial\begin{pmatrix}\mathrm{B}\\ \boldsymbol{\sigma^2}\end{pmatrix}} = \sum_{i=1}^{n}\begin{pmatrix}\left(\frac{y_i-\mathrm{B}'\mathbf{X_i}}{\sigma^2} - \frac{\lambda_i}{\sigma}\right)\mathbf{X_i} \\ -\frac{1}{\mathbf{2\sigma^2}} + \frac{(y_i-\mathrm{B}'\mathbf{X_i})^2}{2\sigma^4} - \frac{\alpha_i\lambda_i}{2\sigma^2}\end{pmatrix} = \sum_{i=1}^{n}\mathbf{g_i} = \mathbf{0} \tag{4.6.5}$$

其中

$$\alpha_i = (a - \mathrm{B}'\mathbf{X_i})/\sigma$$

$$\lambda_i = \phi(\alpha_i)/(1 - \Phi(\alpha_i))$$

求解(6.4.6)即可以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(6.4.6)，例如牛顿法。

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$\Gamma = \mathrm{B}/\sigma$和$\theta = 1/\sigma$，得到：

$$\begin{aligned}\ln L = \tfrac{n}{2}\ln(2\pi) + n\ln\theta - \tfrac{1}{2}\sum_{i=1}^{n}(\theta\, y_i - \Gamma'\mathbf{X_i})^2 \\ - \sum_{i=1}^{n}\ln\Phi(\Gamma'\mathbf{X}_i)\end{aligned}$$

这里利用了$1 - \Phi(-t) = \Phi(t)$。对该对数似然函数极大化，求得到Γ和θ的估计量后再利用$\sigma = 1/\theta$和$\mathrm{B} = \Gamma/\theta$求得原参数估计量。

以上只是介绍了方法思路，读者在建立与应用截断被解释变量数据计量经济学模型时还需要进行深入的研究。

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(6.4.4)，会产生什么样的结果？

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：

$$\begin{aligned}E(y_i\,|y_i > a) = \int_a^{\infty} y_i\phi(y_i\,|y_i > a)dy_i \\ = \mathrm{B}'\mathbf{X_i} + \sigma\tfrac{\phi((a-\mathrm{B}'\mathbf{X_i})/\sigma)}{1-\Phi((a-\mathrm{B}'\mathbf{X_i})/\sigma)}\end{aligned} \tag{4.6.6}$$

(6.4.7)所示的条件均值是解释变量和待估参数的非线性函数。将(6.4.7)记为：

$$E(y_i\,|y_i>a)=\mathrm{B}'\mathbf{X_i}+\sigma\lambda\,(\alpha_i) \tag{4.6.7}$$

其中$\alpha_i=\frac{\alpha-\mathrm{B}'\mathbf{X}_i}{\sigma}$,于是有：

$$\begin{aligned}\frac{\partial\,E(y_i|y_i>a)}{\partial\,\mathbf{X_i}}&=\mathrm{B}+\sigma\left(\frac{d\lambda_i}{d\alpha_i}\right)\frac{\partial\,\alpha_i}{\partial\,\mathbf{X_i}}\\&=\mathrm{B}+\sigma\,(\lambda_i^2-\alpha_i\lambda_i)\left(\frac{-\mathrm{B}}{\sigma}\right)\\&=\mathrm{B}(1-\lambda_{\mathbf{i}}^{\mathbf{2}}\ +\alpha_i\lambda_i)\\&=\mathrm{B}(1-\delta(\alpha_{\mathbf{i}}\ ))\end{aligned} \tag{4.6.8}$$

将(6.4.8)写成：

$$y_i\,|y_i>a=E(y_i\,|y_i>a)+u_i=\mathrm{B}'\mathbf{X_i}+\sigma\lambda\,(\alpha_i)+u_i \tag{4.6.9}$$

其中u_i是被解释变量观测值与条件期望值之差，具有0均值和异方差。其方差为

$$Var(u_i)=\sigma^2(1-\lambda_i^2+\lambda_i\alpha_i)=\sigma^2(1-\delta_i)$$

对比(6.4.4)与(6.4.10)后发现，由于被解释变量数据的截断问题，使得原模型(6.4.4)变换为(6.4.10)式的模型。如果采用普通最小二乘法直接估计(6.4.4)，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，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。

4.6.3 “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：

$$\begin{matrix}y=0 & \text{当}y^*\leqslant 0\\ y=1 & \text{当}y^*>0\end{matrix}\qquad \text{且}Y^*\sim N(\mu,\sigma^2) \tag{4.6.10}$$

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

$$\begin{matrix}y=a & \text{当}y^*\leqslant a\\ y=y^* & \text{当}y^*>a\end{matrix}\qquad \text{且}Y^*\sim N(\mu,\sigma^2)$$

单方程线性“归并”问题的计量经济学模型为：

$$y_i = \mathrm{B}'\mathbf{X_i} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \tag{4.6.11}$$

注意，这里的被解释变量是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Tobin于1958年最早提出的，所以也称为Tobin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(6.4.11)，有

$$\begin{aligned} &P(y=0) = P(y^* \le 0) = \Phi\left(-\tfrac{\mu}{\sigma}\right) = 1 - \Phi\left(\tfrac{\mu}{\sigma}\right) \\ &P(y) = P(y^*) \qquad \text{当} y^* > 0 \end{aligned} \tag{4.6.12}$$

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(6.4.13)，很容易得到模型(6.4.12)对数似然函数：

$$\ln L = \sum_{y_i>0} -\frac{1}{2}\left(\ln(2\pi) + \ln\sigma^2 + \frac{(y_i - \mathrm{B}'\mathbf{X_i})^2}{\sigma^2}\right) + \sum_{y_i=0} \ln\left(1 - \Phi\left(\frac{\mathrm{B}'\mathbf{X_i}}{\sigma}\right)\right) \tag{4.6.13}$$

显然，(6.4.14)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(6.4.14)极大化，就可以求得具有良好性质的参数估计量。如果对(6.4.14)进行再参数化，可使得估计过程更为简单。即令$\Gamma = \mathrm{B}/\sigma$和$\theta = 1/\sigma$，得到

$$\ln L = \sum_{y_i>0} -\frac{1}{2}\left(\ln(2\pi) - \ln\theta^2 + (\theta\, y_i - \Gamma'\mathbf{X_i})^2\right) + \sum_{y_i=0} \ln\left(1 - \Phi\left(\Gamma'\mathbf{X}_i\right)\right) \tag{4.6.14}$$

对(6.4.15)极大化，由于Hessian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Γ和θ的估计量后再利用$\sigma = 1/\theta$和$\mathrm{B} = \Gamma/\theta$求得原参数估计量。

4.7 持续时间数据被解释变量计量经济学模型

持续时间分析早就被用于自然科学领域。例如，电子元器件的使用寿命问题，已经持续使用一段时间的元器件是否需要及时更换，当然是人们十分关心的问题。再如，经过手术的病人的存活时间问题，也是医生最

为关心的问题。但是，在社会经济研究领域引入持续时间分析的历史并不长。本节只简单介绍这类模型的基本思路，有兴趣的读者在实际应用时还需要深入探讨。

4.7.1 计量经济学中持续时间分析问题的提出

"持续"(duration)问题，即以某项活动的持续时间作为研究对象的经济问题。可以从以下两个方面提出这个问题。

1.经济生活中的持续时间问题

例如，失业问题是人们普遍关注的问题。尤其是仍然处于失业状态的失业者，希望知道自己在最近的短期间内能够重新就业的可能性有多大。从直观上看，存在两种完全不同的答案：一种是失业时间越长，在最近的短期间内能够重新就业的可能性就越大，因为你已经等待了足够的时间；另一种是失业时间越长，在最近的短期间内能够重新就业的可能性就越小，因为你在相当长的时间内都没有能够重新就业，说明你不具备重新就业的条件。两种答案，似乎都有道理。

再如，在国外的教科书中经常作为例子的罢工问题。已经持续了一段时间的罢工能否在最近结束？从直观上看，也存在两种完全不同的答案：一种是持续时间越长，在最近的短期间内能够解决问题、结束罢工的可能性就越大，因为罢工已经持续了足够的时间，双方都有尽快结束的愿望；另一种是持续的时间越长，在最近的短期间内能够结束的可能性就越小，因为之所以持续了这么长时间，说明引起罢工的原因是复杂的，在这么长时间内都没有能够解决，近期也很难有进展。两种答案，也都有道理。

其它诸如购买问题，已经使用了相当长时间的物品是否需要在近期内重新购买等等，都属于以持续时间作为研究对象的经济问题。

2.持续时间被解释变量的计量经济学问题

以失业的持续时间分析为例，看看这类计量经济学问题的特征。

以失业的持续时间t作为被解释变量，以年龄、受教育程度、家庭状况、工作经历、健康状况等作为解释变量，建立如下失业模型：

$$T_i = \mathrm{B}'\mathbf{X}_i + \varepsilon_i \qquad i = 1, 2, \cdots, n \tag{4.7.1}$$

对大量的失业者进行调查，以他们的失业持续时间作为被解释变量的观测值，然后估计模型，是经典计量经济学模型中通常的做法。但是，在这个问题中有3个特殊之处。

第一，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。所以必须将样本观测值分为两类，一类是对已经重新工作的失业者进行的调查，一类是对仍处于失业状态的失业者进行的调查。对于前者，失业持续时间的观测值是真实的；而对于后者，样本观测值实际上是上一节中介绍的"归并"数据。

第二，取得解释变量的样本观测值存在困难。因为在失业持续的时间内，年龄、家庭状况、健康状况等是变化的，这是显而易见的。能够得到的仅是失业已经持续到t时刻的解释变量的样本观测值，并不能用来解释整个持续时间。

第三，也是最重要的，失业者关心的不是如何解释失业已经持续的时间，而是希望知道在观测值t时刻之后的最短时间内能够重新就业的可能性为多大。而直接估计模型(6.5.1)，只是解释失业已经持续的时间，不能回答失业者最关心的问题。

基于以上特殊之处，必须特殊研究持续时间被解释变量计量经济学模型。

4.7.2 Hazard比率与Hazard比率模型

首先从上述持续时间被解释变量计量经济学模型的第3个特征入手，并假设解释变量的样本观测值在失业持续的时间内是不变化的，即忽略上述第2个问题。

1.Hazard比率

假设随机变量T具有连续的概率密度函数$f(t)$，t是T的一个观测值，即事件已经持续的时间。那么应该有

$$P(T \le t) = F(t) = \int_0^t f(z)dz$$

因为持续时间非负，所以积分从0开始。不难理解，持续时间不少于t的概率为：

$$S(t) = 1 - F(t) = P(T \ge t)$$

称$S(t)$为生存函数。而人们最关心的事件在t之后的一个短时间Δ内结束的概率可以表示为$P(t \le T \le t + \Delta \,|T \ge t)$，还可以用下列函数来进一步表示：

$$\begin{aligned}\lambda(t) &= \lim_{\Delta\to 0}\tfrac{P(t\le T\le t+\Delta|T\ge t)}{\Delta}\\ &= \lim_{\Delta\to 0}\tfrac{F(t+\Delta)-F(t)}{\Delta S(t)}\\ &= \tfrac{f(t)}{S(t)}\end{aligned} \tag{4.7.2}$$

称(6.4.2)为Hazard比率。事件以该比率在已经持续t时间后结束，显然该比率的大小即是人们关心的问题的答案。

Hazard比率与t的概率密度函数$f(t)$、条件分布函数$F(t)$和生存函数$S(t)$之间存在如下关系：

$$\lambda(t) = \frac{-d\ln S(t)}{dt} \tag{4.7.3}$$

$$f(t)=S(t)\lambda(t) \tag{4.7.4}$$

由(6.4.4)得到：

$$\ln S(t)=-\int_0^t \lambda(z)dz \stackrel{令}{=} -\Lambda(t)$$

则有

$$S(t)=e^{-\Lambda(t)} \tag{4.7.5}$$

2.不考虑外生变量的Hazard比率模型

既然人们更关心事件在t之后的一个短时间Δ内结束的可能性，而该可能性又可以通过Hazard比率来描述，那么可以直接建立Hazard比率模型，估计Hazard比率的参数，然后再通过积分得到生存函数$S(t)$和条件分布函数$F(t)$。这就是持续时间被解释变量计量经济学模型的总的研究思路。

如何构造和求解Hazard比率模型,首先通过两个简单的例子来说明.

(1)Hazard比率为一个常数

假设Hazard比率为一个常数λ，即假设事件在t之后的一个短时间Δ内结束的概率是相同的，与已经持续的时间t无关。这种事件在实际中也是存在的，被称为“无记忆”的过程。那么即有

$$\frac{-d\ln S(t)}{dt}=\lambda$$

该微分方程的解是：

$$\ln S(t)=k-\lambda t$$

$$S(t)=Ke^{-\lambda t} \tag{4.7.6}$$

其中K是积分常数。因为

$$S(0)=P(T\geq 0)=1$$

由(6.4.7)得到$K=1$，进一步得到模型的解为：

$$S(t)=e^{-\lambda t}$$

这就是说t的生存函数$S(t)$服从指数分布。这种分布已经被广泛地用于电子元器件的持续时间模型中。

如何估计常数λ？因为对于指数分布，有

$$E(t)=1/\lambda$$

那么λ的最大似然估计量为：

$$\hat{\lambda} = 1/\bar{t}$$

其中的$\bar{t}$由样本观测值计算得到，于是得到Hazard比率的估计量。至此，该持续时间计量经济学问题研究完毕。

(2)Hazard比率为一个线性函数

假设Hazard比率为一个线性函数，即

$$\lambda(t) = \alpha + \beta t$$

那么有

$$\Lambda(t) = \int_0^t \lambda(z)dz = \alpha t + \tfrac{1}{2}\beta t^2$$

$$f(t) = S(t)\lambda(t) = \lambda(t)e^{-\Lambda(t)} \tag{4.7.7}$$

由t的密度函数和样本观测值，利用最大似然法，可以得到参数α,β的估计量，进而得到Hazard比率的估计量，使该持续时间计量经济学问题得到解决。

如果得到参数β的估计量为正，表示Hazard比率随着持续时间的增长而增大，也表示在t之后的一个短时间Δ内结束事件的概率随着持续时间的增长而增大；如果得到参数β的估计量为负，则反之。

3.几种常用的Hazard比率模型

在上面的描述中，首先对Hazard比率作出假设，导出生存函数$S(t)$和密度函数$f(t)$，然后利用最大似然法估计参数。如果人们并不首先对Hazard比率作出假设，而是直接对生存函数$S(t)$所服从的分布作出假设，然后直接估计该分布的参数，结果是相同的。这就是实际中通常采用的思路。其过程如下：

$$S(t) \xrightarrow{(6.4.4)} \lambda(t) \xrightarrow{(6.4.5)} f(t) \xrightarrow{ML} \text{参数估计量} \longrightarrow \text{Hazard比率}$$

除了指数分布（对应于Hazard比率为常数）外，经常被作为生存函数$S(t)$的分布假设的还有韦伯分布、对数正态分布、对数逻辑分布等。对于韦伯分布，有

$$S(t) = e^{-(\lambda t)^p}$$

对于对数正态分布，有

$$S(t) = \Phi(-p\ln(\lambda t))$$

对于对数逻辑分布，有

$$S(t) = 1/(1 + (\lambda t)^p)$$

对于生存函数$S(t)$的每种分布，都有对应的Hazard比率函数。例如服从韦伯分布的生存函数，其对应的Hazard比率函数是一个单调增函数，形式为：

$$\lambda(t) = \lambda p(\lambda t)^{p-1}$$

模型参数估计的任务就是估计这些分布函数中的λ, p等。

4.考虑两类样本数据的最大似然估计

正如前面提及的，必须将持续时间样本观测值分为两类，一类是对已经结束的事件进行的调查，一类是对仍处于持续过程中的事件进行的调查。对于前者，持续时间的观测值是真实的；而对于后者，样本观测值实际上是“归并”数据。那么，参考§4.5中所介绍的，Hazard比率模型的对数似然函数为：

$$\ln L = \sum_{\text{非归并观察值}} \ln f(t \mid (\lambda, p)) + \sum_{\text{非归并观察值}} \ln S(t \mid (\lambda, p)) \tag{4.7.8}$$

因为$f(t) = S(t)\lambda(t)$，所以上式可以写成：

$$\ln L = \sum_{\text{非归并观察值}} \ln \lambda(t \mid (\lambda, p)) + \sum_{\text{非归并观察值}} \ln S(t \mid (\lambda, p)) \tag{4.7.9}$$

在实际应用中采用哪种形式，依方便而定。

5.考虑外生变量的Hazard比率模型

在上面的描述中，实际上是依赖持续时间样本观测值和分布假设估计模型参数，没有引入影响持续时间的各种因素。所以上述模型不是经典意义上的计量经济学模型，没有揭示事件持续的因果关系。

对事件持续进行因果分析，则要引入影响持续时间的各种因素。但是，人们并不建立(6.5.1)式的模型，而是以Hazard比率为被解释变量，以影响持续时间的各种因素为解释变量建立模型，而且为了估计的方便，对模型的关系类型作出特定的假设。例如，对于生存函数$S(t)$服从韦伯分布的情况，建立如下模型：

$$\lambda_i = e^{-\mathrm{B}'\mathbf{X_i}} \tag{4.7.10}$$

并且假设在持续的时间内，$\mathbf{X_i}$具有不变的观测值。

构造如下与(6.4.11)等价的线性模型：

$$w_i = p\ln(\lambda_i t_t) = p(\ln t_i - \mathrm{B}'\mathbf{X_i}) \tag{4.7.11}$$

对服从韦伯分布的概率密度函数和生存函数进行变量置换，显然有：

$$f(w_i) = p\exp\left(w_i - e^{w_i}\right)$$

$$S(w_i) = \exp\left(-e^{w_i}\right)$$

前者为模型(6.4.12)的概率密度函数，后者为它的生存函数。于是就可以采用最大似然法估计模型(6.4.12)。

考虑到持续时间样本观测值分为两类，令

$$\begin{cases} \delta_i = 1 & \text{如果持续时间观察值取自已经结束事件的个体} \\ \delta_i = 0 & \text{如果持续时间观察值取自尚未结束事件的个体} \end{cases}$$

于是对数似然函数为：

$$\ln L = \sum_{i=1}^{n}\left(\delta_i \ln f(w_i) + (1-\delta_i)\ln S(w_i)\right) \tag{4.7.12}$$

可以进一步化简为：

$$\ln L = \sum_{i=1}^{n}\left(\delta_i(w_i - \ln\sigma) - e^{w_i}\right) \tag{4.7.13}$$

其中$\sigma = 1/p$。

对(6.4.14)求极大，利用$\partial w_i/\partial\sigma = -w_i/\sigma$和$\partial w_i/\partial \mathrm{B} = -\mathbf{X}_i/\sigma$，得到关于参数估计量的方程组，采用牛顿迭代方法求解方程组，即得到参数估计量。

4.8 本章思考题和综合练习题

一、思考题

1.在经典计量经济学模型中，对被解释变量样本数据有哪些假定？

2.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？

3.单方程平行数据模型的3种形式及其经济含义是什么？通过什么方法检验并设定模型的形式？

4.如何引入虚拟变量表示平行数据模型的固定影响变截距？

5.随机影响变截距平行数据模型的一般表示、带来的问题和估计方法的思路是什么？

6.固定影响变系数平行数据模型GLS估计的具体步骤是什么？为什么其估计量比每个截面个体以时序数据为样本的单独估计更有效？

7.固定影响变系数平行数据动态模型工具变量法估计的步骤是什么？如何选择工具变量？

8.离散选择模型的研究思路是什么？为什么一般要将原始模型变换为效用模型？为什么必须选择某种特定的概率分布？

9.重复观测值可以得到和不可以得到情况下的离散选择模型的估计方法是什么？

10.关于多元离散选择模型、多元名义logit离散选择模型、多元条件logit离散选择模型、Nested Logit模型的概念。

11.为什么说受限被解释变量数据计量经济学模型的关键是正确描述受限被解释变量的概率分布？

12."截断随机变量"分布与连续随机变量分布的关系是什么？

13.为什么截断被解释变量模型不能采用OLS方法估计？

14."归并"问题与"截断"问题的主要区别是什么？

15.持续时间数据被解释变量模型的研究对象与应用领域是什么？

二、综合练习题

1.选择一个研究对象，例如以我国各省、市、自治区为截面个体的生产函数模型、消费函数模型等，建立平行数据计量经济学模型。完成如下步骤的工作：

⑴收集不少于10年的数据；

⑵进行模型设定的检验；

⑶建立并估计固定影响的模型；

⑷对模型结果进行经济意义分析。

2.建立一个实际的二元离散选择模型。包括：

⑴选择研究对象并调查样本观测值；

⑵分别估计二元probit和logit模型，比较估计结果；

⑶进行变量显著性检验；

⑷应用模型进行预测；

⑸对二元离散选择模型的建立与应用进行总结。

3.建立一个实际的计量经济学模型并估计参数，然后人为地将它变成一个"截断"问题，再进行估计，对二者的结果进行比较分析。

第五章 非线性计量经济学模型

非线性计量经济学模型理论方法是计量经济学的重要组成部分，尤其是近30年来，由于计算机应用技术的发展，使得计算速度和时间已经不再是建立模型的障碍，人们可以尽可能地从模型质量方面去考虑模型的结构和估计方法，非线性计量经济学模型得到了很快的发展和广泛的应用。本书§.1讨论了传统的非线性计量经济学模型的一些基本问题，在本章中将较为详细地讨论近代非线性计量经济学模型的有关问题。

5.1 可线性化的非线性计量经济学模型

正如在§2.1中介绍的，许多经济变量之间的关系虽然是非线性的，但是可以通过某些简单的变换，转换为线性问题。但这种转换存在一些问题，§2.1从经济学假设的角度提出了这个问题，本节将从计量经济学和几何的角度对这些问题进行探讨。

5.1.1 通过变量变换线性化

从§2.1已知，通过变量变换可以将关于变量的非线性模型和一些关于参数的非线性模型化为线性模型。将这类问题还可以分为如下两种情况进行讨论。

1. 变量变换不涉及模型参数

存在一类模型，经过变量变换后模型参数不发生变化，可表示成如下形式：

$$Y = g(\beta_0 + \beta_1 f_1(X_1, X_2, \cdots, X_l) + \cdots + \beta_k f_k(X_1, X_2, \cdots, X_l)) + u \quad (5.1.1)$$

其中g是一个已知的函数；$f_1, f_2, \cdots, f_k$是$X_1, X_2, \cdots, X_l$的k个已知的函数。令

$$\begin{aligned} M &= g^{-1}(Y) \\ Z_i &= f_i(X_1, X_2, \cdots, X_l) \quad i = 1, \cdots, k \end{aligned}$$

非线性模型(5.1.1)就转换为线性模型：

$$M = \beta_0 + \beta_1 Z_1 + \cdots + \beta_k Z_k + \varepsilon \tag{5.1.2}$$

在经济分析中常用的这类函数列于表5.1.1中。

表 5.1.1: 线性变换函数表

原模型	变换函数
$\frac{1}{y} = a + \frac{b}{x}$	$M = \frac{1}{y} \quad Z = \frac{1}{x}$
$y = a + b \ln x$	$M = y \quad Z = \ln x$
$y = \frac{1}{a+be^{-x}}$	$M = \frac{1}{y} \quad Z = e^{-x}$
$y = a + bx + cx^2$	$M = y \quad Z_1 = x \quad Z_2 = x^2$
$y = \sqrt{a+bx}$	$M = y^2 \quad Z = x$
$y = a + be^{-2x}$	$M = y \quad Z = e^{-2x}$
$y = e^{ax} x^b$	$M = \ln y \quad Z_1 = x \quad Z_2 = \ln x$
$y = e^{a-b/x}$	$M = \ln y \quad Z = -\frac{1}{x}$

以前没有讨论而应该讨论的问题是，对（5.1.1）进行非线性最小二乘估计的结果会与对（5.1.2）进行普通最小二乘估计的结果一致吗？如果答案是否定的，那么在什么条件下，两个结果是一致的呢？

将（5.1.1）的非线性最小二乘估计的正规方程组：

$$\begin{aligned}
&\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} g(\beta_0 + \beta_1 Z_{1i} + \cdots + \beta_k Z_{ki}) \\
&\sum_{i=1}^{n} Z_{1i} Y_i = \sum_{i=1}^{n} Z_{1i} g(\beta_0 + \beta_1 Z_{1i} + \cdots + \beta_k Z_{ki}) \\
&\cdots\cdots \\
&\sum_{i=1}^{n} Z_{ki} Y_i = \sum_{i=1}^{n} Z_{ki} g(\beta_0 + \beta_1 Z_{1i} + \cdots + \beta_k Z_{ki})
\end{aligned} \tag{5.1.3}$$

与（5.1.2）的普通最小二乘估计的正规方程组：

$$\begin{aligned}
&\sum_{i=1}^{n} M_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} Z_{1i} + \cdots + \beta_k \sum_{i=1}^{n} Z_{ki} \\
&\sum_{i=1}^{n} Z_{1i} M_i = \beta_0 \sum_{i=1}^{n} Z_{1i} + \beta_1 \sum_{i=1}^{n} Z_{1i}^2 + \cdots + \beta_k \sum_{i=1}^{n} Z_{1i} Z_{ki} \\
&\cdots\cdots \\
&\sum_{i=1}^{n} Z_{ki} M_i = \beta_0 \sum_{i=1}^{n} Z_{ki} + \beta_1 \sum_{i=1}^{n} Z_{ki} Z_{1i} + \cdots + \beta_k \sum_{i=1}^{n} Z_{ki}^2
\end{aligned} \tag{5.1.4}$$

进行比较，容易得出结论：两个结果是不一样的。只有当$g(x)$为线性函数时，两者相同。另外线性模型（5.1.2）和下列线性模型

$$\frac{M}{f(Z_1,\cdots,Z_k)} = \beta_0\frac{1}{f(Z_1,\cdots,Z_k)} + \beta_1\frac{Z_1}{f(Z_1,\cdots,Z_k)} + \cdots + \beta_k\frac{Z_k}{f(Z_1,\cdots,Z_k)} + v \tag{5.1.5}$$

的普通最小二乘估计的结果也是不同的。其中

$$v = \frac{\varepsilon}{f(Z_1,\cdots,Z_k)} \tag{5.1.6}$$

关键是变量的变换引起了随机扰动的变换，认清这点非常重要。因为这将影响到随机扰动的假设条件，进而影响到参数估计的统计性质。如果假定原模型的随机扰动项是正态独立同方差的，但对于变换后的模型来说，随机扰动项不再是正态独立同方差的，而是独立异方差的。因此，在（5.1.1）的随机扰动项是独立正态同方差下，要么对原非线性模型（5.1.1）采用非线性最小二乘估计，要么对转换为线性的模型（5.1.2）使用加权最小二乘估计，即找到权数$W = 1/f(Z_1,\cdots,Z_k)$,对（5.1.2）进行加权最小二乘估计。这等价于对(5.1.5)采用普通最小二乘估计。若（5.1.2）的随机扰动项是独立同方差，则可直接对（5.1.2）进行普通最小二乘估计。

对于模型(6.5.1)，在随机干扰项与解释变量不相关的条件下，无论随机干扰项同方差、异方差或自相关时，OLS估计都为一致估计。但当随机干扰项存在异方差或自相关时，OLS不如GLS和GMM有效。此时，应采用GLS估计或GMM估计。当采用GLS方法时，由于随机干扰项的协方差阵未知，所以，参数估计应采用可行的广义最小二乘法(FGLS)进行估计。当采用GMM方法时，则参数的GMM估计值与OLS估计值相同，但参数GMM估计$\hat{\beta}$的标准误差的计算公式要根据异方差或自相关进行调整。调整的计算公式请参见§3.6。

2.变量变换涉及模型参数

这类模型变量变换后模型参数发生了变化，对于模型

$$Y = h(X_1, X_2, \cdots, X_l, \alpha) + u \tag{5.1.7}$$

假设经变量变换：$M = f(Y)$，$Z_i = f_i(X_1, X_2, \cdots, X_l)$，$i = 1, \cdots, k$ (其中f是一个已知的函数$f_1, f_2, \cdots, f_k$是$X_1, X_2, \cdots, X_l$的k个已知的函数)和参数变换：$\beta = \varphi(\alpha)$(其中$\varphi(\cdot)$是一个一一映射向量函数)后，非线性模型(5.1.7)就转换为线性模型(6.5.1)。

表5.1.2给出了在经济分析中常用几种变换函数。

问题在于对（5.1.7）进行非线性最小二乘估计的结果与对（5.1.2）进行普通最小二乘估计的结果一致吗？答案是否定的。根本的原因与前面一

表 5.1.2: 常用几种涉及参数变化的变换函数

原模型	变换函数
$y = ax^b$	$M = \ln y \quad Z = \ln x \quad \beta_1 = \ln a \quad \beta_2 = b$
$y = ae^{bx}(a > 0)$	$M = \ln y \quad Z = x \quad \beta_1 = \ln a \quad \beta_2 = b$
$y = K(1 - ae^{-x})^3$	$M = y^{1/3} \quad Z = e^{-x} \quad \beta_1 = \sqrt[3]{K} \quad \beta_2 = -a\sqrt[3]{K}$

样，在于被解释变量作了非线性变换。从几何上看，给定样本容量为n的一组样本，一般非线性模型$y = \eta(x, \beta) + \varepsilon$ (y为被解释变量的数据列向量，x为解释变量的数据矩阵，β为$(k+1)\times 1$参数列向量，$\eta(x,\beta)$为被解释变量的期望函数向量，ε 为残差列向量)的非线性最小二乘估计也就是在n维欧氏空间中的点y到该空间中$k+1$维期望曲面$\eta(x,\beta)$上的投影。被解释变量y变换后，被解释变量空间变了，而不同被解释变量空间上的距离是不一样的，在原被解释变量空间上两点的直线连线在新被解释变量空间上不再是直线。所以，两个估计结果不一样。该问题的处理方法与前问题相同。

5.1.2 通过参数变换线性化

模型通过参数变换后，模型的变量不发生变化，可表示成如下形式：

$$Y = g_0(\alpha_0, \alpha_1, \cdots, \alpha_k) + g_1(\alpha_0, \alpha_1, \cdots, \alpha_k)X_1 + \cdots + g_k(\alpha_0, \alpha_1, \cdots, \alpha_k)X_k + u \tag{5.1.8}$$

令$\beta_i = g_i(\alpha_0, \alpha_1, \cdots, \alpha_k) \quad i = 0, 1, \cdots, k$

模型(5.1.8)化为：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u \tag{5.1.9}$$

显然，(5.1.8)和(6.4.6)代表同一n维空间的同一$(k+1)$维超平面，因而用最小二乘估计得到的结果是一致的。且(5.1.8)代表的超平面的参数坐标是非均匀的，不可直接得到最小二乘估计表达式。而(6.4.6)代表的超平面的参数坐标是均匀的，可直接得到β的最小二乘估计表达式。

5.2 非线性模型估计中的优化计算方法

非线性计量经济学模型参数估计问题可转化为无约束或带约束的最小或最大的优化问题。在计量经济学中常用的两个参数估计方法就是最小化残差平方和和最大化对数似然函数。在某些情况下，需要在对参数施加等式约束或不等式约束下对目标函数(如残差平方和或对数似然函数)求最小或最大值。

5.2.1 无约束优化

设k元函数

$$f: R^k \to R \tag{5.2.1}$$

记$g(x)$、$H(x)$分别是$f(x)$的梯度向量和海塞(Hessian)矩阵，即：

$$g(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_k}\right)_{1\times k} \tag{5.2.2}$$

$$H(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{k\times k} \tag{5.2.3}$$

因为函数$f(x)$的最小化问题等价于函数$[-f(x)]$的最大化问题，即$\min\limits_x f(x)$等价于$\max\limits_x[-f(x)]$。所以，可以只考虑函数的最大化问题。即

$$\max_x f(x) \tag{5.2.4}$$

对于最大化问题，由微积分的知识，有以下定理。

定理5.2.1：x_0为问题(6.4.5)的解$\Leftrightarrow g(x_0) = 0$(称为一阶条件)且$H(x_0)$负定(称为二阶条件)。

如果x_0不是最大值，可由$f(x)$在x_0处的一阶近似$f(x) \approx f(x_0) + g(x_0)(x - x_0)$来改进。当$g_i(x_0) > 0$时，取$\Delta x_i = x_i - x_{0i} > 0$；当$g_i(x_0) < 0$时，取$\Delta x_i = x_i - x_{0i} < 0$，称$\Delta x$为步长，则$f(x) > f(x_0)$。常用的柯西步长和牛顿步长为：

柯西步长：$\Delta x = g(x_0)'$ (5.2.5)

牛顿步长：$\Delta x = -H(x_0)^{-1}g(x_0)'$ (5.2.6)

1. 数值计算：步长的大小和方向

优化问题就是寻找步长，使得优化的一阶和二阶条件得到满足。让$\Delta x = sd$，其中$s > 0$是步长的大小，d是步长的方向。一般地说，步长的方向d是根据梯度向量来选择。让$d = Mg'$，其中M是正定矩阵。

(1)步长大小的确定

给定x_0和d，求$s > 0$，使得$f(x_0 + sd)$达最大值。由$f(x)$在x_0处的二阶近似

$$f(x) \approx f(x_0) + g(x_0)(x - x_0) + \frac{1}{2}(x - x_0)'H(x_0)(x - x_0) \tag{5.2.5}$$

有：

$$f(x_0 + sd) \approx f(x_0) + g(x_0)sd + \frac{1}{2}s^2 d'H(x_0)d \tag{5.2.6}$$

$$f(x_0+sd_0)œŒˆ$$

$$g(x_0)d+sd'H(x_0)d=0 \tag{5.2.7}$$

所以，最优的步长大小为：

$$s=-(gd)/(d'Hd) \tag{5.2.8}$$

柯西最优步长：$s=-(gg')/(gHg')$，$M=I$ (5.2.11)

牛顿最优步长：$s=1$，$M=-H^{-1}$ (5.2.12)

在实践中，最优步长很少使用，而是采用如下双向线性搜寻方法：若$f(x_0+sd)>f(x_0)$，则扩大步长大小$s=1.1s$，直到不能改进为止；若$f(x_0+sd)\le f(x_0)$，则缩小步长大小$s=0.9s$，直到$f(x_0+sd)>f(x_0)$为止。若$f(x)$在此步长大小缩小过程中无法改进，则$f(x)=f(x_0)$已是最优或是鞍点。

(2)步长方向的确定

因为$d=Mg'$，所以正定矩阵M的不同的确定方法就产生不同的优化方法。下面是常见的几种确定步长方向的方法：

a)Steepest ascent 方法：$M=I$

b)Broyden-Fletcher-Goldfarb-Shanno(BFGS)方法

$$\begin{aligned}M^{i+1}=M^i+&\left[1+\frac{\Delta g^iM^i(\Delta g^i)'}{(\Delta x^i)'(\Delta g^i)'}\right]\left[\frac{\Delta x^i(\Delta x^i)'}{(\Delta x^i)'(\Delta g^i)'}\right]\\&-\frac{\Delta x^i\Delta g^iM^i+M^i(\Delta g^i)'(\Delta x^i)'}{(\Delta x^i)'(\Delta g^i)'}\end{aligned} \tag{5.2.9}$$

c)Davidon-Fletcher-Powell(DFP)方法

$$M^{i+1}=M^i+\frac{\Delta x^i(\Delta x^i)'}{(\Delta x^i)'(\Delta g^i)'}+\frac{M^i(\Delta g^i)'\Delta g^iM^i}{\Delta g^iM^i(\Delta g^i)'} \tag{5.2.10}$$

d)Newton-Raphson 方法：$M=-H^{-1}$

e)Greenstadt 方法

$$M=(\sum(|w^i|v^i(v^i)'))^{-1}\sum(|1/w^i|v^i(v^i)' \tag{5.2.11}$$

其中w^i和v^i分别为矩阵H的第i个特征值及其对应的特征向量。

f)Quadratic-hill climing方法：$M=-(H-rI)^{-1}$。

可见，a)、b)和c)不需要矩阵H的计算；而d)、e)和f) 需要矩阵H的计算，且f)中$r\ge 0$的选取应使得$H-rI$为负定。

在给定一个允许误差e下，迭代计算到第i步的收敛标准为：

a)$f(x^{i+1}) \geq f(x^i)$;
b)$f(x^{i+1}) - f(x^i) \leq e$;
c)$||x^{i+1} - x^i|| \leq e$;
d)$||g(x^{i+1})|| \leq e$。

至此，可以得到优化问题(6.4.5)迭代求解的步骤：给定初始值x^0和步长方向d^0,确定步长s^0；计算$x^1 = x^0 + s^0 d^0$, 检查是否收敛，如果收敛，则停止，如果没有收敛，则计算步长方向d^1,并确定步长s^1；如此下去，直到收敛为止。

2.一个例子

例5.2.1 产出Q、资本投入K和劳动投入L的数据如表5.2.1，CES生产函数模型的参数估计就是下列残差平方和的最小化：

$$f(\beta_1, \beta_2, \beta_3, \beta_4) = \sum_{i=1}^{n} \{\ln Q_i - \beta_1 - \beta_4 \ln[\beta_2 L_i^{\beta_3} + (1-\beta_2) K_i^{\beta_3}]\}^2 \quad (5.2.12)$$

表 5.2.1: 例5.2.1数据

L	K	Q	L	K	Q
0.228	0.802	0.256918	0.664	0.129	0.186747
0.258	0.249	0.183599	0.631	0.017	0.020671
0.821	0.771	1.212883	0.059	0.906	0.100159
0.767	0.511	0.522568	0.811	0.223	0.252334
0.495	0.758	0.847894	0.758	0.145	0.103312
0.487	0.425	0.763379	0.05	0.161	0.078945
0.678	0.452	0.62313	0.823	0.006	0.005799
0.748	0.817	1.031485	0.483	0.836	0.72325
0.727	0.845	0.569498	0.682	0.521	0.776468
0.695	0.958	0.882497	0.116	0.93	0.216536
0.458	0.084	0.108827	0.44	0.495	0.541182
0.981	0.021	0.026437	0.456	0.185	0.31632
0.002	0.295	0.00375	0.342	0.092	0.123811
0.429	0.277	0.461626	0.358	0.485	0.386354
0.231	0.546	0.268474	0.162	0.934	0.279431

采用Quadratic Hill-Climbing方法，在给定初始值β_1^0=1,β_2^0=0.5,$\beta_3^0 = \beta_4^0$=-1和允许误差e= 1e-005下，经过36步迭代达到收敛，计算结果如下：

函数最小值为1.7611

参数估计值分别为0.12449、0.33668、-3.0109和-0.33631

梯度向量为(2.6755e-006, 4.6166e-007, 2.5664e-006, 1.7166e-006)

海赛矩阵为

$$\begin{pmatrix} 60.000 & -5.7563 & 35.531 & 295.65 \\ -5.7563 & 19.377 & -3.4569 & -23.595 \\ 35.531 & -3.4569 & 35.461 & 298.10 \\ 295.65 & -23.595 & 298.10 & 2509.4 \end{pmatrix}$$

参数估计的渐近方差和协方差阵为

$$\begin{pmatrix} 0.0055717 & -0.0013729 & -0.065566 & 0.0071194 \\ -0.0013729 & 0.011684 & 0.16893 & -0.019797 \\ -0.065566 & 0.16893 & 5.2462 & -0.61389 \\ 0.0071194 & -0.019797 & -0.61389 & 0.071948 \end{pmatrix}$$

表5.2.2列出了参数估计的标准误差和渐近t统计量。

表 5.2.2: 参数估计的统计量

参数估计	标准误差	渐近t统计量
0.12449	0.074644	1.6678
0.33668	0.10809	3.1147
-3.0109	2.2904	-1.3145
-0.33631	0.26823	-1.2538

5.2.2 约束优化问题

上述无约束优化中，没有附加任何约束条件。但是在实际问题的研究中，经常存在有约束的情况。下面分两种情况作简单介绍。

等式约束问题

设$f: R^k \to R$

等式约束最大化问题为：

$\max\limits_x f(x)$st: $c(x) = 0$ (5.2.17)

其中$c(x)$是J维约束向量。因为最小化问题$\min\limits_x f(x)$,st:$c(x) = 0$等价于最大化问题$\max\limits_x(-f(x))$,st: $c(x) = 0$,所以，我们只考虑约束最大化问题。

借助于拉格朗日函数

$$L(x, \lambda) = f(x) - \lambda' c(x) \tag{5.2.13}$$

将等式约束问题(5.2.17)转化为对x和λ求无约束最大化的问题：

$$\max_{x,\lambda} L(x, \lambda) \tag{5.2.14}$$

并利用下列定理。

定理5.2.2：x_0为等式约束问题(5.2.17)的解的一阶条件为：$c(x_0)=0$，且$g(x_0)-\lambda'\partial c(x_0)/\partial x=0$。

不等式约束问题

不等式约束最大化问题为

$\max\limits_x f(x)$st: $c(x)\le 0$ (5.2.20)

因为最小化问题$\min\limits_x f(x)$,st: $c(x)\le 0$等价于最大化问题$\max\limits_x(-f(x))$,st: $c(x)\le 0$;最小化问题$\min\limits_x f(x)$，st: $c(x)\ge 0$等价于最大化问题$\max\limits_x(-f(x))$，st: $(-c(x))\le 0$;最大化问题$\max\limits_x f(x)$，st: $c(x)\ge 0$等价于最大化问题$\max\limits_x f(x)$，st: $(-c(x))\le 0$。所以，只考虑不等式约束最大化问题(5.2.20)。

