

Essentials of Econometrics (2nd Edition)

经济教材译丛

HZ BOOKS
华章经管

经济计量学 精要

(原书第2版)

(美) 达莫达尔 N. 古亚拉提著

张涛 等译
汪同三 审校



机械工业出版社
China Machine Press



McGraw-Hill

经济教材译丛

经济计量学 精要

(原书第2版)

(2nd Edition)

Essentials of Econometrics

(美) 达莫达尔 N. 古亚拉提 (Damodar N. Gujarati) 著

张涛 等译 汪同三 审校



机械工业出版社
China Machine Press

Damodar N.Gujarati: Essentials of Econometrics ,2nd ed.

Copyright © 1999 by McGraw-Hill Companies, Inc.

All rights reserved.For sale in the China only.

本书中文简体字版由McGraw-Hill公司授权机械工业出版社在中国境内独家出版发行，未经出版者书面许可，不得以任何方式抄袭、复制或节录本书中的任何部分。

版权所有，侵权必究。

本书版权登记号：图字：01-1999-3107

图书在版编目(CIP)数据

经济计量学精要 / (美)古亚拉提(Gujarati,D.N.)著；张涛等译。-北京：机械工业出版社，2000.5

(经济教材译丛)

书名原文：Essentials of Econometrics

ISBN 7-111-07962-0

I .经… II .①古… ②张… III .计量经济学 IV .F224.0

中国版本图书馆CIP数据核字 (2000) 第18450号

机械工业出版社 (北京市西城区百万庄大街22号 邮政编码100037)

责任编辑：张德斌 版式设计：陈子平

北京第二外国语学院印刷厂印刷·新华书店北京发行所发行

2000年7月第1版 2001年4月第2次印刷

787 mm × 1092mm 1/16 · 22.5印张

定价：38.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

译者序

经济计量学(Econometrics)这一术语首先是由挪威学者R.弗瑞希于1926年提出的。经过近80年的发展,经济计量学现已成为经济科学中的一个重要分支,在宏、微观经济领域中得到了越来越广泛的应用。

《经济计量学精要》是一本标准的教科书。全书共分四个部分,第一部分详细介绍了标准统计的内容,这是经济计量方法的理论基础;第二部分着重介绍了经济计量学的基础工具——线性回归模型;第三部分对古典线性回归模型的若干假设进行了详尽地阐述;第四部分简要介绍了联立方程技术。通览本书,具有以下两个突出的特点:一是基础性。本书是一本经济计量学初级教科书。主要以理论分析为主,没有涉及复杂的数学计算、推导和证明,通过宏、微观经济中的典型实例向初学者通俗、形象地介绍经济计量学的基本原理。二是全面性。本书的组织结构是从概率论基础到单方程技术,再到联立方程技术,全面反映了经济计量学研究的主要内容,并介绍了经济计量学研究的最新成果。可以预期,《经济计量学精要》一书将对经济计量学的教学和研究起到有益的作用。

各章译者分别是:第1章至第9章由张涛主译,第10章至第14章由王智勇主译,第15章由张涛主译。此外,王宏伟、彭鹏、赵陵、唐华茂、丁国勇、李克成、桑惊、韩曦、彭俊明、张思奇等同志也帮助翻译了部分章节。汪同三研究员对全书进行了仔细的审校。钟瑛、夏蜀、吕凌、费多益、林新等同志在文字校对方面做了大量的工作。书中的图表、图形由张跃华、汪超负责处理。在本书的翻译过程中,还得到了中国社会科学院数量经济与技术经济研究所的万利华、曹曼株、张杰、刘戈平等同志的帮助,在此深表感谢。

由于译者水平有限,书中误译在所难免,恳请指正!

张涛

1999年12月

前 言

本书第2版在以下方面略作改动：

第1章重写了经济计量学方法论一节，强调指出：在用估计的模型进行假设检验和预测前，首先要检查模型的准确性。在本章附录中提供了一些网站，可供查阅有关经济数据。丰富的经济数据使得经济计量建模阶段的数据收集容易实现。

第2章增加了对峰度和偏度的讨论。它们可用于判断随机变量是否服从正态分布。正态分布是整个经济计量学的核心。

第3章更深入地讨论了如何从正态分布中获得随机样本。在这一章中，还介绍了靴攀抽样法的概念。靴攀抽样法广泛运用于经济计量学的许多领域。本章还对抽样、概率、随机变量的分布、中心极限定理等作了详细的讨论。在讨论过程中，还介绍了蒙特卡罗试验。除此之外，还对正态分布、 t 分布、 χ^2 分布、 F 分布之间的关系作了总结。这些概率分布将在本书中大量使用。

