

应用计量经济学

(原书第5版)

Introductory Econometrics With Applications (5th edition)

(美)

拉姆·拉玛纳山 (Ramu Ramanathan)

加利福尼亚州立大学

著

薛菁睿 译



机械工业出版社
China Machine Press

经济教材译丛

应用计量经济学

(原书第5版)

Introductory Econometrics With Applications (5th edition)

(美)

拉姆·拉玛纳山 (Ramu Ramanathan)

加利福尼亚州立大学

著

薛菁睿 译



机械工业出版社
China Machine Press

很荣幸能够有机会翻译《应用经济计量学》这本书,也非常高兴能够借助译者序这个版面来和广大读者分享一下我的翻译心得。

1. 本书是经济计量学的理论、实际应用和示范的完美结合。经济计量学就是在经济学中应用统计工具的科学。具体而言,就是利用数学方法,根据实际统计资料,为经济理论中阐述的经济关系计量出实际数值,以使用计量结果验证或改进经济理论的文字阐述,并进一步解释过去、预测未来和规划政策。经济计量学因此常被视为经济理论、数学和统计的结合。而《应用经济计量学》并没有纯粹地进行枯燥的理论讲解,而是通过一个个实例把理论和实践生动地结合在一起,能够激发读者的兴趣和发挥读者的想像力,让我们真正地了解经济计量学的魅力所在,让我们体会到它并不是高高在上的,而是实实在在为我们的社会服务的一门优秀学科。

2. 书中提供了大量注释计算机输出内容,以便学生确切了解经济计量方法在实践中的应用。理论最终要服务于现实。实际使用情况的具体介绍和展示有助于读者建立对这门学科的感性认识。

3. 书中收录了大量有关中途失败和意外结果的实例,提供了如何使用诊断检验改善模型设定和估计的建议。任何学科的实践过程都不是一帆风顺的,因此,在学习过程中多了解一些可能出现的失败情况和意外情况,无论对于我们的实践指导,还是心理准备方面,都是大有裨益的。

4. 完整的数学、概率和统计的背景信息介绍,方便读者适时回顾和查看。

《应用经济计量学》一书的作者拉姆·拉玛纳山博士在经济计量学研究领域著作颇丰,作品曾在美国、英国、澳大利亚和印度的主要刊物上发表。相信本书的面世一定会给广大读者提供一个深入、完整地了解经济计量学并与实践相结合的机会。鉴于个人水平有限,翻译中难免有不足之处,欢迎广大读者提出宝贵意见。

薛菁睿

2003年3月1日

CH1-54 03

很荣幸本书能够出到第5版,为此我要感谢国内外的众多教授,他们在经济计量学的理论基础和本书所讲述的实际应用中找到了适用于其课程的平衡点。本版将延续以往各版的传统,即,(1) 理论假设的清楚设定,性质的证明大多放在附录中;(2) 详细描述估计模型和进行改善模型设定的诊断检验和/或估计方法的步骤;(3) 实际应用举例,说明如何把经济计量方法应用到现实的数据中去;(4) 结果的直觉解释。

第1章增加了对各类数据——试验数据、观察数据和样本调查数据——的讨论。由于经济计量学大多数时候涉及的都是对一个或多个变量建模,影响一个或多个变量的其他变量值已知,所以第2章和第3章的理论部分变化最大。在第2章中,联合概率一节的内容有了新的扩展,增加了对条件概率、条件期望和条件方差的讨论。在第3章的理论部分,所有有关误差项的假设都作为对给定 X 的条件。无偏性和高斯-马尔可夫定理的证明开始的时候都假设 X 已知,然后再扩展到其为随机变量的情况。第4章新增了一节,讨论在多元回归模型中生成因变量的预测值和置信区间。同时,还增加了有关解释回归系数,尤其是解释以百分比或比例度量的变量的回归系数的小节。强调对结果的适当解释在第6章中通过讨论和列表总结对包含变量对数的模型系数的解释得到了加强和扩展。第7章就对数线性模型中的虚拟系数的解释进行了修改。第10章的变化也很大,增加了有关向量自回归(VAR)模型和panel数据处理的内容。第11章把预测合并内容扩展到包括序列相关、ARCH影响和时变权重。在第14章中关于实证项目进行步骤的描述上,扩展了对如何计算指数以及妥当处理基期发生的变化的重要性的讨论。