构造拉格朗日函数：

$$L(x,\lambda)=f(x)-\lambda'c(x) \tag{5.2.15}$$

并利用下列定理，可以方便求解该问题。

定理5.2.3 x_0为不等式约束问题(5.2.20)的解的一阶条件为Kuhn-Tucker条件：$g(x_0)-\lambda'\partial c(x_0)/\partial x=0$，$c(x_0)\le 0$且$\lambda_i c_i(x_0)=0$，$i=1,2,\cdots,J$。

例5.2.2（续例5.2.1) 对参数施加如下约束：

$$0\le\beta_2\le 1\quad \beta_3<-1\quad \beta_4=1/\beta_3$$

利用GAUSS软件中NLPSolve，计算结果如下：

在约束下，函数的最小值为1.765907;

参数估计及其标准误差和渐近t统计量如表5.2.3;

表 5.2.3: 参数估计统计量

参数估计	标准误差	渐近t统计量
0.11849	0.070746	1.6748
0.32238	0.10325	3.1225
-3.4403	1.7792	-1.9337

参数估计的方差和协方差阵为：

$$\begin{pmatrix} 0.0050050 & -0.0022037 & -0.096838 \\ -0.0022037 & 0.010660 & 0.097666 \\ -0.096838 & 0.097666 & 3.1655 \end{pmatrix}$$

5.3 非线性回归模型（一）——一般估计方法

许多经济现象中变量之间的因果关系往往不是线性的，例如反映生产投入和产出之间的CES生产函数。本节讨论关于参数是非线性的回归模型的一般估计方法，尽管这类模型的参数估计相当困难，但由于许多应用软件，例如TSP、SPSS、SAS、GAUSS和MATLAB等的可利用，使得非线性回归模型参数的迭代逼近的估计几乎和线性回归模型的参数估计一样容易。本节以及§.4和§5.5节主要讨论不可转换成线性的非线性模型，对于可转换成线性的非线性回归模型也可以采用这三节的方法进行处理。

5.3.1 非线性回归模型

设y为被解释变量，$x=(x_1,x_2,\cdots,x_k)$为解释变量向量，它是影响变量y的k个因素。例如，要建立生产函数，这时研究的经济变量为产出量，影响产出量y的因素有许多，如资本投入、劳动力投入、技术水平等，从影响因素中选择最主要的且可观察的因素，如资本投入x_1，劳动力投入x_2等作为解释变量。

非线性回归模型的一般形式：

$$f(y,x,\beta)=u \tag{5.3.1}$$

其中$f(y,x,\beta)$是非线性函数，β是$l\times 1$未知参数向量，u为随机误差。假设已有样本$(y_1,x_1),(y_2,x_2),\cdots,(y_n,x_n)$，其中$x_i=(x_{1i},\cdots,x_{ki})$，$i=1,\cdots,n$。将模型（5.3.1）写为

$$f(y_i,x_i,\beta)=u_i \quad i=1,\cdots,n \tag{5.3.2}$$

写成矩阵的形式为

$$F(Y,X,\beta)=U \tag{5.3.3}$$

其中

$$Y=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix},\quad X=\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}=\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1}\\ x_{12} & x_{22} & \cdots & x_{k2}\\ \vdots & \vdots & \vdots & \vdots\\ x_{1n} & x_{2n} & \cdots & x_{kn}\end{pmatrix}$$

$$F = \begin{pmatrix} f(y_1, x_1, \beta) \\ f(y_2, x_2, \beta) \\ \vdots \\ f(y_n, x_n, \beta) \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_l \end{pmatrix} \quad U = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

最常见的非线性回归模型是y和x相分离的形式：

$$f(y, x, \beta) = y - g(x, \beta) = u \tag{5.3.4}$$

称$g(x, \beta)$为期望函数，对应的矩阵表示为：

$$Y = G(X, \beta) + U$$

其中

$$G(X, \beta) = \begin{pmatrix} g(x_1, \beta) \\ g(x_2, \beta) \\ \vdots \\ g(x_n, \beta) \end{pmatrix}$$

例如，一元非线性回归模型：

$$y_i = \beta_1 + \beta_2 e^{\beta_3 x_i} + \varepsilon_i \quad i = 1, \cdots, n \tag{5.3.5}$$

和二元非线性回归模型：

$$y_i = \beta_1(\beta_2 x_{1i}^{\beta_3} + (1 - \beta_2) x_{2i}^{\beta_3})^{1/\beta_3} + \varepsilon_i \quad i = 1, \cdots, n \tag{5.3.6}$$

5.3.2 非线性回归模型的非线性最小二乘估计

1．非线性最小二乘估计

模型(5.3.2)参数的非线性最小二乘估计是就是最小化误差平方和：

$$S(\beta | Y, X) = U'U = F(Y, X, \beta)' F(Y, X, \beta) \tag{5.3.7}$$

由微积分知识知，该问题转化为求解方程组(一阶条件)

$$\frac{\partial S(\beta | Y, X)}{\partial \beta} = 2(\frac{\partial U}{\partial \beta'})' U = 0 \tag{5.3.8}$$

其中，$\frac{\partial S}{\partial \beta} = \begin{pmatrix} \frac{\partial S}{\partial \beta_1} \\ \frac{\partial S}{\partial \beta_2} \\ \vdots \\ \frac{\partial S}{\partial \beta_l} \end{pmatrix}$，$\frac{\partial U}{\partial \beta'} = \begin{pmatrix} \frac{\partial U_1}{\partial \beta_1} & \frac{\partial U_1}{\partial \beta_2} & \cdots & \frac{\partial U_1}{\partial \beta_l} \\ \frac{\partial U_2}{\partial \beta_1} & \frac{\partial U_2}{\partial \beta_2} & \cdots & \frac{\partial U_2}{\partial \beta_l} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial U_n}{\partial \beta_1} & \frac{\partial U_n}{\partial \beta_2} & \cdots & \frac{\partial U_n}{\partial \beta_l} \end{pmatrix}_{n \times l}$

并检验下列矩阵在（5.3.8）解$\hat{\beta}$处是否为正定矩阵

$$\frac{\partial^2 S(\beta|Y,X)}{\partial\beta'\partial\beta} = 2\left[\left(\frac{\partial U}{\partial\beta'}\right)'\left(\frac{\partial U}{\partial\beta'}\right) + \sum_{i=1}^{n} u_i \frac{\partial^2 u_i}{\partial\beta'\partial\beta}\right] \tag{5.3.9}$$

其中

$$\frac{\partial^2 S(\beta|Y,X)}{\partial\beta'\partial\beta} = \begin{pmatrix} \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_1^2} & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_1\partial\beta_2} & \cdots & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_1\partial\beta_l} \\ \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_2^2} & \cdots & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_2\partial\beta_l} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_l\partial\beta_1} & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_l\partial\beta_2} & \cdots & \frac{\partial^2 S(\beta|Y,X)}{\partial\beta_l^2} \end{pmatrix}$$

$$\frac{\partial^2 u_i}{\partial\beta'\partial\beta} = \begin{pmatrix} \frac{\partial^2 u_i}{\partial\beta_1^2} & \frac{\partial^2 u_i}{\partial\beta_1\partial\beta_2} & \cdots & \frac{\partial^2 u_i}{\partial\beta_1\partial\beta_l} \\ \frac{\partial^2 u_i}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 u_i}{\partial\beta_2^2} & \cdots & \frac{\partial^2 u_i}{\partial\beta_2\partial\beta_l} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 u_i}{\partial\beta_l\partial\beta_1} & \frac{\partial^2 u_i}{\partial\beta_l\partial\beta_2} & \cdots & \frac{\partial^2 u_i}{\partial\beta_l^2} \end{pmatrix} \tag{5.3.10}$$

例5.3.1 用非线性最小二乘法估计模型（5.3.5）参数的一阶条件为：

$$\begin{aligned} \frac{\partial S(\beta|Y,X)}{\partial\beta_1} &= -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i}) = 0 \\ \frac{\partial S(\beta|Y,X)}{\partial\beta_2} &= -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i})e^{\beta_3 x_i} = 0 \\ \frac{\partial S(\beta|Y,X)}{\partial\beta_3} &= -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 e^{\beta_3 x_i})\beta_2 x_i e^{\beta_3 x_i} = 0 \end{aligned} \tag{5.3.11}$$

2.非线性最小二乘的几何

对于模型（5.3.4），可将被解释变量(响应变量)向量Y视为n维欧氏空间（称为响应空间）的一个点，视函数向量$G(X,\beta)$为n维空间的一个参数向量为β的l维曲面(称为期望曲面)。这样，非线性最小二乘估计就是点Y在期望曲面上的投影的参数坐标$\hat{\beta}$。

3.参数初始值的选择

因为非线性回归模型的期望曲面形状复杂和参数坐标的非均匀性，如果参数初始值随意选择，常会造成非线性优化计算的迭代次数太多，或迭代无法收敛。为了确保非线性优化迭代计算的成功，一个最好的方法是设法获得恰当的参数初始值，以使收敛尽快发生。对于初始值的确定，可尝试如下的方法：

(1)利用期望函数相对于参数的性态

非线性回归的优点之一是期望函数中的参数通常都有经济意义，这对决定参数初始值非常有用。例如，C－D生产函数模型$Q = \beta_1 K^{\beta_2} L^{\beta_3} + u$中，参数$\beta_2$和$\beta_3$分别为资本和劳动对产出的弹性系数，因而可利用历史上的弹性系数作为它们的初始值。有时，研究期望函数在原点或其它特定点的形态能够获得初始值。例如，取$x = 10$,模型$g(x,\beta) = \beta_1 + \beta_2 \exp(-\beta_3(x-10))$的函数值为$\beta_1+\beta_2$；取$x \to \infty$,$g(x,\beta)$有渐近线$\beta_1$(假设$\beta_3 > 0$)。再如，对于模型$g(x,\beta) = \beta_1 x/(\beta_2 + x)$，$\beta_1$表示当$x \to \infty$时$g(x,\beta)$的渐近值$\tilde{y}$；而$\beta_2$是当$y = \tilde{y}/2$时解释变量$x$的数值。

(2)利用期望函数的导函数相对于参数的性态

有时，期望函数在特定点处的变化率能够用来获得参数的初始估计。例如对于双指数模型

$$g(x,\beta) = \beta_1 \exp(-\beta_2 x) + \beta_3 \exp(-\beta_4 x) \tag{5.3.12}$$

假设$\beta_2 > \beta_4 > 0$，函数当x很大时表现得与简单指数项$\beta_3 \exp(-\beta_4 x)$相似，而当x很小时，则表现得与$\beta_3 + \beta_1 \exp(-\beta_2 x)$相似。于是，当$x$很小时，函数的变化率提供了$\beta_2$的一个初始估计；当$x$很大时，变化率提供了$\beta_4$的一个初始估计。

(3)变换期望函数以获得更简单的关系（特别是线性关系）

期望函数的变换经常能够用来获得初始值。例如，对于

$$g(x,\beta) = \beta_1 x/(\beta_2 + x) \tag{5.3.13}$$

可简单地取模型函数的倒数得到线性形式：

$$\frac{1}{g(x,\beta)} = \frac{1}{\beta_1} + \frac{\beta_2}{\beta_1}\frac{1}{x}$$

这样，普通最小二乘方法可应用于数据的倒数$1/y$关于$1/x$(包括常数项)的线性回归，从而获得线性模型参数的估计，由此进一步获得原模型参数的初始值。对于模型

$$g(x,\beta) = \frac{\beta_1\beta_3(x_2 - x_3/1.632)}{1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3} \tag{5.3.14}$$

可变换为线性形式，因为

$$\frac{x_2 - x_3/1.632}{g(x,\beta)} = \frac{1}{\beta_2\beta_3} + \frac{\beta_2}{\beta_1\beta_3}x_1 + \frac{1}{\beta_1}x_2 + \frac{\beta_4}{\beta_1\beta_3}x_3$$

根据$(x_2 - x_3/1.632)/y$关于x_1, x_2, x_3(含常数项)的线性回归可获得参数的初始值。

变量变换也是非常有效的。作$\ln y$对x的图象经常会揭示数据真实的特性或者使人发现模型的某一部分何时是主要因素，进而能够洞悉

到何处能够测量一个比率并与某个参数联系起来。例如，对于双指数模型(6.4.8),当x很大时，模型$g(x,\beta)$近似等价于模型$\ln g = \ln\beta_3 - \beta_4 x$，此时$\ln y$对$x$的图象为一直线，可以简单地凭视觉拟合来获得$\beta_3$和$\beta_4$的初始值。这些值又能够用来计算$\beta_3\exp(-\beta_4 x)$,进而获得残差$\tilde{y} = y - \beta_3\exp(-\beta_4 x)$。然后，再作$\ln\tilde{y}$对$x$的图象，从而进一步得到$\beta_1$和$\beta_2$的初始值。这个过程称为退层；当期望函数是几个指数项之和时，可以采用这一过程。

(4)用数值替换某些参数或者在某些特定点计算函数值

通过降低维数(参数逐次进行估计)以便获得初始值是一个常用的技术，而退层就是这一技术的一个特例。每一参数获得估计后，其它参数会更容易估计。作为降维的另一个例子，考虑模型$g = \beta_1 + \beta_2\exp[-\beta_3(x-10)](\beta_3 > 0)$。当$x\to\infty$时，被解释变量的极限值为$\beta_1$，因而可获得初始值$\beta_1^0$；当$x = 10$时，被解释变量的值为$\beta_1+\beta_2$，可利用差$y(10)-\beta_1^0$得到初始值$\beta_2^0$。最后利用$\ln[(y-\beta_1^0)/\beta_2^0]$关于$x-10$不含截距的线性回归来获得$\beta_3^0$。此外，一旦确定了$\beta_1^0$及$\beta_2^0$，也可将这些值代入期望函数，并在特定的$x$处计算$[1/(x-10)]/\ln[(y-\beta_1^0)/\beta_2^0]$，从而得到$\beta_3^0$。

有时，可以降低模型的维数并由此间接地减少参数的个数。例如，考虑模型

$$g(x,\beta) = \exp[-\beta_1 x_1\exp(-\beta_2/x_2)] \tag{5.3.15}$$

如果有某些x_2的值非常大，此模型近似为$\exp(-\beta_1 x_1)$。这样，对应于这些x_2的值，对被解释变量取对数可以容易地获得β_1的初始值。类似地，对于模型

$$g(x,\beta) = \beta_1\beta_2 x_1/(1+\beta_1 x_1+\beta_2 x_2) \tag{5.3.16}$$

当x_2很小时，利用$1/y$关于$1/x_1$的回归，容易获得参数的初始值。

⑸利用条件线性关系

在许多模型函数中，可能有若干参数是依条件线性的。在给定非线性参数的初始值的条件下，可通过线性回归来获得这些参数的初始值。考虑

$$g(x,\beta) = \beta_1 + \beta_2\exp(-\beta_3 x) \tag{5.3.17}$$

其中β_1,β_2都是依条件线性的参数，所以一旦β_3^0已经获得，可以应用最小二乘法来获得β_1^0及β_2^0。

4.参数变换与参数约束

在非线性模型中，变换参数能够产生准确得多的线性近似。这对于改善近似推断区域和加速最小二乘的收敛都有有益的作用。参数变换还能用来处理对有关参数施加的限制。

参数变换与被解释变量的变换很不相同。被解释变量的变换使响应空间变形，并引出新的期望曲面。因而影响到随机扰动项及其有关假设条件

的有效性。相比之下，参数变换仅仅对参数空间及现行期望曲面上的点给出了新的坐标。因此，参数变换虽然影响到线性近似的有效性以及根据线性近似的推断，但并不影响模型的确定性部分及随机性部分的有关假设。

利用参数变换改善近似推断区域可参见5.4节。这里将集中讨论如何利用变换对参数施加限制。

(1)带限制的参数

在多数非线性模型中，参数必须限制在具有科学意义的区域内。当拟合模型时，经常可以忽略约束条件，简单地考察所收敛的参数估计是否满足约束条件。如果模型很好地拟合了数据，参数估计应当在有意义的区域内。尽管如此，在迭代过程中，允许参数进入非法区域有时是很危险的，因为那样的话，参数值可能剧烈波动或者引起数值溢出。在此情形下，应当在整个估计过程中对参数施加限制。

幸运的是，对于非线性回归模型的参数来说，其约束类型通常简单得足以由参数变换来处理。

①正数约束$\beta > 0$

可通过重新设置参数$\theta = \ln\beta$。这样，整个迭代过程$\beta = \exp(\theta)$恒正。

②区间类的约束$a \le \beta \le b$

可采用形为

$$\beta = a + \frac{b-a}{1+\exp(-\theta)} \tag{5.3.18}$$

的logistic型变换来实现。

③次序类的约束

次序类的约束可以用来保证具有可交换参数的模型有唯一的参数估计值。作为这样一类模型的一个例子，考虑双指数模型

$$g(x,\beta) = \beta_1 \exp(-\beta_2 x) + \beta_3 \exp(-\beta_4 x) \tag{5.3.19}$$

其中参数对(β_1,β_2)与(β_3,β_4)是可交换的；换言之，(β_1,β_2)与(β_3,β_4)互相交换不会改变期望函数的值。可交换参数可能引起棘手的最优化问题，因为线性近似不能说明这种参数的可交换性。要求参数有次序类的约束

$$\beta_2 > \beta_4 > 0$$

能够消除参数的可交换性。而利用变换

$$\beta_4 = \exp(\theta_4)$$

$$\beta_2 = [1+\exp(\theta_2)]\exp(\theta_4) \tag{5.3.20}$$

则可实现上述约束。因为β_1, β_3是依条件线性的参数，它们的最优值当β_2, β_4不等时是唯一确定的。这样，只需保持β_2与β_4的次序约束就能消除参数的可交换性。

5.获得收敛

获得收敛有时是困难的。如果遇到这一困难，不妨检查下述各项：

期望函数的表达式是否正确？

导函数的表达式是否正确？

数据输入是否正确？

参数的初始值是否有合理的数值？是否有正确的符号？是否对应着相应的参数？

如果对于所有上述问题的回答都是肯定的，仔细检查优化迭代过程的输出结果。大多数好的软件都可提供每一步迭代运算的详细结果以帮助找出错误所在。检查初始平方和$S(\beta^0)$是否比被解释变量的总变差(中心化平方和SST)更小。如果不是，函数拟合比没有函数拟合更差。尽管已经核查过，可能还是弄错了期望函数或者是数据，或者是初始值。

检查参数增量相对于参数是否大体上有相同的数量级？当参数向量加上增量后，是否在参数空间的某个不适宜的区域内？例如，有正值约束的参数是否取了负值？有没有参数其值大得不合理或者小得不适宜？如果这样，导函数中是否有错误？

有时，模型可能因为有太多的参数以至于不能收敛。考察某些参数值，看看没有这些参数会不会变成一个更简单模型。也应考虑参数增量的相关性，看看有没有成对的参数倾向于一起变化。如果这样，可能有共线性关系或者参数设置过多。

6.拟合的评估和模型的修正

任何非线性回归分析既要评估模型对于数据的拟合，也要评估有关随机扰动假设的合理性。如果模型不合适，或者模型合理但某些假设不合理，那么模型应当加以修正，直到获得满意的结果为止。拟合的评估和模型的修正可按照下述步骤进行：

a)非线性优化迭代计算所收敛的参数值可能是明显错误的或者值得怀疑的，这是因为有可能收敛到局部极小值点，或者由于期望曲面某种异常性态而使计算搁置。因此，任何模型拟合的评估首先应当仔细考察参数的估计值及其经济意义。如果参数值不合适，察看是否运用了正确的初始值，也要察看计算程序是否并非简单的终止，而是由于缺乏进展或者太多的迭代次数但实际上并未收敛。

b)应当研究迭代进展信息，弄清楚收敛是否平滑。如果迭代计算过程平稳地进行并得到明显合适的收敛点但参数并不合理，可察看期望函数的表达式，察看导函数的表达式，检查参数初始值和数据等。同时，还要检查残差的性态是否得当？

c)如果上述检查是满意的，但参数向量并不令人满意，可尝试取一个相当不同的参数初始值向量。如果收敛到同一个点，数据可能试图表明：期望函数是不合适的。

d)当合理的收敛值得到后，检查参数的近似标准差和近似t值[该值为（参数估计）/(近似标准差)]。如果某个t值不显著，可考虑从期望函数中删除该参数并重新拟合模型。
e)也应当检查参数估计量的近似相关矩阵，弄清楚是否有些参数估计量有过分高的相关值，因为过高的相关值可能表明参数过多(对数据集而言，模型太复杂)。构成高相关的因素确切地依赖于所研究的模型及数据的类型。一般，大约0.8的绝对相关值应当引起注意。以科学合理的方式尝试简化模型或者变量和参数以降低参数估计量的共线性程度。
f)当简单而合适的期望函数找到后，拟合值关于观测值的散点图是评估拟合的一个有用的工具。残差对拟合值及解释变量的散点图也是强有力的工具。也应对残差与其它潜在可能的因素作图以帮助检测模型的合适性。应特别注意残差是否有均匀的分布，因为残差图中任何非随机性的现象都提醒我们可能出现异方差。如果有异方差，可考虑变换数据来获得常数方差；同时，变换模型函数以保持模型的正确性。
g)如果残差对解释变量或其它变量的图象显示出残差有非随机性的表现，则这可能意味着期望函数不合适。在这样的情况下，应当以科学的方法尝试扩展模型以便消除非随机性。
h)应当绘制残差的概率图以便证实随机扰动的正态性假设。如果明显地缺乏正态性，应设法确定这种缺乏是由于少数异常点所引起的还是由于不合适的期望函数所引起。对于明显的异常点，应检查数据记录是否正确，可考虑是否将它们删除。如果残差明显非正态，如果是期望函数不合适，考虑变换数据或模型；如果是存在异常点，应将最小二乘准则改变为稳健估计准则。

7.模型的比较

在某些情况下，可能不止一个函数能够用来作为模型。例如，在一个双指数模型

$$g(x,\beta)=\beta_1\exp\left(-\beta_2 x\right)+\beta_3\exp\left(-\beta_4 x\right) \tag{5.3.21}$$

中，β_4可能是零。在此情形下，模型简化为

$$g(x,\beta)=\beta_1\exp\left(-\beta_2 x\right)+\beta_3 \tag{5.3.22}$$

或者β_3可能是零，此时，模型简化为

$$g(x,\beta)=\beta_1\exp\left(-\beta_2 x\right) \tag{5.3.23}$$

在这种嵌套模型的情况下找出最简单的、能够很好地拟合数据的模型。

在其它情况下，可能要比较非嵌套的模型。例如，模型1

$$g(x,\beta)=\beta_1[1-\exp(-\beta_2 x)] \tag{5.3.24}$$

对模型2

$$f(x,\beta)=\frac{\beta_1 x}{\beta_2+x} \tag{5.3.25}$$

两模型当$x=0$时，都有$g=0$,而当$x\to\infty$时都渐近趋向于β_1。在这些情形下，某一个模型可能对数据给出更为优越的拟合，当然应该选择该模型。

(1)嵌套模型

可能有若干个模型，它们都能很好地拟合数据。为了决定其中最简单的一个，可以象在线性情况下一样，采用一个似然比检验。由于球形正态的假设，可以通过附加平方和进行评估。附加平方和是由于从偏模型扩展到全模型时引入附加参数而引起的。

设S表示平方和，v表示自由度，P表示参数的数目，而下标f和p则用来区别全模型和偏模型，下标e说明附加，附加平方和为偏模型的残差平方和减去全模型的残差平方和。计算总结在表5.3.1中，我们将比值S_e^2/S_f^2与临界值$F(v_e,v_f;a)$进行比较。如果计算得出的比值小于临界值，则接受偏模型；否则，采用全模型。

注意，对于线性最小二乘，附加平方和分析是确切的，因为正是数据向量y在响应空间的线性子空间的投影决定了S_p和S_f。在数学上，偏模型的期望平面是全模型期望平面的一个线性子空间。偏模型的残差向量因此能够分解为全模型的残差向量和与之正交向量，因为全模型的残差平方和S_f服从$\sigma^2\chi^2$分布，其自由度为$N-P_f$；偏模型的残差平方和S_p服从$\sigma^2\chi^2$分布，其自由度为$N-P_p$，所以偏模型的残差向量分解出与全模型残差向量的正交向量的平方和也服从$\sigma^2\chi^2$分布，其自由度等于P_f-P_p。

表 5.3.1: 嵌套模型的附加平方和分析

来源	平方和	自由度	均方	F值
附加参数	$S_e=S_p-S_f$	$v_e=P_f-P_p$	$S_e^2=S_e/v_e$	S_e^2/S_f^2
全模型	S_f	$v_f=N-P_f$	$S_f^2=S_f/v_f$	
偏模型	S_p	$v_p=N-P_p$		

对于非线性模型，可以预料，这个分析仅仅是近似的，因为计算得出的比值并非确切的F分布。然而，比值的分布仅仅受固有非线性的影响，但并不受参数效应非线性的影响。而固有非线性强度一般是小的(参见5.4节)。

(2)非嵌套的模型

当试图确定几个非嵌套的模型中哪一个最好时，首先应当从经济学角度考虑。如果在理论上有理由认为某一个模型优于其它模型，应当特别重

视经济学理论和行为规律所提供的理由，因为数据分析的主要目的是解释或者说明经济现象，而非简单地获得最好的模型拟合。

如果从经济学角度不能提出令人信服的理由来选择某个模型时，那么应该应用统计分析的方法来确定最好的模型。最重要的方法可能是残差分析。一般来说，若模型的残差平方和最小，并且看起来残差最随机化，则这样的模型应当被选择。残差应当对预测值，对解释变量，以及对任何其它潜在可能的解释变量作残差图。

8.我国粮食生产模型

例5.3.2 我国粮食生产模型

$$\frac{Y}{M} = A \cdot (\frac{L}{M})^{\alpha} \cdot (\frac{K}{M})^{1-\alpha} + u \tag{5.3.26}$$

样本观测值为我国1975～1996年粮食产量、农业劳动力、粮食播种面积和化肥施用量的数据，见表5.3.2。参数的非线性最小二乘估计结果见表5.3.3。

我国1975～1996年粮食产量,农业劳动力,粮食播种面积和化肥施用量

5.3.3 非线性回归模型的非线性加权最小二乘估计

上述非线性最小二乘方法假定不同时期观察点的重要程度是一样的，不论是早期的观察点还是近期观察点都一样重要，但通常认为近期观察点比早期的观察点更重要。此时，就要采用非线性加权最小二乘估计。记权数矩阵为$W = W(\beta|Y, X)$是一个对角元非负的对角n阶方阵。定义$U^* = WU$为加权误差项。非线性加权最小二乘方法也就是最小化加权误差平方和：

$$S^*(\beta|Y, X) = (U^*)'(U^*) \tag{5.3.27}$$

例5.3.3(续上例) 设W=diag(1,2,Λ,23)，则参数的非线性加权最小二乘估计结果如表5.3.4。

5.3.4 非线性回归模型的最大似然估计

假定

$$U \sim N(0, \sigma^2 I) \tag{5.3.28}$$

其中I为n阶单位阵，则第i个观察点似然函数为：

$$L_i(\beta, \sigma^2|y_i, x_i) = \frac{J_i(\beta)}{\sqrt{2\pi\sigma^2}} \exp[-u_i^2/(2\sigma^2)]$$

表 5.3.2: 我国1975 1996年粮食产量,农业劳动力,粮食播种面积和化肥施用量

年份（年）	粮食产量 Y（万吨）	农业劳动力 L（万人）	粮食播种面积 M （千公顷）	化肥施用量 K（万公斤）
1975	28452	27561	121062	550000
1976	28631	27965	120743	597000
1977	28273	28124	120400	679000
1978	30477	28373	120587	884000
1979	33212	28692	119263	1086000
1980	32056	29181	117234	1269000
1981	32502	29836	114958	1335000
1982	35450	30917	113463	1513000
1983	38728	31209	114047	1660000
1984	40731	30927	112884	1740000
1985	37911	31187	108845	1776000
1986	39151	31311	110933	1931000
1987	40298	31720	111268	1999000
1988	39408	32308	110123	2141500
1989	40755	33284	112205	2357400
1990	44624	33336.4	113466	2590300
1991	43529	34186.3	112314	2805100
1992	44266	34037	110560	2930200
1993	45649	33258.2	110509	3151900
1994	44510	32690.3	109544	3317900
1995	46662	32334.5	110060	3593700
1996	50454	32260.4	110060	3827900

表 5.3.3: 模型(5.3.25)的非线性最小二乘估计结果

参数	参数估计值	标准差	t-统计量
A	0.4777665	0.0404860	11.800777
α	0.7668588	0.0199345	38.468979

表 5.3.4: 模型(5.3.25)的非线性加权最小二乘估计结果

参数	参数估计值	标准差	t-统计量
A	0.3482549	0.0461061	7.5533395
α	0.6960095	0.0292356	23.806898

对数似然函数为

$$\ln L_i(\beta,\sigma^2|y_i,x_i) = -\frac{1}{2}[\ln(2\pi)+\ln(\sigma^2)+u_i^2/\sigma^2]+\ln(J_i(\beta)) \quad (5.3.29)$$

其中$u_i = f(y_i,x_i,\beta)$且$J_i(\beta) = |\partial u_i/\partial y_i|$是$u_i$到$y_i$变换的雅可比行列式。模型参数的最大似然估计等价于最大对数似然估计，也就是最大化n个观察点的对数似然函数之和：

$$\ln L(\beta,\sigma^2,|y,x) = -\frac{1}{2}n[\ln(2\pi)+\ln(\sigma^2)] - \frac{1}{2}U'U/\sigma^2 + \sum_{i=1}^{n}\ln(J_i(\beta)) \quad (5.3.30)$$

因为

$$\frac{\partial \ln L(\beta,\sigma^2|Y,X)}{\partial\sigma^2} = -\frac{1}{2}\cdot\frac{n}{\sigma^2}+\frac{1}{2}\frac{U'U}{\sigma^4} \quad (5.3.31)$$

将$\sigma^2 = U'U/n$代入（5.3.29）,得到

$$\ln L_c(\beta|Y,X) = -\frac{1}{2}n[1+\ln(2\pi)-\ln n] - \frac{1}{2}n\times\ln(U'U) + \sum_{i=1}^{n}\ln(J_i(\beta)) \quad (5.3.32)$$

称之为中心化对数似然函数。令

$$U^* = U/[(J_1\times\cdots\times J_n)^{1/n}] \quad (5.3.33)$$

(5.3.31)可写成

$$\ln L_c(\beta|y,x) = -\frac{1}{2}n[1+\ln(2\pi)-\ln n] - \frac{1}{2}n\times\ln[(U^*)'(U^*)] \quad (5.3.34)$$

因而，最大化中心化对数似然函数$\ln L_c(\beta|Y,X)$等价于最小化加权误差平方和

$$S^*(\beta|Y,X) = (U^*)'(U^*) \quad (5.3.35)$$

记β的最大似然估计为$\hat{\beta}$，则σ^2的最大似然估计为

$$\hat{\sigma}^2 = \hat{U}'\hat{U}/n$$

很明显，模型的最大对数似然估计一般不等价于非线性最小二乘估计，而是一个加权非线性最小二乘估计。在$f(y,x,\beta) = y - g(x,\beta)$情况下，由于$J_1 = \cdots = J_n = 1$，所以最大对数似然估计才等价于非线性最小二乘估计。

例5.3.4 考虑模型

$$\ln\left(\frac{Y}{M}\right) = \beta_0 + \frac{1}{\beta_1}\cdot\ln(\beta_2\left(\frac{L}{M}\right)^{\beta_1} + \beta_3\left(\frac{K}{M}\right)^{\beta_1}) + u \tag{5.3.36}$$

数据见表5.3.2，则参数的最大对数似然估计结果如表5.3.5。

表 5.3.5: 模型(5.3.35)的最大对数似然估计

参数	参数估计值	标准差	t-统计量
β_0	-0.60875	1.27191	-0.47861
β_1	0.83021	0.5244	1.5832
β_2	1.4941	2.2384	0.66746
β_3	0.015437	0.0026109	5.9126

下面讨论在$f(y,x,\beta) = y - g(x,\beta)$e5Θ´ffiOœC5”

假设

$$U \sim N(0, \sigma^2 I) \tag{5.3.37}$$

性质1：模型(6.4.2)的参数非线性最小二乘估计和最大对数似然估计$\hat{\beta}$是渐近无偏、一致估计且渐近地服从正态分布$N(\beta, Q)$„

其中

$$Q = \left[-E\left(\frac{\partial^2 \ln L_c}{\partial\beta'\partial\beta}\right)\right]^{-1} = \sigma^2\left[\left(\frac{\partial U}{\partial\beta'}\right)'\left(\frac{\partial U}{\partial\beta'}\right)\right]^{-1} \tag{5.3.38}$$

性质2：模型(6.4.2)的参数σ^2最大对数似然估计$\hat{\sigma}^2 = \hat{U}'\hat{U}/n$是渐近无偏和一致估计。

5.3.5 非线性回归模型的工具变量估计

考虑非线性模型

$$y_i = g(x_i, \beta) + u_i \tag{5.3.39}$$

假设$U = (u_1, \cdots, u_n)' \sim N(0, \sigma^2 I)$；解释变量$x_i$可能与干扰项相关；假设存在工具变量$[z_1, \cdots, z_L]$使得

$$p\lim \frac{1}{n} Z'U = 0 \tag{5.3.40}$$

且

$$p\lim \frac{1}{n} Z'X^0 \neq 0 \tag{5.3.41}$$

其中$X^0 = \frac{\partial G(X,\beta)}{\partial \beta}$。

非线性工具变量法的参数估计为下式达最小的解b_{IV}：

$$\begin{aligned} Min_\beta S(\beta) &= \tfrac{1}{2}\{[Y - G(X,\beta)]'Z\}(Z'Z)^{-1}\{Z'[Y - G(X,\beta)]\} \\ &= \tfrac{1}{2}U(\beta)'Z(Z'Z)^{-1}Z'U(\beta) \end{aligned} \tag{5.3.42}$$

并且有

$$b_{IV} \xrightarrow{a} N[\beta, \sigma^2 Q] \tag{5.3.43}$$

其中σ^2和Q的估计分别为：

$$\hat{\sigma}^2 = U(\beta)'U(\beta)/n \,|_{\beta = b_{IV}} \tag{5.3.44}$$

$$\hat{Q} = Z'\hat{X}^0)^{-1}(Z'Z)(\hat{X}^0{\prime}Z)^{-1} \tag{5.3.45}$$

其中

$$\hat{X}^0 = \frac{\partial G(X,\beta)}{\partial \beta} \,|_{\beta\hat{}_{IV}}$$

5.4 非线性回归模型（二）—统计推断

本节首先研究非线性强度的曲率度量，并讨论如何应用它们来指示局部线性近似的合理性。并对67个真实的数据集与模型应用曲率度量，由此说明两类非线性(固有非线性和参数效应非线性)在实用中的重要性，也有力地支持了期望曲面平坦性假设的有效性。其次，将对非线性回归中参数的推断区域和推断区间等统计推断问题进行讨论。

5.4.1 非线性强度的曲率度量

非线性回归模型可以表示为

$$y_i = g(x_i, \beta) + u_i \quad i = 1, \cdots, n \tag{5.4.1}$$

其中g为期望函数，y_i为被解释变量，x_i为$1 \times k$确定性解释变量向量，β为$l \times 1$参数向量。

构造$n \times 1$向量$\eta(\beta)$，其第i个分量为

$$\eta_i(\beta) = g(x_i, \beta) \quad i = 1, \cdots, n$$

并记非线性回归模型为

$$Y = \eta(\beta) + U \tag{5.4.2}$$

假定U服从球形正态分布：

$$U \sim N(0, \sigma^2 I) \tag{5.4.3}$$

要决定期望曲面的平坦程度以及参数线在切平面上的不均匀程度，可以利用期望函数的二阶导数，并由此导出固有非线性和参数效应非线性的曲率度量。曲率还可用来研究期望曲面参数的重新设置，以便获得更为有效的线性近似的参数推断区域。

1. 速度向量和加速度向量

线性模型的基本特征之一是期望函数关于参数的二阶或更高阶的导数为零。因此，通过研究期望函数的二阶导数来度量模型的非线性强度是合乎逻辑的(Bates,1978;Bates和Watts,1980;Beale,1960)。为了清晰起见，用点符号表示一阶导数和二阶导数。这样，对于一个非线性模型，$n\times$的一阶导数矩阵可记为$\dot{V}$：

$$\{\dot{V}\}_{ip} = \frac{\partial g(x_i, \beta)}{\partial \beta_p} \quad i = 1, \cdots, n, p = 1, \cdots, l \tag{5.4.4}$$

$n \times l \times l$的二阶导数立体阵可记为$\ddot{V}$：

$$\{\ddot{V}\}_{ipq} = \frac{\partial^2 g(x_i, \beta)}{\partial \beta_p \partial \beta_q} \quad i = 1, \cdots, n, p, q = 1, \cdots, l \tag{5.4.5}$$

以矩阵的符号表示，则为

$$\dot{V} = \frac{\partial \eta}{\partial \beta'} \tag{5.4.6}$$

其中$\dot{V}$的每一行是$\eta(\beta)$的对应分量相对于β的梯度，也可认为$\dot{V}$由l个列向量（速度向量）$\dot{v}_p(p = 1, \cdots l)$

$$\dot{v}_p = \frac{\partial \eta}{\partial \beta_p} \tag{5.4.7}$$

组成。同样可记

$$\ddot{V} = \frac{\partial^2 \eta}{\partial \beta \partial \beta'} \tag{5.4.8}$$

其中立体阵$\ddot{V}$的每个面（共n个面）是$\eta(\beta)$的一个分量关于β的$l \times l$二阶导数矩阵或者Hessian矩阵。也可认为立体阵$\ddot{V}$由列向量（加速度向量）$\ddot{v}_{pq}(p, q = 1, \cdots l)$组成。

显然，速度向量$\dot{v}$是切向量，它们给出了η相对于参数的变化率。相应地，加速度向量$\ddot{v}$给出了速度向量相对于参数的变化，即

$$\ddot{v}_{pq} = \frac{\partial \dot{v}_p}{\partial \beta_q} \tag{5.4.9}$$

非线性模型为

$$g(x, \beta) = \frac{\beta_1 x}{\beta_2 + x}$$

表5.4.1为变量的数据，表5.4.2为在非线性最小二乘估计$\hat{\beta} = (212.7, 0.0641)'$处的速度向量和加速度向量。

表 5.4.1: 变量的数据

x	0.02	0.02	0.06	0.06	0.11	0.11	0.22	0.22	0.56	0.56	1.10	1.10
y	76	47	97	107	123	139	159	152	191	201	207	200

(1)切向加速度和法向加速度

因为$\ddot{v}_{pq} = \ddot{v}_{qp}$，所以加速度向量在响应空间生成维数最大不超过$l(l+1)/2$的子空间。加速度向量可以分解为切平面上的分量和正交于切平面的分量，分别称为加速度向量的切向分量和法向分量。为了确定加速度向量的切向分量和法向分量，将加速度向量投影到l维切空间，也投影到正交于切空间但由加速度向量生成的l^*维空间。这可以由下述步骤来完成。先将$l(l+1)/2$个加速度向量($\ddot{v}_{pq}, \ddot{v}_{qp}$只取一个)并入矩阵$\ddot{W}$,并与$\dot{V}$的切向量合并，这样可得：

$$D = (\dot{V}, \ddot{W}) \tag{5.4.10}$$

然后，对D进行QR分解（参考本节附录）：$D = QR = (Q_1|Q_2|Q_3)R$，其中Q是$n \times n$阶正交矩阵，$Q'Q = QQ' = I$,Q_1, Q_2, Q_3分

表 5.4.2: 速度向量和加速度向量

x	$\dot{v}_1$	$\dot{v}_2$	$\ddot{v}_{11}$	$\ddot{v}_{12}$	$\ddot{v}_{22}$
0.02	0.237812	-601.458	0	-2.82773	14303.4
0.02	0.237812	-601.458	0	-2.82773	14303.4
0.06	0.483481	-828.658	0	-3.8959	13354.7
0.06	0.483481	-828.658	0	-3.8959	13354.7
0.11	0.631821	-771.903	0	-3.62907	8867.4
0.11	0.631821	-771.903	0	-3.62907	8867.4
0.22	0.774375	-579.759	0	-2.72571	4081.4
0.22	0.774375	-579.759	0	-2.72571	4081.4
0.56	0.897292	-305.807	0	-1.43774	980
0.56	0.897292	-305.807	0	-1.43774	980
1.10	0.944936	-172.655	0	-0.81173	296.6
1.10	0.944936	-172.655	0	-0.81173	296.6

别是$n\times l$，$n\times l^*$和$n\times(n-l-l^*)$矩阵，R为n行的上三角阵（其非零的行数等于D的秩），可表示为$R=\begin{pmatrix} R_1 \\ 0 \end{pmatrix}$。$R$的每一列为$D$的相应列在$Q$坐标下的新表示，其中$R$的每一列前$l$个分量为$D$的相应列在切空间上的投影坐标，接下去的$l^*$个分量为$D$的相应列在正交于切空间但由加速度向量生成的$l^*$维空间上的投影坐标。

将立体阵$\ddot{V}$乘以$(Q_1|Q_2)'$，则有

$$\ddot{A}=[(Q_1|Q_2)'][\ddot{V}] \tag{5.4.11}$$

这给出了紧凑的$(l+l^*)\times l\times l$加速度立体阵$\ddot{A}$。这里，方括号表示乘积中的求和按下述方式进行：设B是$N_1\times N_2$的矩阵，C是$N_2\times N_3\times N_4$的立体阵，则积$A=[B][C]$是$N_1\times N_3\times N_4$的立体阵，A的第i面第p行第q列的元素为

$$\{A\}_{ipq}=\sum_{j=1}^{N_2}\{B\}_{ij}\{C\}_{jpq}$$

例5.4.2(续上例) 利用表5.4.2给出的速度向量和加速度向量来组成矩阵D,对D进行QR分解得到：

$$R_1=[(Q_1|Q_2)'][D]=\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

$$= \begin{bmatrix} -2.44 & 1568.7 & 0 & 7.378 & -16185.7 \\ 0 & 1320.3 & 0 & 6.210 & -25030.4 \\ 0 & 0 & 0 & 0 & 8369.1 \end{bmatrix}$$