第6章用蒙特卡罗试验论证了最小二乘估计量的一些性质。本章还介绍了正态检验的内容，比如正态概率图，J-B检验。除此之外，还增加了一些例子。

第7章增加了有关模型结构稳定性的讨论并介绍了检验模型结构稳定性的Chow检验，而且探讨了虚拟变量是如何应用于Chow检验的。

第11章对异方差的讨论中包括了White的一般异方差检验以及White修正异方差后的标准差和 t 统计量。

第13章增加了模型选择的两个常用标准：Akaike信息标准（AIC）和Schwarz信息标准（SIC）。

第14章讨论了3个新的专题：（1）伪回归现象；（2）非平稳的单位根检验；（3）随机游走模型；

第15章是本书新增内容。本章讨论了联立方程模型，即研究经济变量相互关系的模型。介绍了一些重要概念：联立问题、模型识别问题、间接最小二乘（ILS）、两阶段最小二乘（2SLS）等，并通过美国具体的宏观经济数据来解释这些概念。

对这些新增专题的讨论无疑使第2版的内容比第1版丰富了许多。但是，对这些内容的讨论只是起到抛砖引玉的作用，因为本书的目的是向初学者简单地介绍经济计量学的主要内容。需要更深入了解的读者可参阅《经济计量学基础》（*Basic Econometrics*），以及Willam H.Greene所著的《经济计

量分析》(*Econometric Analysis*, Prentice-Hall, 3rd ed., 1997)、Russell Davidson 与 James MacKinnon 所著的《经济计量学：估计和推断》(*Estimation and Inference in Econometrics*, Oxford University Press, 1993)。

数学要求

本书很少用到矩阵代数和微积分。我一直坚信应该以最直观的方式向初学者介绍经济计量学，而无须用大量的矩阵代数和微积分。而且，也未给出证明过程，除非这些证明很容易理解。当然，教师可以根据需要在适当的地方给出证明。一些证明过程可参见《经济计量学基础》一书。

解决问题的方法

我坚信边学边用。因此，在每章的最后都提供了大量的习题。在本书的最后部分还提供了部分习题的答案。

计算机与经济计量学

许多优秀的统计软件对初学者学习经济计量学大有益处。在本书的例子中使用了EViews、MINITAB、SHAZAM等统计软件。

总之，本书的目的是以一种轻松的方式向初学者介绍经济计量学这门学科。我希望本书能为你将来的学术或专业研究有所帮助。

致谢

在本书的写作过程中得到许多同仁的帮助。感谢波士顿大学(Boston University)的Eli Berman，加利福尼亚州立大学(California State University)的John Eckalbar，波士顿学院(Boston College)的Joseph Quinn，他们通阅了本书的第二版，并提供了大量的宝贵意见；我深深感谢西点军校社会科学系系主任Col. Daniel Kanfman所给予的不遗余力的支持；感谢Major Gregory Wise，他为本书部分专题提出了诸多建设性意见；感谢Major Kevin，他解决了本书中的许多计算机问题；感谢Mark Pettit，是他指引我进入网络世界；感谢我的妹夫Joseph Buschi，他解决了书中出现的诸多硬件及软件问题；感谢我的嫂子Sushila BuschiGidwani，与她有关经济学、政治学的讨论使我受益非浅；感谢Donna Wong，他为本书制作了整个数据盘；最后，我要感谢我的妻子Pushpa及女儿Joan和Diane，她们一直是我灵感和鼓励的源泉。

目 录

译者序 前言

第1章 经济计量学的特征及

研究范围1

- 1.1 什么是经济计量学1
- 1.2 为什么要学习经济计量学1
- 1.3 经济计量学的方法论2
- 1.4 全书结构8
- 习题9
- 附录1A 万维网上的经济数据10

第一部分 概率与统计基础

第2章 基本统计概念的回顾14

- 2.1 一些符号14
- 2.2 试验、样本空间、样本点和事件15
- 2.3 随机变量16
- 2.4 概率17
- 2.5 随机变量与概率密度函数20
- 2.6 多元随机变量的概率密度函数23
- 2.7 概率密度的特征26
- 2.8 从总体到样本34
- 2.9 小结37
- 参考文献38
- 习题38