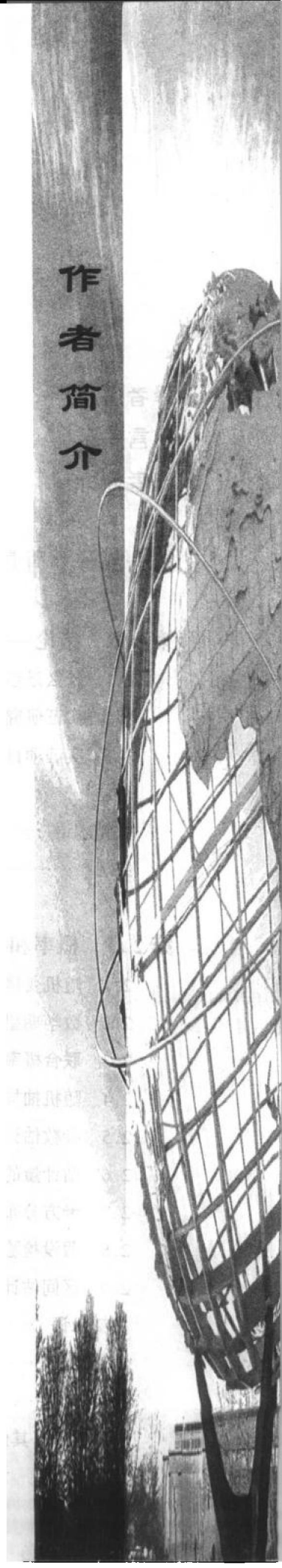
本版增加了若干新的数据集、章末练习和参考书目。所有的数据都有很多格式供读者选择:ASCII(即文本格式)、B34S、Eviews、Excel、GRET、PcGive和SHAZAM。附录D为数据列表,包括变量名称、度量单位和数据源。本书网站 <http://econ.ucsd.edu/~rramanat/embook5.htm> 的变化也很大,包括测试题、本书数据集的链接以及外部数据库、经济计量软件包等。有关这些链接的详细内容见附录C。

欢迎广大读者提出宝贵的意见和建议(e-mail: rramanathan@ucsd.edu),尤其是希望读者们能够指出虽然经过仔细校对但仍难免出现的一些印刷排版错误。

作者简介

作者拉姆·拉玛纳山为马德拉斯大学数学硕士、印度统计学院统计学硕士、明尼苏达州立大学经济学博士。

拉姆·拉玛纳山博士的主要研究领域为应用经济计量学、能源经济学、经济增长和国际贸易。其著作颇丰,作品在美国、英国、澳大利亚和印度的主要刊物上发表。著作包括《经济计量学统计方法》和《经济增长理论入门》,与他人合著的有《圣地亚哥:经济分析》、《固体废物管理外部效应的度量》和《区域负载曲线模型》。他是其自1967年以来所执教的圣地亚哥加州大学的名誉退休教授、加州大学能源研究组顾问委员会会员、国际性杂志《能源》的副主编、数量经济研究协会(QUERI)的主席。同时,拉姆·拉玛纳山博士也是圣地亚哥政府协会能源金融委员会的委员,并在其中发挥主要作用。



目 录

译者序

前言

作者简介

2.A.1 求和的一些实用结果 (42)

2.A.2 最大化和最小化 (43)

2.A.3 更多有关估计的内容 (47)

第一部分 背景

第1章 绪论 (2)

1.1 什么是经济计量学 (2)

1.2 实证研究的基本步骤 (3)

1.3 实证项目 (8)

总结 (9)

关键术语 (9)

练习 (10)

第2章 概率和统计知识回顾 (11)

2.1 随机变量和概率分布 (12)

2.2 数学期望值、均值和方差 (14)

2.3 联合概率、协方差和相关 (18)

2.4 随机抽样和抽样分布 (24)

2.5 参数估计方法 (25)

2.6 估计量的性质 (27)

2.7 卡方分布、 t -分布和 F -分布 (30)

2.8 假设检验 (33)

2.9 区间估计 (36)

关键术语 (37)

参考书目 (39)

练习 (40)

附录 2.A 其他推导 (42)

第二部分 基础篇

第3章 简单线性回归模型 (52)

3.1 基本模型 (53)

3.2 利用最小二乘法估计基本模型 (55)

3.3 估计量的性质 (58)

3.4 估计量的精确度和拟合优度 (61)

3.5 假设检验 (64)

3.6 度量单位的改变 (70)

3.7 应用：用恩格尔曲线来估计
保健支出和收入之间的关系 (71)

3.8 置信区间 (74)

3.9 预测 (74)

3.10 回归模型中的因果关系 (76)

3.11 应用：专利和研发支出之间的关系 (77)

总结 (80)

关键术语 (81)

参考书目 (83)

练习 (83)

附录 3.A 其他推导 (91)

3.A.1 简单线性模型的三维表示 (91)

3.A.2 有关求和的更多结果 (91)

3.A.3 最小二乘法推导正规方程 (91)

3.A.4 最佳线性无偏估计量和
高斯—马尔可夫定理 (92)

3.A.5 最大近似估计量	(93)
3.A.6 估计量方差的推导	(94)
3.A.7 误差项方差的无偏估计量	(94)
3.A.8 方程(3-26)的推导	(95)
3.A.9 方程(3-27a)的推导	(95)
3.A.10 证明 $r_{xy}^2 = R^2$ 为简单回归 模型	(95)
3.A.11 方程(3-29)的推导	(96)
3.A.12 方程(3-30)的推导	(96)

第4章 多元回归模型

4.1 正规方程	(98)
4.2 拟合优度	(100)
4.3 模型选择的一般标准	(102)
4.4 假设检验	(103)
4.5 设定误差	(110)
4.6 应用: 公共汽车乘客人数 的决定因素	(114)
4.7 应用: 女性劳动力参与情况	(117)
4.8 实例: 净移民率和生活质量	(121)
4.9 实证项目	(122)
总结	(122)
关键术语	(123)
参考书目	(124)
练习	(125)
附录 4.A 其他推导	(136)
4.A.1 三变量回归模型	(136)
4.A.2 由于省略相关变量而引起 的偏差	(137)
4.A.3 性质 4.4 的证明	(138)