显然左上角2×2矩阵R_{11}是由$\dot{V}$的QR分解得到的上三角矩阵非零部分，右上角2×3矩阵R_{12}给出了加速度向量在切平面的投影$[Q_1'][\ddot{W}]$,而右下角1×3矩阵R_{22}给出了加速度向量在与切空间正交的加速度空间的投影$[Q_2'][\ddot{W}]$。

重新整理R_{12}和R_{22}的因素，可得一个$3\times 2\times 2$的加速度立体阵$\ddot{A}$，即

$$\ddot{A} = \begin{bmatrix} 0 & 7.378 \\ 7.378 & -16185.7 \end{bmatrix} \begin{bmatrix} 0 & 6.210 \\ 6.210 & -25030.4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 8369.1 \end{bmatrix}$$

加速度向量在切平面外的伸展程度度量了期望曲面与切平面的偏离度，因而也度量了期望曲面的非平坦程度。称这种非平坦程度为固有非线性程度。它并不依赖于期望曲面所选用的参数设置，而仅依赖于期望曲面上点的位置和期望函数的表达式。加速度向量在切平面上的投影必然依赖于参数的选择，因而可以用来度量参数线在期望曲面上的非均匀程度。这种非均匀程度称为参数效应非线性强度，或简称为参数效应。

因为立体阵$\ddot{A}$中的元素提供了有关参数效应和固有非线性的信息，记$\ddot{A}$的前l个面为A^{β}表示参数效应加速度立体阵，而记$\ddot{A}$的后l^*个面为A^{τ}，表示固有加速度立体阵。

(2)任意方向上的加速度

速度向量和加速度向量仅能提供参数在参数空间中沿参数轴变化而引起期望曲面相应的变化的信息。为了度量在参数空间中β在$\hat{\beta}$附近沿任一方向h上的速度和加速度，现引进距离参数b，并取$\beta=\hat{\beta}+bh$，其中h是l维单位向量（$h'h=1$）。将直线$\beta(b)=\hat{\beta}+bh$映射到期望曲面，由此得到通过点$\eta(\hat{\beta})$的曲线$\eta_h(b)=\eta(\hat{\beta}+bh)$。通过对$b$微分并在$b=0$处计值，可得该曲线在$\hat{\beta}$处的切向量和加速度向量。因此

$$\left.\frac{d\eta_h}{db}\right|_{b=0} = \sum\nolimits_{p=1}^{l} \frac{d\eta}{d\beta_p}\left|\hat{\beta}\right. \left.\frac{d\beta_p}{db}\right|_{b=0} = \sum\nolimits_{p=1}^{l} \dot{v}_p h_p$$

以更紧凑的形式，该式可表示为

$$\dot{\eta}_h = \dot{V}h \tag{5.4.12}$$

与此相似，相应于方向h的加速度向量为：

$$\left.\frac{d^2\eta_h}{db^2}\right|_{b=0} = \sum\nolimits_{q=1}^{l} \frac{\partial \sum_{p=1}^{l} \dot{v}_p h_p}{\partial \beta_q}\left|\hat{\beta}\right. \left.\frac{d\beta_q}{db}\right|_{b=0} = \sum\nolimits_{p=1}^{l}\sum\nolimits_{q=1}^{l} \ddot{v}_{pq} h_p h_q \tag{5.4.13}$$

同样，该式可表示为：

$$\ddot{\eta}_h = h'\ddot{V}h \tag{5.4.14}$$

由(6.4.13)可知，方向h上的速度恰好是速度向量$\dot{v}$的线性组合；而由(6.4.14)和(6.4.15)可知，加速度也恰好是加速度向量$\ddot{v}$的线性组合。在(6.4.15)中，$n\times 1$的向量$\ddot{\eta}_h$可由$n\times l\times l$阶的立体阵乘以$1\times l$和$l\times 1$的向量得到。$\ddot{\eta}_h$的每个坐标具有形式$h'\ddot{V}_i h$，其中$\ddot{V}_i$是$\ddot{V}$的第i个面。

因为问题的主要兴趣是切平面上和法平面上的加速度，可将$\ddot{V}$按(6.4.12)式表示为$\ddot{A}$，这样的表示更为简洁。这时，方向h上的加速度变成：

$$(Q_1|Q_2)'\ddot{\eta}_h = h'\ddot{A}h$$

或者

$$(Q_1|Q_2)'\ddot{\eta}_h = \sum_{p=1}^{l}\sum_{q=1}^{l}\{\ddot{A}\}_{pq}h_p h_q \tag{5.4.15}$$

其中$\{\ddot{A}\}_{pq}$是立体阵$\ddot{A}$中$(l+l^*)\times 1$的向量。注意，这些向量可由R_1的列得到，而R_1可由方程(6.4.11)所定义的矩阵D进行QR分解得到。记矩阵

$$\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix} = [r_{11}, r_{12}, r_{22}, \cdots, r_{ll}] \tag{5.4.16}$$

则(6.4.16)可表示为：

$$(Q_1|Q_2)'\ddot{\eta}_h = \sum_{p=1}^{l} r_{pp}h_p^2 + 2\sum_{p=1}^{l}\sum_{q=p+1}^{l} r_{pq}h_p h_q \tag{5.4.17}$$

此式对于计算是相当方便的。

例5.4.3(续上例) 计算方向$h=(0.6,0.8)'$上的加速度。利用上例中的R_1的值，可得：

$$(Q_1|Q_2)'\ddot{\eta}_h = (0.6)^2 r_{11} + (0.8)^2 r_{22} + 2(0.6)(0.8)r_{12}$$

$$= 0.36\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + 0.64\begin{pmatrix} -16185.7 \\ -25030.4 \\ 8369.1 \end{pmatrix} + 0.96\begin{pmatrix} 7.378 \\ 6.210 \\ 0 \end{pmatrix} = \begin{pmatrix} -10351.8 \\ -16013.5 \\ 5356.2 \end{pmatrix}$$

2. 相对曲率

虽然加速度刻划了期望曲面的非线性特征，但是它们并不是非线性强度的有用度量，因为它们依赖于数据和参数的尺度。为了避免这种对于尺度的依赖性，将速度转换为相对曲率。

方向h上的曲率c_h定义为加速度向量的长度与切向量长度的平方之比，即

$$c_h = \frac{||\ddot{\eta}_h||}{||\dot{\eta}_h||^2}$$

对于参数效应曲率，有

$$c_h^\beta = \frac{||\ddot{\eta}_h^\beta||}{||\dot{\eta}_h||^2} \tag{5.4.18}$$

而对于固有效应曲率，则有

$$c_h^\tau = \frac{||\ddot{\eta}_h^\tau||}{||\dot{\eta}_h||^2} \tag{5.4.19}$$

从几何上看，固有曲率是方向η_h上期望曲面最好的近似园的半径的倒数。因此，如果期望曲面的平坦性假设合理，则固有曲率很小。

对加速度向量和切向量乘以Q'并不改变分子和分母的数值。所以采用更紧凑的立体阵A^β并利用

$$||h'A^\beta h||^2 = \sum_{p=1}^{l} (h'\ddot{A}_p h)^2$$

可计算(6.4.19)中的分子，同样采用立体阵A^τ并利用

$$||h'A^\tau h||^2 = \sum_{p=l+1}^{l+l^*} (h'\ddot{A}_p h)^2$$

可计算(6.4.20)中的分子。而分母则利用

$$||\dot{\eta}_h||^2 = ||\dot{V}h||^2 = h'R_{11}'R_{11}h = ||R_{11}h||^2$$

来计算，其中R_{11}由方程(6.4.11)定义的矩阵D的QR分解给出。

为了简化曲率的计算，约定对任何单位向量h都有$||\dot{\eta}_h|| = 1$。这能够通过β的正交参数线性变换来实现，即

$$\phi = R_{11}(\beta - \hat{\beta}) \tag{5.4.20}$$

进一步，因为曲率是由被解释变量倒数的单位来度量的，它们的值依赖于数据的尺度。为了消除这种相依性，将响应变量和曲率都乘

以$s\sqrt{l}[s^2=S(\hat{\beta})/(n-l)]$，由此将曲率转换为无量纲的相对曲率，最后得到相对曲率立体阵($||\dot{\eta}_h||=1$)：

$$C=(R_{11}^{-1})'\ddot{A}R_{11}^{-1}s\sqrt{l} \tag{5.4.21}$$

立体阵C由一个$l\times l\times l$阶的参数效应相对曲率立体阵C^{β}和一个$l^{*}\times l\times l$阶的固有相对曲率立体阵C^{τ}组成，其中C^{β}是C的前l个面，C^{τ}是C的后l^{*}个面。

例5.4.4(续上例) $s=10.93$，l=2，

$$R_{11}^{-1}=\begin{bmatrix}-0.4092 & 0.4861\\ 0 & 0.0007574\end{bmatrix}$$

根据例5.4.1，有

$$\ddot{A}=\begin{bmatrix}0 & 7.378\\ 7.378 & -16185.7\end{bmatrix}\begin{bmatrix}0 & 6.210\\ 6.210 & -25030.4\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & 8369.1\end{bmatrix}$$

所以

$$C=\begin{bmatrix}0 & -0.035\\ -0.035 & -0.059\end{bmatrix}\begin{bmatrix}0 & -0.030\\ -0.030 & -0.151\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & 0.074\end{bmatrix}$$

其中前2个矩阵给出了C^{β}，第三个矩阵给出了C^{τ}。

变换参数能够产生准确得多的线性近似且对改善参数的近似推断区域以及加速参数估计最小二乘的收敛有益。假如要检查参数变换$\theta=G(\beta)$后，期望曲面上的参数坐标的均匀性是否得到改善，则可利用参数变换的参数效应立体阵转换公式

$$C^{\theta}=C^{\beta}-[R_1][(R_1^{-1})'[\dot{G}^{-1}][\ddot{G}]R_1^{-1}] \tag{5.4.22}$$

计算参数变换后的参数效应立体阵。然后，检查参数效应立体阵是否有了改善。

另外，Ratkowsky(1983)提出了用参数估计的Box偏差来衡量新参数的好坏，也是很值得倡导的。

例5.4.5(续上例) 对模型重新设置参数为：

$$\theta=\begin{pmatrix}\frac{\beta_2}{\beta_1}\\ \frac{1}{\beta_1}\end{pmatrix}$$

故有

$$\dot{G} = \begin{bmatrix} -\frac{\beta_2}{\beta_1^2} & \frac{1}{\beta_1} \\ -\frac{1}{\beta_1^2} & 0 \end{bmatrix}$$

且

$$\ddot{G} = \begin{bmatrix} -\frac{2\beta_2}{\beta_1^3} & -\frac{1}{\beta_1^2} \\ -\frac{1}{\beta_1^2} & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\beta_1^3} & 0 \\ 0 & 0 \end{bmatrix}$$

这样，可得

$$[\dot{G}^{-1}][\ddot{G}] = \begin{bmatrix} -\frac{2}{\beta_1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{\beta_1} \\ -\frac{1}{\beta_1} & 0 \end{bmatrix}$$

最后，得到

$$\begin{aligned} C^\theta &= \begin{bmatrix} 0.000 & -0.035 \\ -0.035 & -0.059 \end{bmatrix} \begin{bmatrix} 0.000 & -0.030 \\ -0.030 & -0.151 \end{bmatrix} \\ &\quad - \begin{bmatrix} 0.000 & 0.003 \\ 0.003 & -0.027 \end{bmatrix} \begin{bmatrix} 0.000 & 0.002 \\ 0.002 & -0.005 \end{bmatrix} \\ &= \begin{bmatrix} 0.000 & -0.038 \\ -0.038 & -0.059 \end{bmatrix} \begin{bmatrix} 0.000 & -0.032 \\ -0.032 & -0.146 \end{bmatrix} \end{aligned}$$

可见，参数效应的变化是轻微的。

3. RMS曲率

为了理解模型的非线性，计算相对曲率立体阵是有益的。但是，对于数据分析，经常需要一个简单而综合的非线性强度的指标，使得在特定场合下能够评估线性近似的优良性。为此，采用均方根（RMS）曲率，即所有方向上曲率平方和的均方根，现以c^β表示RMS参数效应曲率，以c^τ表示RMS固有曲率(c^τ不依赖于期望曲面所选用的参数设置，仅依赖于期望曲面上的点和期望函数的表达式)。一般RMS曲率的表达式在下面给出；只要在相应的方程中插入上标β或τ，并选择指标i的适当范围就可以得到参数效应和固有效应的结果。

计算RMS曲率

在切平面上方向h对应的曲率为

$$c_h = ||h'C|| = \sqrt{\sum_i (h'C_i h)^2}$$

其中C_i是C的第i个面。在所有方向h上积分c_h^2并除以l维球面面积A，则得到均方曲率

$$c^2 = \frac{1}{A}\int \sum_i (h'C_i h)^2 dA \tag{5.4.23}$$

其中dA是球面面积的元素。Bates和Watts(1980)给出了最后结果：

$$c^2 = \frac{1}{l(l+2)}\sum_i \left[2\sum_{p=1}^{l}\sum_{q=1}^{l} c_{ipq}^2 + \left(\sum_{p=1}^{l} c_{ipq}\right)^2\right] \tag{5.4.24}$$

在（5.4.23）和(6.2.24)中，对于c^β，i遍取从1到l的整数，对于c^τ，i遍取从$l+1$到至多$l(l+1)/2$的整数。

RMS曲率何时才算是“大”的。Bates和Watts(1988)认为当$c\sqrt{F} < 0.3$时，可接受期望曲面平坦性假设或参数坐标均匀性假设，这里$F = F(l, n-l; \alpha)$。

例5.4.6(续上例) 取$\alpha = 0.05$，计算得到：

$$c^\tau = 0.045 \quad c^\beta = 0.05 \quad c^\tau\sqrt{F} = 0.09 \quad c^\beta\sqrt{F} = 0.21$$

(2)实际中的RMS曲率

现在介绍Bates和Watts(1988)对67个数据集和模型组合计算的RMS曲率。某些数据集不只采用一个期望函数，所以仅有37个不同的数据集和19个期望函数。37个数据集来源和19个期望函数参见他们的书（1988）中附录7。

对于所研究的67个数据集与某些组合，表5.4.3给出了RMS固有$c\sqrt{F}$和参数效应$c\sqrt{F}$的值，其中α选择为0.05。表中n是观测的总数，l是模型中参数的个数。图5.4.1a给出了参数效应$c^\beta\sqrt{F}$关于固有$c^\tau\sqrt{F}$的散点图。图5.4.1b和5.4.1c给出了参数效应$c^\beta\sqrt{F}$的直方图和固有$c^\tau\sqrt{F}$的直方图。图5.4.1a中的实线是等值线。由此容易看出，固有曲率比参数效应曲率更小，几乎在所有情况下都小得多。图5.4.1a还显示出固有非线性强度和参数效应非线性强度之间存在的相关。

参照直方图，可以看到固有曲率(图5.4.1b)一般并不大(大约93%的$c^\tau\sqrt{F}$值小于0.3)。另一方面，参数效应曲率(图5.4.1c)往往很强(大约仅10%的数据分析中$c^\theta\sqrt{F}$的值小于0.3)。

以上分析是根据各种来源获得的真实数据而得出的。这些结果表明，线性近似的推断区域在许多实际情况下可能是非常误导的，因为它要求均匀坐标的假设。上述结果也有力地支持了平坦性假设的有效性。平坦性假设已经在用来导出截面t图和截面配对图。

图5.4.1 曲率图

5.4.2 非线性回归模型的假设检验

假设

表 5.4.3: RMS固有$c\sqrt{F}$和参数效应$c\sqrt{F}$的值

数据集	模型	l	n	$c^{\beta}\sqrt{F}$	$c^{\tau}\sqrt{F}$
1	A	2	7	0.12	0.05
2			11	0.26	0.12
3			12	0.21	0.09
4	B		7	1.02	0.10
5			8	1.04	0.06
6			6	3.09	0.09
7			6	4.55	0.26
8			6	1.33	0.11
9			6	1.63	0.17
10			6	1.17	0.27
11			6	3.50	0.49
12			4	1.91	0.10
13	C		44	0.93	0.10
14	D		38	0.02	0.01
15	E		43	46.28	0.11
16	F		9	93.49	0.13
17			15	11.21	0.14
18			9	2.66	0.17
19			15	30.43	0.14
20			27	1.18	0.16
21			6	2.10	0.32
22			4	25.67	3.11
23			13	1.80	0.04
16	G		9	0.65	0.07
17			15	2.24	0.14
18			9	0.55	0.12
19			15	1.52	0.11
24			7	2.65	0.25
25			7	2.26	0.22
26			7	3.47	0.22
27			7	2.36	0.26
28			7	3.64	0.22
16	H		9	2.27	0.09
17			15	0.62	0.22
18			9	0.64	0.13
19			15	0.76	0.21
24			7	5.11	0.19
25			7	4.80	0.18
26			7	9.35	0.19
27			7	5.02	0.23
28			7	12.51	0.20
24	I		7	42.21	0.17
25			7	88.93	0.18
26			7	410.78	0.19
27			7	147.59	0.21
24	J		7	1.03	0.15

$$U \sim N(0, \sigma^2 I)$$

考虑参数的$J(< l)$个约束，约束方程为：

$$H_0 : c(\beta) = 0 \tag{5.4.25}$$

其中$c(\beta)$为$J \times 1$向量。

1. F检验

在假设(6.2.25)下，

$$F = \frac{(S(b) - S(\hat{\beta}))/J}{S(\hat{\beta})/(n-l)} \tag{5.4.26}$$

渐近地服从于$F(J, n-l)$分布。其中b为在约束(6.2.25)下模型参数β的非线性最小二乘估计，$\hat{\beta}$为在无约束下模型参数β的非线性最小二乘估计。

2. Wald检验

在假设(6.2.25)下，

$$W = c(\hat{\beta})'[Est.Asy.Var(c(\hat{\beta}))]^{-1}c(\hat{\beta})$$

$$= c(\hat{\beta})' \left[\left(\frac{\partial c(\hat{\beta})}{\partial \beta'} \right) Var(\hat{\beta}) \left(\frac{\partial c(\hat{\beta})}{\partial \beta'} \right)' \right]^{-1} c(\hat{\beta}) \tag{5.4.27}$$

渐近地服从于$\chi^2(J)$分布。

3. 拉格朗日乘子检验

在假设(6.2.25)下，

$$LM = (\frac{\partial \ln L(b)}{\partial \beta'}) Var(b) (\frac{\partial \ln L(b)}{\partial \beta'})' \tag{5.4.28}$$

渐近地服从于$\chi^2(J)$分布。其中

$$Var(b) = H^{-1}[I - G'(GH^{-1}G')^{-1}H^{-1}] \tag{5.4.29}$$

$$H = \left(-\frac{\partial^2 \ln L(b)}{\partial \beta' \partial \beta} \right)$$

$$G = \left(\frac{\partial c(b)}{\partial \beta'} \right)$$

LM检验统计量可由下列近似得到：

$$LM = \{[U'(\frac{\partial U}{\partial \beta'})] \cdot [(\frac{\partial U}{\partial \beta'})'(\frac{\partial U}{\partial \beta'})]^{-1}[U'(\frac{\partial U}{\partial \beta'})]'\}/(\frac{U'U}{n}) \quad (5.4.30)$$

其中

$$U = F(Y, X, b)$$

$$\frac{\partial U}{\partial \beta'} = \frac{\partial U(b)}{\partial \beta'} \quad (5.4.31)$$

4. 似然比检验

在假设(6.2.25)下，

$$LR = -2(\ln L(b) - \ln L(\hat{\beta})) = n \ln(S(b)/S(\hat{\beta})) \quad (5.4.32)$$

渐近地服从于$\chi^2(J)$分布。

注意：上述假设检验分析仅是近似的，检验统计量都是渐近地服从于某分布。因而，采用上述检验必须在大样本下进行。

例5.4.7(续5.2.1) 产出Q、资本投入K和劳动投入L的数据如表5.2.1，CES生产函数模型为：

$$f(\beta_1, \beta_2, \beta_3, \beta_4) = \sum_{i=1}^{n} \{\ln Q_i - \beta_1 - \beta_4 \ln[\beta_2 L_i^{\beta_3} + (1-\beta_2) K_i^{\beta_3}]\}^2$$

检验假设：$\beta_4 = 1/\beta_3$。

经计算W=0.071253,LM=0.092612,LR=0.082175。而在显著性水平为5%下，自由度为1的χ^2分布临界值是3.84。所以，Wald检验、拉格朗日乘子检验和似然比检验都在显著性水平为5%下接受了假设。

5.4.3 非线性回归推断的线性近似方法

下面通过导数矩阵进行线性近似来推断非线性模型，并在最小二乘估计处计值。根据

$$\eta(\beta) \approx \eta(\hat{\beta}) + \hat{V}(\beta - \hat{\beta}) \quad (5.4.33)$$

可以导出近似的似然区域和置信域。

在线性回归中，置信水平为$1-\alpha$的参数推断区域可以表示为；

$$(\beta - \hat{\beta})' X' X (\beta - \hat{\beta}) \leq l s^2 F(l, n-l; \alpha) \quad (5.4.34)$$

从几何上看，因为期望曲面是平面，并且残差向量垂直于期望平面，所以参数合理值的区域对应为期望平面上的一个圆盘。通过线性映射可以将期望平面中圆盘上的点映射为参数平面上的点，于是圆盘映射为参数平面上的椭圆。

类似于方程(5.8.34)，非线性回归的近似推断区域定义为

$$(\beta-\hat{\beta})'\hat{V}'\hat{V}(\beta-\hat{\beta})\leq ls^2F(l,n-l;\alpha) \tag{5.4.35}$$

或等价地

$$(\beta-\hat{\beta})'\hat{R}_1'\hat{R}_1(\beta-\hat{\beta})\leq ls^2F(l,n-l;\alpha)$$

其中导数矩阵$\hat{V}\hat{Q}_1\hat{R}_1$在$\hat{\beta}$处计值。这个推断区域的边界为：

$$\{\beta=\hat{\beta}+\sqrt{ls^2F(l,n-l;\alpha)}\hat{R}_1^{-1}d\,|\,||d||=1\} \tag{5.4.36}$$

非线性回归推断的线性近似方法是建立在期望曲面平坦性假设和参数坐标均匀性假设的基础之上的，所以在运用该方法时要考查这两个假设的合理性。

5.4.4 非线性推断域的图示法

不幸的是，在许多非线性分析中线性近似不能够充分地描述非线性分析的推断结果。下面将给出改进的图示法，由此反映非线性分析的推断结果。

联合参数似然区域

对于非线性模型

$$y_i=g(x_i,\theta)+u_i \quad i=1,\cdots,n$$

或

$$y=\eta(\theta)+u$$

在模型中，已假定随机扰动项的球形正态性，这样就可把似然方法的统计推断与期望空间中期望曲面的几何性质联系起来。对于线性和非线性模型，如果都假定球形正态性，那么似然边界线是由所有满足$\eta(\theta)$与y的距离等于定长的θ组成，也就是说，所有使$S(\theta)$为常数的θ组成。为了联系“置信水平”与边界线，类似于线性模型，假定概率保证程度为$1-\alpha$的联合似然区域是由所有满足：

$$S(\theta)\leq S(\hat{\theta})+\left[1+\frac{l}{n-l}F(l,n-l;\alpha)\right] \tag{5.4.37}$$

的θ组成，与线性模型一样，它是期望曲面与以y为中心的圆球的交线。然而，现在的期望曲面已不是一个平面，即使能够确定交线上的点，也不能简单地把期望曲面上的点映回到参数空间。

当$l=2$时，可通过计算网格上θ值对应的$S(\theta)$，然后利用网格内的线段来近似边界线。当$l>2$时，应描绘截面t图、截面迹图和截面配对图，以反映似然边界线和非线性程度。

例5.4.8(续例5.4.1) 图5.4.2的+号处表示模型参数的非线性最小二乘估计，实线显示了模型参数的置信水平为80%和95%的似然边界线，虚线表示线性近似椭圆。此时，线性近似的效果不错。

图5.4.2 例5.4.8图示图5.4.3 例5.4.9图示

表5.4.4 例5.4.9变量数据

x	1	2	3	4	5	7
y	8.3	10.3	19.0	16.0	15.6	19.8

例5.4.9 非线性模型

$$y=\theta_1[1-\exp(\theta_2 x)]$$

表5.4.4为变量数据，图5.4.3的+号处表示模型参数的非线性最小二乘估计，实线显示了模型参数的置信水平为80%和95%的似然边界线，虚线表示线性近似椭圆。此时，线性近似的椭圆完全不合适。

2. 截面t图

为了得到非线性模型参数的边缘似然区间，首先叙述线性模型置信区间与平方和函数的关系。对于一个线性模型，θ_p的$1-\alpha$边缘区间可以通过参数的学生化：

$$\frac{\theta_p-\hat{\theta}_p}{se(\hat{\theta}_p)}=\delta(\theta_p) \tag{5.4.38}$$

表示为：

$$-t(n-l;\alpha/2)\le\delta(\theta_P)\le t(n-l;\alpha/2) \tag{5.4.39}$$

但是学生化参数也可表示为：

$$\frac{\theta_p-\hat{\theta}_p}{se(\hat{\theta}_p)}=sign(\theta_p-\hat{\theta}_p)\sqrt{\tilde{S}(\theta_p)-S(\hat{\theta})}/s \tag{5.4.40}$$

其中

$$\tilde{S}(\theta_P)=\min_{\theta_{-p}}S((\theta_p,\theta'_{-p})')=S((\theta_p,\tilde{\theta}'_{-p})') \tag{5.4.41}$$

为子集平方和函数，$\tilde{\theta}_{-p} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_{p-1}, \tilde{\theta}_{p+1}, \cdots, \tilde{\theta}_n)'$为$\theta_{-p}$对应于$\theta_p$的条件最小二乘估计。记号$(\theta_p, \tilde{\theta}'_{-p})'$表示元素为$(\tilde{\theta}_1, \cdots, \tilde{\theta}_{p-1}, \theta_p, \tilde{\theta}_{p+1}, \cdots, \tilde{\theta}_n)'$的向量。

对于非线性模型，利用相同的记号来定义截面t函数：

$$\tau(\theta_p) = sign(\theta_p - \hat{\theta}_p)\sqrt{\tilde{S}(\theta_p) - S(\hat{\theta})}/s \tag{5.4.42}$$

类似于线性模型，参数θ_p的置信水平为$1-\alpha$的似然区间可以定义为所有满足

$$-t(n-l;\alpha/2) \leq \tau(\theta_P) \leq t(n-l;\alpha/2) \tag{5.4.43}$$

的θ_p组成的集合。称$\tau(\theta_p)$关于θ_p或$\delta(\beta_p)$的图象为截面t图。

截面t函数$\tau(\theta_p)$关于β_p的图象给出了单参数的精确似然区间，并揭示了估计的非线性程度。为了说明这一点，假定模型是线性的。于是$\tau(\theta_p)$关于β_p的图象为一条直线。$\tau(\theta_p)$关于学生化参数$\delta(\theta_p)$的图象是一条过原点的直线，其斜率为1。对非线性模型来说，$\tau(\beta_p)$关于$\delta(\theta_p)$的图象是一条曲线，而曲率的大小提供了模型非线性的信息。曲线越弯曲(曲率越大)，非线性程度越高。

例5.4.10(续 例5.4.8) 图5.4.4为模型的截面t图，分别关于β_1(图5.4.4a)和θ_2(图5.4.4b)的截面t图。实线为截面t图，直线为线性近似。点虚线为99%边缘似然区间。参数θ_1的置信水平为99%的似然区间为[191.1,236.7],其线性近似区间[190.7,234.7]是一个很好的近似。图5.4.4a中截面t图略有弯曲，说明非线性强度很小。参数β_2的置信水平为99%的似然区间为[0.0408,0.0972],其线性近似区间[0.0379,0.0907]也是一个很好的近似。图5.4.4b中截面t图略有弯曲，与θ_1的非线性相比，β_2的非线性强度较大，但它仍然较小。

图5.4.4 例5.4.10的截面t图

3. 截面迹图

作出条件极大值$\tilde{\theta}_{-p}$的分量关于θ_p的函数图象，可以得到另一个实用图——截面迹图。例如，计算完θ_1的子集似然，可以描出$\tilde{\theta}_2$关于θ_1的图象，等等，直至$\tilde{\theta}_p$关于θ_1的图象。接着，求解θ_2的子集似然，并作出$\tilde{\theta}_1$关于θ_2的图象，$\tilde{\theta}_3$关于θ_2的图象，等等，直至$\tilde{\theta}_3$关于θ_2。从参数θ_3到θ_p,我们仍然继续上述过程，计算其它参数的条件最小值，并作出截面迹图。最后，综合$\tilde{\theta}_q$关于θ_p和$\tilde{\theta}_p$关于θ_q的图象，这样得到配对的截面迹图。

对于非线性模型，截面迹图是弯曲的。如果似然边界线很狭长，那么截面迹图会靠得很近；如果似然边界线很宽大，那么截面迹图趋于垂直；如果似然边界线近乎于椭圆，那么截面迹图很平直。

例5.4.11(续例5.4.10) 图5.4.5为模型参数的似然截面迹图。图5.4.5给出了$\tilde{\theta}_2$关于θ_1 (实线)和$\tilde{\theta}_1$关于θ_2 (虚线)的截面迹图，并添加了置信水平为80%

和95%的似然边界线(点虚线)。图形表明，实线与边界线相交，交点处切线垂直，虚线也与边界线相交，交点处切线水平。

图5.4.5 例5.4.11似然截面迹图图5.4.6 例5.4.12似然截面迹图

例5.4.12(续例5.4.9) 图5.4.6为模型参数的似然截面迹图。给出了$\tilde{\theta}_2$关于θ_1 (实线)和$\tilde{\theta}_1$关于θ_2 (虚线)的截面迹图，并添加了置信水平为50%、80%和95%的似然边界线(点虚线)。图形表明，实线与边界线相交，交点处切线垂直，虚线也与边界线相交，交点处切线水平。

4.截面配对图

对于多个参数的模型，很难确定似然边界线并作图。但利用子集平方和以及截面迹图可以准确近似似然区域的二维投影。于是，对于区域的范围可以有一个直观的了解。比方说，为了确定置信水平为95%的似然边界线在(θ_1,θ_2)平面上的投影，利用θ_1的子集平方和可以找出似然边界线与$\tilde{\theta}_2$关于θ_1的迹图的交点，这样就确定了似然边界线上两个点。此外，似然边界线上这两个点的切线必定相互平行，因为这两点表示似然边界线在θ_1方向上的边界。类似地，根据θ_2的子集平方和和$\tilde{\theta}_2$关于θ_1的迹图，可以确定似然边界线上另外两个点。同时知道似然边界线在这两点处也有水平的切线。

利用所有这些信息——截面t图，似然边界线上的点，边界线上这些点的切线方向，以及边界线界于这些点的参数值之间等事实——可以给出似然边界线的一个精确的补插，称这些补插曲线为截面配对图。补插的步骤如下：

首先，利用三次样条曲线将变换到τ_p,τ_q。

其次，将τ的坐标标准化，即把τ除以$\sqrt{lF(l,n-l;\alpha)}$。这样，在τ的标准化坐标下，正方形-1$\leq$ τ_p,τ_q $\leq$ 1界定了置信水平为1-α的联合似然边界线。

第三，在τ的标准化坐标下，置信水平为1-α的联合似然边界线可以表示为如下参数形式：

$$\tau_p = \cos(a+d/2)$$

$$\tau_g = \cos(a-d/2) \tag{5.4.44}$$

这里角度a由$-\pi$增至π，相位d为正数。当边界线是椭圆时，相位为常数；当边界线不是椭圆时，相位将变动。因此，在一组给定点$(\theta_{p,i},\theta_{q,i})$,$i=1,\cdots,4$的情形下，可以计算反余弦$s_{p,i}=\arccos\theta_{p,i}$和$s_{q,i}=\arccos\theta_{q,i}$并计算反余弦的平均值和差值。再根据反余弦变换只能得到[0, π]之间的值且$\cos(-x)=\cos x$来确定角度和相位。

第四，在角度和相位图上，进行周期样条曲线的补插。

第五，利用(5.8.44)变换回到τ_p,τ_q。最后，从τ_p,τ_q变换回到。

例5.4.13(续例5.4.11) 图5.4.7 给出了τ坐标下模型参数的截面迹图。表5.4.5给出了τ坐标下，由截面迹图导出的模型参数的置信水平为95%的

边界线上点的坐标及其反余弦值，同时也给出了反余弦的平均值和差值及调整后的角度和相位。

图5.4.7 例5.4.13截面迹图

表5.4.5 模型参数的边界线上点的坐标及其反余弦值

标准化		反余弦				角度	相
τ_1	τ_2	1	2	平均	差	a	d
1.000	0.801	0.000	0.641	0.321	-0.641	-0.321	0.6
0.795	1.000	0.651	0.000	0.326	0.651	0.326	0.6
-1.000	-0.762	3.142	2.437	2.789	0.704	2.789	0.7
-0.769	-1.000	2.448	3.142	2.795	-0.693	-2.795	0.6

图5.4.8 例5.4.13图图5.4.9 例5.4.13图

图5.4.8给出了模型参数的95%似然边界线的补插。图5.4.8a给出了对应于截面迹上点的相位$(\cdot)$与补插曲线，图5.4.8b给出了对应点$(\cdot)$和补插值$a+d/2$和$a-d/2$。图5.4.8c和图5.4.8d分别给出了τ坐标下和θ坐标下迹上点$(\cdot)$与补插边界线。

图5.4.9a给出了模型参数的对应于置信水平为50，80，95和99%的迹图上点的坐标，以及由这些点的得到的补插边界线（截面配对图）。图5.4.9b给出了精确边界线图。该图表明，在所有置信水平上补插截面配对图都很准确。

当参数不只两个时，可以按照类似的步骤得到配对图。

5.4.5 附录：QR分解

为了计算$N\times P$矩阵X的QR分解，我们使用Householder变换(参见Householder，1958)，它是平面上反射的一种推广。记

$$H_u = I - 2uu'$$

这里I是$N\times N$单位阵，u为一个$N\times 1$的单位向量($u'u=1$)。容易验证，H_u是对称正交矩阵。

将H_u乘以$N\times 1$向量y，可得

$$H_u y = y - 2uu'y$$

它对应y关于过原点且垂直于u的直线进行反射得到的向量，见图5.4.10。选取

$$u = \frac{y-||y||e_1}{||y-||y||e_1||} \tag{5.4.45}$$

或

$$u = \frac{y + ||y||e_1}{||y + ||y||e_1||}$$

其中$e_1 = (1, 0, \cdots, 0)'$，利用Householder变换可以使向量y变换到x轴上；也就是说，在新坐标系下，向量$H_u y$的坐标为$(\pm||y||, 0, \cdots, 0)'$。

图5.4.10 QR分解

例A：为了计算

$$X = \begin{bmatrix} 1 & 1.26 \\ 1 & 1.82 \\ 1 & 1.22 \end{bmatrix}$$

的QR分解，采用(5.8.45)，取

$$u_1 = \frac{(1, 1, 1)' - \sqrt{3}(1, 0, 0)'}{1.5925} = \frac{(-0.7321, 1, 1)'}{1.5925}$$

使X的第一列变换到x轴上。

$$X_1 = H_{u_1} X = \begin{bmatrix} 1.7321 & 3.0600 \\ 0 & -0.6388 \\ 0 & -0.2388 \end{bmatrix}$$

现在进行第二次旋转，但它要垂直于第一次旋转。这时选择的u_2要使H_{u_2}不改变X_1的第一列，并能使X_1的第二列主对角线下的元素为零。为了保证不改变X_1的第一列，选取u_2的第一个元素为0。于是，向量u_2选定为

$$u_2 = \frac{(0, -0.6388, -0.2388)' - 0.6820(0, 1, 0)'}{1.3422}$$

由此可得

$$R = H_{u_2} H_{u_1} X = \begin{bmatrix} 1.7321 & 3.0600 \\ 0 & 0.6820 \\ 0 & 0 \end{bmatrix}$$

正如上例所述，在得到Householder 反射向量$u_1, \cdots, u_p$的同时，也得到了矩阵：

$$R = H_{u_P} \cdots H_{u_1} X$$

根据$R = Q'X$，可以推得：

$$Q = H_{u_1} \cdots H_{u_P}$$

5.5 非线性回归模型（三）—专门问题

本节讨论非线性回归模型中的几个专门问题，包括著名的Box-Cox变换和在金融市场计量经济学中得到广泛应用的自回归条件异方差模型（ARCH，Auto Regressive Conditional Hoteroskedasticity)、广义自回归条件异方差模型（GARCH，General Auto Regressive Conditional Hoteroskedasticity)等。

5.5.1 因变量的参数变换

1. 因变量的参数变换

(1) 因变量的参数变换模型

假定模型是

$$h(y_i,\theta)=g(x_i,\beta)+u_i \quad i=1,\cdots,n \tag{5.5.1}$$

其中$h(\cdot)$和$g(\cdot)$是非线性函数，y_i是因变量（被解释变量），$x_i=\begin{pmatrix} x_{1i} & x_{2i} & \cdots & x_{ki}\end{pmatrix}$是解释变量向量，$\theta$和$\beta$是参数，$u_i$是随机干扰项。并假定

$$(u_1,\cdots,u_n)'\sim N(0,\sigma^2 I) \tag{5.5.2}$$

(2)参数估计

模型参数的一种估计方法是最小二乘法，即最小化

$$S(\theta,\beta)=\sum_i[h(y_i,\theta)-g(x_i,\beta)]^2 \tag{5.5.3}$$

不过，对于这种回归模型，最大似然估计量是更有效的并且通常并非更麻烦。由假设(6.4.2),易推得y_i的密度函数是：

$$\left|\frac{\partial u_i}{\partial y_i}\right|(2\pi\sigma^2)^{-1/2}\exp\left(-\frac{[h(y_i,\theta)-g(x_i,\beta)]^2}{2\sigma^2}\right) \tag{5.5.4}$$

变换的雅可比行列式是：

$$J(y_i,\theta)=\left|\frac{\partial u_i}{\partial y_i}\right|=\left|\frac{\partial h(y_i,\theta)}{\partial y_i}\right|=J_i$$

再由假设(6.4.2)，可得因变量样本的对数似然函数为：

$$\ln L=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2+\sum_i\ln J(y_i,\theta)$$

$$-\frac{1}{2\sigma^2}\sum_i [h(y_i,\theta)-g(x_i,\beta)]^2 \tag{5.5.5}$$

应注意对数似然的两个方面。第一，很明显若没有雅可比行列式，参数的非线性最小二乘估计将是最大似然估计，然而，如果雅可比行列式包括θ，最小二乘法不是最大似然法。第二，至于，与无因变量参数变换的非线性回归模型类似，的最大似然估计量是：

$$\hat{\sigma}^2=\frac{1}{n}\sum_i [h(y_i,\hat{\theta})-g(x_i,\hat{\beta})]^2 \tag{5.5.6}$$

其中$\hat{\theta},\hat{\beta}$为最大似然估计。最大化对数似然函数的一阶条件为：

$$\frac{\partial \ln L}{\partial \beta}=\frac{1}{\sigma^2}\sum_i u_i\frac{\partial g(x_i,\beta)}{\partial \beta}=0$$

$$\frac{\partial \ln L}{\partial \theta}=\sum_i \frac{1}{J_i}\left(\frac{\partial J_i}{\partial \theta}\right)-\left(\frac{1}{\sigma^2}\right)\sum_i u_i\frac{\partial h(y_i,\theta)}{\partial \theta}=0$$

$$\frac{\partial \ln L}{\partial \sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_i u_i^2=0 \tag{5.5.7}$$

将$\sigma^2=\frac{1}{n}\sum\limits_i u_i^2$代入(6.4.6)，得到中心化对数似然函数：

$$\ln L_c=\sum_i \ln J(y_i,\theta)-\frac{n}{2}[1+\ln(2\pi)]-\frac{n}{2}\ln\left[\frac{1}{n}\sum_i u_i^2\right] \tag{5.5.8}$$

最大化(6.4.9)，就得到参数θ和β的最大似然估计$\hat{\theta}$和$\hat{\beta}$。

(3)响应系数和弹性系数

$$u_i=h(y_i,\theta)-g(x_i,\beta)=f(y_i,x_i,\theta,\beta)$$

$$\frac{\partial y_i}{\partial x_{ji}}=-\frac{\partial f}{\partial x_{ji}}\bigg/\frac{\partial f}{\partial y_i}=\frac{\partial g}{\partial x_{ji}}\bigg/\frac{\partial h}{\partial y_i}$$

$$\varepsilon_{yj}=\frac{\partial y}{\partial x_{ji}}\cdot\frac{x_{ji}}{y_i}=(\frac{\partial g}{\partial x_{ji}}\bigg/\frac{\partial h}{\partial y_i})\cdot(\frac{x_{ji}}{y_i}) \tag{5.5.9}$$

2. Box-Cox变换

建立模型最重要的就是确定变量之间的函数形式，但经济学家通常不知道变量之间的确切函数形式。Box和Cox（1964）引入Box-Cox变换，且由Zarembka(1968,1974)在经济学中作为一种建模手段得到一定的普及。