第3章 一些重要的概率分布42

- 3.1 正态分布42
- 3.2 样本均值 \bar{X} 的抽样分布或概率分布48
- 3.3 χ^2 分布53
- 3.4 t 分布54

- 3.5 F 分布57
- 3.6 t 分布、 F 分布、 χ^2 分布与正态分布的
关系59
- 3.7 小结59
- 习题59

第4章 统计推断：估计与假设

检验62

- 4.1 统计推断的含义62
- 4.2 估计和假设检验：统计推断的
两个孪生分支62
- 4.3 参数估计63
- 4.4 点估计量的性质65
- 4.5 统计推断：假设检验68
- 4.6 小结76
- 习题76

第二部分 线性回归模型

第5章 线性回归的基本思想：

双变量模型80

- 5.1 回归的含义80
- 5.2 总体回归函数：一个假设的例子81
- 5.3 总体回归函数误差的设定82
- 5.4 随机误差项的性质83
- 5.5 样本回归函数84
- 5.6 “线性”回归的特殊含义86
- 5.7 从双变量回归到多元线性回归87
- 5.8 参数的估计：普通最小二乘法88
- 5.9 实例92
- 5.10 小结94
- 习题95
- 附录5A 最小二乘估计量的推导98

第6章 双变量模型：假设检验	99	附录7A.2 式(7-31)的推导	151
6.1 古典线性回归模型	99	附录7A.3 式(7-53)的推导	151
6.2 普通最小二乘估计量的方差与标准差	101	附录7A.4 EViews计算结果 (抵押贷款债务一例)	152
6.3 普通最小二乘估计量的性质	103	第8章 回归方程的函数形式	153
6.4 OLS估计量的抽样分布或概率分布	104	8.1 如何度量弹性：对数线性模型	154
6.5 假设检验	105	8.2 线性模型与对数线性模型比较	156
6.6 拟合优度的检验：判定系数 r^2	110	8.3 多元对数线性回归模型	157
6.7 回归分析结果的报告	113	8.4 如何测度增长率：半对数模型	160
6.8 正态性检验	114	8.5 线性对数模型：解释变量是对数形式	163
6.9 关于计算：回归分析的软件	115	8.6 双曲函数模型	165
6.10 实例：美国进口支出	116	8.7 多项式回归模型	167
6.11 预测	119	8.8 不同函数形式模型小结	168
6.12 实例	120	8.9 小结	169
6.13 小结	122	习题	169
习题	122	附录8A 对数	174
第7章 多元回归：估计与假设检验	127	第9章 包含虚拟变量的回归模型	176
7.1 三变量线性回归模型	127	9.1 虚拟变量的性质	176
7.2 多元线性回归模型的若干假定	129	9.2 包含一个定量变量，一个两分定性变量的回归模型	179
7.3 多元回归参数的估计	130	9.3 虚拟变量有多种分类的情况	182
7.4 实例：未偿付抵押贷款债务	132	9.4 包含一个定量变量，两个定性变量的回归模型	183
7.5 估计的多元回归方程的拟合优度： 多元判定系数 R^2	133	9.5 模型的推广	184
7.6 多元回归的假设检验：一般的解释	134	9.6 回归模型中的结构稳定性： 虚拟变量法	185
7.7 对回归参数进行假设检验	135	9.7 虚拟变量在季节分析中的应用	189
7.8 对联合假设的检验	137	9.8 小结	192
7.9 从多元回归模型到双变量模型： 设定误差	140	习题	192
7.10 两个不同的 R^2 的比较：校正的判定系数	140	第三部分 实践中的回归分析	
7.11 什么时候增加新的解释变量	141	第10章 多重共线性	200
7.12 回归模型的结构稳定性检验： Chow检验	141	10.1 多重共线性的性质：完全多重共线性的情况	200
7.13 实例	144		
7.14 小结	147		
习题	147		
附录7A.1 OLS估计量的推导	151		