第5章 多重共线性

5.1 多重共线性的例子	(141)
5.2 完全多重共线性	(142)

5.3 接近多重共线性	(143)
5.4 应用	(147)
总结	(150)
关键术语	(150)
参考书目	(151)
练习	(151)
附录 5.A 通过方程(5-6)推导 方程(5-4)	(152)

第三部分 扩充篇

第6章 选择函数形式和检验

模型设定

6.1 指数函数和对数函数基础 知识复习	(157)
6.2 线性对数关系	(158)
6.3 倒数转换	(160)
6.4 多项式曲线拟合	(160)
6.5 交互作用项	(162)
6.6 行为滞后(动态模型)	(163)
6.7 应用: 专利和研发支出关系的 再分析	(164)
6.8 对数线性关系(或半对数模型)	(168)
6.9 模型间 R^2 值的比较	(170)
6.10 双对数(对数对数)模型	(171)
6.11 应用: 估计乘公共汽车的弹性	(172)
6.12 其他模型	(173)
6.13 “从一般到简单”的亨德瑞/LSE 模型建立方法	(174)
6.14 使用拉格朗日乘数检验的 “从简单到一般”的模型设立方法	(175)
6.15 拉姆齐检验关于回归设定错误 的 RESET 方法	(180)
总结	(181)
关键术语	(181)

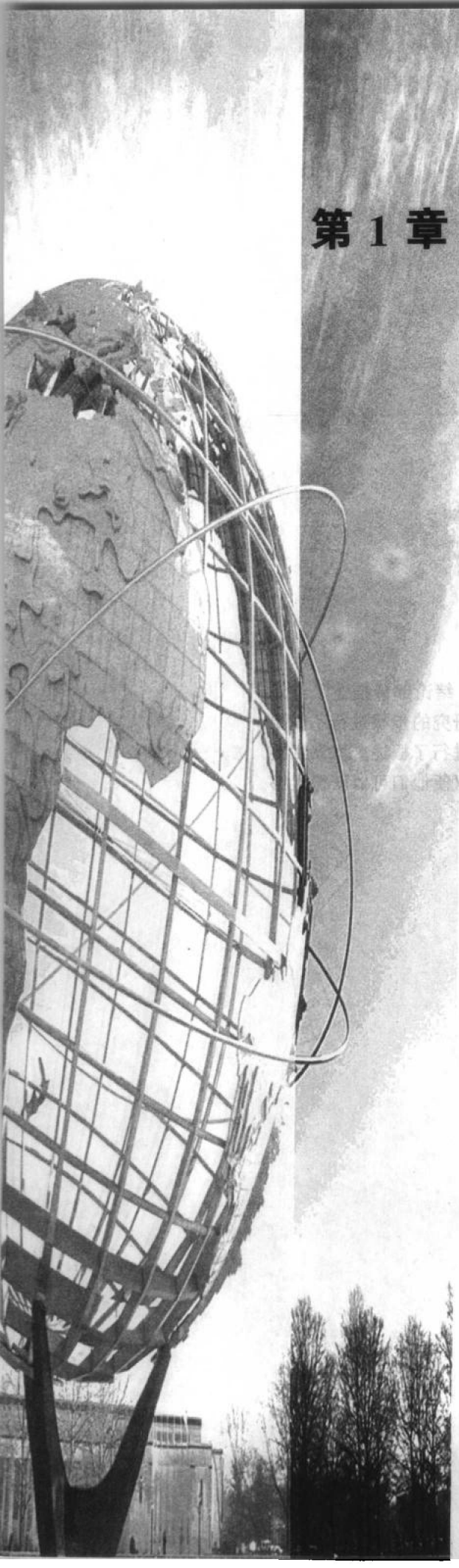
参考书目	(182)	关键术语	(248)
练习	(184)	参考书目	(249)
附录 6.A 有关 LR、瓦尔德和 LM 检验 的更多内容	(190)	练习	(250)
6.A.1 似然率检验	(191)	附录 8.A 异方差存在情况下 OLS 估计量 的性质	(257)
6.A.2 瓦尔德检验	(192)		
6.A.3 拉格朗日乘数检验	(192)	第 9 章 序列相关	(259)
第 7 章 定性(或虚拟)自变量	(194)	9.1 一阶序列相关	(261)
7.1 仅有两种分类的定性变量	(195)	9.2 忽略序列相关的后果	(261)
7.2 具有多种分类的定性变量	(200)	9.3 一阶序列相关检验	(263)
7.3 定性变量对斜率项的影响 (协方差分析)	(203)	9.4 序列相关的处理	(265)
7.4 应用: 工资模型的协方差分析	(205)	9.5 高阶序列相关	(271)
7.5 估计季节影响	(209)	9.6 恩格尔的 ARCH 检验	(272)
7.6 检验结构变化	(211)	9.7 应用: 用电需求	(275)
7.7 实例: 输送机解除管制	(214)	总结	(282)
7.8 应用: 建筑用密封剂的需求	(215)	关键术语	(283)
7.9 实证项目	(217)	参考书目	(284)
总结	(217)	练习	(285)
关键术语	(218)	附录 9.A 其他推导	(293)
参考书目	(218)	9.A.1 DW 的 d 近似等于 $2(1 - \hat{\rho})$ 的证明	(293)
练习	(219)	9.A.2 u_t 为 AR(1) 时的性质	(293)
		9.A.3 AR(1) 存在情况下第一个观察值 的处理	(294)
第四部分 横截面数据和时间序列数据的一些特殊问题		第 10 章 分布滞后模型	(295)
第 8 章 异方差	(234)	10.1 滞后自变量	(296)
8.1 忽略异方差的后果	(236)	10.2 滞后因变量	(300)
8.2 异方差检验	(236)	10.3 滞后因变量和序列相关	(303)
8.3 估计方法	(241)	10.4 滞后因变量模型的估计	(305)
8.4 应用: 美国医疗保健的支出模型	(246)	10.5 应用: 英国消费支出的动态模型	(308)
8.5 实证项目	(247)	10.6 应用: 再论小时电力负荷模型	(308)
总结	(248)	10.7 单位根和迪基-富勒检验	(310)
		10.8 误差纠正模型	(313)
		10.9 应用: 美国国防支出的误差纠正	