(1)Box-Cox非线性回归模型

Box和Cox（1964）提出了变换：

$$x^{(\lambda)}=\frac{x^{\lambda}-1}{\lambda} \tag{5.5.10}$$

要求变量x为正值。尽管λ的取值可以是整个实数域但多数应用有意义的取值范围为[-2，2]，当λ=2，是二次变换；当λ=05，是平方根变换；当λ=1，是线性变换；当λ=-1，是倒数变换；当λ=0，是对数变换（$\lim\limits_{\lambda\to 0}x^{(\lambda)}=\lim\limits_{\lambda\to 0}\frac{x^{\lambda}-1}{\lambda}=\ln(x)$）。

假设被解释变量y_i，解释变量$x_{1i},x_{2i},\cdots,x_{ki}$被都取正值，对每个被解释变量和解释变量分别进行Box-Cox变换建立非线性回归模型：

$$u_i=f(y_i,x_i,\lambda,\beta)=y_i^{(\lambda_0)}-(\beta_0+\beta_1 x_{1i}^{(\lambda_1)}+\cdots+\beta_k x_{ki}^{(\lambda_k)}) \tag{5.5.11}$$

(2)Box-Cox非线性回归模型的参数估计

假设

$u=(u_1,\cdots,u_n)'\sim N(0,\sigma^2 I)$

Kø.(6.4.12)æ)C¸Æœq,

$\ln L(\lambda,\beta,\sigma^2|y,x)=-\frac{1}{2}n[\ln(2\pi)+\ln(\sigma^2)]-\frac{1}{2}(u'u/\sigma^2)+(\lambda_0-1)\sum\limits_{i=1}^{n}\ln y_i$

(5.5.12)

中心化对数似然函数

$\ln L_c(\lambda,\beta|y,x)=-\frac{1}{2}n[1+\ln(2\pi)-\ln n]-\frac{1}{2}n\ln(u'u)+(\lambda_0-1)\sum\limits_{i=1}^{n}\ln y_i$

$$=-\frac{1}{2}n[1+\ln(2\pi)-\ln n]-\sum_{i=1}^{n}\ln y_i-\frac{1}{2}n\ln(u^{*\prime}u^{*}) \tag{5.5.12}$$

其中$u^{*}=u/[(y_1\times\cdots\times y_n)^{\lambda_0/n}]$

最大化(6.4.13)，就得到参数λ和β的最大似然估计$\hat{\lambda}$和$\hat{\beta}$。σ^2的最大似然估计为：

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}f^2(y_i,x_i,\hat{\lambda},\hat{\beta}) \tag{5.5.13}$$

(3)$AX\Sigma 5X$

$$\frac{\partial y_i}{\partial x_{ji}}=-\frac{\partial f}{\partial x_{ji}}\Big/\frac{\partial f}{\partial y_i}=\beta_j x_{ji}^{\lambda_j-1}/y_i^{\lambda_0-1}$$

$$\varepsilon_{yj}=\frac{\partial y_i}{\partial x_{ji}}\cdot\frac{x_{ji}}{y_i}=-(\frac{\partial f}{\partial x_{ji}}\Big/\frac{\partial f}{\partial y_i})\cdot(\frac{x_{ji}}{y_i})=\beta_j x_{ji}^{\lambda_j}/y_i^{\lambda_0}$$

表5.5.1中y的数据是用下述模型生成的

$$y = \beta_1^* + \beta_2 x + \beta_3 z + u \tag{5.5.14}$$

式中$\beta_1^* = 2, \beta_2 = \beta_3 = 1$,$u$是i.i.d.$N(0,\sigma^2)$,$\sigma^2 = 2.25$。(6.4.15)可看作

$$y^{(\lambda_1)} = \beta_1 + \beta_2 x^{(\lambda_2)} + \beta_3 z^{(\lambda_3)} + v \tag{5.5.15}$$

式中$\lambda_1 = \lambda_2 = \lambda_3 = 1$,$\beta_2 = \beta_3 = 1$,$\beta_1 = \beta_1^* + \beta_2 + \beta_3 - 1 = 3$。方程(6.4.16)用最大似然法估计，首先假定$\lambda$相同，结果列在表5.5.2中。然后考虑$\lambda$不同，结果列在表5.5.3中。

表5.5.1 生成数据表

x	z	y	x	z	y
5.381719	3.336064	11.00728	5.331691	5.716792	10.80478
3.009112	5.232138	9.48510	4.800605	5.066528	11.58843
4.375199	3.546423	6.444172	3.941124	3.141868	9.877543
5.472562	5.590764	13.44912	5.720727	2.194553	8.04193
4.701641	2.12414	11.50133	4.536456	2.789461	13.27586
4.047996	3.869655	12.35815	3.084652	4.212297	9.501099
5.855000	4.078571	12.99527	4.508164	5.634752	10.30391
4.820906	3.933097	14.13140	5.632088	4.536204	14.21714
2.453591	5.121045	8.881935	4.889667	4.324151	9.392368
2.453289	3.521771	5.953811	4.028880	3.201167	11.79774
2.512200	2.927539	7.790439	3.825508	5.426684	13.31844
3.462947	4.706172	10.50841	5.364851	4.563735	10.90918
3.816963	3.639600	10.39946	4.445447	3.734707	13.38010
2.723284	4.589414	9.679147	3.971793	4.092285	8.883878
4.203508	2.221079	7.571837	5.899375	3.708031	7.875557
3.493063	5.657767	12.43048	5.514231	3.824409	11.76839
4.294265	3.375917	13.01008	4.328222	2.222048	9.82212
3.345298	5.094239	10.55083	2.191853	5.540067	13.19992
5.096056	3.493393	9.798799	5.982851	4.974421	14.95182
3.257751	2.634803	9.098704	3.787909	4.886206	10.57826

在λ相同的假定下获得的第一组估计值是合理的，虽然λ的估计值相当低，使得95%置信区间(0.607,0.951)不能包含真值λ=1。β_2,β_3的估计值和也比它们的真值稍低，但对应的95%置信区间确实包含了真值。

在λ不同这一假定下获得的估计值相当差。虽然样本容量为40对于其它许多类型的模型来说会被认为相当大了，但对于这个模型来说显然不够大。直观地看，这个困难在于，以x为例，存在两个参数β_2和λ_2，

二者都左右着x对y (或对$y^{(\lambda_1)}$)的影响。可能存在范围很宽的β_2和λ_2值，它们导致的x对y的影响类似。例如，高的β_2和低的λ_2具有的影响可能类似于低的β_2和高的λ_2的影响。这种可能性似乎可由β_2和λ_2的估计值分别为25.846和-1.536以及标准误差分别为73.34和2.30来证实。显然，在试图估计这类模型时必须极其谨慎，没有很大的样本容量不应该试图去做。

表5.5.2 λ相同时的估计结果

	ﬄO	IOﬀ	tO
β_1	3.380	1.128	2.99645
β_2	0.746	0.278	2.68345
β_3	0.574	0.278	2.06475
λ	0.779	0.088	8.85227
σ^2	1.240	0.392	3.16327

表5.5.3 λ不同时的估计结果

	ﬄO	IOﬀ	tO
β_1	-8.471	26.127	-0.3242
β_2	25.846	73.338	0.35245
β_3	1.610	5.264	0.30585
λ_1	0.969	0.089	10.8876
λ_2	-1.536	2.298	-0.6684
λ_3	0.391	2.613	0.14964
σ^2	2.876	0.910	3.16044

5.5.2 异方差性的非线性方法

对于非线性模型

$$f(y_i, x_i, \beta) = u_i \quad i = 1, \cdots, n$$

假设

$$U = (u_1, \cdots, u_n)' \sim N(0, \sigma^2\Omega(\alpha)) \tag{5.5.16}$$

其中$\Omega(\alpha)$为对角元为正的对角方阵，且依赖于参数$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)$.即$u_i$不存在序列相关但存在异方差现象。

被解释变量样本的对数似然函数为：

$$\ln L(\beta, \alpha, \sigma^2 | y, x) = -\frac{1}{2}n[\ln(2\pi) + \ln(\sigma^2)]$$

$-\frac{1}{2}\ln(|\Omega|)-\frac{1}{2}U'\Omega^{-1}U/\sigma^2+\sum\limits_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|)$ (5.5.18)

将σ^2用$\sigma^2=U'\Omega^{-1}U/n$代替得到中心化对数似然函数为：

$$\ln L_c(\beta,\alpha|y,x)=-\frac{1}{2}n[1+\ln(2\pi)-\ln(n)]$$

$-\frac{1}{2}\ln(|\Omega|)-\frac{1}{2}n\ln(U'\Omega^{-1}U)+\sum\limits_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|)$ (5.5.19)

下面以异方差的不同结构对Ω的参数α及模型参数β同时进行最大似然估计。

1.$u_i\sim N(0,\sigma_i^2)$

对数似然函数为：

$$\ln L(\beta,\sigma_1^2,\cdots,\sigma_n^2|y,x)=-\frac{1}{2}n\ln(2\pi)$$

$-\frac{1}{2}\sum\limits_{i=1}^{n}\ln(\sigma_i^2)-\frac{1}{2}\sum\limits_{i=1}^{n}(u_i^2/\sigma_i^2)+\sum\limits_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|)$ (5.5.20)

2.$\sigma_i^2=\sigma^2h_i(\alpha)$

对数似然函数为：

$$\ln L(\beta,\alpha,\sigma^2|y,x)=-\frac{1}{2}n[\ln(2\pi)+\ln(\sigma^2)]-\frac{1}{2}\sum_{i=1}^{n}\ln(h_i)$$
$$-\frac{1}{2}(1/\sigma^2)\sum_{i=1}^{n}(u_i^2/h_i)+\sum_{i=1}^{n}\ln(|\frac{\partial u_i}{\partial y_i}|) \qquad (5.5.17)$$

让$u_i^*=u_i/h_i^{1/2}$，$U^*=(u_1^*,\cdots,u_n^*)'$。将$\sigma^2=(U^*)'(U^*)/n$代入(6.4.18)，得到中心化对数似然函数：

$$\ln L_c(\beta,\alpha|y,x)=-\frac{1}{2}n[1+\ln(2\pi)-\ln(n)]$$

$-\frac{1}{2}\sum\limits_{i=1}^{n}\ln(h_i)-\frac{1}{2}n\ln[(U^*)'(U^*)]+\sum\limits_{i=1}^{n}\ln\left|\frac{\partial u_i}{\partial y_i}\right|$ (5.5.22)

下面是$h_i=h(\alpha|z_i)=h(z_i\alpha)$的常见形式，其中$z$是一组解释变量，可以是$x$也可不是。

(1)$\sigma_i^2=\sigma^2(z_i\alpha)$，$z_i\alpha>0$

(2)$\sigma_i^2=\sigma^2(z_i\alpha)^2$

(3)$\sigma_i^2=\sigma^2\exp(z_i\alpha)$

(4)$\sigma_i^2=\sigma^2\prod\limits_{j=1}^{m}z_{ij}^{\alpha_j}$

例5.5.2 1979年美国各州的公立学校平均开支Y和人均收入X的数据见表5.5.4。建立模型：

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u \tag{5.5.18}$$

表5.5.4 数据表

州	开支	收入	州	开支	收入	州	开支	收入
AL	275	6247	AK	821	10851	AZ	339	7374
AR	275	6183	CA	387	8850	CO	452	8001
CT	531	8914	DE	424	8604	DC	428	10022
FL	316	7505	GA	265	6700	HI	403	8380
ID	304	6813	IL	437	8745	IN	345	7696
IA	431	7873	KS	355	8001	KY	260	6615
LA	316	6640	ME	327	6333	MD	427	8306
MA	427	8063	MI	466	8442	MN	477	7847
MS	259	5736	MO	274	7342	MT	433	7051
NB	294	7391	NV	359	9032	NH	279	7277
NJ	423	8818	NM	388	6505	NY	447	8267
NC	335	6607	ND	311	7478	OH	322	7812
OK	320	6951	OR	397	7839	PA	412	7733
RI	342	7526	SC	315	6242	SD	321	6841
TN	268	6489	TX	315	7697	UT	417	6622
VT	353	6541	VA	356	7624	WA	415	8450
WV	320	6456	WI	NA	7597	WY	500	9096

下面考虑两种随机扰动项的异方差结构,并分别用最大似然方法同时估计出模型和异方差结构中的参数。

(1)$\sigma_i^2 = \sigma^2 X_i^\alpha$

最大似然估计的结果列于表5.5.5。

表5.5.5 (1)情况的估计结果

	fflO	IOff	tO
β_0	831.02	354.36	2.3451
β_1	-1833.2	936.60	-1.9573
β_2	1588.7	611.87	2.5964
α	3.0100	1.3644	2.2061

(2)$\sigma_i^2 = \sigma^2 \exp(\alpha X_i)$

最大似然估计的结果列于表5.5.6。

表5.5.6 (2)情况的估计结果

	ﬄO	IOﬀ	tO
β_0	831.21	362.81	2.2911
β_1	-1833.1	961.84	-1.9058
β_2	1588.5	630.65	2.5189
α	3.9094	1.7291	2.2610

5.5.3 序列相关性的非线性方法

考虑非线性模型：

$$u_t = f(y_t, x_t, \beta) t = 1, 2, \cdots, T \tag{5.5.19}$$

假定u_t同方差不存在异方差现象，但存在序列相关。下面分序列相关的不同类型进行讨论。

1.AR（1）：$u_t = \rho u_{t-1} + \varepsilon_t$

假定$\varepsilon \sim N(0, \sigma^2 I)$，且$|\rho| < 1$（模型稳定的条件）。很明显有

$\sigma^2 = Var(\varepsilon_t) = (1-\rho^2) Var(u_t)$

让

$$\varepsilon_1 = (1-\rho^2)^{1/2} u_1, \varepsilon_t = u_t - \rho u_{t-1} \quad t = 2, 3, \cdots, T \tag{5.5.20}$$

$$|\partial\varepsilon_t/\partial y_t| = \begin{cases} (1-\rho^2)^{1/2} \left|\frac{\partial u_1}{\partial y_1}\right| & t = 1 \\ \left|\frac{\partial u_t}{\partial y_t}\right| & t > 1 \end{cases} \tag{5.5.21}$$

被解释变量样本的对数似然函数为：

$$\ln L(\beta, \sigma^2, \rho | y, x) = -\frac{1}{2} T[\ln(2\pi) + \ln(\sigma^2)]$$

$$-\frac{1}{2}\varepsilon'\varepsilon/\sigma^2 + \frac{1}{2}\ln(1-\rho^2) + \sum_{t=1}^{T} \ln\left|\frac{\partial u_t}{\partial y_t}\right| \quad 5.5.32$$

将$\sigma^2 = \varepsilon'\varepsilon/T$代入(5.5.32)，得到中心化对数似然函数：

$$\ln L_c(\beta, \rho | y, x) = -\frac{1}{2} T[1 + \ln(2\pi) - \ln T]$$

$$-\frac{1}{2} T \times \ln(\varepsilon'\varepsilon) + \frac{1}{2}\ln(1-\rho^2) + \sum_{t=1}^{T} \ln\left|\frac{\partial u_t}{\partial y_t}\right| \tag{5.5.22}$$

最大化(5.5.33)，就得到参数β和ρ的最大似然估计$\hat{\beta}$和$\hat{\rho}$。σ^2的最大似然估计为：

$$\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/T \tag{5.5.23}$$

2．MA（1）：$u_t = \varepsilon_t - \theta\varepsilon_{t-1}$

假定$\varepsilon \sim N(0, \sigma^2 I)$，且$|\theta| < 1$（模型稳定的条件）。令

$$\varepsilon_0 = E(\varepsilon) = 0 \quad \varepsilon_t = u_t + \theta\varepsilon_{t-1} \quad t = 1, 2, \cdots, T \tag{5.5.24}$$

$$|\partial\varepsilon_t/\partial y_t| = \left|\frac{\partial u_t}{\partial y_t}\right| \tag{5.5.25}$$

$$\ln L(\beta, \sigma^2, \theta|y, x) = -\frac{1}{2}T[\ln(2\pi) + \ln(\sigma^2)] - \frac{1}{2}\varepsilon'\varepsilon/\sigma^2 + \sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right|$$

5.5.37）

中心化对数似然函数为：

$$\ln L_c(\beta, \rho|y, x) = -\frac{1}{2}T[1 + \ln(2\pi) - \ln T] - \frac{1}{2}T \times \ln(\varepsilon'\varepsilon) + \sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right|$$

(5.5.38)

最大化(5.5.38)，就得到参数β和ρ的最大似然估计$\hat{\beta}$和$\hat{\rho}$。σ^2的最大似然估计为：

$$\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/T$$

3．ARMA（1，1）：$u_t = \rho u_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$

假定$\varepsilon \sim N(0, \sigma^2 I)$，且$|\rho| < 1, |\theta| < 1$（模型稳定的条件）。令

$$\varepsilon_1 = (1 - \rho^2)^{1/2}u_1 \quad \varepsilon_t = u_t - \rho u_{t-1} + \theta\varepsilon_{t-1} \quad t = 2, \cdots, T \tag{5.5.26}$$

$$|\partial\varepsilon_t/\partial y_t| = \begin{cases} (1-\rho^2)^{1/2}\left|\frac{\partial u_1}{\partial y_1}\right| & t = 1 \\ \left|\frac{\partial u_t}{\partial y_t}\right| & t > 1 \end{cases} \tag{5.5.27}$$

$$\ln L(\beta, \sigma^2, \rho, \theta|y, x) = -\frac{1}{2}T[\ln(2\pi) + \ln(\sigma^2)]$$

$$-\frac{1}{2}\varepsilon'\varepsilon/\sigma^2 + \frac{1}{2}\ln(1-\rho^2) + \sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right| \quad 5.5.41$$

中心化对数似然函数为：

$$\ln L_c(\beta,\rho,\theta|y,x) = -\frac{1}{2}T[1+\ln(2\pi)-\ln T]$$

$$-\frac{1}{2}T\times\ln(\varepsilon'\varepsilon) + \frac{1}{2}\ln(1-\rho^2) + \sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right| \qquad (5.5.28)$$

最大化(5.5.42)，就得到参数β和ρ的最大似然估计$\hat{\beta}$和$\hat{\rho}$。σ^2的最大似然估计为：

$$\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/T$$

例5.5.3 投资Y、利率、GNP和价格指数的数据见表5.5.7。建立模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \qquad (5.5.29)$$

其中Y为实际投资=（名义）投资/价格指数，X_1为实际利率=（名义）利率-价格指数变化率，X_2为实际GNP=（名义）GNP/价格指数。

表5.5.7 数据表

下面考虑两种随机扰动项的序列相关结构,并分别用最大似然方法同时估计出模型和序列相关结构中的参数。

AR（1）：$u_t = \rho u_{t-1} + \varepsilon_t$

最大似然估计的结果列于表5.5.8。

表5.5.8 情况（1）的估计结果

(2)MA（1）：$u_t = \varepsilon_t - \theta\varepsilon_{t-1}$

最大似然估计的结果列于表5.5.9。

表5.5.9 情况（2）的估计结果

5.5.4 条件异方差性的非线性方法

通常横截面数据会产生异方差，而一般时间序列没有异方差现象。在对宏观经济数据的分析中，特别在外汇和股票市场等投机市场，发现有些经济现象时间序列数据模型，其扰动项在较大幅度的波动之后伴随着较大幅度波动在较小幅度的波动之后伴随着较小幅度波动，即随机扰动项的无条件方差是常量，但条件方差却是变化的量。作为资产持有者，并不会对报酬率的无条件方差感兴趣，而是感兴趣于它的条件方差。Engle(1982)为研究这类现象提出了条件异方差模型。

年份	GNP	投资	价格指数	利率
1963	596.7	90.0	0.7167	3.23
1964	637.7	97.4	0.7277	3.55
1965	691.1	113.5	0.7436	4.04
1966	756.0	125.7	0.7676	4.50
1967	799.6	122.8	0.7906	4.19
1968	873.4	133.3	0.8254	5.16
1969	944.0	149.3	0.8679	5.87
1970	992.7	144.2	0.9145	5.95
1971	1077.6	166.4	0.9601	4.88
1972	1185.9	195.0	1.0000	4.50
1973	1326.4	229.8	1.0575	6.44
1974	1434.2	228.7	1.1508	7.83
1975	1549.2	206.1	1.2579	6.25
1976	1718.0	257.9	1.3234	5.50
1977	1918.3	324.1	1.4005	5.46
1978	2163.9	386.6	1.5042	7.46
1979	2417.8	423.0	1.6342	10.28
1980	2631.7	401.9	1.7842	11.77
1981	2954.1	474.9	1.9514	13.42
1982	3073.0	414.5	2.0688	11.02

	ﬄO	IOﬀ	tO
β_0	-14.861	29.747	-0.49959
β_1	-0.78573	2.5254	-0.31114
β_2	0.17027	0.024658	6.9050
ρ	0.30096	0.25994	1.1578

	ﬄO	IOﬀ	tO
β_0	-13.912	29.103	-0.47802
β_1	-0.88807	2.5289	-0.35117
β_2	0.16943	0.024133	7.0207
θ	-0.43252	0.29298	-1.4763

一般认为，投资者的投资决策是基于对股票报酬分布的认识，因此，股票报酬方差或标准差是影响投资行为的重要因素。同时，深刻了解股票市场波动(风险)，分析引起波动的原因，并预测波动，政府可采取相应措施降低市场波动。所以，探讨股票市场的波动以及与预期股票报酬的关系，对于政府制定股市干预政策及金融政策，投资者评价股票价值和参与市场交易策略具有十分重要的意义。

1. ARCH（q）模型

设$u_t = f(y_t, x_t, \beta)$是一个实数域上的离散时间随机过程，ψ_t是直到t时间的所有信息集(即由$u_1, \cdots, u_t$产生的σ一域)。ARCH(q)过程由下面给出：

$$u_t|\psi_{t-1} \sim N(0, h_t), \tag{5.5.30}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 = \alpha_0 + \alpha(B) u_t^2 \tag{5.5.31}$$

其中$q > 0; \alpha_0 > 0, \alpha_i \geq 0, i = 1, \cdots, q$。

由Bollerslev(1986)可知：当且仅当$\alpha(1) < 1$时,由(5.4.44)和(5.4.45)定义的ARCH(q)过程是宽平稳的,其均值$E(u_t) = 0$,无条件方差$V(u_t) = \alpha_0(1 - \alpha(1))^{-1}$且协方差$cov(u_t, u_s) = 0, t \neq s$。

设样本有n个观察个数，则对数似然函数为：

$$\ln L(\beta, \alpha|y, x) = -\frac{1}{2} T \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln(h_t) - \frac{1}{2} \sum_{t=1}^{T} (u_t^2/h_t) + \sum_{t=1}^{T} \ln \left| \frac{\partial u_t}{\partial y_t} \right|$$

(5.5.46)

使该函数达最大值，就可得到参数β, α的估计。

2. GARCH（p,q）模型

GARCH(p,q)过程由下面给出：

$$u_t|\psi_{t-1} \sim N(0, h_t), \tag{5.5.32}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i u_{t-i}^2 + \sum_{i=1}^{p} \theta_i h_{t-i} = \alpha_0 + \alpha(B) u_t^2 + \theta(B) h_t \tag{5.5.33}$$

其中$p \geq 0, q > 0; \alpha_0 > 0, \alpha_i \geq 0, i = 1, \cdots, q; \theta_i \geq 0, i = 1, \cdots, p$。

由Bollerslev(1986)可知：当且仅当$\alpha(1) + \theta(1) < 1$时,由(5.8.28)和(5.8.31)定义的GARCH(p,q)过程是宽平稳的,其均

值$E(u_t)=0$,无条件方差$V(u_t)=\alpha_0(1-\alpha(1)-\theta(1))^{-1}$且协方差$cov(u_t,u_s)=0,t\neq s$。

对数似然函数为：

$$\ln L(\beta,\alpha,\theta|y,x)=-\frac{1}{2}T\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\ln(h_t)-\frac{1}{2}\sum_{t=1}^{T}(u_t^2/h_t)+\sum_{t=1}^{T}\ln\left|\frac{\partial u_t}{\partial y_t}\right|$$

(5.5.48)

使该函数达最大值，就可得到参数β,α,θ的估计。

3. ARCH-M(q）模型

ARCH-M(q）模型为：

$$u_t=f(y_t,x_t,\beta,h_i)=y_i-x_i\beta-\gamma h_i$$

$$u_t|\psi_{t-1}\sim N(0,h_t)$$

$$h_t=\alpha_0+\sum_{i=1}^{q}\alpha_i u_{t-i}^2=\alpha_0+\alpha(B)u_t^2 \tag{5.5.34}$$

其中$q>0$;$\alpha_0>0,\alpha_i\geq 0,i=1,\cdots,q$,$\alpha(1)<1$。

对数似然函数为：

$$\ln L(\beta,\gamma,\alpha|y,x)=-\frac{1}{2}T\ln(2\pi)-\frac{1}{2}\sum_{t=1}^{T}\ln(h_t)-\frac{1}{2}\sum_{t=1}^{T}(u_t^2/h_t)$$

4. GARCH-M(p,q) 模型

GARCH-M(p,q) 模型为

$$u_t=f(y_t,x_t,\beta,h_i)=y_i-x_i\beta-\gamma h_i$$

$$u_t|\psi_{t-1}\sim N(0,h_t)$$

$$h_t=\alpha_0+\sum_{i=1}^{q}\alpha_i u_{t-i}^2+\sum_{i=1}^{p}\theta_i h_{t-i}=\alpha_0+\alpha(B)u_t^2+\theta(B)h_t \tag{5.5.35}$$

其中$p\geq 0,q>0$;$\alpha_0>0,\alpha_i\geq 0,i=1,\cdots,q$;$\theta_i\geq 0,i=1,\cdots,p$,$\alpha(1)+\theta(1)<1$ 。

对数似然函数为：

$$\ln L(\beta,\gamma,\alpha,\theta|y,x) = -\frac{1}{2}T\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln(h_t) - \frac{1}{2}\sum_{t=1}^{T}(u_t^2/h_t)$$

5. 拉格朗日乘子检验

可以采取由Engle(1982)提出的拉格朗日乘子检验(Lagrange multiplier test)对ε_t是否有ARCH或GARCH现象进行检验。首先，对ε_t平方，得到ε_t^2。然后，对ε_t^2进行AR(q)自回归估计得到拟合优度R^2。最后,利用结果：在不存在ARCH或GARCH的原假设下，统计量nR^2服从于自由度为q的χ^2分布。若显著性水平为5%，当nR^2值大于χ^2分布的临界值时，则拒绝ε_t不存在ARCH或GARCH的假设,即认为存在ARCH或GARCH现象。

例5.5.4 我国通货膨胀的GARCH(1,1)模型及其估计。通货膨胀的变量采用100*ln(与前年同期比较的居民消费价格指数),本文采用1993年4月到1998年11月68个月的月度资料，有关居民消费价格指数资料来自《中国物价》。

首先，用普通最小二乘法得到通货膨胀自回归的最佳模型如下：

$$\pi_t = \beta_1\pi_{t-1} + \beta_2\pi_{t-3} + \varepsilon_t \tag{5.5.36}$$

然后，对ε_t是否有GARCH 现象进行拉格朗日乘子检验，计算出NR^2＝4.45221 (R^2为ε_t^2进行AR(q)自回归估计的拟合优度) ,大于χ^2分布的临界值3.84(在显著性水平为5%下)，于是拒绝ε_t不存在GARCH的假设。

最后，对ε_t建立GARCH(1,1)模型，回归模型(5.8.34) 的普通最小二乘法估计和GARCH模型的最大似然估计的计算结果如表5.5.10，参数估计值的下方为其t统计量。

表5.5.10 估计结果

	β_1	β_2	α_0	α_1	θ_1
回归模型(最小二乘法)	1.179823 (17.48936)	−0.194254 (−2.94192)	0.88937	0	0
GARCH模型(最大似然估计)	1.1893 (18.576)	−0.2095 (−3.4233)	0.36193 (2.7923)	0.296 (2.8598)	0.5337 (4.8159)

可见α_1和θ_1都显著不为0, 且$\alpha_1+\theta_1$=0.8297¡1，因而GARCH(1,1)模型较好地拟合了通货膨胀序列。图5.5.1和图5.5.2分别为在95%显著性水平下，回归模型和GARCH(1,1)模型对通货膨胀的预测置信区间。从图中可看出，通货膨胀回归模型的预测置信区间的长度为常量，而GARCH模型的预测置信区间长度是变化的。93－94年通货膨胀GARCH模型的预测置信区间大于回归模型,96－98年通货膨胀GARCH模型的预测置信区间小于

回归模型。这正好与93－94年波动幅度较大的高通货膨胀且自95年底我国经济实现了“软着陆”后，96－98年波动幅度较小的低通货膨胀的情况相符合。

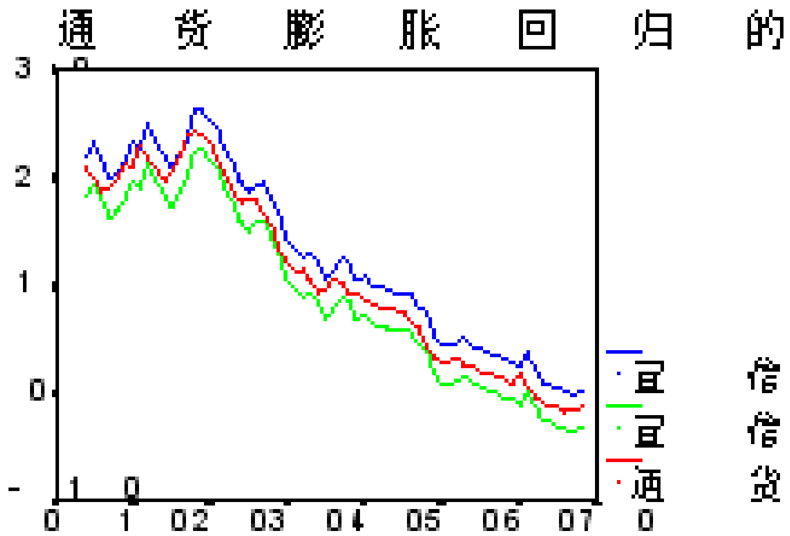


图 5.1: 回归模型的预测置信区间

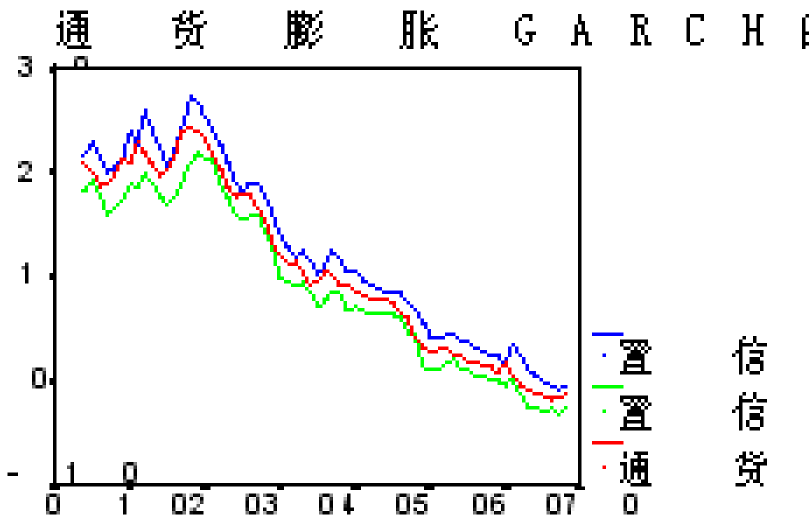


图 5.2: GARCH(1,1)的预测置信区间

5.6 广义指数分布模型

利用广义指数分布建立的非线性模型非常灵活、有力，尤其能对有较大离散跳跃的内生经济变量建立较理想的模型，而这类变量不适合采用哑变量对线性模型修正的方法来建立模型，因为这需要精确确定经济变量何时发生跳跃。在证卷市场或外汇市场，股价或汇率的跳跃并不伴随影响因素的较大变化。即相对大的被解释变量的变化与相对小的解释变量的变化相伴随。广义指数分布非线性模型为研究这类现象的规律性提供了强有力的工具。

该方法的一个非常重要特征是它提供了简单经济模型和经济变量分布形式的联系。该分布形式可直接从模型的结构形式得到而不是建立在估计阶段所作的分布形式的主观假设基础之上。该方法的灵活性在于它包含了很广泛一类的模型，因而可在这个更大的模型范围内比较对同一经济现象建立的不同模型的优劣。

5.6.1 广义指数分布模型

广义指数分布模型的一般形式为：

$$y_t = \int z_t f(z_t;\beta)dz_t + \varepsilon_t \tag{5.6.1}$$

其中ε_t为随机干扰项，$f(\cdot)$是在给定解释变量$x_t = (x_{1t},\cdots,x_{lt})$下，被解释变量$y_t$的条件分布函数，假定它具有如下形式：

$$f(y_t;\beta) = \exp(\theta_{1,t}y_t + \theta_{2,t}y_t^2/2 + \cdots + \theta_{k,t}y_t^k/k - \eta_t) \quad t = 1,\cdots,T$$

(5.6.2)

$$\eta_t = \ln\int \exp(\theta_{1,t}y_t + \theta_{2,t}y_t^2/2 + \cdots + \theta_{k,t}y_t^k/k)dy_t \tag{5.6.2}$$

其中$\theta_{j,t}$为解释变量的函数，可以是线性的，也可以是非线性的。线性的表达式为

$$\theta_{j,t} = \beta_{j,0} + \beta_{j,1}x_{1t} + \cdots + \beta_{j,l}x_{lt} = X_t\beta_j \tag{5.6.3}$$

其中$X_t = \begin{pmatrix} 1, & x_{1t}, & \cdots, & x_{lt} \end{pmatrix}$，$t = 1,\cdots,T$为常数项和解释变量构成的$1\times(l+1)$向量，$\beta_j = \begin{pmatrix} \beta_{j,0} & \beta_{j,1} & \cdots & \beta_{j,l} \end{pmatrix}'$，$j = 1,\cdots,k$为$(l+1)\times 1$参数向量。

下面将看到线性和非线性模型都是广义指数分布模型的特例。

对于线性模型

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_l x_{lt} + u_t \tag{5.6.4}$$

假设随机干扰项u_t是独立同分布且$u_t \sim N(0,\sigma^2)$，则在给定解释变量下，y_t的条件分布为：

$$\begin{aligned} f(y_t;\beta) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_t - \beta_0 - \beta_1 x_{1t} - \cdots \beta_l x_{lt})^2\right] \\ &= \exp\left(\frac{\beta_0 + \beta_1 x_{1t} + \cdots + \beta_l x_{lt}}{\sigma^2} y_t - \frac{1}{\sigma^2} y_t^2/2 - \eta_t\right) \end{aligned} \tag{5.6.5}$$

其中

$$\eta_t = \ln(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2}(\beta_0 + \beta_1 x_{1t} + \cdots \beta_l x_{lt})^2 \tag{5.6.6}$$

可见，线性模型是广义指数分布模型的特例。

对于非线性模型

$$y_t = h(x_t, \beta) + u_t \tag{5.6.7}$$

其中y_t为被解释变量，x_t为解释变量向量，β为参数向量，u_t为随机变量。假设随机干扰项u_t是独立同分布且$u_t \sim N(0,\sigma^2)$，则在给定解释变量下，y_t的条件分布为：

$$\begin{aligned} f(y_t;\beta) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}[y_t - h(x_t,\beta)]^2\right] \\ &= \exp\left(\frac{h(x_t,\beta)}{\sigma^2} y_t - \frac{1}{\sigma^2} y_t^2/2 - \eta_t\right) \end{aligned} \tag{5.6.8}$$

其中

$$\eta_t = \ln(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2} h^2(x_t,\beta) \tag{5.6.9}$$

可见，非线性模型是广义指数分布模型的特例。

传统的线性和非线性模型通常假定被解释变量在解释变量给定下的条件分布都是单峰的，即使假设条件分布不是正态分布时也如此，且具有相同的函数形式。而广义指数分布模型允许被解释变量的条件分布随着时间的变化而不同，有时是单峰的，有时是多峰的。条件分布是单峰的被解释变量的数值一般取值于山峰的附近，所以所取的值不会出现大的跳跃性的变化。条件分布是多峰的被解释变量的数值一般在各个山峰附近取值，所以所取的值会出现较大的跳跃，可从这个山峰跳到另一个山峰。此时，即使是解释变量没有什么变化，被解释变量的条件分布也没有什么变化，但

由于被解释变量的条件分布是多峰的，被解释变量的数值却可以出现较大的跳跃。

5.6.2 广义指数分布模型多峰的识别

非常重要的一点是，广义指数分布非线性模型允许条件分布的形式随时间变化，因而有时的条件分布可能是多峰的。识别广义指数分布是否在任何时期都为多峰的，是十分重要的。这个信息的价值在于，多峰可能与被解释变量的较大离散跳跃相联系，意味着从一个峰到另一个峰运动，这样的跳跃并不伴随解释变量的较大变化。

对模型进行变换有利于识别多峰。考虑(5.6.2)($k=4$)，可利用

$$w_t = y_t(-\theta_{4,t})^{0.25} + \theta_{3,t}(-\theta_{4,t})^{-0.75}/3 \tag{5.6.10}$$

经过繁琐的代数运算，有

$$g(w_t;\beta) = \exp(\tau_{0,t}w_t + \tau_{1,t}w_t^2/2 - w_t^4/4 - \eta_t^*) \tag{5.6.11}$$

其中

$$\eta_t^* = \ln\int \exp(\tau_{0,t}w_t + \tau_{1,t}w_t^2/2 - w_t^4/4)dw_t \tag{5.6.12}$$

其中$\tau_{0,t}$和$\tau_{1,t}$是θ的函数。注意到$w_t^3/3$的系数为0，$w_t^4/4$的系数为1。

变换(6.4.11)的重要性在于，变换后的分布(6.4.12)和变换前的分布(5.6.2)有相同的形状特征。特别地，如果(5.6.2)的分布是双峰的，则(6.4.12)的分布也是双峰的。消除立方项的好处是容易得到检验分布双峰性的条件。分布(6.4.12)多峰的充分必要条件是卡当(Cardan)判别式

$$\delta_t = \frac{\tau_{0,t}^2}{4} - \frac{\tau_{1,t}^3}{27} \tag{5.6.13}$$

为负。可推出分布(6.4.12)多峰的必要条件是$\tau_{1,t}$为正。当判别式从正值变为负值时，意味着分布(6.4.12)从单峰变为多峰。

5.6.3 广义指数分布非线性模型的估计和预测

采用最大似然方法进行估计，对数似然函数为

$$\ln L(\beta) = \sum_{t=1}^{T} \ln f(y_t;\beta) \tag{5.6.14}$$

参数β的最大似然估计为使对数似然函数到达最大的参数值，即

$$\hat{\beta} = \arg\max_{\beta} \ln L(\beta) \tag{5.6.15}$$

它的协方差阵为：
$\mathrm{COV}(\hat{\beta}) = \left[-E\left(\frac{\partial^2 \ln L}{\partial\beta'\partial\beta}\right)\right]^{-1}$ (5.6.17)
被解释变量的拟合值或预测值为：

$$\hat{y}_t = \int z_t f(z_t;\hat{\beta}) dz_t \tag{5.6.16}$$

对于线性模型，被解释变量的拟合值或预测值为：

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \cdots + \hat{\beta}_l x_{lt} \tag{5.6.17}$$

对于非线性模型，被解释变量的拟合值或预测值为：

$$\hat{y}_t = h(x_t, \hat{\beta}) \tag{5.6.18}$$

5.6.4 广义指数分布非线性模型的应用

以被解释变量价格p_t和影响它的两个解释变量x_t和y_t的关系为例。变量的数据序列如图(6.5.1)。

图5.6.1 p_t、x_t和y_t的数据序列

首先，建立线性模型

$$p_t = \alpha_0 + \alpha_1 x_t + \alpha_2 y_t + u_t \tag{5.6.19}$$

其中u_t为随机干扰项，假设它是独立同分布且$u_t \sim N(0, \sigma_u^2)$。用普通最小二乘法进行估计的结果如下：

$$\hat{p}_t = -\underset{(5.068)}{1.746} + \underset{(59.166)}{0.185}\, x_t - \underset{(0.663)}{0.020}\, y_t \tag{5.6.20}$$

且$\hat{\sigma}_u = 0.462$；$\bar{R}^2 = 0.947$；$DW = 1.327$；$BP(2) = 69.142$；$ARCH(1) = 45.018$。

其中估计的参数值下括号里的是t统计量值。$BP(2)$为检验异方差的Breusch-Pagan统计量,服从于自由度为2的χ^2分布（参见Greene(1997)）。$ARCH(1)$为检验一阶自回归条件异方差的统计量，服从于自由度为1的χ^2分布(参见§5.5)。从传统的拟合优度的标准来看，线性模型拟合得很好。$\bar{R}^2 = 0.947$表明被解释变量的94.7%的变差由解释变量所解释。图(5.6.2)为被解释变量的实际值和拟合值的对比图，从图可知，在绝大多数的时间，对被解释变量的一般数值波动能被很好地拟合。

图5.6.2 p_t实际值和OLS拟合值对比图

然而，仔细考察图(6.5.1)，就会发现线性模型不能捕捉在样本数据开头、中间和尾巴的较大幅度的波动。线性模型的这个识别的错误也可从$DW = 1.327$和$BP(2) = 69.142$反映出来，分别说明了线性模型的随机干扰项存在严重的序列相关和异方差。这些结果表明变量之间有些基本的特征是无法由线性模型来反映。$ARCH(1) = 45.018$也表明在线性模型的随机干扰项中有明显的非线性结构。

检验随机干扰项u_t是正态随机变量的方法是构造普通最小二乘的残差直方图(图5.6.3)，从直方图看出没有理由拒绝u_t的分布是单峰的假设。检验正态性更正式的方法是Jarque-Bera检验（参见Greene(1997)），计算出该检验统计量的数值是114.071，与自由度为2的χ^2分布的临界值表比较可知在显著性水平1%下，应该拒绝随机干扰项u_t为正态分布的假设。