10.2	接近或者不完全多重共线性的情形	202	13.5	小结	274
10.3	多重共线性的理论后果	204	习题		275
10.4	多重共线性的实际后果	204	第14章 单方程回归模型：几个补充专题		280
10.5	多重共线性的测定	206	14.1	有限最小二乘法	280
10.6	多重共线性必定不好吗	208	14.2	动态经济模型：自回归模型和分布滞后模型	283
10.7	一个扩充例子：1960至1982年期间美国的鸡肉需求	209	14.3	用于分布滞后模型的夸克方法	287
10.8	如何对付多重共线性：补救措施	211	14.4	解释变量是虚拟变量的情形	289
10.9	小结	215	14.5	分对数模型	292
习题		215	14.6	伪回归现象	296
第11章 异方差		219	14.7	小结	303
11.1	异方差的性质	219	习题		304
11.2	异方差的后果	224	第四部分 联立方程模型简介		
11.3	异方差的检验：如何知道存在异方差问题	225	第15章 联立方程模型		310
11.4	观察到异方差该怎么办：补救措施	231	15.1	联立方程模型的性质	311
11.5	White异方差校正后的标准差和t统计量	235	15.2	联立方程的偏误：普通最小二乘估计量的不一致性	312
11.6	实例	236	15.3	间接最小二乘法	313
11.7	小结	238	15.4	实例	314
习题		238	15.5	模型识别问题	315
第12章 自相关		244	15.6	模型识别的判定规则：识别的阶条件	320
12.1	自相关的性质	244	15.7	对过度识别方程的估计：两阶段最小二乘法	320
12.2	自相关的后果	247	15.8	2SLS：一个数字例子	321
12.3	自相关的诊断	247	15.9	小结	323
12.4	补救措施	253	习题		323
12.5	如何估计 ρ	254	附录15A OLS估计量的一致性		325
12.6	小结	257	附 录		
习题		258	附录A 统计表		328
第13章 模型选择：标准与检验		262	附录B 部分习题答案		343
13.1	“好的”模型具有的特性	262	参考文献		347
13.2	设定误差的类型	263			
13.3	诊断设定误差：设定误差的检验	269			
13.4	用于预测的模型选择	273			

第 1 章

经济计量学的特征及研究范围

在经济学、金融学、管理学、营销学以及一些相关学科的研究中，定量分析用得越来越多，对于这些领域的初学者来说，掌握一至两门经济计量方面的课程是必要的——这个领域的研究变得十分流行。本章的目的旨在给初学者一个经济计量学的概貌。

1.1 什么是经济计量学

简单地说，经济计量学(Econometrics)就是经济的计量。虽然，对诸如国民生产总值(GNP)、失业、通货膨胀、进口、出口等经济概念的定量分析十分重要，但从下面的定义中，我们不难看出经济计量学的研究范围更为宽泛：

经济计量学是利用经济理论、数学、统计推断等工具对经济现象进行分析的一门社会科学。¹

经济计量学运用数理统计知识分析经济数据，对构建于数理经济学基础之上的数学模型提供经验支持，并得出数量结果。²

1.2 为什么要学习经济计量学

从上述定义我们知道经济计量学涉及经济理论、数理经济学、经济统计学(即经济数据)，以及数理统计学等相关学科，但它是一门有其自己研究方向的一门独立学科。

从本质上说，经济理论所提出的命题和假说，多以定性描述为主。例如，微观经济理论中提到的：在其他条件不变的情况下(经济学中著名的 *Ceteris paribus* 从句)，一种商品价格的上升会引起该商品需求量的减少。因而得出结论：商品的价格与该商品的需求量呈反方向变动——这就是著名的向下倾斜的需求曲线，简称需求法则。但是，该理论本身却无法度量价格和需求量这两个变量之间的数量关系，也就是说，它不能告诉我们商品的价格发生某一变动时，该商品的需求量增加或减少了多少。经济计量学家的任务就是提供这样的数量估计。换一种说法，经济计量学是依据观测和试验，对大多数经济理论给出经验的解释。如果在研究或试验中发现，当每单位商品的价格上升一美元，引起该商品需求量的下降，比如说下降100个单位，那么，我们不仅验证了需求法则，而且还提供了价格和需求量这两个变量之间的数量估计。

1 Arthur S. Goldberger, *Econometric Theory*, Wiley, New York, 1964, p.1.

2 P.A. Samuelson, T.C. Koopmans, and J.R.N. Stone, "Report of the Evaluative Committee for *Econometrica*", *Econometrica*, vol. 22, no. 2, April 1954, pp.141-146.