模型	(315)	关键术语	(367)
10.10 协积	(319)	参考书目	(368)
10.11 因果关系	(322)	练习	(368)
10.12 合并横截面和时间序列数据 (或 panel 数据)	(323)	第 13 章 联立方程模型	(370)
10.13 实证项目	(325)	13.1 联立方程模型的结构和简化形式 ..	(371)
总结	(325)	13.2 忽略联立的后果	(372)
关键术语	(327)	13.3 识别问题	(373)
参考书目	(328)	13.4 估计方法	(376)
练习	(331)	13.5 实例: 隐形眼镜业的规定	(379)
第五部分 专题		13.6 应用: 简单的凯恩斯模型	(380)
第 11 章 预测	(336)	总结	(382)
11.1 拟合值、事后和事前预测值	(337)	关键术语	(383)
11.2 模型评估	(337)	参考书目	(383)
11.3 条件预测值和无条件预测值	(338)	练习	(384)
11.4 时间趋势预测	(339)	附录 13.A OLS 估计值极限的推导	(385)
11.5 组合预测	(342)	第六部分 实践	
11.6 经济计量模型预测	(346)	第 14 章 实证项目研究	(388)
11.7 时间序列模型预测	(349)	14.1 选题	(389)
总结	(356)	14.2 文献回顾	(392)
关键术语	(357)	14.3 建立一般模型	(393)
参考书目	(358)	14.4 收集数据	(393)
练习	(359)	14.5 实证分析	(396)
第 12 章 定性和有限因变量	(360)	关键术语	(398)
12.1 线性概率(或二元选择)模型	(361)	参考书目	(399)
12.2 概率单位模型	(361)	附录 A 统计表	(401)
12.3 对数单位模型	(363)	附录 B 部分练习答案	(427)
12.4 有限因变量	(365)	附录 C 计算机练习部分	(445)
总结	(367)	附录 D 数据描述和计算机练习部分 ..	(448)

第一部分

背 景

第一部分由两章组成,旨在为下文提供背景资料。绪论部分描述了经济计量学的定义并给出了具体的应用实例,接着对实证研究的步骤进行了简要说明。第2章对经济计量学中使用的概率和统计概念进行了概述。本章主要针对那些具有初级统计知识但却不懂微积分的读者,以便他们可以获得足够的知识来理解以后的内容。



第 1 章

绪 论

1.1 什么是经济计量学

简单来说,经济计量学就是在经济学中应用统计工具的科学。和经济统计学不同的是,经济计量学是经济理论、数学工具和统计方法的综合体,而非主要为统计数据。从广义的角度来说,经济计量学涉及:(1)估计经济关系;(2)经济理论实例和涉及经济行为的假设检验;(3)预测经济变量的行为。在以下各节中,我们将通过众多简短的实例分别对这些行为进行说明。

1.1.1 估计经济关系

实证经济学提供了人们尝试根据数据估计经济关系的大量实例。例子包含如下内容:

(1) 政府和私营部门分析人员的兴趣点在于估计各种产品和服务的供求;

(2) 私企的兴趣点在于估计不同层次的广告对销售和利润的影响;

(3) 股票市场分析人员的兴趣点在于研究股票价格和股票发行公司的特点以及与整个经济态势的关系;

(4) 联邦政府和州政府也许想估计货币和财政政策对就业或失业、收入、进出口、利率、通货膨胀率和预算赤字等重要变量的影响;

(5) 地方政府关心收入与税率和人口等影响收入的各种因素之间的关系;

(6) 市政府可能会关心位于其所辖地区的某家公司的影响,尤其是对住房需求、就业水平、销售和财产收入以及学校、污水处理设施、电力等公共服务需求的影响。

1.1.2 假设检验

和其他科学一样,经济学中的大量内容都涉及有关经济行为的**假设检验**。举例说明如下:

- (1) 某家快餐连锁店想确定其最新的广告活动是否促进了销售;
- (2) 政府和私营部门的分析人员都想知道,对于某一社区或服务来说,需求与价格和收入之间是否具有弹性关系;
- (3) 几乎所有的公司都想知道收入是否随着经营规模的扩大而增长;
- (4) 政府决策者、烟草公司和医疗调查人员想知道卫生局局长关于吸烟与肺癌(和其他呼吸道疾病)的报告是否已导致香烟消费量的剧减;
- (5) 宏观经济学家想度量各种政府政策的效果;
- (6) 公共事业委员想知道要求加强住宅和建筑保温程度的规定是否大大降低了能源消耗;
- (7) 执法机构和州议会想度量加大对酒后开车的立法力度对减少人员死伤的效果。

1.1.3 预测

变量确认完毕并度量它们对研究课题的明显影响后,我们也许想利用估计关系式来预测未来值。有关**预测**的示例如下所示:

- (1) 公司预测销售、利润、生产成本和库存要求;
- (2) 公共事业项目对能源提出要求,以便建设足够的发电设备和/或安排从外地购买;
- (3) 众多公司预测股票市场指数和某些股票的价格;
- (4) 联邦政府规划收入、支出、通货膨胀、失业、预算和贸易赤字等;
- (5) 市政府例行预测当地人口、就业、商业和工业建设等领域的增长对学校、道路、警察局、消防局及公共事业等的数量需求。

由于本章开始所确认的三个一般步骤通常是基于样本数据而不是完整的调查基础,所以这些标准调查中会存在一定的不确定性;尤其是:(1) 估计关系不精确;(2) 假设检验的结论受到弃真或取伪的影响;(3) 基于估计关系式的预测值几乎从来都不是精确的。为了减少不确定性,经济计量人员通常会估计出研究变量的若干不同关系,然后进行一系列检验来判定哪种关系最能描述或预测变量的实际行为。

也正是这种不确定性使得统计方法在经济计量学中具有非常重要的地位。下一章将总结本书中将要用到的基本的统计概念,在需要的时候可供参考。现在让我们一起看一下进行实证研究的基本步骤。

1.2 实证研究的基本步骤

调查人员进行实证研究的基本步骤如下:(1) 建立模型;(2) 收集数据;(3) 估计模型;(4) 假设检验;(5) 解释结果。图 1-1 为这些步骤的流程图。在本节中我们将对这些步骤进行简单的说明。第 14 章将对这些步骤进行详细的描述。如果老师计划把实证项目作为教学课程的一部分,则应提早讲述第 14 章。

1.2.1 建立模型

任何经济、社会、政治或自然科学系统的分析都是建立在对系统中的动因的行为进行描述的基本逻辑结构(被称为模型)的基础上的,这是分析的基本框架。同自然学一样,经济学模型也是以方程的形式建立的。在此例中,方程主要用于描述经济和相关变量的行为。调查人员所建立的模型可能只包含一个方程或是一个包含若干方程的系统。

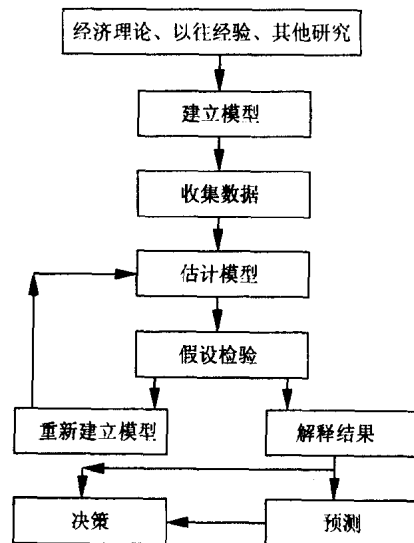


图 1-1 实证研究流程图

1. 单一方程模型

在单一方程的设定中,分析人员选择调查人员想对其行为进行解释的单一变量(用 Y 表示)。 Y 的名称很多,最常用的是**因变量**,但同时也被称为从属变量和左手边变量。然后调查人员会确认影响因变量的若干变量(用 X 表示)。这些变量的名称也很多,**自变量**是最常用的名称,但同时也被称为外生变量、解释变量、回归量和右手边变量。自变量的选择可以依据经济理论、经验、其他研究或直觉判断进行。举例来说,假如一家公司想对人手需求做出判断。这家公司的经济分析人员可能就会使用利润最大化的微观经济学理论来确定公司要雇多少人。公司的利润取决于其产品的价格和数量、员工人数、工资水平、利率、资本成本和原材料成本,等等。利润最大化的原则就要涉及到研究人数(或工作时间)和上述变量之间的理论关系。在这个例子中, Y 指要雇佣的人数(或工作时间), X 指产品价格、工资、利率和原材料成本等。目的则是要对该理论关系进行估计并将该分析结果用于政策制定上。

2. 联立方程模型

在一些经济计量学研究中,调查人员可能对多个变量感兴趣,从而同时建立若干方程,这些方程被称为**联立方程模型**。供求方程和宏观经济模型就是其中的两个例子。方程之一可能是有关总消费与可支配收入和利率的消费函数,另外一个可能是有关投资与可支配收入和利率的投资函数。同时也可能存在一个研究货币需求与收入和利率之间关系的货币需求函数。其他方程可能是有关总需求与总供给以及货币需求和货币供给的平衡条件。