解决序列相关、异方差和ARCH的标准方法是广义最小二乘法GLS。但在被解释变量的条件分布是单峰的假设下，GLS估计是有偏估计。事实上，由Jarque-Bera检验拒绝u_t的正态性暗示着应考虑分布是多峰的情形。

图5.6.3 残差直方图

对被解释变量p_t建立广义指数分布模型并利用最大似然估计，结果如下：

$$f(p_t;\hat{\beta}) = \exp(\hat{\theta}_{1,t}p_t + \hat{\theta}_{2,t}p_t^2/2 + \hat{\theta}_{3,t}p_t^3/3 + \hat{\theta}_{4,t}p_t^4/4 - \eta_t)$$

$$\hat{\theta}_{1,t} = \underset{(3.333)}{3.437}$$

$$\hat{\theta}_{2,t} = \underset{(2.160)}{-5.663} - \underset{(0.803)}{0.174}\,y_t + \underset{(1.805)}{0.220}\,x_t$$

$$\hat{\theta}_{3,t} = \underset{(4.945)}{3.805} + \underset{(6.684)}{0.304}\,x_t - \underset{(1.073)}{0.043}\,y_t$$

$$\hat{\theta}_{4,t} = -\underset{(7.343)}{2.261} \tag{5.6.21}$$

且$\hat{\sigma}_u = 0.328$；$\bar{R}^2 = 0.973$；$DW = 2.051$；$BP(2) = 24.844$；$ARCH(1) = 0.946$。可见拟合优度已有改进，从线性模型的94.7%上升到非线性模型的97.3%。检查t统计量，发现在$\theta_{3,t}$中y_t的系数不显著，在$\theta_{2,t}$中x_t和y_t的系数都不显著。

面对这样的结果，消去$\theta_{3,t}$中y_t后，重新进行估计，结果如下：

$$\hat{\theta}_{1,t} = \underset{(3.192)}{3.374}$$

$$\hat{\theta}_{2,t} = \underset{(1.861)}{-4.013} - \underset{(3.812)}{0.367}\,y_t + \underset{(1.357)}{0.186}\,x_t$$

$$\hat{\theta}_{3,t} = \underset{(4.307)}{3.260} + \underset{(6.083)}{0.294}\, x_t$$

$$\hat{\theta}_{4,t} = \underset{(6.746)}{-2.203}$$

这时，$\theta_{2,t}$中y_t的系数变得显著，但x_t的系数仍不显著。
消去$\theta_{2,t}$中x_t后，再进行估计，结果如下：

$$\hat{\theta}_{1,t} = \underset{(3.260)}{2.819}$$

$$\hat{\theta}_{2,t} = \underset{(3.622)}{-5.563} - \underset{(3.878)}{0.380}\, y_t$$

$$\hat{\theta}_{3,t} = \underset{(5.251)}{2.633} + \underset{(7.940)}{0.254}\, x_t$$

$$\hat{\theta}_{4,t} = \underset{(8.117)}{-1.982}$$

且$\hat{\sigma}_u = 0.379$；$\bar{R}^2 = 0.965$；$DW = 2.022$；$BP(2) = 26.572$；$ARCH(1) = 0.657$。图(6.4.4)是该模型的被解释变量的拟合值与实际值的比较图。

图5.6.4 p_tML拟合值与实际值的比较图

可见：拟合优度96.5%比线性模型的94.7%仍有改进；所有系数都显著不为0；$DW = 2.022$说明没有序列相关的现象；$ARCH(1) = 0.657$表明广义指数分布模型的随机干扰项不存在非线性结构；$BP(2) = 26.572$说明仍有一定程度的异方差，但比线性模型的异方差程度有明显的下降；比较图(6.4.4)和(5.6.2)，知道广义指数分布模型在时间的开头、中间和尾巴也能拟合得较好。

更重要的，由卡当(Cardan)判别式可得：在时间的开头、中间和尾巴，价格的条件分布为多峰的，在其它时间为单峰的，如图(6.4.5)所示。从图(6.5.1)可知，价格在时间的开头、中间和尾巴与其它时段相比有明显的跳跃性的变化，而解释变量x_t和y_t都没有明显的跳跃性的变化。因而采用广义指数分布模型就能较好地拟合它们之间的关系。

图5.6.5 价格条件分布的卡当识别图

5.7 非线性联立方程模型

在经典计量经济学模型的联立方程模型中，关于联立方程模型的一些基本概念已经为读者所熟悉。而且，关于变量非线性的模型，在宏观经济计量模型系统中是经常出现的。本节对这些问题不再介绍，只是简单讨论一般的非线性问题。

5.7.1 非线性方程组

现在考虑如下非线性方程组：

$$y_{it} = g_i(X_t, \beta) + u_{it}, \quad i = 1, \cdots, m, \quad t = 1, \cdots, T \tag{5.7.1}$$

其中y_{it}，$i = 1, \cdots, m$为被解释变量,$X_t = (x_{1t}, \cdots, x_{kt})$为$1 \times k$解释变量向量,$\beta = (\beta_1, \cdots, \beta_l)'$为$l \times 1$参数向量,$g_i(\cdot)$，$i = 1, \cdots, m$为解释变量和参数的非线性函数。如果应用内生变量与外生变量的概念，该方程组的每个方程都没有内生解释变量。

假定随机干扰项的均值为零，同一方程不同时间的随机干扰项不相关，不同方程的随机干扰项相关，协方差阵为$\Sigma = (\sigma_{ij})_{m \times m}$，即

$$Eu_{it} = 0 i = 1, \cdots, m, \quad t = 1, \cdots, T$$

$$cov(u_{it}, u_{js}) = \sigma_{ij} i, j = 1, \cdots, m t, s = 1, \cdots, T$$

$$\Sigma^{-1} = \left(\sigma^{ij}\right)_{m \times m}$$

将(6.5.1)写成矩阵形式为：

$$Y_i = G_i(X, \beta) + U_i i = 1, \cdots, m \tag{5.7.2}$$

其中

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T \times 1} \qquad G_i(X, \beta) = \begin{pmatrix} g_i(X_1, \beta) \\ g_i(X_2, \beta) \\ \vdots \\ g_i(X_T, \beta) \end{pmatrix}_{T \times 1}$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{pmatrix}_{T \times k} \qquad U_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}_{T \times 1}$$

还可将(6.4.2)写成：

$$Y = G(X, \beta) + U \tag{5.7.3}$$

$$Y=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_m\end{pmatrix}_{mT\times 1}\qquad G(X,\beta)=\begin{pmatrix}G_1(X,\beta)\\G_2(X,\beta)\\\vdots\\U_m(X,\beta)\end{pmatrix}_{mT\times 1}\qquad U=\begin{pmatrix}U_1\\U_2\\\vdots\\U_m\end{pmatrix}_{mT\times 1}$$

1. GLS估计

当Σ已知时，β的GLS估计为极小化

$$U(\beta)'\Omega^{-1}U(\beta)=\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma^{ij}[Y_i-G_i(X,\beta)]'[Y_j-G_j(X,\beta)] \tag{5.7.4}$$

的解向量。其中$\Omega=\Sigma\otimes I_T$。

当Σ未知时，需要Σ的一个一致估计。为此，取ΩE_{mT},极小化(6.4.5)，即极小化

$$U(\beta)'U(\beta)=\sum_{i=1}^{m}[Y_i-G_i(X,\beta)]'[Y_i-G_i(X,\beta)]$$

$=\sum_{i=1}^{m}\sum_{t=1}^{T}[y_{it}-g_i(X_t,\beta)]^2$ (5.7.5)

得到β的一致估计$\tilde{\beta}$。然后。代入(6.4.2)得到σ_{ij}的一致估计：

$$s_{ij}=\frac{1}{T}U_i'(\tilde{\beta})U_j(\tilde{\beta}) \tag{5.7.5}$$

$$S^{-1}=\left(s^{ij}\right)_{m\times m}$$

将(6.4.5)中σ^{ij}用s^{ij}代替，就可得到参数β的可行广义最小二乘（FGLS）估计$\hat{\beta}$，它的渐近协方差阵为：

$$Est.Asy.Var[\hat{\beta}]=[\sum_{i=1}^{m}\sum_{j=1}^{m}s^{ij}X_i^0(\beta)'X_j^0(\beta)]^{-1} \tag{5.7.6}$$

其中

$$X_i^0(\beta)=\frac{\partial G_i(X,\beta)}{\partial\beta'}=\begin{pmatrix}\frac{\partial G_i(X,\beta)}{\partial\beta_1} & \cdots & \frac{\partial G_i(X,\beta)}{\partial\beta_l}\end{pmatrix} \tag{5.7.7}$$

2. ML估计

假定随机干扰项为正态分布，则Y的对数似然函数为

$$\ln L = -\frac{mT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m}u_{ti}u_{tj}\sigma^{ij} \tag{5.7.8}$$

易证对数似然函数为（Greene(1997)）：

$$\ln L = -\frac{T}{2}[M\ln(2\pi) + \ln|\Sigma| + tr(\Sigma^{-1}W)] \tag{5.7.9}$$

其中$W = (W_{ij})_{m\times m}$，$W_{ij} = \frac{1}{T}U_i'(\beta)U_j(\beta)$。所以

$$\frac{\partial \ln L}{\partial \Sigma} = -\frac{T}{2}\Sigma^{-1}(\Sigma - W)\Sigma^{-1}$$

于是，中心化对数似然函数为：

$$\ln L_c = -\frac{T}{2}[m(1+\ln(2\pi)) + \ln|W|] \tag{5.7.10}$$

因而，参数β的最大似然估计为：

$$\hat{\beta}_{ML} = \min_{\beta}\ln|W| \tag{5.7.11}$$

另一种得到最大似然估计的途径是迭代的FGLS：第一步，首先得到参数β的一致估计；直到第$l+1$步，利用参数β的第l步估计β_l，得到Σ的估计$\Sigma(\beta_l)$，再进行参数β的第$l+1$步GLS估计β_{l+1}；如此下去，直到β和Σ的估计都收敛为止。这样得到的参数估计渐近等价于最大似然估计。

3. GMM估计

假设解释变量与随机干扰项相关，此时，非线性FGLS和MLE估计不是一致估计。假设有工具变量向量$z_t = (z_{1t}, \cdots, z_{pt})$,$p \geq l$,使得

$$E[z_t u_{it}] = 0 \quad t = 1, \cdots, T \quad i = 1, \cdots, m \tag{5.7.12}$$

对应的样本矩为：

$$\frac{1}{T}\sum_{t=1}^{T} z_t[y_{it} - g_i(X_t, \beta)] = 0, i = 1, \cdots, m \tag{5.7.13}$$

让$W_{ij} = \frac{1}{T^2}E[Z'U_iU_j'Z]$

其中

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{pmatrix} = \begin{pmatrix} z_{11} & z_{21} & \cdots & z_{p1} \\ z_{12} & z_{22} & \cdots & z_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{1T} & z_{2T} & \cdots & z_{pT} \end{pmatrix}_{T\times p}$$

则GMM估计为最小化

$$q=\sum_{i=1}^{m}\sum_{j=1}^{m}[U_i(\beta)'Z]W^{ij}[Z'U_j(\beta)] \tag{5.7.14}$$

其中W^{ij}为第i行块第j列块矩阵为W_{ij}的矩阵(W_{ij})的逆矩阵$(W_{ij})^{-1}$的第i行块第j列块矩阵。

为了得到W^{ij}的一致估计，首先，最小化

$$\sum_{i=1}^{m}\sum_{j=1}^{m}[U_i(\beta)'Z](Z'Z)^{-1}[Z'U_j(\beta)]$$

得到参数β的一个一致估计$\tilde{\beta}$。其次，根据同一方程不同时间的随机干扰项不相关，不同方程的随机干扰项相关$cov(u_{it},u_{js})=\sigma_{ij}$的假定，得到$W_{ij}$的一致估计为：

$$\tilde{W}_{ij}=\frac{1}{T^2}Z'U_i(\tilde{\beta})U_j'(\tilde{\beta})Z$$

记$\tilde{W}^{ij}$为第i行块第j列块矩阵为$\tilde{W}_{ij}$的矩阵$(\tilde{W}_{ij})$的逆矩阵$(\tilde{W}_{ij})^{-1}$的第i行块第j列块矩阵。则$\tilde{W}^{ij}$是W^{ij}的一致估计。将(6.4.15)中W^{ij}用$\tilde{W}^{ij}$代替，最小化(6.4.15)的一阶条件为：

$$\frac{\partial q}{\partial\beta}=\sum_{i=1}^{m}\sum_{j=1}^{m}(X_i^0(\beta)'Z)\tilde{W}^{ij}(Z'U_j(\beta))=0$$

参数的GMM估计的渐近协方差阵为：

$$V_{GMM}=\left[\sum_{i=1}^{m}\sum_{j=1}^{m}(X_i^0(\beta)'Z)\tilde{W}^{ij}(Z'X_j^0(\beta))\right]^{-1} \tag{5.7.15}$$

5.7.2 非线性联立方程模型

非线性联立方程模型的一般表达式为：

$$\begin{cases} f_1(Y_t,X_t,\beta_1)=u_{1t} \\ f_2(Y_t,X_t,\beta_2)=u_{2t} \\ \vdots \\ f_m(Y_t,X_t,\beta_m)=u_{mt} \end{cases},t=1,\cdots,T \tag{5.7.16}$$

其中$Y_t=(y_{1t},y_{2t},\cdots,y_{mt})$是m个内生变量第$t$期的数据向量；$X_t=(x_{1t},x_{2t},\cdots,x_{kt})$是k个先决变量（包括外生变量和滞后内生变量）第$t$期的数据向量；$\beta_i,u_{it}$分别是第$i$个方程的参数和随机扰动项。

非线性二阶段最小二乘法

这是一种单方程估计方法。可将(6.4.17)写成向量形式：

$$\begin{cases} F_1(Y,X,\beta_1)=U_1 \\ F_2(Y,X,\beta_2)=U_2 \\ \vdots \\ F_m(Y,X,\beta_m)=U_m \end{cases} \tag{5.7.17}$$

其中

$$F_i(Y,X,\beta_i)=\begin{pmatrix} f_i(Y_1,X_1,\beta_i) \\ f_i(Y_2,X_2,\beta_i) \\ \vdots \\ f_i(Y_T,X_T,\beta_i) \end{pmatrix} \quad U_i=\begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}, (i=1,\cdots,m)$$

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}=\begin{pmatrix} y_{11} & y_{21} & \cdots & y_{m1} \\ y_{12} & y_{22} & \cdots & y_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ y_{1T} & y_{2T} & \cdots & y_{mT} \end{pmatrix}_{T\times m} \quad X=\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix}=\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{pmatrix}_{T\times k}$$

为了估计第i个方程

$$F_i(Y,X,\beta_i)=U_i \tag{5.7.18}$$

的参数，由Amemiya提出的非线性2SLS法是极小化：

$$U_i'X(X'X)^{-1}X'U_i \tag{5.7.19}$$

在正规的条件下(邹至庄,1988)，所得的估计量是一致估计且渐近正态的。

非线性三阶段最小二乘法

这是一种系统估计方法。可以(6.4.18)表示成：

$$F(Y,X,\beta)=U \tag{5.7.20}$$

其中

$$F(Y,X,\beta)=\begin{pmatrix} F_1(Y,X,\beta_1) \\ F_2(Y,X,\beta_2) \\ \vdots \\ F_m(Y,X,\beta_m) \end{pmatrix}_{mT\times 1} \quad U=\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{pmatrix}_{mT\times 1}$$

为了估计模型(6.4.21)的参数，由Jorgenson Laffont(1974)提出的非线性3SLS法是极小化：

$$F'[S^{-1} \otimes X(X'X)^{-1}X']F \tag{5.7.21}$$

其中S是Σ的一致估计。非线性3SLS估计是一致估计且渐近正态的(邹至庄,1988)。

充分信息极大似然法（FIML）

将(6.4.17)表示成：

$$G_t(Y,X,\beta) = U_t \quad , \quad t = 1,\cdots,T \tag{5.7.22}$$

其中

$$G_t(Y_t,X_t,\beta) = \begin{pmatrix} f_1(Y_t,X_t,\beta_1) & f_2(Y_t,X_t,\beta_2) & \cdots & f_m(Y_t,X_t,\beta_m) \end{pmatrix}$$

$$U_t = [u_{1t},u_{2t},\cdots,u_{mt}]$$

还可将(6.2.22)表示成：

$$G(Y,X,\beta) = U \tag{5.7.23}$$

其中

$$G(Y,X,\beta) = \begin{pmatrix} G_1(Y_1,X_1,\beta) \\ G_2(Y_2,X_2,\beta) \\ \vdots \\ G_T(Y_T,X_T,\beta) \end{pmatrix}_{T\times m} \qquad U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_T \end{pmatrix} = \begin{pmatrix} u_{11} & u_{21} & \cdots & u_{m1} \\ u_{12} & u_{22} & \cdots & u_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ u_{1T} & u_{2T} & \cdots & u_{mT} \end{pmatrix}_{T\times m}$$

假定同一方程不同时间的随机扰动项不相关且同方差，不同方程的随机扰动项相关$cov(u_{it},u_{jt}) = \sigma_{ij}$，协方差阵为$\Sigma = (\sigma_{ij})_{m\times m}$，则$U_t = [u_{1t},u_{2t},\cdots,u_{mt}]$的联合概率密度为：

$$(2\pi)^{-m/2}|\det(\Sigma)|^{-1/2}\exp(-\frac{1}{2}U_t\Sigma^{-1}U_t')$$

于是，$Y_t = [y_{1t},y_{2t},\cdots,y_{mt}]$的联合概率密度为：

$$(2\pi)^{-m/2}|\det(J_t)||\det(\Sigma)|^{-1/2}\exp(-\frac{1}{2}U_t\Sigma^{-1}U_t') \tag{5.7.24}$$

其中$J_t(\beta) = \det(\partial U_t/\partial Y_t)$。所以，对数似然函数为：

$$\begin{aligned}\ln L(\beta,\Sigma|Y,X) = &-\tfrac{mT}{2}\ln(2\pi) + \sum_{t=1}^{T}\ln|\det(J_t)| - \tfrac{T}{2}\ln|\det(\Sigma)| \\ &-\tfrac{1}{2}\sum_{t=1}^{T}U_t\Sigma^{-1}U_t'\end{aligned} \tag{5.7.25}$$

将$\Sigma = U'U/T$代入(6.2.25)，得到中心化对数似然函数：

$$\begin{aligned}\ln L_c(\beta|Y,X) = &-\tfrac{mT}{2}\ln(1+2\pi) + \sum_{t=1}^{T}\ln|\det(J_t)| \\ &-\tfrac{T}{2}\ln|\det(U'U/T)|\end{aligned} \tag{5.7.26}$$

最大化(5.8.26)，得到参数β的最大似然估计$\hat{\beta}$。

Σœq,ﬄO

$$\hat{\Sigma} = U(\hat{\beta})'U(\hat{\beta})/T \tag{5.7.27}$$

5.8 非均衡计量经济学模型

非均衡计量经济学模型在70年代初期由Fair和Jaffee提出。无论是微观经济还是宏观经济，均衡是暂时的、相对的，而非均衡才是长期的、绝对的。非均衡模型较均衡模型更客观地描述了经济运行的现实。非均衡模型的理论研究发展过程，很大程度上体现在参数估计技术的不断提高，单一市场的非均衡模型向多市场的非均衡模型的发展。在应用方面,非均衡模型的应用几乎遍及了从微观到宏观的所有领域。如劳动力市场、生产要素市场、信贷市场，而最多的还是消费品市场。目前，非均衡模型的理论和应用研究，在国际经济学界正处于蓬勃发展时期，大致在两个方向进行。其一是以Gourieroux、Ito等为代表对多市场非均衡模型的研究，以此来分析市场之间相互溢出效应，从而将整个宏观经济问题作为一个非均衡系统予以研究；其二是以Portes、Winter、David、Yeo等为代表的关于中央计划经济的应用研究。

5.8.1 非均衡计量经济学基本模型

1. 基本模型

考虑需求和供给市场的模型且假定市场处于非均衡状态。与均衡模型相比，非均衡模型放松了需求量和供给量都等于实际交易量的假定。换言之，非均衡意味着市场交易是以不能清理市场的价格进行，使得某些购买

者和供给者不能交易愿意购买和愿意供给的量。一个标准的假定是实际的交易量由市场短缺的一方决定(Clower(1965))，这意味着实际交易量等于需求量和供给量的最小者。

Fair和Jaffee在70年代提出了下述联立方程系统：

$$D_t = \alpha_1 P_t + X_{1t}\beta_1 + u_{1t} \tag{5.8.1}$$

$$S_t = \alpha_2 P_t + X_{2t}\beta_2 + u_{2t} \tag{5.8.2}$$

$$Q_t = \min(D_t, S_t) \quad t = 1, \cdots, T \tag{5.8.3}$$

这里，X_1和X_2分别代表决定需求和供给的外生变量向量，D_t和S_t为系统内生的需求和供给变量，P_t为系统内生的价格变量。与通常意义下的联立方程系统不同的是，这里的D_t和S_t是不能直接观察的内生变量，称为潜在变量，能直接观察的是市场的实际交易量Q_t。如果市场的供给和需求均衡，即$D_t = S_t$，那么实际交易量Q_t既代表需求，也代表供给，即$Q_t = D_t = S_t$。否则，如果市场的供需不均衡，即$D_t \neq S_t$，则$Q_t = \min(D_t, S_t)$。

非均衡模型中的方程(6.4.4)，即$Q_t = \min(D_t, S_t)$称为极小化条件，是非线性的，因而整个模型为一个非线性模型。

很容易将基本模型推广为随机方程是非线性的情形：

$D_t = m_1(P_t, X_{1t}, \beta_1) + u_{1t}$(6.5.1)*

$S_t = m_2(P_t, X_{2t}, \beta_2) + u_{2t}$(6.4.2)*

$$Q_t = \min(D_t, S_t) \quad t = 1, \cdots, T \tag{5.8.4}$$

和随机方程是无参数的情形：

$D_t = m_1(P_t, X_{1t}) + u_{1t}$(6.5.1)**

$S_t = m_2(P_t, X_{2t}) + u_{2t}$(6.4.2)**

$$Q_t = \min(D_t, S_t) \quad t = 1, \cdots, T \tag{5.8.5}$$

2. 最大似然估计

假设同一随机方程的随机扰动项在不同时间是独立同分布，不同随机方程的随机扰动项相关。要对模型进行估计，最大的困难在于极小化条件和D_t和S_t是不可直接观察的潜在变量。由于D_t和S_t没有直接的样本观察值，因而就无法比较他们观察值的大小，这就使极小化条件通过D_t和S_t的观察值来转化成为不可能。记D_t¡S_t的概率为：

$$\lambda_t = \Pr ob(D_t < S_t)$$

$$= \Pr ob(u_{1t} - u_{2t} < X_{2t}\beta_2 + \alpha_2 P_t - X_{1t}\beta_1 - \alpha_1 P_t) \tag{5.8.6}$$

以$f(u_1, u_2)$表示u_1和u_2的联合密度函数，则

$$g(D_t, S_t) = Jf(D_t - \alpha_1 P_t - X_{1t}\beta_1, S_t - \alpha_2 P_t - X_{2t}\beta_2)$$

为D_t和的联合密度函数。其中J是由方程(5.7.1)和(5.7.2)所确定的雅可比行列式的绝对值(J=1)。如果观察值Q_t是来自需求曲线上，应有$S_t > D_t = Q_t$的条件密度函数：

$$h(Q_t|Q_t = D_t) = h(Q_t|S_t > D_t)$$

$$= \int_{Q_t}^{+\infty} g(Q_t, S)dS/\lambda_t \tag{5.8.7}$$

同理

$$h(Q_t|Q_t = S_t) = h(Q_t|S_t < D_t)$$

$$= \int_{Q_t}^{+\infty} g(D, Q_t)dD/(1 - \lambda_t) \tag{5.8.8}$$

由概率论和数理统计的知识，有Q的无条件密度函数：

$$h(Q_t) = h(Q_t|Q_t = D_t)\lambda_t + h(Q_t|Q_t = S_t)(1 - \lambda_t)$$

$$= \int_{Q_t}^{+\infty} g(Q_t, S)dS + \int_{Q_t}^{+\infty} g(D, Q_t)dD \tag{5.8.9}$$

于是，Q的样本的似然函数为：

$$L = \prod_t h(Q_t)$$

$$= \prod_t [\int\limits_{Q_t}^{+\infty} f(Q_t - \alpha_1 P_t - X_{1t}\beta_1, S - \alpha_2 P_t - X_{2t}\beta_2)dS$$

$$+ \int\limits_{Q_t}^{+\infty} f(D - \alpha_1 P_t - X_{1t}\beta_1, Q_t - \alpha_2 P_t - X_{2t}\beta_2)dD] \tag{5.8.10}$$

(6.4.11)式包含了模型的所有参数。参数的最大似然估计为使(6.4.11)达最大的参数估计，等价于使对数似然函数达最大，即最大化

$$\ln L=\sum_t \ln h(Q_t) \tag{5.8.11}$$

若假定u_1,u_2是相互独立的随机变量，则

$$f(u_{1t},u_{2t})=f_1(u_{1t})f_2(u_{2t})$$

于是

$$g(D_t,S_t)=g_1(D_t)g_2(S_t)$$

其中$g_1(D_t)=f_1(D_t-\alpha_1P_t-X_{1t}\beta)$和$g_2(S_t)=f_2(S_t-\alpha_2P_t-X_{2t}\beta_2)$。所以

$$h(Q_t)=g_1(Q_t)\int\limits_{Q_t}^{+\infty}g_2(S)dS+g_2(Q_t)\int\limits_{Q_t}^{+\infty}g_1(D)dD \tag{5.8.12}$$

对于随机方程是非线性的模型，若假定u_1,u_2是相互独立的随机变量，同样可得到对数似然函数：

$$\ln L=\sum_{t=1}^{T}\ln[g_1(Q_t)\int_{Q_t}^{\infty}g_2(S)dS+g_2(Q_t)\int_{Q_t}^{\infty}g_1(D)dD] \tag{5.8.13}$$

其中$g_1(D_t)$和$g_2(S_t)$分别是D_t和S_t的密度函数。若$f_1(u_{1t})$和$f_2(u_{2t})$分别是u_{1t}和u_{2t}的密度函数，则

$$g_1(D_t)=f_1[D_t-m_1(P_t,X_{1t},\beta_1)]$$

$$g_2(S_t)=f_2[S_t-m_2(P_t,X_{2t},\beta_2)] \tag{5.8.14}$$

对于随机方程是无参数的模型，很容易将局部回归的思想应用于似然估计。采用局部线性回归，首先，对两个随机方程在第t期局部进行线性近似：

$$D_i=(1,P_i-P_t,X_{1i}-X_{1t})\alpha_t+v_{1i} \tag{5.8.15}$$

$$S_i=(1,P_i-P_t,X_{2i}-X_{2t})\beta_t+v_{2i} \tag{5.8.16}$$

假定v_{1i}、v_{2i}是相互独立均值为零的正态随机变量，则

$$f(v_{1i}, v_{2i}) = f_1(v_{1i}) f_2(v_{2i})$$

于是

$$g(D_i, S_i) = g_1(D_i) g_2(S_i)$$

其中

$$g_1(D_i) = f_1[D_i - (1, P_i - P_t, X_{1i} - X_{1t})\alpha_t]$$

$$g_2(S_i) = f_2[S_i - (1, P_i - P_t, X_{2i} - X_{2t})\beta_t]$$

模型的局部对数似然函数为：

$$\ln L = \sum_{i=1}^{T} \ln[g_1(Q_i) \int_{Q_t}^{\infty} g_2(S)dS + g_2(Q_i) \int_{Q_i}^{\infty} g_1(D)dD] \tag{5.8.17}$$

模型的局部核权对数似然函数定义为：

$$\ln \tilde{L} = \sum_{i=1}^{T} \ln[\tilde{g}_1(Q_i) \int_{Q_t}^{\infty} \tilde{g}_2(S)dS + \tilde{g}_2(Q_i) \int_{Q_i}^{\infty} \tilde{g}_1(D)dD]$$

其中函数$\tilde{g}_1(D)$和$\tilde{g}_2(S)$已用核权函数进行了加权：

$$\tilde{g}_1(D) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{[D - (1, P_i - P_t, X_{1i} - X_{1t})\alpha_t]^2 * kw_{1i}}{2\sigma_1^2}\right) \tag{5.8.18}$$

$$\tilde{g}_2(S) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{[S - (1, P_i - P_t, X_{2i} - X_{2t})\beta_t]^2 * kw_{2i}}{2\sigma_2^2}\right) \tag{5.8.19}$$

其中$kw_{1i} = K_h(P_i - P_t, X_{1i} - X_{1t})$和$kw_{2i} = K_h(P_i - P_t, X_{2i} - X_{2t})$分别是需求方程和供给方程在$t$期的核权函数(参见§3.4)。使得(6.4.18)达最大，就得到局部参数的α_t和β_t的局部核权最大似然估计。

将局部回归的思想应用于非均衡模型特别有意义，因为局部斜率的局部核权最大似然估计反映了影响需求和供给的因素对需求和供给的影响程度，因而由这些局部斜率的变化情况可反映经济运行的非均衡调整过程。

5.8.2 非均衡计量经济总量模型

基本模型如用于宏观总量数据，极小化条件可能需要进行改进。因为所有宏观数据，如部门或行业数据，均应看成是有限个微观市场的加总，极小交易法则在微观市场上成立，但在宏观市场上，有一些微观市场超需求，有些微观市场超供给，而微观市场充分多，可能导致宏观市场的超额需求和超额供给并存，由此使得总量交易方程不能采用极小法则。

1. 总量交易法则

Spencer总量交易法则为总量需求和总量供给的凸组合：

$$Q = \lambda D + (1-\lambda)S \tag{5.8.20}$$

其中λ为待估参数。

Siebrand将交易曲线设定为：

$$Q = [\delta(D)^{-\rho} + (1-\delta)(S)^{-\rho}]^{-\frac{1}{\rho}} \tag{5.8.21}$$

这里δ和$(1-\delta)$分别表示需求和供给的权重，ρ为非线性参数。

如果允许市场同时存在超额需求和超额供给，Burkett(1988)建议的双曲交易法则：

$$Q = \frac{1}{2}(D+S) - \frac{1}{2}\sqrt{(D-S)^2 + 4\gamma^2 DS} \tag{5.8.22}$$

$\gamma\Omega|$œŒX”dø.Vø.”$\Phi\gamma 0\Lambda I\Lambda U\Delta$¨$T$ø.Æø.(离散开关模型)一致。

2. 双曲模型的局部核权最大似然估计

采用局部线性回归，首先，对需求和供给方程在局部进行线性近似：

$$D \approx (1, P-P_t, X_1 - X_{1t})\alpha_t = \tilde{D} \tag{5.8.23}$$

$$S \approx (1, P-P_t, X_2 - X_{2t})\beta_t = \tilde{S} \tag{5.8.24}$$

在估计时，误差项应加进局部线性近似后的双曲线交易函数中。因而有：

$$u_i = Q_i - [\frac{1}{2}(\tilde{D}_i + \tilde{S}_i) - \frac{1}{2}\sqrt{(\tilde{D}_i - \tilde{S}_i)^2 + 4\gamma^2 \tilde{D}_i \tilde{S}_i}] \tag{5.8.25}$$

其中$\tilde{D}_i$和$\tilde{S}_i$分别是(6.2.23)和(6.2.24)的总需求和总供给的近似函数。

假定随机变量u_i服从于正态分布，则双曲模型的局部核权中心化对数似然函数为：

$$\ln \tilde{L}_c = -\frac{T}{2}[1+\ln(2\pi)] - \frac{T}{2}\ln[\frac{1}{T}\sum_{i=1}^{T}(u_i^2 * kw_i)] \tag{5.8.26}$$

其中$kw_i = K_h(P_i - P_t, X_{1i} - X_{1t}, X_{2i} - X_{2t})$是基于需求和供给所有解释变量在局部$t$期的核权函数在第$i$个观察值的取值。最大化(5.8.26)，就得到局部参数α_t和β_t的估计。

5.8.3 试例：我国消费品市场非均衡计量经济模型

1. 模型设定

基于中央计划经济消费品市场的Portes和Winter模型（基本模型的无参数情形），采用双曲交易法则。中国消费品市场需求和供给函数的局部线性近似为：

$$\tilde{D} = A_t^D + \alpha_{1t}(SAV1 - SAV1_t) + \alpha_{2t}(DYD - DYD_t) + \alpha_{3t}(YD1 - YD1_t)$$

$$\tilde{S} = A_t^S + \beta_{1t}(CT - CT_t) + \beta_{2t}(NMPX - NMPX_t) + \beta_{3t}(CZX - CZX_t)$$

(5.8.25)

其中A_t^D和A_t^S分别是第t个观察点D和S的估计。α_t是斜率系数或是D对应于它的解释变量SAV1、DYD和YD1的响应系数。类似地，β_t是斜率系数或是S对应于它的解释变量CT、NMPX和CZX的响应系数。变量表如下：

D=总需求

S=总供给

C=个人消费

SAV1=前一期个人储蓄

DYD＝个人可支配收入的变化量

YD1＝前一期个人可支配收入

CT＝个人消费的二阶指数时间趋势

NMPX=(CT/NMPT)(NMP-NMPT)

NMP=国民收入

NMPT＝NMP的二阶指数时间趋势

CZX＝[(Z/NMP)-(ZT/NMPT)]*NMP

Z=在固定资产、生成成本和公共消费的投资

ZT＝Z的二阶指数时间趋势

在估计时误差项应加进双曲线交易函数中。因而有：

$$u_i = Q_i - [\frac{1}{2}(\tilde{D}_i + \tilde{S}_i) - \frac{1}{2}\sqrt{(\tilde{D}_i - \tilde{S}_i)^2 + 4\gamma^2\tilde{D}_i\tilde{S}_i}] \tag{5.8.27}$$

假定随机变量u_i服从于正态分布。

2. 模型估计

双曲模型的局部核权中心化对数似然函数为：

$$\ln \tilde{L}_c = -\frac{T}{2}[1+\ln(2\pi)] - \frac{T}{2}\ln[\frac{1}{T}\sum_{i=1}^{T}(u_i^2 * kw_i)] \tag{5.8.28}$$

采用K近邻核权函数，邻域的大小由$K = rT$确定，取r=0.5。更高的r值将导致局部回归时使用太多的观察点而使得影响不大的观察点也用于回归。太小的r值将导致太少的观察点用于局部回归而使得计算无法收敛。

最大化(5.8.28)，得到局部参数α_t和β_t的估计。表5.8.1列出了需求和供给函数的斜率系数以及双曲交易函数的磨擦系数。数字是系数在不同期间的平均值，括号里的数字是它们对应的标准误差。

表5.8.1 需求和供给函数的有关系数

	α_1	α_2	α_3	β_1	β_2	β_3	γ
54-57	-0.847 (0.010)	0.701 (0.009)	1.018 (0.001)	1.113 (0.000)	0.736 (0.001)	-0.784 (0.002)	0.000 (0.000
58-66	-0.749 (0.145)	0.742 (0.088)	1.025 (0.017)	1.091 (0.043)	0.690 (0.060)	-0.737 (0.055)	0.002 (0.005
67-77	-0.878 (0.152)	0.614 (0.061)	1.003 (0.035)	0.968 (0.010)	0.876 (0.087)	-0.970 (0.113)	0.000 (0.000
78-79	-0.613 (0.003)	0.603 (0.028)	0.974 (0.000)	0.941 (0.045)	0.813 (0.121)	-0.816 (0.220)	0.000 (0.000
80-84	-0.253 (0.229)	0.567 (0.016)	0.918 (0.033)	0.982 (0.004)	1.025 (0.104)	-0.896 (0.039)	0.000 (0.000
85-87	-0.044 (0.016)	0.484 (0.017)	0.909 (0.004)	0.992 (0.005)	1.185 (0.011)	-1.003 (0.028)	0.000 (0.000
88-91	0.041 (0.006)	0.176 (0.001)	0.828 (0.001)	1.032 (0.001)	0.965 (0.004)	-0.449 (0.007)	0.000 (0.000
54-79	-0.801 (0.152)	0.672 (0.089)	1.009 (0.028)	1.033 (0.075)	0.776 (0.094)	-0.836 (0.120)	-0.001 (0.003
80-91	-0.084 (0.177)	0.403 (0.175)	0.884 (0.046)	1.003 (0.026)	1.049 (0.104)	-0.787 (0.257)	0.000 (0.000
54-91	-0.575 (0.373)	0.587 (0.174)	0.970 (0.068)	1.023 (0.065)	0.862 (0.160)	-0.820 (0.173)	0.000 (0.003

在表5.8.1，γ在所有时期都为零，超额需求和超额供给不会同时存在。系数的值基本上落在理论范围（$-1 < \alpha_1 < -1/3, 0 < \alpha_2 < 1, \alpha_3 = 1, \beta_1 > 0, \beta_2 > 0$和$\beta_3 < 0$）中（Portes和Winter[1980]）。储蓄对需求的响应系数α_1的数值变化表明在改革开放前后储蓄对需求的影响明显不同。这是由于改革开放前收入只有工资，工作的铁饭碗且通货膨胀很低，具有传统观念的父母趋向于为下一代储蓄。然而，自1980年以后，更多的商品可

获得，收入水平的提高，通货膨胀也增高，下一代更富有且在经济上更不依赖于父母。所以，1980年后，储蓄不再是决定消费的重要因素。

基于双曲交易模型，表5.8.2列出了需求、供给和交易量的局部最大似然估计值$\hat{D}$、和$\hat{C}$及其绝对超额需求估计，相对超额需求估计$(\hat{D}-\hat{C})/\hat{C}\%$和相对超额供给估计$(\hat{S}-\hat{C})/\hat{C}\%$。"*"号表明两个标准误差显著的超额需求和超额供给。绝对量都以1952年的人民币亿元为计量单位。表5.8.2中的超额需求和超额供给的类型与中国经济的历史相当吻合。1980年前的结果与Portes和Santorum(1988)对中央计划国家消费品供求所作的研究结果基本一致。从1957到1960年，大跃进运动打乱了各种生产部门的正常生产，扭曲了资源的合理配置。因而，不仅大跃进的目标没有达到，反而造成消费品的短缺。然而，表5.8.2没有反映在60年代的早期严重的自然灾害同时以货币和实物支付的巨大债务所造成的消费者的艰辛。这可能是模型的问题，没有将实物支付债务作为解释变量(缺数据)。在1966年到1976年文化大革命期间，超额需求和供给估计量的波动反映了这段动荡的历史。由于个人可支配收入增长的下滑造成购买力的下降。在70年代后期，"革命"热情下降，生产活动部分得以恢复，在没有额外固定资产投资的情况下，国民收入递增。

表5.8.2 局部最大似然估计值

5.8.4 市场经济条件下的非均衡计量经济学模型

1. 非均衡价格调整模型

市场经济条件下消费品市场有一个基本的特征，即价格对供求的有效调节，反之，供求压力又影响价格的运动。如果消费品市场是一个完备的市场，价格对供求的调节将是充分有效的，供求压力完全决定了价格的运动，市场由价格来结清。始于70年代中期，西方学者才开始将价格调节方程纳入非均衡系统之中。

价格调整方程最初的设定形式为$\Delta P_t = r(D_t - S_t)$，即供给和需求的非均衡压力引起价格的运动。在系统中加上这一调节方程, 有：

$$D_t = \alpha_1 P_t + X_{1t}\beta_1 + u_{1t}$$

$$S_t = \alpha_2 P_t + X_{2t}\beta_2 + u_{2t}$$

$$Q_t = \min(D_t, S_t)$$

$$\Delta P_t = r(D_t - S_t) \tag{5.8.29}$$

系统的内生变量为P_t、Q_t、D_t和S_t，其中D_t和S_t仍然是不能直接观测的潜在变量，系统的外生变量为X_1和X_2。

c	C	$\hat{C}$	$\hat{D}$			$(\hat{D}-\hat{C})/\hat{C}\%$	$(\hat{S}-\hat{C})/$
54	49.13	49.28	49.28	49.68	-0.40	0.00	0.81
55	53.42	52.89	52.89	53.58	-0.69	0.00	1.31
56	56.90	57.10	57.87	57.10	0.78	1.36	0.00
57	59.76	59.76	59.76	60.61	-0.85	0.00	1.42
58	62.11	61.89	63.79	61.89	1.90	3.07*	0.00
59	57.07	57.43	59.74	57.43	2.30	4.01*	0.00
60	53.77	54.23	54.23	60.44	-6.21	0.00	11.45*
61	50.04	51.42	51.42	61.31	-9.89	0.00	19.23*
62	52.69	50.29	50.29	63.72	-13.43	0.00	26.70*
63	59.54	60.29	60.29	66.15	-5.86	0.00	9.71*
64	64.71	65.05	65.05	69.46	-4.41	0.00	6.78*
65	72.39	71.98	72.10	71.98	0.12	0.17	0.00
66	77.47	77.10	79.43	77.10	2.33	3.02*	0.00
67	82.07	82.04	82.54	82.04	0.49	0.60*	0.00
68	81.42	81.44	84.50	81.44	3.05	3.75*	0.00
69	85.89	85.59	85.59	87.19	-1.60	0.00	1.87*
70	91.92	91.85	91.85	93.63	-1.78	0.00	1.94*
71	95.52	95.57	98.38	95.57	2.81	2.94*	0.00
72	100.65	100.62	10.62	101.32	-0.70	0.00	0.70*
73	108.42	108.10	108.10	108.10	0.00	0.00	0.00
74	110.50	110.77	111.72	110.77	0.95	0.86*	0.00
75	114.53	114.67	114.67	115.87	-1.20	0.00	1.05*
76	118.40	118.70	118.70	121.78	-3.08	0.00	2.59*
77	121.26	121.02	121.02	123.92	-2.90	0.00	2.40*
78	129.12	128.85	128.85	128.85	0.00	0.00	0.00
79	139.58	139.57	140.23	139.57	0.66	0.47*	0.00
80	154.72	154.77	154.77	158.14	-3.37	0.00	2.18*
81	167.52	167.52	167.52	172.34	-4.82	0.00	2.88*
82	178.72	178.97	178.97	185.09	-6.12	0.00	3.42*
83	193.74	194.54	194.54	201.20	-6.66	0.00	3.42*
84	217.82	218.82	218.82	221.45	-2.63	0.00	1.20*
85	249.94	248.39	248.39	248.39	0.00	0.00	0.00
86	264.48	266.97	269.12	266.97	2.15	0.81*	0.00
87	284.09	282.61	294.64	282.61	12.03	4.26*	0.00
88	308.49	309.06	309.06	317.54	-8.48	0.00	2.74*
89	310.61	309.74	309.74	326.12	-16.38	0.00	5.29*
90	322.24	322.07	322.07	329.87	-7.80	0.00	2.42*
91	353.58	353.57	376.39	353.57	22.82	6.45*	0.00