数理经济学(mathematical economics)主要关心的是用数学公式或数学模型来描述经济理论,而不考虑对经济理论的度量和经验解释。而经济计量学家感兴趣的却是对经济理论的经验确认。下面我们将会讲到,经济计量学家通常采用数理经济学家提供的数学模型,但把它们用于经验检验。经济统计学家主要关心的是收集、处理经济数据并将这些数据绘制成图表的形式。这是经济统计学家的工作:他或她收集GNP、失业、就业、价格等数据,这些数据就成为经济计量分析的原始数据。但经济统计学家却不关心用这些收集到的数据来检验经济理论。

虽然,数理统计学提供了许多分析工具,但由于经济数据独特的性质,即许多数据的生成并非可控制试验的结果,因此,经济计量学经常需要使用特殊的方法。类似于气象学,经济计量学所依据的数据不能直接控制。所以,由公共和私人机构收集的消费、收入、投资、储蓄、价格等方面的数据从本质上说是非试验性的。这就产生了数理统计学不能正常解决的一些特殊问题。而且,这些数据很可能包含了测量的误差,或是遗漏数据或是丢失数据。这就要求经济计量学家去运用特殊的方法来处理这些测量误差。

对于主修经济学和商业专业的学生来说,学习经济计量学有实用性。毕业以后,在其工作中,或许被要求去预测销售量、利息率、货币供给量或是估计商品的需求函数、供给函数以及价格弹性等等。在经济学家以专家的身份出现在联邦政府调节机构中之前,通常代表当事人或公众。而汽油和电的价格是由政府调节机构规定的,因此,这就要求经济学家能估计提议的价格的上涨对需求量(如用电量)的冲击。在这种情况下,经济学家需要建立一个关于用电量的需求函数,并根据这个需求函数估计需求的价格弹性,即,价格变动的百分比所引起需求量改变的百分比。掌握经济计量学知识对于估计这些需求函数是很有帮助的。

客观地说,在经济学和商科专业的学习与培训中,经济计量学已成为不可或缺的一部分。

1.3 经济计量学的方法论

一般说来,用经济计量方法研究经济问题可分为如下步骤:

- (1) 理论或假说的陈述;
- (2) 收集数据;
- (3) 建立数学模型;
- (4) 建立统计或经济计量模型;
- (5) 经济计量模型参数的估计;
- (6) 检查模型的准确性:模型的假设检验;
- (7) 检验来自模型的假说;
- (8) 运用模型进行预测。

为了阐明经济计量学的方法论,我们来考虑这样一个问题:经济形势会影响人们进入劳动力市场的决定吗?也就是说,经济形势是否对人们的工作意愿有影响?假设用失业率(Unemployment Rate, UNR)来度量经济形势,用劳动力参与率(Labor Force Participation Rate, LFPR)来度量劳动力的参与,UNR和LFPR的数据由政府按时公布,那么,如何回答这个问题呢?我们按上述步骤进行分析。

1.3.1 理论或假说的陈述

首先要了解经济理论对这一问题的阐述是怎样的。在劳动经济学中,关于经济形势对人们工作意愿的影响有两个相对立的假说。一个是受挫-工人假说[discouraged-worker hypothesis (effect)],该假说提出当经济形势恶化时,表现为较高的失业率,许多失业工人放弃寻找工作

的愿望并退出劳动市场。另一个是**增加-工人假说**[added-worker hypothesis (effect)], 该假说认为当经济形势恶化时, 许多目前并未进入劳动市场的二手工人(比如带孩子的母亲)可能会由于养家的人失去工作而决定进入劳动市场, 即使这些工作的报酬很低, 只要可以弥补由于养家人失去工作而造成的收入方面的一些损失就行。

劳动力参与率的增加或减少依赖于增加工人和受挫工人的力量对比。如果增加工人的影响占主导地位, 则LFPR将升高, 即使是在失业率很高的情况下。相反地, 如果是受挫工人的影响占主导力量, 那么LFPR将会下降。我们是如何发现这一结果的呢? 这只是一个实践问题。

1.3.2 收集数据

由于实验的目的, 我们需要这两个变量的数量信息。一般来说, 有三种统计数据可用于实践分析:

- (1) 时间序列数据
- (2) 横截面数据
- (3) 合并数据(时间序列数据与横截面数据的联合)