例 1.1

为了使大家能够更好地理解经济计量学的基本结构,我们将给出一个有关因变量(Y)和自变量(X)的简单例子。举例来说,一个房地产代理商想研究房屋售价和房屋特征之间的关系,这些特征包括停车场大小、居住面积、卧室和浴室的数量、是否可观景等等。该代理商最想知道的是,房屋的这些特征对售价会有什么影响。这就是**嗜好价格指数模型**的一个例子。在这种模型中,商品的价格取决于它的特征(另一个例子是,一辆车的售价和它的特征之间的关系)。

虽然列出的所有特征在解释房屋在售价上的差别方面很重要,但为了举例说明,我们只考虑其中的一个特征,如居住面积。假设 PRICE 是房屋的售价, SQFT 是以平方英尺^①度量的居住面积。为简单起见,假设这两个变量之间的关系是线性的,从而可得方程 $PRICE = \alpha + \beta SQFT$, 式中 α 为截距, β 为直线的斜率。假设两座房屋的居住面积相同,但它们的售价却不一定相同,可能纯属偶然或更可能是由于两座房屋在某些特征上的区别没有被考虑进去(如庭院的大小)。因此该关系可能不会很精确,会存在一定的误差。考虑到这些误差,经济计量模型将设立如下:

$$PRICE = \alpha + \beta SQFT + u \quad (1-1)$$

式中, u 为未观察的随机变量,称为**误差项**(同时也称为干扰项或随机项),其统计性质我们将在以后进行描述。误差项随观察值的不同而不同。方程(1-1)称为**线性回归模型**或**简单线性回归模型**。直线 $\alpha + \beta SQFT$ 称为模型的判定部分,误差项 u 称为随机部分。

接着房地产的例子来说,假设我们把 SQFT 定为 5 级,从 1 500、1 750、2 000、2 250 到 2 500,列举某一给定区域 SQFT 位于(或非常接近)这些等级的所有房屋并获得它们的售价。^②如前所述,即使两座房屋的居住面积相同,其售价也可能不同。在这里我们想要度量的是价格的变动在多大程度上是由因素“SQFT”造成的。如果在坐标轴上把成组的 PRICE 和 SQFT 的值绘制成图,所得图形就如图 1-2 所示,其中的圆圈表示标点。由于大房子的价格较高,所以图中的点会随着横轴上向右移动而向上偏移。

接下来我们分别计算在不同 SQFT 等级上的平均价格。在图 1-2 中,这些点用 $\times$ 表示。方程(1-1)假设这些平均点落在直线 $\alpha + \beta SQFT$ 上。判定部分可以视做所研究地区的房屋整个总体的因变量和自变量之间的“统计平均”关系式。为此, α 和 β 被称为**总体参数**(有时也称为**真实参数**)。我们永远也不会知道“真实的”平均关系(指**总体回**

① 1 平方英尺 = 0.092 903 0 平方米。

② 在实践中,由于费用太高,所以我们不可能对房屋总体进行完整的调查。相反,我们会抽取样本并取得其观察值。

归),但在第3章中我们将从一个样本中推出“估计”关系(称为**样本回归**)。未观察误差项 u 代表省略变量的影响(如庭院大小、房屋已使用的年限以及其他影响售价的特征)以及一些内在不可预测因素的影响。

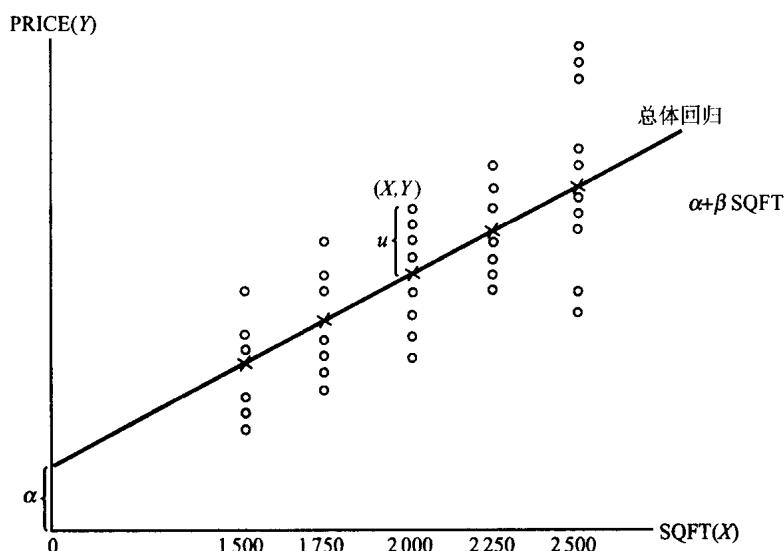


图 1-2 基于 SQFT 的 PRICE 图

由于为了确定 α 和 β 的值而对某一地区的所有房屋进行调查的费用将会非常昂贵,所以调查人员通常会抽取一个房屋样本,利用所获得的有关该样本的信息就 α 和 β 的总体值和方程(1-1)线性回归假设的充分性得出推论。由于结论是建立在房屋样本基础之上的,所以会存在误差。因此对这些误差进行研究以求改善公式或巩固结论就变得非常重要了。