因为在需求大于供给时，价格对于供需的调节速度不同于供给大于需求时价格的调节速度。Laffont和Garcia(1977)对调节方程作如下改进：

$$\Delta P_t = \begin{cases} r_1(D_t - S_t) & D_t - S_t > 0 \\ r_2(D_t - S_t) & D_t - S_t < 0 \end{cases} \tag{5.8.30}$$

调节方程(5.8.29)和(5.8.30)的设定是一种较简单的情况，即价格运动受供求影响而没有其它外生因素。

2. 模型估计

(1)最大似然估计

1974年，Amemiya导出了系统(5.8.29)的最大似然估计(简记为MLE)。同年，Fair和Kelejian导出了这个系统的另外一种最大似然估计。下面用Amemiya方法导出了系统(5.8.29)的最大似然估计，对于系统(5.8.30)可类似地进行而得到参数的最大似然估计。

记$Q_t = D_t$时，内生变量Q_t和P_t的联合密度函数为$f_1(Q_t, P_t)$；类似地，记$Q_t = S_t$时，内生变量Q_t和P_t的联合密度函数为$f_2(Q_t, P_t)$。

当$Q_t = D_t$时，有

$$Q_t = \alpha_1 P_t + X_{1t}\beta_1 + u_{1t}$$

$$\Delta P_t = r(D_t - S_t) = r(Q_t - \alpha_1 P_t - X_{2t}\beta_2 - u_{2t})$$

上述两个方程，在给定了u_1,u_2的联合密度函数$g(u_1, u_2)$以后，可以求出

$$f_1(Q, P) = J_1 g(u_1, u_2)$$

这里J_1是由上述两个方程所确定的雅可比行列式的绝对值。同理，当$Q_t = S_t$时，

$$Q_t = \alpha_2 P_t + X_{2t}\beta_2 + u_{2t}$$

$$\Delta P_t = r(D_t - S_t) = r(\alpha_1 P_t + X_{1t}\beta_1 + u_{2t} - Q_t)$$

因此有$f_2(Q, P) = J_2 g(u_1, u_2)$，其中$J_2$是由上述两个方程所确定的雅可比行列式的绝对值。仍记$\lambda_t = \Pr ob(D_t < S_t)$，则$Q_t$、$P_t$在条件$\Delta P_t < 0$的条件密度函数为：

$$f_1(Q_t, P_t | \Delta P < 0) = f_1(Q_t, P_t | D_t < S_t) = f_1(Q_t, P_t)/\lambda_t$$

同理

$$f_2(Q_t, P_t | \Delta P > 0) = f_2(Q_t, P_t | D_t < S_t) = f_2(Q_t, P_t)/1 - \lambda_t$$

类似(6.4.11)式的推导，由Amemiya导出的极大似然函数为：

$$L = \prod_{\Delta P_t<0} f_1(Q_t, P_t) \prod_{\Delta P_t>0} f_2(Q_t, P_t) \tag{5.8.31}$$

该式包括了系统的全部待估参数，通过对该式取对数后迭代法求极值解，可以得到系统的全部待估参数的MLE。但是式(5.8.31)仍基于用ΔP_t的符号来确定Q_t的来源。

(2)两阶段最小二乘估计

下面导出系统(5.8.30)的两阶段最小二乘估计（TSLS）。

由于当$\Delta P_t > 0, D_t > S_t = Q_t$，故

$$Q_t = \alpha_2 P_t + X_{2t}\beta_2 + u_{2t} \tag{5.8.32}$$

而由方程$\Delta P_t = r_1(D_t - S_t)$，有

$$Q_t = D_t - \frac{1}{r_1}\Delta P_t$$

于是便有

$$Q_t = \alpha_1 P_t + X_{1t}\beta_1 - \frac{1}{r_1}\Delta P_t + u_{1t} \tag{5.8.33}$$

同理，当$\Delta P_t < 0, Q_t@_t < S_t$,故

$$Q_t = \alpha_1 P_t + X_{1t}\beta_{12} + u_{1t} \tag{5.8.34}$$

而由方程$\Delta P_t = r_2(D_t - S_t)$，有

$$Q_t = S_t + \frac{1}{r_2}\Delta P_t = \alpha_2 P_t + X_{2t}\beta_2 - \frac{1}{r_2}\Delta P_t + u_{2t} \tag{5.8.35}$$

定义

$$Z_{1t} = \begin{cases} \Delta P & \Delta P > 0 \\ 0 & \Delta P < 0 \end{cases}$$

将(5.8.33)和(5.8.34)二式结合，便有

$$Q_t = \alpha_1 P_t + X_{1t}\beta_1 - \frac{1}{r_1} Z_{1t} + u_{1t} \tag{5.8.36}$$

类似地，定义

$$Z_{2t} = \begin{cases} \Delta P & \Delta P < 0 \\ 0 & \Delta P > 0 \end{cases}$$

将(5.8.32)和(5.8.35)二式结合起来，便有

$$Q_t = \alpha_2 P_t + X_{2t}\beta_2 - \frac{1}{r_2} Z_{2t} + u_{2t} \tag{5.8.37}$$

以上述推导过程不难看出，系统的极小化条件通过ΔP_t的符号而转化，且调节方程也一并转化。这种转化的结果，使系统转化为由式(5.8.36)和(5.8.37)组成的一个联立方程系统，而且每个方程是利用Q_t全部的信息。由于P_t是系统的内生变量，故$E(P_t u_t) \neq 0$，所以$E(Z_{1t}, u_{1t}) \neq 0, i = 1, 2$。

由此可以得到采用TSLS的具体步骤如下：

第一阶段：用全部观察值，分别作P_t, Z_{1t}, Z_{2t}对所有外生变量的OLS，得到P_t, Z_{1t}, Z_{2t}的拟合值$\hat{P}_t, \hat{Z}_{1t}, \hat{Z}_{2t}$，将这些值替代(5.8.36)和(5.8.37)式中的相应部分，这样就得到：

$$Q_t = \alpha_1 \hat{P}_t + X_{1t}\beta_1 - \frac{1}{r_1} \hat{Z}_{1t} + u_{1t} \tag{5.8.38}$$

$$Q_t = \alpha_2 \hat{P}_t + X_{2t}\beta_2 - \frac{1}{r_2} \hat{Z}_{2t} + u_{2t} \tag{5.8.39}$$

第二阶段：对上述两个方程分别作OLS，由此而得到参数估计量是一致估计量。

5.8.5 多市场非均衡计量经济学模型

以T.Ito、C.Gourieroux、A.Monfort,、J.J.Laffont等为先驱、从70年代后期就开始对多市场非均衡模型进行研究。多市场非均衡计量经济模型的一般结构式为(为简记，将价格P和工资W作为解释变量向量X和Z的一个分量)：

$$D_{1t} = X_{1t}\alpha_1 + \mu_1(Q_{2t} - S^*_{2t}) + u_{1t}$$

$$S_{1t} = Z_{1t}\beta_1 + \omega_1(Q_{2t} - D^*_{2t}) + v_{1t}$$

$$Q_{1t} = \min(D_{1t}, S_{1t})$$

$$D_{2t} = X_{2t}\alpha_2 + \mu_2(Q_{1t} - S^*_{1t}) + u_{1t}$$

$$S_{2t} = Z_{2t}\beta_2 + \omega_2(Q_{1t} - D^*_{1t}) + v_{2t}$$

$$Q_{2t} = \min(D_{2t}, S_{2t}) \tag{5.8.40}$$

这里，D_1和S_1分别表示商品市场的需求与供给，Q_1为其交易量，D_2和S_2分别表示劳动市场的需求与供给，Q_2为其交易量，X和Z为解释变量，包括内生变量及其滞后，u_{1t}, u_{2t}, v_{1t}和v_{2t}是均值为零、方差有限的正态随机变量。

$$D_{1t}^* = X_{1t}\alpha_1$$

$$S_{1t}^* = Z_{1t}\beta_1$$

$$D_{2t}^* = X_{2t}\alpha_2$$

$$S_{2t}^* = Z_{2t}\beta_2 \tag{5.8.41}$$

由方程(5.8.40)定义的“有效”需求和供给考虑到了其它市场的约束。方程(5.8.41)代表无约束的需求和供给，并不考虑其它市场的溢出。Portes(1977)提出另外的模型设定，即将S_{2t}^*,D_{2t}^*,S_{1t}^*和D_{1t}^*用S_{2t},D_{2t},S_{1t}和D_{1t}代替使得其它市场的溢出由其有效需求和供给而不是无约束需求和供给来代表。无论那一种设定，当两市场都出清时，需求和供给的行为方程都是与那个特定市场相关的外生因素的简单函数，无约束和有效需求和供给相等，即市场间无溢出。

Ito(1980)建议该模型的价格调整机制为：

$$\Delta P_{1t} = \begin{cases} \lambda_1(D_{1t} - S_{1t}) & D_{1t} > S_{1t} \\ \delta_1(D_{1t} - S_{1t}) & D_{1t} < S_{1t} \end{cases}$$

$$\Delta P_{2t} = \begin{cases} \lambda_2(D_{2t} - S_{2t}) & D_{2t} > S_{2t} \\ \delta_2(D_{2t} - S_{2t}) & D_{2t} < S_{2t} \end{cases} \tag{5.8.42}$$

其中P_{1t}和P_{2t}是两个市场可观察的价格，$\Delta P_{it} = P_{i,t+1} - P_{it}$或$\Delta P_{it} = P_{i,t} - P_{it-1} (i = 1,2)$，使得当前价格是外生或内生。Ito(1980)证明了(5.8.40)和(5.8.42)模型的参数可通过对下列联立方程组应用两阶段最小二乘法而得到一致估计：

$$Q_{1t} = X_{1t}\alpha_1 - \mu_1 Z_{2t}\beta_2 + \mu_1 Q_{2t} - \lambda_1^* d_{1t} + (u_{1t} - \mu_1 v_{2t})$$

$$Q_{1t} = Z_{1t}\beta_1 - \omega_1 X_{2t}\alpha_2 + \omega_1 Q_{2t} + \delta_1^* s_{1t} + (v_{1t} - \omega_1 u_{2t})$$

$$Q_{2t} = X_{2t}\alpha_2 - \mu_2 Z_{1t}\beta_1 + \mu_2 Q_{1t} - \lambda_2^* d_{2t} + (u_{2t} - \mu_2 v_{1t})$$

$$Q_{2t} = Z_{2t}\beta_2 - \omega_2 X_{1t}\alpha_1 + \omega_2 Q_{1t} + \delta_2^* s_{2t} + (v_{2t} - \omega_2 u_{1t}) \tag{5.8.43}$$

其中$\lambda_i^* = 1/\lambda_i$，$\delta_i^* = 1/\delta_i$，且

$$d_{it} = \begin{cases} \Delta P_{it} & \Delta P_{it} > 0 \\ 0 & \end{cases}$$

$$s_{it} = \begin{cases} \Delta P_{it} & \Delta P_{it} < 0 \\ 0 & \end{cases} \qquad i = 1, 2 \tag{5.8.44}$$

在第一阶段，如果P_{1t}, P_{2t}没出现在X_{1t}, X_{2t}, Z_{1t}和Z_{2t}中，$Q_{1t}, Q_{2t}, d_{1t}, d_{2t}, s_{1t}$和$s_{2t}$用外生变量$X_{1t}, X_{2t}, Z_{1t}$和$Z_{2t}$进行回归，得到拟合值$\hat{Q}_{1t}, \hat{Q}_{2t}, \hat{d}_{1t}, \hat{d}_{2t}, \hat{s}_{1t}$和$\hat{s}_{2t}$，并在第二阶段$Q_{1t}, Q_{2t}, d_{1t}, d_{2t}, s_{1t}$和$s_{2t}$用拟合值代替；如果$P_{1t}, P_{2t}$出现在$X_{1t}, X_{2t}, Z_{1t}$和$Z_{2t}$中，且$\Delta P_{it} = P_{i,t} - P_{it-1}$，使得$P_{1t}, P_{2t}$是内生变量，则$P_{1t}, P_{2t}, Q_{1t}, Q_{2t}, d_{1t}, d_{2t}, s_{1t}$和$s_{2t}$用外生变量进行回归，得到拟合值，并在第二阶段用拟合值代替原变量。

通常，在非均衡模型中，两个相互影响的市场的需求和供给函数或是受其它市场的超额需求或是受超额供给的影响。这样得到如下的模型：

$$D_{it} = X_{it}\alpha_i + \mu_i(Q_{jt} - S_{jt}^*) + \theta_i(Q_{jt} - D_{jt}^*) + u_{it}$$

$$S_{it} = Z_{it}\beta_i + \omega_i(Q_{jt} - S_{jt}^*) + \phi_i(Q_{jt} - D_{jt}^*) + v_{it} \tag{5.8.45}$$

其中$i, j = 1, 2$且$i \neq j$。这个模型的参数可通过下列联立方程组的两阶段最小二乘估计而得到一致估计：

$$Q_{it} = X_{it}\alpha_i - \mu_i Z_{jt}\beta_j - \theta_i X_{jt}\alpha_j + (\mu_i + \theta_i)Q_{jt}$$

$$-\lambda_i^* d_{it} + [u_{it} - (\mu_i + \theta_i)v_{jt}]$$

$$Q_{it} = Z_{it}\beta_i - \omega_i X_{jt}\alpha_j - \phi_i Z_{jt}\beta_j + (\omega_i + \phi_i)Q_{jt}$$

$$+ \delta_i^* s_{it} + [v_{it} - (\omega_i + \phi_i)u_{jt}] \tag{5.8.46}$$

其中$i, j = 1, 2$且$i \neq j$。

模型(5.8.45)可在两市场模型产品－产品、要素－产品和要素－要素市场得到应用。可依赖于两市场之间的关系，对溢出系数$\mu_i, \theta_i, \omega_i$和$\phi_i$施加约束。例如，在劳动－商品市场，可假定$\theta_i$和$\phi_i$都为零，以反映两个市场是要素－产品市场。

很容易将两个相互影响市场的非均衡模型推广到n个相互影响市场的非均衡模型：

$$D_{it}=X_{it}\alpha_i+\sum_{j=1}^{n}\mu_{ij}(Q_{jt}-S_{jt}^*)+\sum_{j=1}^{n}\theta_{ij}(Q_{jt}-D_{jt}^*)+u_{it}$$

$$S_{it}=Z_{it}\beta_i+\sum_{j=1}^{n}\omega_{ij}(Q_{jt}-S_{jt}^*)+\sum_{j=1}^{n}\phi_{ij}(Q_{jt}-D_{jt}^*)+v_{it}$$

$$Q_{it}=\min(D_{it},S_{it})$$

$$\Delta P_{it}=\begin{cases}\lambda_i(D_{it}-S_{it}) & D_{it}>S_{it}\\ \delta_i(D_{it}-S_{it}) & D_{it}<S_{it}\end{cases} \qquad (5.8.47)$$

其中$i=1,\cdots,n$且$i\neq j$。模型参数可通过如下联立方程组的两阶段最小二乘估计而得到一致估计：

$$Q_{it}=X_{it}\alpha_i-\sum_{j=1}^{n}\mu_{ij}Z_{jt}\beta_j-\sum_{j=1}^{n}\theta_{ij}X_{jt}\alpha_j+\sum_{j=1}^{n}(\mu_{ij}+\theta_{ij})Q_{jt}$$

$$-\lambda_i^*d_{it}+[u_{it}-\sum_{j=1}^{n}(\mu_{ij}+\theta_{ij})v_{jt}]$$

$$Q_{it}=Z_{it}\beta_i-\sum_{j=1}^{n}\omega_{ij}X_{jt}\alpha_j-\sum_{j=1}^{n}\phi_{ij}Z_{jt}\beta_j+\sum_{j=1}^{n}(\omega_{ij}+\phi_{ij})Q_{jt}$$

$$+\delta_i^*s_{it}+[v_{it}-\sum_{j=1}^{n}(\omega_{ij}+\phi_{ij})u_{jt}] \qquad (5.8.48)$$

其中$\lambda_i^*,\delta_i^*,d_{it}$和$s_{it}$如前一样定义，$i=1,\cdots,n$且$i\neq j$。

与两市场模型的情况一样，可根据市场间的经济关系对溢出系数$\mu_{ij},\theta_{ij},\omega_{ij}$和$\phi_{ij}$施加约束。

5.9 本章思考题和综合练习题

一、思考题

1.理解为什么对（5.1.1）进行非线性最小二乘的结果会与对（5.1.2）进行普通最小二乘的结果不一致。在什么条件下，两者相同?

2.熟悉无约束优化问题求解的一阶条件和二阶条件，熟悉优化问题求解的数值迭代计算的步骤。

3.非线性计量经济模型参数估计非线性最小二乘法和最大似然法的基本原理各是什么？参数估计优化数值迭代求解时参数初始值如何选择？参数估计的渐近协方差阵如何计算？

4.理解固有非线性和参数效应非线性的曲率度量概念，写出相对曲率立体阵、参数变换后的参数效应立体阵、RMS参数效应曲率和RMS固有曲率的计算步骤。

5.掌握异方差、序列相关和条件异方差的参数最大似然估计方法。如何检验模型的随机扰动项存在条件异方差现象。

6.理解广义指数分布非线性模型的应用价值。

7.写出非线性联立方程模型参数估计的充分信息极大似然法（FIML）的计算步骤。

8.写出无参数非均衡计量经济学模型(5.8.1)**、(5.8.2)**和(5.8.3)的局部核权最大似然估计的计算步骤。

二、综合练习题

自己选择研究对象，收集样本数据，应用计量经济学软件，完成建立非线性计量经济学模型的全过程，并写出详细研究报告。

第六章 动态计量经济学模型

70年代末80年初，以英国计量经济学家D.F.Hendry为代表，在误差修正模型和协整理论的基础上，提出了动态计量经济学模型的理论与方法。80年代末以来，在我国有关学术刊物上，出现了一些介绍“Hendry学派”的文章，对D.F.Hendry为代表的关于计量经济学模型理论的研究成果进行了介绍，对该学派的建模理论给予很高的评价，甚至认为是计量经济学模型理论的革命。但是在西方，它并没有被公认为一个新的学派。不管如何评价，Hendry所倡导的动态建模理论，确实具有与传统的经典建模理论不同的思路。本章试图用通俗的语言，对它进行简单的概念性介绍，讲清楚动态计量经济学模型的方法思路，而不求理论上的完整和数学上的严谨。关于它的系统理论，有兴趣的读者可阅读《动态经济计量学》（D.F.Hendry、秦朵著，上海人民出版社，1998年4月）一书。

6.1 问题的提出

为了方便读者学习本章内容，作为开始，对传统的计量经济学建模理论与动态计量经济学建模理论进行最简单的、然而是最核心的比较，是完全必要的。关于它们的详细比较，将在本章最后一节进行。

6.1.1 传统的计量经济学模型建模理论

1.基本设定理论

计量经济学模型的基本理论形成于40年代，大部分基础性工作是由美国考尔斯经济研究委员会完成的。其基本理论可以概括为以下几点：

⑴依据某种已经存在的经济理论或者已经提出的对经济行为规律的某种解释设定模型的总体结构和个体结构，即模型是建立在已有的经济理论和经济行为规律假设的基础之上的；

⑵引进概率论思想作为模型研究的方法论基础，选择随机联立线性方程组作为模型的一般形式；

⑶模型的识别、参数的估计、模型的检验是主要的技术问题；

⑷以模型对样本数据的拟合优度作为检验模型的主要标准。

2.模型设定方法

在模型设定的方法上，传统宏观计量经济模型经历了“从简单到复杂”向“从一般到简单”的转变。

⑴从简单到复杂

从40年代到60年代，在计量经济学模型理论模型的设定上，大体遵循从简单到复杂的原则，既以一个简单的模型为起点，在这个简单的模型中只包含按照已有的经济理论和经济行为规律假设而选择的少数被认为最主要的变量。然后对该模型进行参数估计和检验，如果具有比较高的对样本数据的拟合优度，就将它作为最终模型。如果模型对样本数据的拟合优度比较低，则说明模型的解释能力不够，于是增加解释变量，再进行估计与检验，直到达到满意的拟合优度为止，得到最后的比较复杂的模型。每次增加变量的依据仍然是已有的经济理论和经济行为规律假设。在当时计算技术尚不发达的情况下，从简单到复杂是一条可行建模方法。

问题在于，拟合优度成为唯一的检验标准，不同的研究者对于同样的研究对象，由于对拟合优度的标准要求不同，或者在设定初始模型时变量选择不同，或者在增加变量时变量选择不同，都可以得到不同的最终模型。

⑵从一般到简单

随着计算技术的发展，到了60、70年代，模型的估计已经很方便了，变量多一点或者少一点其计算工作量已经没有区别。人们以另一种思路设定模型，即开始时建立一个一般的模型，将根据已有的经济理论和经济行为规律假设而认为对被解释变量具有影响的变量都作为解释变量，不管它们的影响是否显著。然后在模型的估计过程中逐渐剔除不显著的变量，最后得到一个比较简单的模型。

与从简单到复杂的建模思路相比较，从一般到简单是一个很大的进步。不同的研究者对于同样的研究对象，如果他们对经济理论和经济行为规律假设的理解是相同的，应该具有同样的起点，那么最后得到的模型也应该是相同的。

3.评价

传统计量经济学模型的设定理论是在计量经济学模型的发展过程中逐渐形成的，反过来又极大地推动了计量经济学模型的发展。现在世界各国无数的宏观计量经济模型，都是在该设定理论的指导下建立起来的。尤其是采用从一般到简单的建模思路，在很大程度上消除了建模过程中的主观性。

但是，这种设定理论是以某种既定的经济理论和经济行为规律假设为基础的，那么对于同样的研究对象，不同的研究者只要对理论假设理解不同，仍然可以建立不同的模型。另外，从应用的方面看，按照这种设定理论建立的模型，只能起到检验理论的作用，不能从中发现理论。

6.1.2 动态计量经济学—Hendry学派建模理论简介

传统的计量经济学在建立模型时多采用理论先导的途径，即首先从先验的经济理论出发设定结构模型，再由数据估计模型所包含的参数。这种建模途径对先验的经济理论具有很强的依赖性。与此相对立的是统计学家所采用的数据先导的途径，将描述数据特征作为建模的主要准则，使模型对样本数据具有很强的依赖性。而Hendry所倡导的动态建模理论，将上述二者兼容，交替运用经济理论和经济数据提供的信息，将建模过程认为是认识的循序渐进的过程。

Hendry建模理论认为，建模过程应该是首先建立一个能够代表数据生成过程(DGP, Data Generation Process)的自回归分布滞后模型(ADL, Autoregressive Distributed Lag)，然后逐步简化，最后得到包含变量间长期稳定关系的简单的模型。

数据生成过程(DGP)是所有变量联合概率分布的一般表达式。即

$$\prod_{t=1}^{T} D(x_t \,|x_{t-1};\theta\,)$$

其中x_t为t时刻所有变量观测值向量，D为概率密度函数，θ是概率密度函数D的未知参数向量。建模者所关心的参数ψ是θ的函数。这里的"所有变量"并不是任意选择的，仍然需要先验的经济理论的指导。

对DGP在一系列假设下进行约化处理，即得到一个自回归分布滞后模型(ADL)：

$$y_t = \beta_0 + \sum_{i=0}^{q} \delta_i z_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \quad \varepsilon_t{*}N(0,\sigma^2)$$

该ADL是由DGP约化而来的，如果每步约化都有效，即关于人们关注的参数ψ无信息损失，那么它可以近似地代表DGP，也就是说它能够提供DGP所能提供的关于参数ψ的信息。这就是通常所说的"从一般到简单"的"一般"，是建立模型的起点。显然，这个起点，与前面所讲的传统建模理论的"一般"的起点具有很大的差异，它更多地依赖于数据和数据生成过程。

对ADL中包含的变量进行单整和协整检验，进行逐步回归剔除明显不显著的变量，并将简化后的模型写成误差修正模型的形式，即得到了包含变量间长期稳定关系的简单的模型。

由于采用了不同的建模思路，模型的评价准则以及对模型中出现的现象的解释都不同于传统的模型。

6.1.3 两点说明

一些文章指出，"从一般到简单"是Hendry学派建模理论的基本点，其

实不然。如上所说，随着计算技术的发展，传统的建模理论也已经突破了“从简单到复杂”的束缚，走上了“从一般到简单”的道路。Hendry学派建模理论与传统的建模理论的区别在于“一般”的起点不同。

也有一些文章指出，Hendry学派建模理论是数据先导，而传统的建模理论是经济理论假设先导。如此理解Hendry学派建模理论也是片面的，它并不排除理论假设的先导作用，而是将在传统的建模理论中被忽视的数据的作用突出出来。

6.2 分布滞后模型

正如上节提及的，Hendry建模理论认为，建模过程应该是首先建立一个能够代表数据生成过程(DGP)的自回归分布滞后模型(ADL,Autoregressive Distributed Lag)，然后逐步约化，最后得到包含变量间长期稳定关系的简单的模型。所以，本节将对分布滞后模型(DLM,Distributed Lag Models)进行较为详细的介绍。实际上，分布滞后模型也是一类模型结构非经典的计量经济学问题。

6.2.1 经济分析中的分布滞后问题

在经济分析中人们发现，一些经济变量，它们的数值是由自身的滞后量或者其它变量的滞后量所决定的，表现在计量经济模型中，解释变量中经常包含某些滞后变量。

1.以投资函数为例

分析中国的投资问题发现，当年的投资额除了取决于当年的收入（即国内生产总值）外，由于投资的连续性，它还受到前1个、2个、3个、...时期投资额的影响，已经开工的项目总是要继续下去的。而每个时期的投资额又取决于每个时期的收入，所以可以建立如下关于投资的计量经济学方程：

$$I_t = \alpha + \beta_0 Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \mu_t \quad 6.2.1$$

其中I表示投资额，Y表示国内生产总值。

2.以固定资产方程为例

分析中国投入使用的固定资产发现，当年投入使用的固定资产由两部分构成，一是前一年投入使用的固定资产扣除报废后的数量，二是新形成并投入使用的部分。由于投资的时滞，当年投资并不全部形成当年投入使用的固定资产，而某年形成并投入使用的固定资产额除了取决于当年的投资外，还受到前1个、2个、3个、...时期投资额的影响。所以可以建立如下关于固定资产的计量经济学方程：

$$K_t = \alpha + \beta_0 K_{t-1} + \beta_1 I_{t-1} + \beta_2 I_{t-2} + \cdots + \mu_t \quad 6.2.2$$

其中I表示投资额，K表示固定资产额。

3.以消费函数为例

(1)从理性预期理论假设出发建立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C_t是收入预期值Y_t^e的函数，即

$$C_t = \alpha + \beta Y_t^e \tag{6.2.1}$$

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(6.5.1)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：

$$C_t = \alpha(1-\lambda) + \lambda C_{t-1} + \beta(1-\lambda)Y_t + \mu_t \quad t = 1, 2, \cdots, T \tag{6.2.2}$$

(2)从适应预期理论假设出发建立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，就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

$$C_t^e = \alpha + \beta Y_t \tag{6.2.3}$$

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：

$$C_t - C_{t-1} = \lambda(C_t^e - C_{t-1})$$

$$\lambda N`X"T\breve{}$$

$$C_t^e = \frac{1}{\lambda}C_t + \frac{\lambda - 1}{\lambda}C_{t-1}$$

代入(6.4.4) 即可求得消费函数模型，其计量形态为：

$$C_t = \lambda\alpha + (1 - \lambda)C_{t-1} + \lambda\beta Y_t + \mu_t \quad t = 1, 2, \cdots, T \tag{6.2.4}$$

上述（6.2.1）、（6.2.2）、（6.2.4）和（6.2.6）所示的模型都属于分布滞后模型。

6.2.2 多项式分布滞后模型

1.多项式分布滞后模型及其参数估计

在（6.2.1）式的分布滞后模型中，方程右边只包含外生解释变量及其滞后量。对于有限滞后长度的情形，其一般形式为：

$$y_t = \alpha + \sum_{i=0}^{p} \beta_i x_{t-i} + \mu_t \tag{6.2.5}$$

该模型中解释变量与随机误差项不相关，可以直接应用OLS估计参数。但是，一个显然的问题是解释变量之间存在共线性，而共线性问题的一个直接后果是参数估计量失去意义，不能揭示x的各个滞后量对被解释变量的影响。所以必须寻求另外的估计方法。

在一般情况下，β_i可以表示为：

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \cdots + \alpha_q i^q \quad (q < p) \tag{6.2.6}$$

即将每个参数用一个多项式表示，于是将(6.4.6)所示的分布滞后模型称为多项式分布滞后模型，也称为Almon分布滞后模型。将(6.4.7)代入(6.4.6)，得到

$$y_t = \alpha + \alpha_0 \sum_{i=0}^{p} x_{t-i} + \alpha_1 \sum_{i=1}^{p} i x_{t-i} + \alpha_2 \sum_{i=1}^{p} i^2 x_{t-i} + \cdots + \alpha_q \sum_{i=1}^{p} i^q x_{t-i} + \mu_t$$

变量置换后有：

$$y_t = \alpha + \alpha_0 z_{0t} + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \cdots + \alpha_q z_{qt} + \mu_t \tag{6.2.7}$$

对(6.4.8)直接应用OLS估计参数，得到$\hat{\alpha}, \hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_q$，然后通过它们与(6.4.6) 中参数之间的关系计算$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$,实现对原模型的估计。即

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^q \\ 1 & 3 & 9 & \cdots & 3^q \\ \vdots \\ 1 & p & p^2 & \cdots & p^q \end{pmatrix} \cdot \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \vdots \\ \hat{\alpha}_q \end{pmatrix} \tag{6.2.8}$$

写成矩阵形式为：

$$\hat{\mathrm{B}} = \mathbf{H}\hat{\mathrm{A}} \tag{6.2.9}$$

2.多项式项数的决定

关于模型(6.4.6)的滞后阶数的选择，更多的是依靠经验判断。一些研究建议，首先对(6.4.6)进行OLS或者GLS估计，根据变量的显著性检验以确定滞后的阶数。

一旦模型(6.4.6)的滞后阶数已经知道，(6.4.7)中q的选择将对估计结果产生影响。选择的方法是从一个可能的最高阶q^*开始，如果没有关于q^*的其它信息，不妨取$q^* = p$，逐次减1进行试验。构造如下统计量：

$$F = \frac{[(\mathbf{e}'\mathbf{e}\,|q) - (\mathbf{e}'\mathbf{e}\,|q^*)]/(q^* - q)}{(\mathbf{e}'\mathbf{e}\,|q^*)/(n - q^* - 1)} \tag{6.2.10}$$

其中n为样本容量，$\mathbf{e}'\mathbf{e}\,|q$和$\mathbf{e}'\mathbf{e}\,|q^*$分别为当多项式的阶数取$q$和$q^*$时的残差平方和。可以证明，$F$统计量服从自由度为$(q^* - q, n - q^* - 1)$的$F$分布。在给定显著性水平下从$F$分布表中查得临界值，如果计算得到的$F$统计量小于临界值，则接受关于$q$值的假设，需要再进行$q-1$的估计和检验；如果计算得到的$F$统计量大于临界值，则拒绝关于$q$值的假设，而接受已经通过检验的$q+1$为多项式的阶数。从直观上看，如果计算得到的$F$统计量比较小，则意味着$\mathbf{e}'\mathbf{e}\,|q$比较接近于$\mathbf{e}'\mathbf{e}\,|q^*$，即是说多项式的阶数取$q$时的结果与取最大值$q^*$时无大差别，当然可以取较小的$q$值。

6.2.3 几何分布滞后模型

1.几何分布滞后模型

在例(6.5.1)的消费函数模型中得到：

$$C_t = \alpha + \beta(1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots) \tag{6.2.11}$$

即消费是由当期和各滞后期的收入所决定，但是各滞后期收入对当期消费的影响权重是越来越小，这是经济生活中的一个常识。称(6.4.12)式的模型为几何分布滞后模型。

几何分布滞后模型的一般形式为：

$$y_t = \alpha + \beta \sum_{i=0}^{\infty} (1-\lambda)\lambda^i x_{t-i} + \mu_t \quad |\lambda| < 1 \tag{6.2.12}$$

即模型的参数为：

$$\beta_i = \beta(1-\lambda)\lambda^i \tag{6.2.13}$$

引入滞后算子L：

$$Lx_t = x_{t-1},\ L(Lx_t) = L^2x_t = x_{t-2},\ L(L(Lx_t)) = L^3x_t = x_{t-3},\ \cdots$$

将(6.4.13)表示为：

$$y_t = \alpha + \beta B(L)x_t + \mu_t$$

其中

$$B(L) = (1-\lambda)(1+\lambda L+\lambda^2L^2+\cdots) = \frac{1-\lambda}{1-\lambda L}$$

因为当$|\lambda| < 1$时存在：

$$\sum_{i=0}^{q} \lambda^i = \frac{1-\lambda^{q+1}}{1-\lambda} \qquad \sum_{i=0}^{\infty} \lambda^i = \frac{1}{1-\lambda}$$

所以有：

$$\beta(1-\lambda)\sum_{i=0}^{\infty}\lambda^i = \beta$$

$$\bar{y} = \alpha + \beta\bar{x} \tag{6.2.14}$$

2.几何分布滞后模型的另一种表示

正如例(6.5.1)的消费函数模型可以表示为

$$C_t = \alpha(1-\lambda) + \lambda C_{t-1} + \beta(1-\lambda)Y_t + \mu_t \quad t = 1,2,\cdots,T$$

一样，几何分布滞后模型的一般形式(6.4.13)可以表示为如下形式：

$$y_t = \alpha\,(1-\lambda\,) + \lambda\, y_{t-1} + \beta\,(1-\lambda\,)x_t + \mu_t \quad t = 1, 2, \cdots, T \tag{6.2.15}$$

实际应用中的几何分布滞后模型都采用这种形式。

2.几何分布滞后模型的估计

如果(6.4.16)中的随机误差项满足基本假设，即不存在异方差和序列相关，并且与解释变量不相关，可以很方便地采用OLS估计模型，得到原模型中参数$\alpha\,,\beta\,,\lambda$的估计量。

但是实际的几何分布滞后模型中的随机误差项往往不满足基本假设。还以(6.5.1)的消费函数模型为例，将随机误差项引入模型，有：

$$C_t = \alpha + \beta(1-\lambda)(Y_t + \lambda Y_{t-1} + \lambda^2 Y_{t-2} + \cdots) + \mu_t$$

$$C_{t-1} = \alpha + \beta(1-\lambda)(Y_{t-1} + \lambda\, Y_{t-2} + \lambda^2 Y_{t-3} + \cdots) + \mu_{t-1}$$

$$C_t - \lambda\, C_{t-1} = \alpha\,(1-\lambda\,) + \beta\,(1-\lambda\,)Y_t + \mu_t - \lambda\, \mu_{t-1}$$

得到：

$$C_t = \alpha\,(1-\lambda\,) + \lambda\, C_{t-1} + \beta\,(1-\lambda\,)Y_t + \mu_t - \lambda\, \mu_{t-1} \quad t = 1, 2, \cdots, T \tag{6.2.16}$$

显然，模型(6.4.17)的随机误差项$\varepsilon_t = \mu_t - \lambda\, \mu_{t-1}$既与解释变量$C_{t-1}$相关，又存在序列相关性（主要是一阶自相关）。

为了克服随机误差项与解释变量C_{t-1}相关的问题，通常采用工具变量方法估计模型。在这个模型中，经常选择Y_{t-1}作为C_{t-1}的工具变量。

为了同时克服随机误差项的序列相关，通常采用工具变量法和广义差分法结合的估计方法。步骤如下：

第一步，采用工具变量方法估计模型，得到残差估计量序列$\hat{\varepsilon}_t$。

第二步，估计$\hat{\varepsilon}_t = \rho\hat{\varepsilon}_{t-1} + \nu_t$，得到随机误差项的相关系数的估计量$\hat{\rho}_t$。

第三步，估计广义差分模型：

$$(C_t - \hat{\rho}\, C_{t-1}) = \alpha\,(1-\lambda\,)(1-\hat{\rho}) + \lambda\,(C_{t-1} - \hat{\rho}\, C_{t-2}) + \beta\,(1-\lambda\,)(Y_t - \hat{\rho}\, Y_{t-1}) + \nu_t$$

得到模型的参数估计量。

6.2.4 自回归分布滞后模型

几何分布滞后模型的优点是滞后权重逐渐衰减，致使模型与数据有较好的一致性；多项式分布滞后模型的优点是滞后分布的模式比较灵活。所以它们是应用较为普遍的两类分布滞后模型。但是，它们也存在各自的缺点，例如多项式分布滞后模型必须选择模型的滞后阶数和多项式的阶数，而几何分布滞后模型的滞后分布模式缺少灵活性。在后来的研究中，人们致力于将二者结合起来，于是Jorgenson(1966)提出了合理滞后模型（Rational Lag Model），在更多的场合被称为自回归分布滞后模型（ARDL或ADL, Autoregressive Distributed Lag Model）。

1.自回归分布滞后模型

自回归分布滞后模型是这样导出的。设如下的模型形式：

$$y_t = u + \frac{B(L)}{C(L)} x_t + \mu_t \tag{6.2.17}$$

其中$B(L)$和$C(L)$是滞后算子多项式，二者之比能够产生各种分布滞后模式。将(6.4.18)写成：

$$C(L)y_t = \alpha + B(L)x_t + C(L)\mu_t \tag{6.2.18}$$

例如，当$B(L)$和$C(L)$是二次多项式时，(6.4.19)即为：

$$\begin{aligned} y_t = u(1-\gamma_1-\gamma_2) &+ \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} \\ &+ \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \mu_t - \gamma_1 \mu_{t-1} - \gamma_2 \mu_{t-2} \end{aligned}$$

在(6.4.19)中，y_t和μ_t被施加了相同的滞后结构，这并不是必须的，也不是所希望的。它们可以具有各自的灵活的滞后结构。于是模型可以写成：

$$C(L)y_t = \alpha + B(L)x_t + D(L)\mu_t \tag{6.2.19}$$

这是人们所熟悉的ARMAX模型，它的一种特殊情况就是本书第二章介绍的ARMA模型。在ARMA模型中，未出现外生变量。更进一步，如果在模型中有1个以上的外生变量，每个外生变量都存在有限的滞后项，模型可以写为：

$$\begin{aligned} y_t = \alpha &+ \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \mathrm{B}'\mathbf{X}_t \\ &+ \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \qquad t = 1, 2, \cdots, T \end{aligned} \tag{6.2.20}$$

这就是(p, q)阶自回归分布滞后模型的一般表达式，在本章中应用的即是这类模型。

2.自回归分布滞后模型的非线性最小二乘估计

如果模型(6.4.21)中不存在移动平均项，则可以采用OLS估计该线性回归模型的参数

$$\phi = (\alpha, \gamma_1, \gamma_2, \cdots, \beta_1, \beta_2, \cdots, \beta_k)$$

模型的解释变量是：

$$\mathbf{Z_t} = (1, y_{t-1}, y_{t-2}, \cdots, y_{t-p}, \mathbf{X'_t})'$$

其中$\mathbf{X}_t$可以包含滞后项。OLS估计量服从渐近正态分布，其渐近协方差矩阵为：

$$Asy.Var(\hat{\phi}) = \frac{1}{T-p} p\lim \sigma_\mu^2 \left(\frac{\mathbf{Z'Z}}{T-p}\right)$$

随机误差项的方差为：

$$\sigma_\mu^2 = p\lim \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\mu}^2$$

当模型中存在移动平均项，线性OLS估计将是非一致性估计。此时应该采用非线性最小二乘估计。如果随机误差项服从正态分布，非线性最小二乘估计等同于最大似然估计，其估计量将是一致性估计量，并服从渐近正态分布。

Greene(1997)提出如下的估计过程。对

$$S(\alpha, \gamma, \beta, \phi) = S(\phi) = \sum_{t=1}^{T} \mu_t^2$$

最小化，采用Gauss-Newton迭代方法，迭代过程为：

$$\hat{\phi}^{s+1} = \hat{\phi}^s - (\mathbf{G'}_s \mathbf{G}_s)^{-1} \mathbf{G'}_s \hat{\mu}_s \tag{6.2.21}$$