1. 时间序列数据

这种数据是按时间序列排列收集得到的。比如GNP、失业、就业、货币供给、政府赤字等。数据是按照一定的时间间隔收集的——每日(比如股票), 每周(比如货币供给), 每月(比如失业率), 每季度(比如GNP), 每年(比如政府预算)。这些数据可能是**定量的**(quantitative)(比如价格、收入、货币供给等), 也可能是**定性的**(qualitative), (比如男或女, 失业或就业, 已婚或未婚, 白人或黑人等)。我们将会发现, 定性的变量(又称为虚拟变量)与定量的变量同样重要。

2. 横截面数据

横截面数据(cross-sectional data)是指一个或多个变量在某一时点上的数据的集合。例如美国人口调查局每10年进行的人口普查数据(最近的一次是在1990年4月1日), 以及密执安大学进行的夏季居民开支调查数据。这些民意调查的结果由Gallup、Harris 和其他的一些调查机构处理。

3. 合并数据

合并数据(pooled data)中既有时间序列数据又有横截面数据。例如, 如果我们收集20年间10个国家有关失业率方面的数据, 那么, 这个数据集合就是一个合并数据, 每个国家的20年间的失业率数据是时间序列数据, 而20个不同国家每年的失业率数据又组成横截面数据。

在合并数据中有一类特殊的数据, 称为**panel数据**(panel data), 又称**纵向数据**(longitudinal or micropanel data)。即同一个横截面单位, 比如说, 一个家庭或一个公司, 在不同时期的调查数据。例如, 美国商业局在一定时期间隔内对住房的调查。在每一时期的调查中, 同样的(或居住在同一地区的)家庭被调查, 以观察自上一次调查以来, 其住房和经济状况是否有变化。纵向数据就是通过重复上述过程而得到的, 它可对研究家庭行为的动态化提供非常有用的信息。

4. 数据来源

成功的经济计量研究需要大量高质量的数据。幸运的是国际互联网为我们提供了大量详实的数据。附录1A列出了一些网址, 提供了各类微观和宏观的经济数据。学生必须熟悉这些网站并学会下载数据。当然, 这些数据会不断更新, 因此可得到最新的数据。

为了便于分析, 这里给出一组时间序列数据。表1-1给出了美国1980~1996年间城市劳动力参与率(Civilian Labor Force Participation Rate, CLFPR)和城市失业率(Civilian Unemployment Rate, CUNR)数据。城市失业率是指城市失业人口占城市劳动力的百分比。¹

与物理学不同, 许多收集的经济数据(比如GNP、货币供给、道-琼斯指数、汽车销售量等)

1 我们仅考虑集合城市劳动力参与率及城市失业率, 还有其他可用的数据, 比如年龄, 性别, 种族构成等。

是非试验性的，因为数据收集机构(比如政府)或许并不直接监控这些数据。因而，劳动力参与和失业的数据来源于劳动力市场上参与者提供给政府的信息。在某种意义上，政府是数据的消极收集者。在收集数据的过程中，政府或许并不知道受挫-工人、增加-工人假说及其他有关的一些假说。因此，收集到的数据可能是几种因素综合的结果，会影响不同个人劳动力参与的决策。也就是说，同样的数据适用于不止一个理论。

表1-1 城市劳动力参与率(CLFPR)，城市失业率(CUNR)与真实的小时平均工资(AHE82)^①

年	CLFPR(%)	CUNR(%)	AHE82/\$
1980	63.8	7.1	7.78
1981	63.9	7.6	7.69
1982	64.0	9.7	7.68
1983	64.0	9.6	7.79
1984	64.4	7.5	7.80
1985	64.8	7.2	7.77
1986	65.3	7.0	7.81
1987	65.6	6.2	7.73
1988	65.9	5.5	7.69
1989	66.5	5.3	7.64
1990	66.5	5.6	7.52
1991	66.2	6.8	7.45
1992	66.4	7.5	7.41
1993	66.3	6.9	7.39
1994	66.6	6.1	7.40
1995	66.6	5.6	7.40
1996	66.8	5.4	7.43

①AHE82代表1982年的平均小时工资(用美元计算)

资料来源: *Economic Report of the President, 1997*, CLFPR from Table B-37, p.343, CUNR from Table B-40, p.346, and AHE82 from Table B-45, p.352.