如前所述,居住面积以外的其他变量也会影响房屋的售价。上述模型的扩展模型就是**多元回归模型**,例子如下(由于使用了若干参数,所以标准惯例是使用带有下标的希腊符号 β)。

$$\text{PRICE} = \beta_1 + \beta_2 \text{SQFT} + \beta_3 \text{YARD} + \beta_4 \text{BATHS} + \beta_5 \text{BEDRMS} + u \quad (1-2)$$

式中, YARD 为庭院的面积, BATHS 为浴室的数量, BEDRMS 为卧室的数量。第4章将对该模型的估计和解释进行详细的讨论。包括非线性的该模型的扩展模型将在第6章中讨论。

例 1.2

假设我们对一座城市中的所有家庭及其月收入(Y)和总消费支出(C)进行调查。如果我们对 C 和 Y 绘图,就会得到如图 1-2 所示的图形,但 X -轴为收入, Y -轴为消费。接下来我们对收入为 500 美元的所有家庭进行讨论,然后对相应的消费支出取平均值。对收入为 1000 美元、1500 美元、2000 美元等的家庭重复该过程并计算对应的平均消费水平。然后对这些平均值绘图。同样,我们假设平均点位于直线($\alpha + \beta Y$)上。由于收入相同的家庭其消费水平也会不同,所以一个典型的观察值可能不会正好落在直线上。因此,对应于此例的简单线性回归模型的形式如下:

$$C = \alpha + \beta Y + u \quad (1-3)$$

在实践中我们不会对所有的家庭进行调查,而是会从中抽取样本并利用观察值对未知参数 α 和 β 进行估计,以及进行检验和对消费和收入呈线性关系的假设的有效性进行验证。

例 1.3

在金融领域的研究中,资本资产定价模型(CAPM)为分析各种资产的风险—收益关系提供了一个框架。假设 r 为某种证券的投资收益(如公司股票), r_m 为市场投资组合的收益(如标准普尔综合指数), r_f 为无风险投资收益(如 30 天的美国短期国库券)。设 $Y = r - r_f$ 为某种给定证券的超额收益, $X = r_m - r_f$ 为市场平均投资组合的超额收益,则标准 CAPM 的

第一部分 背景

公式如下所示:

$$Y = \beta X + u$$

注意,该模型中没有截距项。这是因为收益被表示为与无风险收益的偏差。如果我们有证券收益的历史数据,我们就可以对模型进行估计。金融杂志经常报道证券中的“betas”,所指的就是上述模型中 β 的估计值。估计 β 值大于 1 的证券的表现被认为是强劲的或较不稳定的,估计 β 值小于 1 的证券被认为是较保守的或较稳定的。

例 1.4

在消费者需求理论中,分析人员通常建立“效用函数”并在预算约束下求取最大值,从而推导出商品的需求函数。尤其是我们可以证明在一定假设条件下,某种商品的支出(E)和收入(Y)成比例。这种关系就是我们所知的恩格尔曲线。由此就产生了下列经济计量学模型,其中 α 的理论取值为零。

$$E = \alpha + \beta Y + u$$

系数 β 被解释为相对于收入来说商品消费支出的边际倾向。即,收入每增加 1 美元,商品消费支出就会增加 β 美元。根据所掌握的单个家庭的数据,我们就可以对前面的支出函数进行估计,并对截距项为零的假设进行检验。

3. 其他例子

虽然本书主要考虑的是经济关系和假设检验,但这些技术同样适用于其他学科。下面我们举一些其他领域的例子。

例 1.5

经充分证明,抽烟是导致肺癌致死的主要原因。简单线性回归模型如下:

$$\text{DEATHS} = \alpha + \beta \text{SMOKING} + u$$

式中,DEATHS 为某个地区某一指定期间中每百万人口中患肺癌致死的人数;SMOKING 为以磅计算的人均烟草消费量。因为吸烟人数增加会导致死亡人数的增加,所以 β 为正。同房地产的例子一样,调查人员将得出其他变量也会对肺癌的致死人数有影响的结论(如空气污染)。

例 1.6

许多社会学家和犯罪学家都在争论死刑对暴力犯罪来说是否起到了一个极大的威慑作用。为对这个观点进行检验,我们可设立如下模型(同样忽略其他因素的影响):

$$\text{CRIMES} = \alpha + \beta \text{PUNISHMENT} + u$$

式中,CRIMES 为每千人中暴力犯罪的人数;PUNISHMENT 为最终判死刑的定罪百分比。由于死刑数量的增加会遏制犯罪的增加,所以 β 的期望符号为负。

例 1.7

当限制抽烟的法律失败时,我们经常会把原因归结为烟草业的游说活动。为对这种影响做出估计,我们建立如下模型:

$$\text{VOTE} = \alpha + \beta \text{EXPENSE} + u$$

式中,VOTE 为反对该法律的投票人的百分比;EXPENSE 为烟草业为游说每张投票所支出的费用。由于花费的增加会导致反对该法律的投票人数量的增加,所以 β 的期望符号为正。