其中$\mathbf{G}_s$是残差关于参数的导数的$(T-p-q) \times (1+p+k+q)$阶矩阵。其元素由下式计算得到：

$$\frac{\partial \mu_t}{\partial \alpha} = 1 - \theta_1 \left(\frac{\partial \mu_{t-1}}{\partial \alpha}\right) - \cdots - \theta_q \left(\frac{\partial \mu_{t-q}}{\partial \alpha}\right)$$

$$\frac{\partial \mu_t}{\partial \gamma_i} = y_{t-i} - \theta_1 \left(\frac{\partial \mu_{t-1}}{\partial \gamma_i}\right) - \cdots - \theta_q \left(\frac{\partial \mu_{t-q}}{\partial \gamma_i}\right) \qquad i = 1, 2, \cdots, p$$

$$\frac{\partial \mu_t}{\partial \beta_i} = x_{jt} - \theta_1 \left(\frac{\partial \mu_{t-1}}{\partial \beta_j}\right) - \cdots - \theta_q \left(\frac{\partial \mu_{t-q}}{\partial \beta_j}\right) \qquad j = 1, 2, \cdots, k$$

$$\frac{\partial \mu_t}{\partial \theta_l} = \mu_{t-l} - \theta_1 \left(\frac{\partial \mu_{t-1}}{\partial \theta_l} \right) - \cdots - \theta_q \left(\frac{\partial \mu_{t-q}}{\partial \theta_l} \right) \qquad l = 1, 2, \cdots, q$$

这是一组微分方程，在给出μ_t的初值后可以递推求解。用矩阵形式表示为：

$$\mathbf{G}_{\alpha,t} = \left(\frac{\partial \mu_{t-s}}{\partial \alpha} \right) \qquad s = 1, \cdots, q(1 \times q)$$

$$\mathbf{G}_{\gamma,t} = \left(\frac{\partial \mu_{t-s}}{\partial \gamma} \right) \qquad s = 1, \cdots, q(p \times q)$$

$$\mathbf{G}_{\beta,t} = \left(\frac{\partial \mu_{t-s}}{\partial \beta} \right) \qquad s = 1, \cdots, q(k \times q)$$

$$\mathbf{G}_{\theta,t} = \left(\frac{\partial \mu_{t-s}}{\partial \theta} \right) \qquad s = 1, \cdots, q(q \times q)$$

于是，$\mathbf{G}_s$的第t行是：

$$\mathbf{g}_s^{(t)} = \left(\; \left(\tfrac{\partial \mu_\mathbf{t}}{\partial \alpha} \right) \; \left(\tfrac{\partial \mu_\mathbf{t}}{\partial \gamma'} \right) \; \left(\tfrac{\partial \mu_\mathbf{t}}{\partial \beta'} \right) \; \left(\tfrac{\partial \mu_\mathbf{t}}{\partial \theta'} \right) \; \right)$$

其中

$$\frac{\partial \mu_\mathbf{t}}{\partial \alpha} = 1 - \mathbf{G}_{\alpha,t}\theta \quad \frac{\partial \mu_\mathbf{t}}{\partial \gamma} = \mathbf{y}_{lags} - \mathbf{G}_{\gamma,t}\theta$$

$$\frac{\partial \mu_\mathbf{t}}{\partial \beta} = \mathbf{x}_t - \mathbf{G}_{\beta,t}\theta \quad \frac{\partial \mu_\mathbf{t}}{\partial \theta} = \mu_{lags} - \mathbf{G}_{\theta,t}\theta$$

关于初始值的确定。可以采用工具变量法得到$\phi = (\alpha, \gamma_1, \gamma_2, \cdots, \beta_1, \beta_2, \cdots, \beta_k)$估计量的初始值，然后对残差进行自回归估计，得到$(\theta_1, \theta_2, \cdots, \theta_q)$估计量的初始值。

方差的估计量为：

$$\hat{\sigma}_\mu^2 = \frac{1}{T - p - 1 - k - q} \sum_{t=p+q+1}^{T} \hat{\mu}_t^2 \tag{6.2.22}$$

其中

$$\begin{aligned} \hat{\mu}_t = {} & y_t - \hat{\alpha} - \hat{\gamma}_1 y_{t-1} - \hat{\gamma}_2 y_{t-2} - \cdots - \hat{\gamma}_p y_{t-p} - \hat{\mathrm{B}}'\mathbf{X}_t \\ & + \hat{\theta}_1 \hat{\mu}_{t-1} + \hat{\theta}_2 \hat{\mu}_{t-2} + \cdots + \hat{\theta}_q \hat{\mu}_{t-q} \qquad t = 1, 2, \cdots, T \end{aligned} \tag{6.2.23}$$

6.2.5 向量自回归模型

将ARMA模型扩展到一组变量的情形，即

$$\mathbf{y}_t = \alpha + \Delta_1 \mathbf{y}_{t-1} + \cdots + \Delta_p \mathbf{y}_{t-p} + \mu_t + \Theta_1 \mu_{t-1} + \cdots + \Theta_q \mu_{t-q} \tag{6.2.24}$$

其中

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{pmatrix} \quad \mathbf{y}_{t-1} = \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{Mt-1} \end{pmatrix} \quad \cdots \quad \mathbf{y}_{t-p} = \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{Mt-p} \end{pmatrix}$$

$$\Delta_1 = \begin{pmatrix} \delta_{111} & \delta_{112} & \cdots & \delta_{11M} \\ \delta_{121} & \delta_{122} & \cdots & \delta_{12M} \\ \vdots & & & \\ \delta_{1M1} & \delta_{1M2} & \cdots & \delta_{1MM} \end{pmatrix} \quad \cdots \quad \Delta_p = \begin{pmatrix} \delta_{p11} & \delta_{p12} & \cdots & \delta_{p1M} \\ \delta_{p21} & \delta_{p22} & \cdots & \delta_{p2M} \\ \vdots & & & \\ \delta_{pM1} & \delta_{pM2} & \cdots & \delta_{pMM} \end{pmatrix}$$

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{pmatrix} \quad \mu_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{Mt} \end{pmatrix} \quad \mu_{t-1} = \begin{pmatrix} \mu_{1t-1} \\ \mu_{2t-1} \\ \vdots \\ \mu_{Mt-1} \end{pmatrix} \quad \cdots \quad \mu_{t-q} = \begin{pmatrix} \mu_{1t-q} \\ \mu_{2t-q} \\ \vdots \\ \mu_{Mt-q} \end{pmatrix}$$

$$\Theta_1 = \begin{pmatrix} \varphi_{111} & \varphi_{112} & \cdots & \varphi_{11M} \\ \varphi_{121} & \varphi_{122} & \cdots & \varphi_{12M} \\ \vdots & & & \\ \varphi_{1M1} & \varphi_{1M2} & \cdots & \varphi_{1MM} \end{pmatrix} \quad \cdots \quad \Theta_q = \begin{pmatrix} \varphi_{q11} & \varphi_{q12} & \cdots & \varphi_{q1M} \\ \varphi_{q21} & \varphi_{q22} & \cdots & \varphi_{q2M} \\ \vdots & & & \\ \varphi_{qM1} & \varphi_{qM2} & \cdots & \varphi_{qMM} \end{pmatrix}$$

称(6.2.24)为向量自回归移动平均模型。在实际应用中，经常应用没有移动平均项的简单形式，即

$$\mathbf{y}_t = \alpha + \Delta_1 \mathbf{y}_{t-1} + \cdots + \Delta_p \mathbf{y}_{t-p} + \mu_t \tag{6.2.25}$$

称(6.2.25)为向量自回归模型。

向量自回归模型有两个重要应用。一是用于Granger检验，另一是用于描述脉冲响应函数。这里不作介绍，读者可阅读其它有关教科书。

6.3 从数据生成过程到自回归分布滞后模型

动态计量经济模型的起点是反映数据生成过程（DGP, Data Generation Process）的自回归分布滞后模型（ADL），那么，从DGP到ADL，成为动态计量经济模型理论方法的一个重要部分。

6.3.1 数据生成过程(DGP)

数据生成过程，顾名思义，是描述已经得到的变量观测值是如何生成的。这是一个一般化的概念。可以用所有变量（包括内生变量和外生变量）观测值的联合概率密度函数来表示。设x_t是t时期所有变量观测值向量，用$\mathbf{X}_{t-1}$表示从第1时期到第$t-1$时期所有变量观测值矩阵，即$\mathbf{X}_{t-1}=(x_{t-1},x_{t-2},\cdots,x_1)'$，那么观测值$x_t$的联合概率密度函数，即DGP，可以表示为：

$$\prod_{t=1}^{T} D(x_t\,|\mathbf{X}_{t-1};\Theta) \tag{6.3.1}$$

其中Θ是概率密度函数D的未知参数向量。(6.5.1)的意义是指，x_t的发展变化只依赖于自身的历史$\mathbf{X}_{t-1}$，而与其它因素无关。

如果x_t可以分为内生变量y_t和外生变量z_t，$\mathbf{X}_{t-1}$也作如下相应分解：

$$\mathbf{X}_{t-1}=(\mathbf{Y}_{t-1}:\mathbf{Z}_{t-1})$$

$$\mathbf{Y}_{t-1}=(y_{t-1},y_{t-2},\cdots,y_1)'$$

$$\mathbf{Z}_{t-1}=(z_{t-1},z_{t-2},\cdots,z_1)'$$

则(6.5.1)的联合概率密度函数可以分解为y_t对z_t的条件密度$D_{y|z}$与z_t的边际密度D_z的乘积：

$$D(x_t\,|\mathbf{X}_{t-1}\,;\Theta)=D_{y|z}(y_t\,|z_t,\mathbf{Y}_{t-1},\mathbf{Z}_{t-1};\lambda_1)\cdot D_z(z_t\,|\mathbf{Y}_{t-1},\mathbf{Z}_{t-1};\lambda_2) \tag{6.3.2}$$

其中λ_1和λ_2分别为条件密度和边际密度的参数向量，是由是联合概率密度函数D的原未知参数向量Θ分解而来。如果Θ中包括n个参数，λ_1和λ_2中分别包括n_1和个参数，则应有$n=n_1+n_2$。

6.3.2 弱外生性、强外生性和超外生性

在(6.4.2)中，如果能够将外生变量z_t的边际密度D_z合理地约去，则可以使得描述数据生成过程的联合概率密度函数简化。其条件是外生变量z_t具有弱外生性。

1.弱外生性的条件

弱外生性的概念由Richard于1980年首先提出，后来由Engle等人于1983年进行了系统的定义。

在计量经济分析中，人们往往并不关注所有的参数Θ，而只关注它的一个子集Ψ，并且在Ψ中包括的参数数目$k \le n_1$，因为这些参数描述了内生变量y_t与相关联的变量之间的关系。子集Ψ被称为关注参数。从这里得出一个重要条件：如果能够将外生变量z_t的边际密度D_z从(6.4.2)中合理地约去，那么Ψ应该仅为λ_1的函数而与λ_2无关。即

$$\Psi = g(\lambda_1) \tag{6.3.3}$$

这就构成z_t具有弱外生性的第1个条件。

在满足条件(6.4.4)的情况下，如果λ_1对λ_2存在某种依存性，那么仍然不能将外生变量z_t的边际密度D_z从(6.4.2)中合理地约去。于是引出了z_t具有弱外生性的第2个条件，即λ_1与λ_2相互无关。该条件实际上是对参数λ_1与λ_2的子空间Λ_1与Λ_2的某种约束，要求该两组参数之子空间的积满足：

$$\Lambda_1 \times \Lambda_2 = \{(\lambda_1, \lambda_2) : \lambda_1 \in \Lambda_1 \lambda_2 \in \Lambda_2\} \tag{6.3.4}$$

即该两组参数是来自两个子空间之积空间的，而不是来自该积空间的任意一约束子空间。而且，任何属于Λ_1的λ_1都可以和任何属于Λ_2的λ_2偶合得出联合分布之参数集。通俗地讲，参数子集λ_1与λ_2之间不含任何跨集的约束。

至此，可以给出关于弱外生性的严格的定义：如果条件(6.4.4)和(6.4.5)成立，z_t即为关注参数Ψ的弱外生变量。该定义表明，变量的弱外生性必须是相对于给定的关注参数而言的。

由于条件(6.4.4)和(6.4.5)断绝了关注参数Ψ直接或者间接依存于λ_2的可能性，因此保证了关于z_t的边际密度D_z的信息对于估计关注参数Ψ的无关性，所以z_t的弱外生性是从(6.4.2)中合理地约去z_t的边际密度D_z的充要条件。

弱外生性不是变量本身的性质，是相对于关注参数而言的。下面通过一个例子专门说明这个关系。对于一个简单的粮食供求模型：

$$\begin{cases} p_t = \alpha_0 + \alpha_1 q_t + \mu_t & \mu_t \sim N(0, \sigma_\mu^2) \\ q_t = \beta_0 + \beta_1 p_{t-1} + \varepsilon_t & \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \end{cases}$$

$$E(\mu_t \varepsilon_t) = 0 \quad \forall t, s$$

需求方程的参数集为：$\lambda_1 = (\alpha_0, \alpha_1, \sigma_\mu^2)$；供给方程的参数集为：$\lambda_2 = (\beta_0, \beta_1, \sigma_\varepsilon^2)$。$q_t$是否是弱外生性变量？如果关注参数为需求弹性，即价格与需求量之间的关系，可以用$1/\alpha_1$近似表示（当然，$1/\alpha_1$表示的是绝对变化量之间的关系，需求弹性是相对变化量之间的关系），显然只是λ_1的子集，q_t是弱外生性变量。如果关注参数为系统的稳定性，即需要将供给方程代入需求方程，得到

$$p_t = \alpha_0 + \alpha_1\beta_0 + \alpha_1\beta_1 p_{t-1} + \nu_t$$

表示为：

$$p_t = \gamma + \rho\, p_{t-1} + \nu_t$$

由经济理论知道，参数ρ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需要$\alpha_1 \in \{\lambda_1\}$，而且需要$\beta_1 \in \{\lambda_2\}$，所以$q_t$不是弱外生性变量。

2.强外生性的条件

与外生性有关的概念除了弱外生性外，还有强外生性和超外生性。弱外生性仅相对于利用模型进行关注参数的统计推断的情形，强外生性是相对于利用模型进行预测的情形，超外生性则是相对于利用模型进行政策分析的情形。

直观地看，如果要利用(6.4.2)进行预测并能够合理地约去z_t的边际密度D_z，仅有z_t的弱外生性是不够的，还必须限定z_t不受$\mathbf{Y}_{t-1}$的反馈影响，这就是z_t的强外生性。于是可以给出如下定义：在z_t为关注参数Ψ的弱外生变量的基础上，如果下述条件：

$$D_z(z_t\,|\mathbf{X}_{t-1}\,;\lambda_2) = D_z(z_t\,\Big|\mathbf{Z}_{t-1}^1, \mathbf{X}_0\,;\lambda_2) \tag{6.3.5}$$

成立，则z_t为关注参数Ψ的强外生变量。

3.超外生性条件

如果要利用(6.4.2)进行政策模拟并能够合理地约去z_t的边际密度D_z，仅有z_t的弱外生性是不够的。在弱外生性条件中，条件密度$D_{y|z}$与边际密度D_z分布的参数之间的相互变化五无关属于内生变量y_t和外生变量z_t联合分布体系之内的性质。至于z_t的边际密度D_z的参数集λ_2在某种体系外因素作用下发生了变动，是否会使得$D_{y|z}$的参数集λ_1也随之变动的问题，弱外生性条件并没有给于回答。这就需要在λ_2有可能发生体系外变动的前提下，对λ_1和λ_2之间的关系作出更加明确的限制。于是提出了超外生性条件问题。

设Ω为对数据生成过程(6.5.1)的可行干预集，并设$C^\Theta \in \Omega$为能够引起Θ变动的一类干预集，其中可包括现实中的各种政策干预。仅考虑C^Θ的一个子集C^{λ_2}，即仅考虑直接对边际密度D_z的参数集λ_2产生影响的一类可行干预，显然C^{λ_2}大大约束了所涉及的政策干预的类型。

现在给出如下定义：如果λ_1为独立与子集C^{λ_2}的常数，则称λ_1对于属于子集C^{λ_2}的政策干预具有抗变性。很明显，λ_1对于政策干预的抗变性是由条件模型（条件密度$D_{y|z}$）推断结构模型（用于进行政策模拟的模型）的一个基本条件。如果一个模型的参数经不起任何外部干预的影响，就很难称之为结构模型。

至此，可以给出超外生性条件为：在z_t为关注参数Ψ的弱外生变量的基础上，如果λ_1对于属于子集C^{λ_2}的政策干预具有抗变性，则z_t为关注参数Ψ的超外生变量。

在上述定义中发现，超外生性是以弱外生性为基础，并没有以强外生性为基础。即是说超外生性包含弱外生性，但不包含强外生性。从直观上看，如果条件模型中的解释变量具有强外生性，那么被控制变量（内生变量）的滞后项对它没有影响。然而，许多政策干预是政府根据被控制变量的滞后信息而制定的。这说明，具有强外生性的解释变量一般不应被选择为实施政策干预的超外生变量。

关于弱外生性、强外生性和超外生性的检验，读者可参阅《动态经济计量学》（D.F.Hendry、秦朵，上海人民出版社，1998年）。在更多的情况下，是根据对经济行为的理解作出判断。

6.3.3 约化

约化理论是动态计量经济学模型理论的重要组成部分。由(6.5.1) 或者(6.4.2)表示的数据生成过程经过约化，得到作为应用计量经济学模型起点的自回归分布滞后模型，只有在约化过程中没有关于关注参数的信息损失，自回归分布滞后模型才能代表数据生成过程。由此可见约化在理论上的重要性。但是，在实际建模中，人们往往直接从自回归分布滞后模型开始，并不实际地进行数据生成过程的约化，所以，在这里只对它们作简单的概念性介绍。对动态计量经济学模型理论有兴趣的读者可以参阅《动态经济计量学》一书。

1.条件化—分布的约化

上述将x_t分为内生变量y_t和外生变量z_t，并将(6.5.1)的联合概率密度函数表示为y_t对z_t的条件密度$D_{y|z}$与z_t的边际密度D_z的乘积的过程，称为条件因子分解过程，也称为条件约化或条件化。重要的是在该过程中使得条件(6.4.4)和(6.4.5)成立，即z_t为关注参数Ψ的弱外生变量。于是可以从(6.4.2)中合理地约去z_t的边际密度D_z，使得描述数据生成过程的联合概率密度函数简化为y_t对z_t的条件密度：

$$D_{y|z}(y_t\,|z_t,\mathbf{Y}_{t-1},\mathbf{Z}_{t-1};\lambda_1) \tag{6.3.6}$$

2.新生化—误差项的约化

设(6.4.7)的条件期望为：

$$u_t = E(y_t\,|z_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}; \lambda_1)$$

再设$\varepsilon_t = y_t - u_t$ (6.3.7)

称ε_t为新生误差。在(6.3.7)两边，关于信息集$(\mathbf{Z}_t : \mathbf{Y}_{t-1})$求条件期望，得到：

$$E(\varepsilon_t\,|\mathbf{Z}_t, \mathbf{Y}_{t-1}) = E(y_t\,|\mathbf{Z}_t, \mathbf{Y}_{t-1}) - u_t = 0 \tag{6.3.7}$$

显然，如果上述条件化是合理的，该新生误差应该是白躁声。

3.常数化—参数的约化

严格讲，对于每个数据观测点都存在一组参数，(6.4.7)应该写为：

$$D_{y|z}(y_t\,|z_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}; \lambda_{1t}) \tag{6.3.8}$$

但是，如果数据生成过程的分布随着每个观测点而改变，所导出的应用模型的参数不具有常数性。所以，在样本数据为时间序列数据时，需要对参数施加时间齐次性约束。这是数据生成过程向应用模型靠拢的关键一步。当模型的参数具有常数性时，有

$$\lambda_{1t} = \lambda_1 \qquad \forall t; \quad \lambda_1 \in \Lambda_1 \tag{6.3.9}$$

常数化过程是通过变量的选择来实现的。正如在§2.2中所介绍的，变参数模型可以通过引入政策变量使之变为常参数模型。当然这里引入的变量应该具有超外生性。

4.截尾化—滞后项的约化

对(6.4.7)中的滞后项进行滞后项截尾，得到：

$$D_{y|z}(y_t\,|z_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}; \lambda_1) = D_{y|z}\left(y_t\,\middle|z_t, \mathbf{Y}_{t-1}^{t-p}, \mathbf{Z}_{t-1}^{t-q}; \lambda_1\right) \tag{6.3.10}$$

这一约化过程合理性的条件是由(6.3.7)得到的新生误差

$$\varepsilon_t = y_t - E\left(y_t\,\middle|z_t, \mathbf{Y}_{t-1}^{t-p}, \mathbf{Z}_{t-1}^{t-q}; \lambda_1\right) \tag{6.3.11}$$

是白躁声。

5.线性化—函数形式的约化

如果约化后的数据生成过程(6.4.11)服从正态分布，则y_t基于z_t的条件模型便是线性的。因此，在参数满足常数性条件下，函数形式的约化旨在寻求线性模型与正态分布的统一。

具体地，函数形式的约化相当于对变量y_t和z_t的某种函数转换，即

$$y_t^* = h(y_t) \qquad z_t^* = g(z_t)$$

使得转换后的条件分布近似服从同方差正态分布，即

$$D_{y|z}(y_t\,|z_t, \mathbf{Y}_{t-1}^{t-p}, \mathbf{Z}_{t-1}^{t-q}; \lambda_1) = D_{y^*|z^*}(y_t^*\,|z_t^*, \mathbf{Y}_{t-1}^{*t-p}, \mathbf{Z}_{t-1}^{*t-q}; \gamma) \underset{app}{\sim} N(\eta_t, \sigma^2) \tag{6.3.12}$$

$$labeleq : 060316. \tag{6.3.13}$$

在(6.3.13)成立的条件下，y_t^*基于z_t^*的条件模型便是线性的。

6.3.4 自回归分布滞后模型和数据生成过程

数据生成过程经过上述约化后，得到它的衍生模型，如果每一步约化都是有效的，即没有关于关注参数的信息损失，那么该衍生模型代表了数据生成过程。

利用滞后算子L来表示衍生模型：

$$\mathbf{A}(L)\mathbf{y}_t^* = \mathbf{B}(L)\mathbf{z}_t^* + \varepsilon_t \qquad \varepsilon_t \underset{app}{\sim} N_{n1}(0, \Sigma_\varepsilon) \tag{6.3.14}$$

其中$\mathbf{A}(L)$和$\mathbf{B}(L)$是维数为p和q的多项式矩阵。可见，约化得到的衍生模型实际上是向量自回归分布滞后模型。而且，误差项不是源生的，而是模型衍生的，其定义为：

$$\varepsilon_t = \mathbf{A}(L)\mathbf{y}_t^* - \mathbf{B}(L)\mathbf{z}_t^* \tag{6.3.15}$$

于是，(6.4.13)的模型就成为代表数据生成过程的动态计量经济学模型的起点。在后面的模型比较中，还将对该模型及其误差项的性质进行讨论。

正如前面已经反复说明的，在实际建模时，直接以自回归分布滞后模型为起点，并不讨论上述约化过程。那么，如何才能保证所选择的ADL模型满足上述约化的要求呢？关于线性化，采用经典模型的方法处理；关于截尾化，下节还要简单介绍；关于常数化，更多的是一种假设；关于新生化，主要依靠变量选择，尽可能将对被解释变量具有影响的变量选入模型；关于条件化，即变量关于参数的弱外生性，主要依靠对经济行为的理解加以确定。

6.4 从自回归分布滞后模型到误差修正模型

动态计量经济学模型的一般形式是§2.5中介绍的误差修正模型，那么从反映数据生成过程的自回归分布滞后模型到误差修正模型，成为动态模型的主要技术内容。

6.4.1 自回归分布滞后模型的阶数的决定

从§6.3已经知道，经过有效的约化，数据生成过程可以表示为自回归分布滞后模型。所以，在实际建立动态计量经济学模型时，就以自回归分布滞后模型作为起点。

自回归分布滞后模型中变量的选择，仍然需要从经济理论和对经济行为规律的认识出发，人们不会也不应该将毫无关系的变量选入模型。正是在这点上，体现了动态计量经济学模型离不开理论导向。

反映数据生成过程的自回归分布滞后模型的一般表达式为：

$$y_t = \beta_0 + \sum_{i=0}^{q} \delta_i z_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \quad \varepsilon_t * N(0, \sigma^2) \tag{6.4.1}$$

理论上讲，决定模型中变量滞后的最大阶数(p, q)的依据是误差项$\{\varepsilon_t\}$为一新生过程，或者称为新生误差，新生误差除了必须是白躁声外，还具有其它重要性质。

但是，在实际中，人们更多的是从经验出发判断阶数(p, q)。例如，如果采用季节数据，那么滞后阶数至少大于4；如果采用年度数据，那么滞后阶数至少大于1。如果y表示消费，z表示收入，那么$(p, q) = (2, 2)$较为合适，因为消费的调整比较容易；如果y表示投资，z表示收入，那么(p, q)应该比较高，因为投资的调整比较困难，开工的项目很难停下来，而一般的大项目具有比较长的建设期。在实践中，也可以开始设定较长的滞后阶数，然后对(6.5.1)采用OLS回归，根据变量的显著性检验确定选择的阶数。

6.4.2 正交化变换

在确定了模型中变量滞后的最大阶数后，对一般模型施加某种变换，使解释变量之间近似正交。

下面以简单线性模型为例，说明正交变换过程。设有线性模型：

$$y_t = \beta' \mathbf{Z}_t + \varepsilon_t \tag{6.4.2}$$

设$\mathbf{H}$为任意一个已知的$(k \times k)$阶非奇异矩阵，满足$\mathbf{H}^{-1}\mathbf{H} = \mathbf{I}$，$\gamma$为一个已知的$(k \times 1)$阶向量。首先从(6.4.2)两边减去$\gamma' \mathbf{Z}_t$：得到：

$$y_t - \boldsymbol{\gamma}'\mathbf{Z}_i = (\boldsymbol{\beta} - \boldsymbol{\gamma})'\mathbf{Z}_t + \varepsilon_t \tag{6.4.3}$$

然后，变换(6.4.3)为：

$$(y_t - \gamma' \mathbf{Z}_t) = \{(\beta - \gamma)'\mathbf{H}^{-1}\}\mathbf{H}\mathbf{Z}_t + \varepsilon_t \tag{6.4.4}$$

即

$$y_t^* = \beta^{*\prime} \mathbf{Z}_t^* + \varepsilon_t \tag{6.4.5}$$

其中，$\beta^* = \mathbf{H}^{-1}(\beta - \gamma)$，$y_t^* = (y_t - \gamma' \mathbf{Z}_t)$，$\mathbf{Z}_t^* = \mathbf{HZ}_t$，实际上是对原始模型(6.4.2)进行了再参数化，仍然有相同数目的参数、相同数目的回归元和相同的误差项。至于是首先估计(6.4.2)得到$\hat{\beta}$然后导出$\hat{\beta}^*$，还是直接估计变换后的模型(6.4.5)得到$\hat{\beta}^*$，是无所谓的。因为

$$\begin{aligned}\hat{\beta}^* &= (\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}y^* = (\mathbf{H}'\mathbf{Z}'\mathbf{ZH})^{-1}\mathbf{H}'\mathbf{Z}'(y - \gamma'\mathbf{Z}) \\ &= \mathbf{H}^{-1}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(y - \gamma'\mathbf{Z}) = \mathbf{H}^{-1}(\hat{\beta} - \gamma)\end{aligned} \tag{6.4.6}$$

从(6.4.2)到(6.4.5)的过程，就是正交变换的过程。

经过正交变换的模型的解释变量是近似正交的，该模型有一个显著的优点是，在模型简化的过程中，当去掉系数估计值小的变量或者去掉显著性差的变量时，在数值上和统计上，几乎不改变剩余变量的系数估计值。这一点已经在本书§3.2中部分回归估计中得到证明。所以首先对模型进行正交变换，然后再对模型进行简化，应该是一条正确的建模路线。

对于一个实际的自回归分布滞后模型进行正交变换，关键是选择合适的**H**矩阵。这并非容易的事情。所以在已经见到的动态模型中，一般都将正交变换过程与引入误差修正机制结合起来，最后得到解释变量具有近似正交性的误差修正模型。

6.4.3 协整检验

在本书§2.5中，已经介绍了协整理论。协整理论是动态计量经济学模型的理论基础，正是由于变量之间可能存在的协整关系，才可能将自回归分布滞后模型变换为具有误差修正机制的误差修正模型。

在§2.5中已经介绍了单整的DF检验和ADF检验，以及检验两个同阶单整变量之间是否存在协整的EG检验（Engle和Granger于1987年提出两步检验法）。这里着重介绍多重协整检验，即检验多个具有同阶单整变量之间是否存在协整关系。

也可以用EG检验法检验多个具有同阶单整变量之间是否存在协整关系。在其第一阶段，需要设计许多线性模型进行OLS估计，应用很不方便。Johansen于1988年，以及与Juselius一起于1990年提出了一种用向量自回归模型进行检验的方法，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。

1.JJ检验的原理

§2.5中(6.2.27)所示的没有移动平均项的向量自回归模型表示为：

$$\mathbf{y}_t = \alpha + \Pi_1 \mathbf{y}_{t-1} + \cdots + \Pi_p \mathbf{y}_{t-p} + \mu_t \tag{6.4.7}$$

其中

$$\mathbf{y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Mt} \end{pmatrix} \quad \mathbf{y}_{t-1} = \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{Mt-1} \end{pmatrix} \quad \cdots \quad \mathbf{y}_{t-p} = \begin{pmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{Mt-p} \end{pmatrix}$$

$$\Pi_1 = \begin{pmatrix} \pi_{111} & \pi_{112} & \cdots & \pi_{11M} \\ \pi_{121} & \pi_{122} & \cdots & \pi_{12M} \\ \vdots & & & \\ \pi_{1M1} & \pi_{1M2} & \cdots & \pi_{1MM} \end{pmatrix} \quad \cdots \quad \Pi_p = \begin{pmatrix} \pi_{p11} & \pi_{p12} & \cdots & \pi_{p1M} \\ \pi_{p21} & \pi_{p22} & \cdots & \pi_{p2M} \\ \vdots & & & \\ \pi_{pM1} & \pi_{pM2} & \cdots & \pi_{pMM} \end{pmatrix}$$

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{pmatrix} \quad \mu_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{Mt} \end{pmatrix}$$

为了简便，改写为：

$$\mathbf{y}_t = \alpha + \sum_{j=1}^{p} \Pi_j \mathbf{y}_{t-j} + \mu_t \tag{6.4.8}$$

如果$\mathbf{y}_t$表示M个$I(1)$过程构成的向量，对(6.4.8)进行差分变换可以得到下式表示的模型：

$$\Delta\mathbf{y}_t = \sum_{j=1}^{p} \Gamma_j \Delta\mathbf{y}_{t-j} + \Pi\mathbf{y}_{t-1} + \varepsilon_t \tag{6.4.9}$$

由于$I(1)$过程经过差分变换将变成$I(0)$过程，即(6.4.9)中的$\Delta\mathbf{y}_t$、$\Delta\mathbf{y}_{t-j}(j = 1, 2, \cdots, p)$都是$I(0)$变量构成的向量，那么只有$\Pi\mathbf{y}_{t-1}$是$I(0)$变量构成的向量，即$y_{1\,t-1}, y_{2\,t-1}, \cdots, y_{M\,t-1}$之间具有协整关系，才能保证新生误差是平稳过程。

如果$R(\Pi) = M$，显然只有$y_{1\,t-1}, y_{2\,t-1}, \cdots, y_{M\,t-1}$都是$I(0)$变量，才能保证新生误差是平稳过程。而这与已知的$\mathbf{y}_t$为$I(1)$过程相矛盾。所以必然存在$R(\Pi) < M$。

如果$R(\Pi) = 0$，意味着$\Pi = \mathbf{0}$，因此(6.4.9)仅仅是个差分方程，各项都是$I(0)$变量，不需要讨论$y_{1\,t-1}, y_{2\,t-1}, \cdots, y_{M\,t-1}$之间是否具有协整关系。

如果$R(\Pi) = r\,(0 < r < M)$，表示存在r个协整组合，其余$M - r$个关系仍为$I(1)$关系。在这种情况下，Π可以分解成两个$(M \times r)$阶矩阵α和β的乘积：

$$\Pi = \alpha\beta' \tag{6.4.10}$$

其中$R(\alpha) = r$，$R(\beta\,) = r$。将(6.4.10)代入(6.4.9)，得到：

$$\Delta \mathbf{y}_t = \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{y}_{t-j} + \alpha\beta'\,\mathbf{y}_{t-1} + \varepsilon_t \tag{6.4.11}$$

该式要求$\beta'\,\mathbf{y}_{t-1}$为一个$I(0)$向量，其每一行都是一个$I(0)$组合变量，即每一行所表示的$y_{1\,t-1}, y_{2\,t-1}, \cdots, y_{M\,t-1}$的线性组合都是一种协整形式。所以矩阵$\beta'$决定了$y_{1\,t-1}, y_{2\,t-1}, \cdots, y_{M\,t-1}$之间协整向量的个数与形式。所以$\beta'$称为协整向量矩阵，r为系统中协整向量的个数。矩阵$\alpha$的每一行$\alpha_j$是出现在第j个方程中的$r$个协整组合的一组权重，故称为调整参数矩阵。当然容易发现，α和β并不是唯一的。

于是，将$\mathbf{y}_t$中的协整检验变成对矩阵Π的分析问题。这就是JJ检验的基本原理。

2.JJ检验的预备工作

Johansen于1988年提出的检验方法必须进行如下步骤的预备工作：

第一步：用OLS分别估计

$$\Delta \mathbf{y}_t = \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{y}_{t-j} + \mu_t \tag{6.4.12}$$

中的每一个方程，计算残差，得到残差矩阵$\mathbf{S}_0$，为一个$(M \times T)$阶矩阵。

第二步：用OLS分别估计

$$\mathbf{y}_{t-1} = \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{y}_{t-j} + \mu_t \tag{6.4.13}$$

中的每一个方程，计算残差，得到残差矩阵$\mathbf{S}_1$，也为一个$(M \times T)$阶矩阵。

第三步：构造上述残差矩阵的积矩阵：

$$\mathbf{R}_{00} = T^{-1}\mathbf{S}_0\mathbf{S'}_0 \quad \mathbf{R}_{01} = T^{-1}\mathbf{S}_0\mathbf{S'}_1$$

$$\mathbf{R}_{10} = T^{-1}\mathbf{S}_1\mathbf{S'}_0 \quad \mathbf{R}_{11} = T^{-1}\mathbf{S}_1\mathbf{S'}_1 \tag{6.4.14}$$

第四步：计算$\mathbf{R}_{10}\mathbf{R}_{00}^{-1}\mathbf{R}_{01}$关于$\mathbf{R}_{11}$的有序特征值和特征向量。特征值即为特征方程

$$\left|\lambda\,\mathbf{R}_{11} - \mathbf{R}_{10}\mathbf{R}_{00}^{-1}\mathbf{R}_{01}\right| = 0 \tag{6.4.15}$$

的解，$1 \geq \lambda_1 \geq \cdots \geq \lambda_r \geq \cdots \geq \lambda_M \geq 0$，构成对角矩阵$\Lambda$；对应的特征向量构成的矩阵为$\mathbf{B}$，则有

$$\mathbf{R}_{11}\mathbf{B}\Lambda = \mathbf{R}_{10}\mathbf{R}_{00}^{-1}\mathbf{R}_{01}\mathbf{B} \tag{6.4.16}$$

其中$\mathbf{B}$由下式正规化：

$$\mathbf{B}'\mathbf{R}_{11}\mathbf{B} = \mathbf{I}$$

第五步：设定似然函数。当Π无约束时，(6.4.15)的M个特征值都保留，其对数似然函数依赖于：

$$-\frac{1}{2}T\sum_{i=1}^{M}\ln(1-\lambda_i) \tag{6.4.17}$$

但当$R(\Pi) = r\,(0 < r < M)$时，对数似然函数是r个最大的特征值的函数：

$$-\frac{1}{2}T\sum_{i=1}^{r}\ln(1-\lambda_i) \tag{6.4.18}$$

3.JJ检验之一——特征值轨迹检验

如果r个最大的特征值给出了协整向量，对其余$M-r$个非协整组合来说，$\lambda_{r+1},\cdots,\lambda_M$应该为0。于是设零假设为：$H_r$：有$M-r$个单位根，即有r个协整关系。备择假设为无约束。

检验统计量为：

$$\eta\,(M-r) = -T\sum_{i=r+1}^{M}\ln(1-\lambda_i) \qquad r = 0,1,2,\cdots,M-1 \tag{6.4.19}$$

服从Johansen分布。当$r = 0,1,2,\cdots,M-1$时可以得到一系列统计量值：$\eta\,(M), \eta\,(M-1), \cdots, \eta(1)$。

依次检验这一系列统计量的显著性。

当$\eta\,(M)$不显著时（即$\eta\,(M)$值小于某显著性水平下的Johansen分布临界值），不拒绝H_0（即不拒绝r=0），说明有0个协整向量（即不存在协整关系）；当$\eta\,(M)$显著时（$\eta\,(M)$值大于某显著性水平下的Johansen分布临界值），拒绝H_0而接受H_1，此时至少有1个协整向量，必须接着检验$\eta\,(M-1)$的显著性。

当$\eta\,(M-1)$不显著时（即$\eta\,(M-1)$值小于某显著性水平下的Johansen分布临界值），不拒绝H_1（即不拒绝r=1），说明有1个协整向量（即存在1种协整关系）；当$\eta\,(M-1)$显著时（$\eta\,(M-1)$值大于某显著性水平下

的Johansen分布临界值），拒绝H_1而接受H_2，此时至少有2个协整向量，必须接着检验$\eta(M-2)$的显著性。

...，一直检验下去，直到出现第一个不显著的$\eta(M-r)$为止，说明存在r个协整向量。这r个协整向量就是对应于最大的r个特征值的经过正规化的特征向量。

(6.4.19)的检验统计量被称为特征值轨迹统计量，于是上述检验被称为特征值轨迹检验。特征值轨迹检验临界值见表6.4.1。

4.JJ检验之一——最大特征值检验

另外一个类似的检验的零假设为：H_r：有$M-r$个单位根，即有r个协整关系。备择假设为有$M-r-1$个单位根。检验统计量为基于最大的特征值$\{\lambda_r\}$的：

$$\varsigma(r-1) = -T\ln(1-\lambda_r) \tag{6.4.20}$$

该统计量被称为最大特征值统计量。于是该检验被称为最大特征值检验。

检验从下往上进行，即首先检验统计量$\varsigma(0)$。如果统计量$\varsigma(0)$不显著，即$\varsigma(0)$值小于某显著性水平下的Johansen分布临界值，则不拒绝H_0（即不拒绝r=0），说明有0个协整向量（即不存在协整关系）；如果统计量$\varsigma(0)$显著，即$\varsigma(0)$值大于某显著性水平下的Johansen分布临界值，则拒绝有0个协整向量的H_0，接受至少有1个协整向量的备择假设，必须接着检验$\varsigma(1)$的显著性。

如果统计量$\varsigma(1)$不显著，即$\varsigma(1)$值小于某显著性水平下的Johansen分布临界值，则不拒绝H_0（即不拒绝r=1），说明有1个协整向量；如果统计量$\varsigma(1)$显著，即$\varsigma(1)$值大于某显著性水平下的Johansen分布临界值，则拒绝有1个协整向量的H_0，接受至少有2个协整向量的备择假设，必须接着检验$\varsigma(2)$的显著性。

...，一直检验下去，直到出现第一个不显著的$\varsigma(r-1)$为止，说明存在（r−1）个协整向量，拒绝至少有r个协整向量的备择假设。这（r−1）个协整向量就是对应于最大的（r−1）个特征值的经过正规化的特征向量。

最大特征值检验临界值见表6.4.1。注意临界值的选取与M、r有关。例如，如果M=2，即变量数为2；T=40。求得到的两个特征值（按照从大到小的顺序排列）为：$\lambda_0 = 0.50, \lambda_1 = 0.10$。由(6.4.20)求出的最大统计量为：

$$\varsigma(0) = -40\ln(1-0.50) = 27.73$$
$$\varsigma(1) = -40\ln(1-0.10) = 4.21$$

首先检验统计量$\varsigma(0)$，此时应该选择对应于M−0=2的临界值14.595（给定显著性水平为95%）。因为27.73¿14.595，所以统计量$\varsigma(0)$显著，则拒绝有0个协整向量的H_0，接受至少有1个协整向量的

备择假设，必须接着检验$\varsigma(1)$的显著性。选择对应于M−1=1的临界值8.083（给定显著性水平为95%），因为4.21¡8.083，表示统计量$\varsigma(1)$不显著，则不拒绝H_0（即不拒绝r=1），说明有1个协整向量。

如果M=3，即变量数为3，则在进行统计量$\varsigma(0)$显著性检验时，应该选择对应于M−0=3的临界值21.279（给定显著性水平为95%）。

表6.4.1由Johansen和Juselius于1990年计算得到。

表6.4.1 Johansen分布临界值表

统计量		50%	80%	90%	95%	97.5%	99%	均值
最大特征值	1	2.415	4.905	6.691	8.083	9.658	11.576	3.030
	2	7.474	10.666	12.783	14.595	16.403	18.782	8.030
	3	12.707	16.521	18.959	21.279	23.362	26.154	13.278
	4	17.875	22.341	24.917	27.341	29.599	32.616	18.451
	5	23.132	27.953	30.818	33.262	35.700	38.858	23.680
特征值轨迹	$\eta_\alpha(1)$	2.415	4.095	6.691	8.083	9.658	11.576	3.030
	$\eta_\alpha(2)$	9.335	13.038	15.583	17.844	19.611	21.962	9.879
	$\eta_\alpha(3)$	20.188	24.445	28.436	31.256	34.062	37.291	20.809
	$\eta_\alpha(4)$	34.873	41.623	45.248	48.419	51.801	55.551	35.475
	$\eta_\alpha(5)$	53.373	61.566	65.956	69.977	73.031	77.911	53.949