两种模型设立的方法区别很大。一种方法是是根据经济理论、直觉、其他研究和过去的经验设立一个基本模型(如方程 1-1),然后诊断检验是否还有其他比较复杂的、更适合一些的模式。这种方法称为从简单到一般的模型建立方法,主要在北美地区使用。相反地,从一般到简单的模型建立方法则是从一般公式开始,再逐步对数据进行削减,从而得到一个简单的模型,这种方法称为亨德瑞/LSE,主要在英国和其他欧洲国家使用。以上两种方法各有利

弊,我们将在第6章的6.13节和6.14节中进行详细讨论。我的建议是,大家对任何一种方法都不要过于教条,而是要取长补短为己所用。

1.2.2 收集数据

为了对调查人员设定的经济计量学模型进行估计,我们需要收集因变量和自变量的相关数据。从广义角度来说,调查人员能收集的数据共有三种:**试验数据**、**样本调查数据**和**观察数据**。试验数据通过进行控制试验获得。举例来说,一家生产化肥的化学公司想度量不同剂量化肥的不同效用。调查人员会选择若干块特征接近的土地,如土壤肥沃程度、灌溉量、使用的杀虫剂等,并使用不同剂量的化肥。另外一个例子是,一家医药公司想研究某种新药的效用,该公司会选择两组特征接近的病人,对一组病人用药而对另一组病人只用安慰剂。再举一个有关用电的例子,用电实行按不同时段收费的办法。电力公司通常会确定具有相应特征的不同家庭或公司,然后度量它们对不同价格的不同反应。在上述所有例子中,试验者选择特定的输入,然后观察相应的反应情况。

样本调查数据的获得需要调查人员设计调查表,列出需要调查的问题,然后对部分人群进行调查。有一些调查是定期进行的,如今日商业调查、今日人口调查、劳动统计局的月调查等,调查人员也会根据特定的需要设计专门的调查表。

观察数据(有时称为**非试验数据**)指不是从样本调查或控制试验中获得的数据,如国内生产总值、通货膨胀、失业、股票市场指数等。

如果分析人员想解释因变量在不同时间的变动情况,就必须获得不同时间点的度量值(称为**时间序列数据**)。举例来说,市政府想预测未来5~10年的住房需求情况。这就需要确定对该地区过去的住房需求有影响的变量,获得过去若干年来的时间序列数据并应用到相应的模型中以对未来需求做出预测。时间的区间或**周期性**可以为年、季度或月,可以根据需要而定。

时间序列数据代表不同时间期间的观察值,而**横截面数据**则代表在某一给定时间点上的度量值。举例来说,某州的房产部想解释为什么各地对住房的需求不同,为此就需要获得某一期间内不同地区特征值的观察值。

大多数数据都可以从公共或私人渠道获得(详见第14章),但通常从这些渠道所获得的数据都不足以解决我们面临的问题,或有些数据根据就找不到。在这种情况下,可以进行一种特殊的调查来收集相关数据。举例来说,几年前,公共事业委员想研究消费者对供电按时段定价的反应。电费按时段定价指时段不同则电费不同,高峰时段价格较高,非高峰时段的价格较低。为了获得相关数据,委员会选择了一些居民用户并给他们安装了电表来记录每天不同时间的用电情况。这种监控将持续一年,以便委员会能够记录不同季节的反应情况。该种数据收集产生的是横截面家庭的时间序列数据,称为**合并横截面或时间序列数据**或为我们所熟知的 **panel 数据**。为此我们需要特别的经济计量技术来处理这种数据。

研究人员经常会遇到的另外一种数据涉及到**总量**。举例来说,食品支出和收入之间的关系所涉及的数据可以分为几个类型:(1)所选某个家庭在各个时间的数据(即时间序列数据);(2)一组家庭在某一时间点的数据(即横截面数据);(3)一组家庭在不同期间的数据(即 panel 数据);(4)若干城市或国家全部居民的总支出和总收入(即若干地理区域居民的总横截面数据);(5)某一地理区域内按时间序列收集的所有居民的总支出和总收入(即所有居民的总时间序列数据)。调查人员所要回答问题的本质就是指定所要收集数据的类型及总量水平。

在获取数据的过程中,调查人员必须考虑到所掌握的数据或许并不能很好地满足自己的需要。举例来说,大量的经济理论都涉及到利率,但不可能仅仅与利率相关。如果分析人员想研究住房需求,就会用到抵押率。但如果重点是新厂房和新设备的资本支出,“优惠利率”或其他与之相关的借款利率将会是很好的利率度量工具。

由于政府定期改变“**基期**”,所以我们在使用**指数**形式的数据时要非常小心。举例来说,**消费物价指数(CPI)**是给定时间点(比如说2000年12月)的固定货物与其在固定年份(如基期为1995年1月)价值的平均度量指标。如果基期变为2000年1月,则以1995年1月为基期度量的所有数据必须转化为新的基期。第14章将详细描述如何计算指数以及如何进行基期转化。

在处理数据时需要进行大量的判断并且要格外小心。由于结论的有效性取决于数据的准确性,所以调查人员