6.4.4 协整与误差修正模型

Engle 和Granger于1987年在协整与误差修正模型之间建立了同构。向量自回归模型(6.4.8)的每个方程都是一个自回归分布滞后模型，如果$\mathbf{y}_t$所包含的M个$I(1)$过程存在协整关系，(6.4.11)

$$\Delta \mathbf{y}_t = \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{y}_{t-j} + \alpha\beta' \mathbf{y}_{t-1} + \varepsilon_t$$

中的每个方程的新生误差都具有平稳性。一个协整体系有多种表示形式，用误差修正模型(ECM)表示是当前处理这种问题的普遍方法。即

$$\Delta \mathbf{y}_t = \sum_{j=1}^{p} \Gamma_j \Delta \mathbf{y}_{t-j} + \theta\, \mathbf{ECM}_{t-1} + \varepsilon_t \qquad (6.4.21)$$

其中每个方程都是一个误差修正型方程，可以剔除一些不显著的滞后差分项。误差修正项反映变量之间的关系偏离长期均衡状态对短期变化的影响，所有作为解释变量的差分项反映各变量短期变化对作为被解释变量的短期变化的影响。

至此，完成了从反映数据生成过程的自回归分布滞后模型

$$y_t = \beta_0 + \sum_{i=0}^{q} \delta_i z_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \quad \varepsilon_t * N(0, \sigma^2)$$

到典型的动态计量经济学模型—误差修正模型

$$\Delta y_t = \tilde{\beta}_0 + \sum_{i=0}^{q} \tilde{\delta}_i \Delta z_{t-i} + \sum_{i=1}^{p} \tilde{\gamma}_i \Delta y_{t-i} + \alpha\, ecm_{t-1} + \varepsilon_t \quad \varepsilon_t * N(0, \sigma^2) \quad (6.4.22)$$

的变换。

6.5 动态计量经济学模型与经典计量经济学模型的比较

本章的第一、二、三、四节和第二章第五节“协整理论和误差修正模型”一起，介绍了动态计量经济学模型理论方法的主要概念和思路。在本节中，将从建模起点、对一些重要概念的解释和模型的评价准则等几个方面，对动态计量经济学模型和经典计量经济学模型进行概念性的比较。

6.5.1 关于建模起点的比较

以单方程模型为例，经典计量经济学模型的起点是如下的理论模型：

$$y_i = f(\mathbf{X}_i, \mathrm{B}, \mu_i) \qquad i = 1, 2, \cdots, n \quad 6.5.1$$

其中y为被解释变量，$\mathbf{X}$是包含k个变量的解释变量向量，B为待估计参数向量，μ为随机误差项。对于线性模型，其形式为：

$$y_i = \mathbf{X}_i \mathrm{B} + \mu_i \qquad i = 1, 2, \cdots, n \quad 6.5.2$$

从§1.3中已经知道，建模者的主要任务是选择变量、确定变量之间的关系形式，而主要依据是经济理论和对所研究的经济对象行为规律的认识。也就是说，是以理论为导向建立作为模型起点的理论模型。

而动态计量经济学模型的起点是如下形式的反映数据生成过程的自回归分布滞后模型：

$$y_t = \beta_0 + \sum_{i=0}^{q} \delta_i z_{t-i} + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \quad \varepsilon_t * N(0, \sigma^2) \qquad (6.5.1)$$

尽管自回归分布滞后模型中变量的选择，仍然需要从经济理论和对经济行为规律的认识出发，体现了动态计量经济学模型离不开理论导向；但是，由于它必须反映数据生成过程，所以数据在该起点模型建立中的作用是无法忽视的。

二者相比较，经典计量经济学模型以理论为导向，不同的建模者，由于对经济理论的理解不同，对于同一个对象，可以建立不同的作为起点的理论模型；而动态模型是从数据生成过程出发，经过约化得到一个一般的自回归分布滞后模型，那么不同的建模者，对于同一个研究对象，应该得到相同的作为起点的理论模型。

6.5.2 关于模型解释的比较

与经典计量经济学模型相比较，动态模型对模型中许多问题有新的解释。

1.关于变量的解释

经典计量经济学模型将变量分为两类：内生变量和外生变量（在单方程模型中为解释变量和被解释变量），内生变量是随机变量，而外生变量则是确定性变量。

动态计量经济学模型认为所有变量都是随机变量，它们共同具有一个联合分布，即数据生成过程。即使相对于关注参数而言是外生变量，仍然是随机变量。

2.关于误差项的解释

经典计量经济学模型认为，随机误差项是由模型中没有包含的因素对被解释变量的影响、模型设定误差、样本观测误差等内容组成的，并假定它服从正态分布。在估计模型时，还需要对误差作出一些假设，这些假设并不反映误差的性质，很难被满足。

动态计量经济学模型认为误差是推导出来的。由约化理论可知，误差是模型的非系统分量，等于模型(6.5.1)左边的变量过程减去右边的系统分量。因此，它的分布是由变量的分布和系统的分布决定的，它具有的良好性质是由约化过程决定的，不是随便假设的。

3.关于外生性概念的解释

经典计量经济学模型认为，外生性的概念是绝对的（对于已经确定的模型系统而言），是变量本身的性质。外生变量是由系统外的因素决定的，只对系统产生影响，而本身不受系统的影响。

动态计量经济学模型认为外生性的概念是相对的，并不是变量本身的性质。对于已经确定的模型系统而言，模型的外生变量将随着关注参数的变化而变化。

4.关于多重共线性的处理方式

经典计量经济学模型认为多重共线性是变量本身之间所具有的性质，因而处理的方式都是从变量出发的。例如去掉部分产生共线性的变量，用变量的差分形式代替原有变量，用主分量代替原有变量，等等。

动态计量经济学模型认为多重共线性不是变量本身之间所具有的性质，而是由不适当的参数化引起的。因此，经过再参数化，将模型变换为误差修正型模型，可以使得变量之间近似正交，从而消除共线性。

5.关于对变量数据的时间序列特性的重视程度

经典计量经济学模型对于变量数据的时间序列特性不进行研究，容易出现伪回归现象。即将实际上没有影响关系的变量，由于都是时间序列数据而存在数据上的相关性，被错误地引入模型之中。

动态计量经济学模型特别重视研究变量数据的时间序列特性，认为只有变量之间存在协整关系，它们之间才具有长期的相互影响关系，可以有效地避免伪回归现象。

6.关于对理论模型和经验模型的概念的理解

经典计量经济学模型忽略理论模型与经验模型之间的区别，认为二者是内在一致的。理论模型描述了研究对象的理论行为，通过收集样本观测值，估计模型与检验模型，得到描述研究对象实际行为的经验模型，二者必须是一致的。如果不一致，即经验模型不能通过检验，则要修改理论模型。

动态计量经济学模型明确指出，理论模型与经验模型是不同的两个概念。前者由经济理论设定，描述经济活动中的长期均衡关系；后者根据理论模型和观测数据共同设定和估计，描述经济活动的短期非均衡性质。二者并不一定是一致的。经验模型通过其长期均衡解与理论模型发生联系。因为动态模型认为，经济理论描述的是理论变量之间的长期均衡关系，而观测数据往往与理论变量是不一致的，所反映的是短期非均衡行为。

6.5.3 关于模型检验的比较

任何模型，在进入应用之前，必须通过检验。经典计量经济学模型与动态计量经济学模型在模型检验的理论与实践方面都存在较大的差异。

1.关于模型检验的指导思想

经典计量经济学模型认为，每个模型都可能是正确的，因而只需要对模型的几个性质进行检验，如果这些检验通过，则接受这个模型。所以在经典计量经济学模型的应用实践中，经常出现这样的现象：同样的研究对象，不同的研究者，依据不同的经济理论建立模型，最后都声称模型通过了检验，理论得到了证实，而这些理论却是不同的，甚至是相互矛盾的。

而动态计量经济学模型则认为，所有模型都可能是错误的，因而必须经过严格的全面的检验，Hendry本人就曾经一再强调，“检验，检验，再检验。”为此，提出了许多准则，要求一开始就建立一个通过所有检验、满足所有准则的“吻合性”模型，以此为基础进行简化。于是，对于同样的研究对象，尽管有不同的研究者，但是在建立的模型中能够被称为“吻合性”模型的只能是一个。

2.关于模型检验的准则

经典计量经济学模型认为，一个成功的模型，必须很好地拟合样本观测值。所以，拟合优度成为最重要的检验准则。关于这个方面，读者已经很熟悉了，不再赘述。

动态计量经济学模型则认为，拟合优度并不是一条好的准则。因为增加更多的"解释"变量，总是可以得到一个高的拟合优度，而且某些实际上是"伪回归"的模型，也可以有很高的拟合优度。动态模型提出了如下模型检验准则：

⑴包容性准则

动态计量经济学模型认为，一个满意的模型必须包容所有从同一数据生成过程约化而来的模型。因为所有模型都是数据生成过程的一种简化，在从数据生成过程到最终模型的约化过程中，会产生各种不同的模型，但是如果该最终模型的每步约化都是有效的，即没有关于关注参数的信息损失，那么它应该包容约化过程中的每一个模型，直致包容数据生成过程。

⑵凝聚性准则

动态计量经济学模型认为，一个满意的模型必须具备数据的凝聚性。即利用模型计算的数据序列$\hat{y}_t$与实际序列y_t之间的差应该是服从均值为0 的正态分布的随机误差。这是由新生化的要求和新生误差的性质所决定的。

⑶因果性准则

动态计量经济学模型认为，一个满意的模型必须包含明确的经济行为上的因果关系。因为数据生成过程是所有变量联合概率分布的一般表达式，所以经过约化的最终模型除了包含变量y之外，还应该包含变量集$\mathbf{Z}$，而$\mathbf{Z}$是序列y_t数据过程的条件和原因。

⑷参数不变性准则

动态计量经济学模型认为，一个满意的模型必须具备参数的不变性。这既是模型可以用于预测和政策分析的必备条件，也是模型约化过程所应该保证的。

⑸长期均衡方程与理论的一致性准则

虽然动态计量经济学模型认为理论模型与经验模型是不同的两个概念，但是，它还认为经验模型通过其长期均衡解与理论模型发生联系。这就是说，长期均衡方程与经济理论之间应该具有一致性，如果存在多种理论，至少必须与某一种理论保持一致。

针对这些准则，设计了相应的检验方法和相应的应用软件，在一些专门的教科书中有系统的介绍。例如，根据凝聚性准则所要求的对模型残差的检验，是必须进行的一项检验。其中主要的是残差的正态性检验，使用的统计量是描述正态性的偏斜度的三阶矩统计量SK：

$$SK = \frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})/(\frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})^2)^{1.5}$$

和描述正态性的峰度的四阶矩统计量EK：

$$EK = [\frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})^4/(\frac{1}{T-1}\sum_{t=1}^{T}(y_t-\bar{y})^2)^2]^{-3}$$

由此构成检验正态性的统计量：

$$\chi^2 = \frac{(T-k)}{6}(SK^2+\frac{1}{4}EK^2)$$

其中k为模型中变量的数目。在正态假设下，该统计量服从自由度为k的χ^2分布。给定显著性水平，查得临界值，如果计算得到的统计量值小于临界值，接受正态分布的假设。

专门用于动态计量经济学模型估计与检验的软件包有PC-GIVE6.0及其改进版本，例如PC-GIVE9.0、PC-FIML（用于联立动态模型的估计）等。PC-GIVE是由D.F.Hendry发展的计量经济学软件包，除了具有一般的计量经济学模型和时间序列分析模型的估计与检验功能外，还为动态模型的估计与检验设计了专门功能。例如，在该软件中，上述检验残差正态性的统计量值可以被自动计算；如果用该软件估计长期均衡方程，用于进行方程显著性检验的WALD检验统计量也能够被自动计算。

6.6 本章思考题和综合练习题

思考题

1.经典计量经济学模型基本设定理论和设定方法的要点。

2.动态计量经济学模型基本设定理论和设定方法的要点。

3.实际经济研究中分布滞后模型的普遍性。

4.多项式分布滞后模型中"多项式"的含义是什么？估计中的主要问题及如何解决？

5.几何分布滞后模型为什么一般采用工具变量法和广义差分法相结合的方法进行估计？其主要步骤是什么？

6.自回归分布滞后模型的一般形式及主要优点是什么？

7.向量自回归模型表达式的含义。

8.数据生成过程的含义及表达式是什么？

9.在动态计量经济学模型中变量外生性的意义与经典计量经济学模型有什么不同？

10.关于弱外生性的条件。

11.约化过程的主要内容和目的是什么？在实际建立动态计量经济学模型时如何保证约化过程的有效性？

12.多重协整检验JJ方法的基本原理和工作步骤。

13.误差修正模型的形式、经济学上的意义以及其主要优点。

15.动态计量经济学模型与经典计量经济学模型在模型解释和模型检验方面的主要区别。

二、综合练习题

1.建立一个含有多重协整关系的实际动态计量经济学模型。包括：

⑴模型变量的选择；

⑵变量滞后阶数的确定；

⑶变量的单整检验；

⑷多重协整检验

⑸长期均衡方程的估计；

⑹误差修正模型的估计；

⑺利用模型进行经济分析。

附录A　Gauss使用说明

A.1　简介

Gauss是一种高级计算机语言程序，适用于解决数学和矩阵等方面的问题，可以用来解决多种数学和计量经济模型。它是多功能的，但是与其他的计量经济学应用程序如SHAZAM和TSP相比，学习起来相对较困难。在初学时，最好是先学习《GAUSS: A beginner's guide》，在运用时可以参考在线帮助。

Gausss是一种很有特色的语言程序，它的每一个变量是矩阵。它能知道矩阵的大小和内容，并能进行矩阵运算。在许多经济学和计量经济学研究中，需要进行含有矩阵的计算，Gauss是一种非常有效的工具。

矩阵的操作说明

Gauss程序的每一行结束时必须用分号"；"。Gauss程序中的表达式：C=A+B的含义是：如果A与B都是一个数，即单值矩阵（1×1矩阵），那么C也是一个单数值矩阵且等于它们的和。如果A和B是具有同样长度的行向量，则C也为同样长度的行向量，且每一个元素等于A和B中相应元素之和。如果A和B都是3×4矩阵，则C也是，且每一元素等于A和B相应元素之和。符号"+"代表矩阵的加法，如果A和B的大小不一样，则Gauss将出现提示，如：

: error G0036 : matrices are not conformable

类似的规则适用于"—"和"*"等运算。

赋值语句

GAUSS对变量赋予一个表达式的值的形式与许多语言程序相同。变量名称在等号左边，表达式在右边。在GAUSS中没有必要在赋值前先用语句说明变量。若一个变量在前面已赋值,则新的值将覆盖前面的值。如：A和B 都是20x30的矩阵，且

C = A + B;

则C 也是20x30的矩阵，它的内部元素是A 和B中相应元素之和。如果C已经存在，是一个20x15的矩阵，则它将自动被20x30的矩阵代替。所以在GAUSS中矩阵变量在大小上是不固定的，不象其它应用程序。

变量名称

GAUSS不能区别字母的大小写，除非用双引号包括。一个变量可

以大到32位，包括字母、数字和横杠。第一个位置必须是字母或横杠。变量PersonalDisposableIncome 与personaldisposableincome是相同的。应用长名称的变量容易阅读，但有时易出错。如果输入有误，则GAUSS将值赋予一个新的变量，而不是所需赋值的变量。

即时和延迟执行

GAUSS启动后进入一个互动环境，即运行窗口（command mode）。进入后，按回车键，GAUSS立即运行。一般情况下，还需要编制一些应用程序。编制程序在编辑窗口，编好程序后，键入

run do_it;

则立即进入运行窗口，运行程序，显示结果。如果结果不满意，可以再进入编辑窗口，修改程序再运行。

显示变量

如果想看某一变量的值，只需键如它的名称。如果这个变量程序中没给定，则显示：

Undefined symbols: ...

GAUSS一般不显示变量的赋值，除非使用者需要。如；若输入

C = A + B;

则不会出现C的值，但如果输入

C;

或print C;

则GAUSS 显示C的值。如果想看A + B的结果，输入

A + B;

如果一个表达式没有任何等号，则被看作一个print语句。如果没有简洁的要求，GAUSS显示整数后8位。

建立矩阵

如果矩阵很小，可以直接输入。如：

a = 3;
b = {1 2 3};
c = {4 , 5 , 6};
d = {1 2 3 , 4 5 6};

这里，a是一个数字，b是一个1x3行向量，c是一个3x1列向量，d是一个2x3矩阵。这样，输入

d;

则出现

1.0000000 2.0000000 3.0000000
4.0000000 5.0000000 6.0000000

在GAUSS中，用空格分开的数字在同一行；用逗号分开的数字在不同的行；所有数字用大括号包括。

从矩阵生成矩阵

一般情况下建立矩阵可以用矩阵元素的运算方法。如：

a = {1 2 3};

b = {4 5 6};
c = a~b;
print c;

则得

1.0000000 2.0000000 3.0000000 4.0000000 5.0000000 6.0000000

而d = a|b;
print d;

则得到

1.0000000 2.0000000 3.0000000
4.0000000 5.0000000 6.0000000

矩阵可以通过这种方法合并，但必须相协调。如不协调，则有提示。如：
c = {a b};

或d = {a,b};

则提示：

(0): error G0008 : Syntax error

使用矩阵的一部分

如果仅用矩阵的一部分，在矩阵名称后用中括号表示，中括号内表示所需要的部分。如：
d[1,2];

显示：

2.0000000

d[2,1];

显示：

4.0000000

d[1,.];

显示：

1.0000000 2.0000000 3.0000000

d[.,2];

显示

2.0000000
5.0000000

d[2, 2:3];

显示

5.0000000 6.0000000

d[2, 3:2];

显示

6.0000000 5.0000000

也可以用变量的形式表示。如：
z = {2,3};
d[2,z];

显示

5.0000000 6.0000000

d[1:2, 2:3];

显示

2.0000000 3.0000000

5.0000000 6.0000000

文字表示式

GAUSS变量有两种：数字和文字。建立一个文字变量后用双引号包括。如：

stg = "This is a string";

print stg;

显示一个文字变量矩阵，矩阵名称前必须有符号"$"。如：

x = {"ab" , "cd"};

print $x;

显示：

ab

cd

x = {"ab" "cd"};

print $x;

显示ab cd

矩阵和排列操作（Matrix and Array Operations）

矩阵操作指的是矩阵的乘法；而排列操作是指矩阵元素之间的运算。

矩阵操作（Matrix Operations）

矩阵的转置只需在矩阵的名称的右上角加上一撇。如：

x = {1 2 3};

print x';

显示

1.0000000

2.0000000

3.0000000

矩阵的加法用(+)号，减法用(-)号。如果A 是3x4矩阵，B是4x3矩阵，则

C = A + B;

显示

(0) : error G0036 : matrices are not conformable

C = A + B';

将得到一个新矩阵。

矩阵的乘法采用星号(*)。两个运算对象的内部大小必须相协调，除了一个矩阵是单值矩阵。

排列运作（Array Operations）

排列操作指的是元素之间的乘法、除法和乘方。在标准乘法符号前有一点（.）。如：

x = {1 2 3};
y = {4 5 6};
x.*y;

显示

4.0000000 10.000000 18.000000

对于除法，同样使用。如：

C = A ./ B;

对于乘方，原理同样。如：

x = {1 2 3};
x. 2;

显示

1.0000000 4.0000000 9.0000000

GAUSS的排列操作包括乘法(.*)、除法(./)和乘方(.)。

矩阵的逻辑和关系运算

GAUSS有六种关系运算：

¡ or LT :小于

¡= or LE : 小于或等于

¿ or GT :大于

¿= or GE : 大于或等于

== or EQ : 等于

/= or NE : 不等于

等于数字与文字变量间比较，这些符号后应加上”$”。

注意：A==B 表示A 等于B；A=B 表示将B的值赋予A。前者是逻辑关系，后者是赋值语句。

当GAUSS进行关系运算时，如果表达式正确，其值为1；如果表达式不正确，其值为0。如：

x = 1 ¡ 3;

显示

1.0000000

x = 1 ¿ 3;

显示

0.0000000

关系运算适用于元素之间协调的矩阵。对于两个矩阵的元素之间的比较，运算符号后必须有”.”。若没有”.”，则比较两个矩阵中的所有元素，其结果总是1或0。若有”.”，运算按各个元素之间的比较结果得到一个矩阵，在正确的地方为1，在不正确的地方为0。如：

A = {1 2, 3 4};

A ¿ 2;

显示0.0000000

A .¿ 2;
显示
0.0000000 0.0000000
1.0000000 1.0000000
同样
A = {1 2, 3 4};
B = {3 1, 2 2};
A ¿ B;
显示
0.0000000
A .¿ B;
显示
0.0000000 1.0000000
1.0000000 1.0000000
经常使用的逻辑运算有：
not
and
or

如果逻辑运算符号后有”.”，结果是由1和0组成的矩阵。0或1是基于两个矩阵的元素之间的逻辑关系的比较，正确时值为1；不正确时值为0。

控制执行语句（Controlling Execution Flow）

控制执行语句（Controlling Execution Flow）包括循环语句（Do Loops ）和条件语句（If statements）。

循环语句（Do Loops）

有两种循环语句：do while语句和do until语句。其区别在于：在条件满足时，前者继续进行循环运行，而后者是在条件不满足时进行循环运行。循环语句以endo语句结束。

一个do while循环是在条件满足时，执行循环。

如：
do while x ¡= 0.5;
....
endo;

同样，一个do unti循环重要条件不满足，就执行循环。如：
do until x ¿ 0.5;
....
endo;

在循环语句中，break 和continue用于控制程序的执行。若遇上break ，程序将跳出循环执行endo下面的程序；若遇上continue，程序将跳到循环的开始，并评估do while 和do until表达式。若条件始终不满足，程序将反复循环执行。

如果do while (or do until) 在条件不满足（满足）时，停止执行循环是很重要的。否则，将进入一个无限的循环，直至关机。

在Do Loop中若设置一个可以自动增加的记值变量，开始进入时设为1，则它在循环程序中不断增加。这种变量是很重要的。如：

true = 1==1;
false = 1==0;
.....
done = false;
do while not done;
....
endo;

当然，有时设定done等于true。

条件语句（If Statements）

如果条件成立，条件语句执行确定的程序；如果不成立，执行其它程序。复杂一些的条件语句含有elseif 和else语句，但结果总以endif语句结束。如：

if x ¿ 0.5;
....
elseif x ¿ 0;
....
else;
....
endif;

简单的条件语句省略了elseif部分。如：

if x ¿ 0.5;
....
else;
....
endif;

更简单的形式省略了else 部分。如：

If x ¿ 0.5;
....
endif;

重叠（Nesting）

所有这些控制语句允许重叠，这样，一种结构处于另一种结构之中。如：

j = 1;
do until j ¿ n;
k = 1;
do while k ¡= n;
if x[j,k] ¿ 0.5;

```
x[j,k] = 1.5;
endif;
k=k+1;
endo;
j=j+1;
endo;
```

数据的输入和输出（Data Input and Output）

数据输入（Data Input）

最直接的方法是在运行窗口（command mode）直接输入，这对少量数据很有用。如：

```
prices = {12.50 37.875 12.25};
assets = {"cash", "bonds", "stocks"};
holdings = {100 200,
    300 400,
    500 600};
```

对于大量数据，最好建立一个数据文件，并在编辑窗口输入数据。在这种文件中，具有数字文本的特征。每一行、一列中元素要分开。如果在一个文件中输入30x7矩阵中的每一个元素，取名为test.tx。为将数据引入GAUSS环境，可用语句：load x[30,7] = test.txt;

数据输出（Data Output）

最简单的方法是用print语句。如：

```
print x;
```

还可用format语句来控制输出的数据和矩阵的形式。在GAUSS中，printfm能用不同的形式输出矩阵的每一行或列。若需要得到屏幕上显示的所有结果，可用语句：

```
output file = [filename] [on or reset];
```

这里[filename]代表一个新的文件名，它储存数据输出结果。On是打开这个文件。若文件已存在，它将打开并补充；若不存在，它将建立。Reset类似于on，除非建立一个新文件。若已存在该文件，它将破坏并以该文件名重建新文件。

当通过直接输出时，应加上一个语句：

```
output off;
```

如：

```
output file = test.out reset;
print x;
output off;
```

将在文件名为test.out的文件中输出名为x的矩阵。

将结果输出到打印机很容易，如：

```
output file = lpt1 reset;
```

注意：如果在编辑窗口写程序文本，在结束程序时应写语句：end;

评价行（Comment Lines）

在程序中写入评价行，是很有帮助的。评价写在一对符号(@)之间或/* 和*/之间。符号内所有内容，GAUSS都忽略。评价行可以超过一行。两者的区别如下：

/* This kind of

/* comment */

can be nested */

@ This kind of comment can not be nested @

多赋值语句

多赋值语句的功能如下：

output1=functionName(input1);

{output1,output2,...}=functionName(input1,input2,...);

数据生成功能

下列功能对建立一个新的矩阵有关。如：

Ones matrix

ones(2,4);

Zeros matrix

zeros(4,4);

Identity matrix

eye(3);

Uniform random numbers

rndu(6,3);

Normal random numbers

rndn(6,3);

Additive sequence

seqa(0,0.1,10);

Multiplicative sequence

seqm(2,2,10);

改变一个已存在的矩阵为大小不同的新矩阵，用reshape功能。如：

x=seqa(1,1,5);

print x;

y=reshape(x,5,5);

print y;

建立一个附矩阵（sub-matrix），用selif 和delif功能。如：

x=rndn(100,4);

y=selif(x, x[.,1] .¿ 0.5);

print y;

同样

y=delif(x, x[.,1] .¡= 0.5);

print y;

其它有用的功能还有：

vec

将矩阵的列变为列向量

vech

将矩阵的下三角部分变为一个列向量

xpnd

将一个列向量扩展为一个对称矩阵

submat

减去一个附矩阵

diag

一个矩阵的对角元素

diagrv

取代一个矩阵的对角元素

矩阵描述功能

为描述一个矩阵，下列功能需要应用。

一个矩阵的行的数量

```
x=rndu(10,4);
rows(x);
```

一个矩阵的列的数量

```
cols(x);
```

一个矩阵每一列的最大元素

```
maxc(x);
```

一个矩阵每一列的最小元素

```
minc(x);
```

找出一个矩阵的最大最小值

```
maxc(maxc(x));
minc(minc(x));
```

还有许多功能，如：

```
x = {1, 2, 3};
y = sumc(x) + 10;
print y;
```

sumc是计算一个矩阵每一列的和

显示

16.000000

如：

```
x = {1 2 3, 4 5 6};
sumc(x);
```

显示

5.0000000
7.0000000
9.0000000

计算一个矩阵每一列的累计和，用cumsumc功能。如：

```
cumsumc(x);
```

类似的还有：

prodc

计算一个矩阵每一列所有元素的产量

cumprodc

计算一个矩阵每一列所有元素的累计产量

meanc

计算一个矩阵每一列的平均值

median

计算一个矩阵每一列的中位数c

stdc

计算一个矩阵每一列标准偏差。

矩阵的排序功能（Matrix Sorting Functions）

sortc

将一个矩阵中每一行的元素按从大到小排序并得出矩阵。

sorthc

将一个矩阵中每一行的元素按从小到大排序并得出矩阵

基本的矩阵计算

下列功能与基本的矩阵计算有关

det

一个方阵的行列式

inv

一个方阵的逆阵

invpd

i一个对称、正定方阵的逆阵

corrx

计算相关矩阵

vcx

计算方差-协方差矩阵

cond

一个矩阵的条件数

rank

一个矩阵的秩

解决线性方程系统

若A 为一个nxn矩阵，b是一个nx1 (or nxn)矩阵，A*x = b,则x与b同样的大小。若A为非齐次矩阵，则$X = A^{-1}b$。如：

a = { 6 8, -2 4};

b = (2, 1};

x = inv(a)*b;

print x;

若矩阵A是对称正定的，运用GAUSS功能solpd可以解出x。如：

a = {40 40,

40 72}; @ a is symmetric positive definite @
b = {2, 1};
x = solpd(b,a);
print x;

这里，如果A是nxk矩阵且n ¿= k，从A*x = b解出x等同于从(A'A)*x = (A'*b)中解出x。即：x = invpd(A'A)*(A'b) 用solpd功能如下：

a = { 6 8, -2 4};
b = (2, 1};
x = solpd(a'b,a'a);
print x;

用GAUSS的除法"/"可求出A*x = b的最小二乘答案。

x = b/a;
print x;

solpd的另一个功能是求出对称正定矩阵的逆阵，它等同于invpd。

x = solpd(eye(2),a'a);
print x;

特征根和特征向量（Characteristic Roots and Vectors）

给定一方阵A，从特征方程(A - lI) x = 0求出x有两步：

第一，解|A - lI| = 0求出特征根l；第二，解(A - lI) x = 0求出特征向量x。假定A是对称矩阵，两个GAUSS功能用于计算对称矩阵的特征根和特征向量。

eigrs

计算对称矩阵的特征根。

eigrs2

计算对称矩阵的特征根和特征向量。如：

a = { 41 -23, -23 13}; @ a is real and symmetric @
r = eigrs(a);
{r, v} = eigrs2(a);
print r~v;

注意eigrs2有两个值，一个是特征根的向量，一个是特征向量的矩阵。结果列在等号的左边，用大括号包括。特征根的一个用途是一个数字矩阵的条件数，这是一个最小二乘回归模型多元共线性的标志。

令X是回归的数据矩阵，X的条件数计算如下：

/* assuming the data matrix X is available ... */
xx = X'X; @ xx is real symmetric @
r = eigrs(xx); @ eigenvalues of xx @
cn = sqrt(maxc(r)./minc(r)); @ condition number of X @
print cn;

与GAUSS 的在线功能cond相比，同样结果是:

print cond(x);

没有列出，但很有用的功能有：

pdfn 标准正态概率密度函数
cdfn 正态分布的积分-低尾部分
cdftc t-分布的cdf余值
cdffc F-分布的cdf余值
cdfchic chi-square分布的cdf余值
gradp 一个函数的一解导数
hessp 一个函数的二解导数
intsimp一个函数用Simpson方法的积分
ols 最小二乘回归
eqsolve 解决非线性方程系统
sqpsolve 用连续二次规划方法解决非线性规划问题

以上这些功能在GAUSS在线帮助中都可以找到。

自定义功能

GAUSS最大的优势是使用者可以利用自定义功能来提高该语言程序的能力。在GAUSS中有两类自定义功能：单语句功能可用fn语句，多语句功能可用一组GAUSS语句。最好是建立一个新的程序段以备将来使用。段落程序用proc语句。单语句包括一个语句，而多语句包括多个语句。一旦一个自定义的段落程序写好，可以放入library中，这为使用者的GAUSS文本的将来应用。

单语句功能以fn 定义，其后为文件名，涉及的变量在括号中，功能的内容在等号的右边。如：

fn value(p, q) = p'*q;

这个函数有两个变量p和q，其结果可能为一个数字、向量和矩阵。当然，进行矩阵乘法式，p和q必须相协调。一个更复杂的多语句段落程序能检查p和q的矩阵大小，并用转置，求得结果。

注意：这里的变量和结果的名称都是严格的局部变量，仅用于这一局部。在整个程序中，必须有一语句用这一变量。如：

cost = value(price,quantity);

GAUSS中的一个**段落程序**是使用者的自定义功能，它多于一行，且较复杂。GAUSS中的在线功能和使用者自定义的功能均可应用该段落程序。程序中有全局变量和局部变量。一个GAUSS段落程序的基本结果如下：

Proc语句：定义段落程序
local 语句：定义局部变量
段落程序的主体
retp语句：跳出段落程序
endp 语句：结束段落程序

程序中可以不只一个local和retp语句，但不能重叠。

A.2 关于在GAUSS程序中应用极大似然法估计的说明

在GAUSS程序中应用极大似然法估计模型，必需在GAUSS程序中装入软件包NLOPT(non-linear optimization package for GAUSS 3.2)。在经济管理学院的计算机机房里的计算机上都已经在GAUSS程序中装入了nlopt和lsq两软件包。

应用极大似然法估计模型，在编制程序时，第一行必须是语句：
use nlopt;

接着，应是数据输入语句和运行结果输出语句。数据输入语句一般是直接从a盘或c盘的某一文本文件直接读入。结果输出也一般直接输出到a盘或c的某一个文件。

接着是定义余值函数和Jacobian函数。注意：定义时，需要用到局部变量，每一个局部变量仅用数据矩阵的一列；自定义的段落程序应以endp结束。

接着是reset语句。由于nlopt是以线性迭代的形式，故该语句必不可少。

接着是对估计的要求语句。这要用到nlopt的功能语句。下面具体说明如下：

Table 1. Global Control Variables in NLOPT

Variable Description Default

_method MAXIMIZE and MINIMIZE method: 0
0 = Steepest-ascent (descent) method
1 = Quasi-Newton BFGS update method
2 = Quasi-Newton DFP update method
3 = Greenstadt method
4 = Newton-Raphson method
5 = Quadratic hill-climbing method
6 = Modified quadratic hill-climbing method
NLSQ estimation method:
0 = Gauss-Newton method
1 = Gauss-Newton BFGS update method
2 = Gauss-Newton DFP update method
3 = Levenberg-Marquardt method
4 = Newton-Raphson method
MAXLIK estimation method:

0 = Berndt-Hall-Hall-Hausman (BHHH) method
1 = Quasi-Newton BFGS update method
2 = Quasi-Newton DFP update method
3 = Greenstadt method
4 = Newton-Raphson method
5 = Quadratic hill-climbing method
6 = Modified quadratic hill-climbing method
_search Line search option 0
0 = bi-directional line search
1 = linear line search
2 = quadratic line search
_iter Maximum number of iterations 1
_tol Tolerance 0.001
_conv Convergence criteria: 0
0 = Convergence in function value only
1 = Convergence in function value and solution
2 = Convergence in function value, solution, and gradient
All convergence criteria are checked with _tol
_dn Divisor for variance and covariance matrix 0
Used in NLSQ estimation methods only
0 = Divisor N-K is used
1 = Divisor N is used
_vcov Report Variance-Covariance matrix 0
_riter Iteration limit for R-value 10
Used in quadratic hill-climbing methods only

接着，应用赋值语句赋予变量的初始值。另外，在程序中应有要求极大似然估计的语句：

call maxlik(...);

程序的结束应以语句：

end;

如果结果不满意，可以通过改变初始值、迭代次数、方法种类等重新拟合。

操作方法是：启动计算机的Window后，在屏幕上有GAUSS程序的图标，双击图标，就启动了GAUSS；也可以从左下端“开始”的菜单中选“程序”，在“程序”中选GAUSS。

启动GAUSS后进入运行窗口。这时，从上面的“窗口”菜单中，选择“编辑窗口”，即进入了编辑窗口。或在运行窗口的上栏输入文件名，按下栏的箭头，即进入编辑窗口。在编辑窗口编辑程序时，首先在上栏输入并确定文件名，再编写程序。程序编好后，按“run”键，GAUSS运行。运行时，GAUSS首先检查使用者的程序。如果有错误，就提示出来。如果没

有错误，就运行并在运行窗口显示结果。如果结果不满意，按“edit”键，又回到编辑窗口，修改程序。如此反复试算，直至满意。

A.3 GAUSS应用实例

数据输入

下面用GAUSS程序及相应软件包估计C-D 函数。数据按如下方式输入，并存入a盘或c盘。

Data File : CJX.TXT

```
YEAR X L1 L2 K1 K2
1929 189.8 173.3 44.151 87.8 888.9
1930 172.1 165.4 41.898 87.8 904.0
1931 159.1 158.2 36.948 84.0 900.2
1932 135.6 141.7 35.686 78.3 883.6
1933 132.0 141.6 35.533 76.6 851.4
1934 141.8 148.0 37.854 76.0 823.7
1935 153.9 154.4 39.014 77.7 805.3
1936 171.5 163.5 40.765 79.1 800.4
1937 183.0 172.0 42.484 80.0 805.5
1938 173.2 161.5 40.039 77.6 817.6
1939 188.5 168.6 41.443 81.4 809.8
1940 205.5 176.5 43.149 87.0 814.1
1941 236.0 192.4 46.576 96.2 830.3
1942 257.8 205.1 49.010 104.4 857.9
1943 277.5 210.1 49.695 110.0 851.4
1944 291.1 208.8 48.668 107.8 834.6
1945 284.5 202.1 47.136 102.1 819.3
1946 274.0 213.4 49.950 97.2 812.3
1947 279.9 223.6 52.350 105.9 851.3
1948 297.6 228.2 53.336 113.0 888.3
1949 297.7 221.3 51.469 114.9 934.6
1950 328.9 228.8 52.972 124.1 964.6
1951 351.4 239.0 55.101 134.5 1021.4
1952 360.4 241.7 55.385 139.7 1068.5
1953 378.9 245.2 56.226 147.4 1100.3
1954 375.8 237.4 54.387 148.9 1134.6
1955 406.7 245.9 55.718 158.6 1163.2
1956 416.3 251.6 56.770 167.1 1213.9
1957 422.8 251.5 56.809 171.9 1255.5
1958 418.4 245.1 55.023 173.1 1287.9
```

1959 445.7 254.9 56.215 182.5 1305.8
1960 457.3 259.6 56.743 189.0 1341.4
1961 466.3 258.1 56.211 194.1 1373.9
1962 495.3 264.6 57.078 202.3 1399.1
1963 515.5 268.5 57.540 205.4 1436.7
1964 544.1 275.4 58.508 215.9 1477.8
1965 579.2 285.3 60.055 225.0 1524.4
1966 615.6 297.4 62.130 236.2 1582.2
1967 631.1 305.0 63.162 247.9 1645.3

应用lsq软件

应用lsq软件估计C-D生产函数，编制的程序为：

```
/*
** estimating a C-D Production Function
** using lsq package
*/
use lsq;
output file = a:\lesson1.out reset;
load data[40,6] = a:\cjx.txt;
year = data[2:40,1];
X = ln(data[2:40,2]);
L = ln(data[2:40,3]);
K = ln(data[2:40,5]);
names = {"X","L","K"};
call reset;
call estimate(X,L~K,names);
end;
```

应用最小二乘法的方法运行的结果为：

```
Least Squares Estimation
————————————————
Dependent Variable = X
Estimation Range = 1 39
Number of Observations = 39
Mean of Dependent Variable = 5.6874
Standard Error of Dependent Variable = 0.46096
R-Square = 0.99463 R-Square Adjusted = 0.99433
Standard Error of the Estimate = 0.034714
Log-Likelihood Function Value = 77.286
Log Ammemiya Prediction Criterion (APC) = -6.6471
Log Akaike Information Criterion (AIC) = -6.6474
Log Schwarz Bayesian Information Criterion (BIC) = -6.5195
Sum of Squares SS DF MSS F Prob¿F
```

Explained 8.0310 2 4.0155 3332.2 1.3921E-041
Residual 0.043382 36 0.0012051
Total 8.0744 38 0.21248
Variable Estimated Standard t-Ratio Prob Partial
Name Coefficient Error 36 DF ¿|t| Regression
L 1.4508 0.083228 17.431 3.9260E-019 0.89407
K 0.38381 0.048018 7.9930 1.7130E-009 0.63960
CONSTANT -3.9377 0.23700 -16.615 1.8332E-018 0.88464

估计的模型为：

ln(X)	=	-3.94	+	1.45 ln(L)	+	0.38 ln(K)

应用NLOPT软件

应用NLOPT软件编制的程序为：

```
/*
** Estimating a C-DProduction Function
** Using NLOPT Package: MAXLIK
*/
use nlopt;
output file = a:\lesson2.out reset;
load data[40,6]=a:\cjx.txt;
proc ces(data,b); @ residual function @
local l,k,q;
l=data[.,3];
k=data[.,5];
q=data[.,2];
retp(ln(q)-b[1]-b[2]*ln(l)-b[3]*ln(k));
endp;
proc jcf(data,b); @ jacobian function @
local e,n;
e=ces(data,b);
n=rows(e);
retp(ones(n,1));
endp;
call reset;
    _iter=100;
    _tol=1.0e-5;
    _conv=1;
    _vcov=1;
b0={-4.0,1.5,0.5};
    _method=5;
```

```
    _jacob=&jcf;
call maxlik(&ces,data,b0);
end;
```

应用极大似然法估计的结果为：

Iterations = 100 Evaluations = 98960

Sum of Squares = 2.1225

Log Likelihood = 1.9685

Gradient of Log Likelihood =

7.1791 -811.43 -787.83

Parameter Variance

0.30446 -0.023743

0.96346 -0.083277

0.043801 -0.083436

Asymptotic Variance-Covaraiance Matrix

-0.023743 0.045731 -0.045773

0.045731 -0.083277 0.083356

-0.045773 0.083356 -0.083436

用最小二乘法和极大似然法估计CES生产函数

用最小二乘法和极大似然法估计CES生产函数的程序

```
    /*
    ** Estimating a CES Production Function
    ** Using NLOPT Package: NLSQ and MAXLIK
    */
use nlopt;
load x[30,3]=a:\judge.txt;
proc ces(data,b); @ residual function @
local l,k,q;
l=data[.,1];
k=data[.,2];
q=data[.,3];
retp(ln(q)-b[1]-b[4]*ln(b[2]*l b[3]+(1-b[2])*k b[3]));
endp;
proc jcf(data,b); @ jacobian function @
local e,n;
e=ces(data,b);
n=rows(e);
retp(ones(n,1));
endp;
call reset;
    _iter=100;
    _tol=1.0e-5;
```

```
    _conv=1;
    _vcov=1;
    _method=0;
b0={1.0,0.5,-1.0,-1.0};
call nlsq(&ces,x,b0);
    _method=5;
    _jacob=&jcf;
call maxlik(&ces,x,b0);
end;
```