1.3.3 建立劳动力参与的数学模型

为了观察CLFPR与CUNR的变动关系，首先我们作散点图(scatter diagram, or scattergram)，见图1-1。从图上可以看出，CLFPR与CUNR呈反方向变动，(将一切情形都考虑到，或许还说明受挫工人效果比增加工人效果强)¹。第一次估计，不妨通过这些散点做一直线并写出两者之间的简单数学模型：

$$CLFPR = B_1 + B_2 CUNR \quad (1-1)$$

注：Y = CLFPR X = CUNR

式(1-1)表明城市劳动力参与率与城市失业率呈线性关系。 B_1 和 B_2 为线性函数的参数(parameters)。 B_1 为截距(intercept)，其值为当CUNR为零时CLFPR的值。 B_2 为斜率(slope)，它是每一单位CUNR的变动所引起的CLFPR的变动率。更一般地，斜率度量了式右边变量每变动

1 关于这一点，参见Shelly Lungerg “The Added Worker Effect,” *Journal of Labor Economics*, vol.3, January 1985, pp.11-37.

2 一般说来，参数是一个不确定的量，可能取一组不同的值。在统计学中，通常用均值和方差等参数来描述随机变量的概率分布函数，在本书第2章中将详细地进行讨论。

3 在第5章，我们将给出回归分析中截距这一概念的一个更为准确的解释。

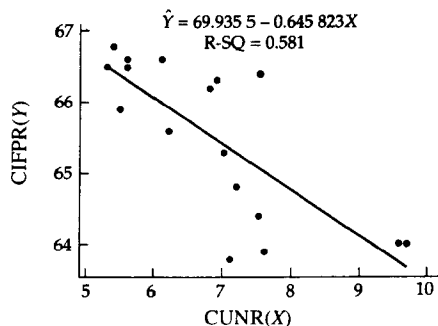


图1-1 城市劳动力参与率(百分比)与城市失业率(百分比)回归分析

一单位所引起的式左边变量的改变量。斜率值可正(若增加工人的效果大于受挫工人效果的影响)可负(若受挫工人效果的影响占主导力量)。在图1-1中,此时的斜率为负。

1.3.4 建立劳动力参与的统计或经济计量模型

式(1-1)给出了描述城市劳动力参与率与城市失业率关系的纯数学模型,数理经济学家或许对它感兴趣,但它对经济计量学家的吸引力却是有限的。因为这样一个模型假设了变量之间的关系是精确的和确定的,也就是说,给定一个CUNR的值,有惟一一个CLFPR的值与之对应。在现实中,很难发现经济变量之间存在如此精确的关系。更一般地,变量之间的关系往往是不确切的或是统计的。

我们可以通过图1-1的散点图清楚地看到这一点。虽然两变量之间存在着反方向变动关系,但两变量并非准确的或完全的线性关系,如果我们通过这17个数据点作一条直线,并不是所有的点都准确地落在这条直线上——别忘了两点确定一条直线。¹为什么这17个数据点没有全都落在这条由数学模型所设定的直线上呢?记住这些数据是非试验性收集到的。因此,如前面所提到的那样,除了增加和受挫工人假说外,还有其他许多因素影响劳动力进入市场的决定。所以,我们所观察到的城市劳动力参与率与城市失业率之间的关系很可能是不确切的。

我们把所有其他影响劳动力参与率的因素都包括在变量 u 中,于是有:

$$CLFPR = B_1 + B_2 CUNR + u \quad (1-2)$$

其中, u 代表随机误差项(random error term),简称误差项(error term)。² u 包括了所有影响城市劳动力参与率但并未在模型中具体给出的因素(除了城市失业率)以及其他的随机因素。在本书第二部分中,我们将会看到经济计量学中的误差项与纯数理经济学的误差项有区别。

式(1-2)就是一个统计的、或经验的、或经济计量模型。更准确地说,它是一个**线性回归模型**(linear regression model),这也正是本书所讨论的主题。在这个模型中,式左边的变量称为**应变量**(dependent variable),式右边的变量称为**自变量**(independent variable)或**解释变量**(explanatory variable)。线性回归分析的主要目标就是解释一个变量(应变量)与其他一个或多个变量(解释变量)之间的行为关系,当然这种关系并非完全准确。

值得注意的是式(1-2)所描述的经济计量模型是来自于式(1-1)所表示的数学模型。这表明数理经济学和经济计量学是互相补充的学科。这一点可以从本书开始给出的经济计量学的定义中清楚地看到。

在继续下文之前,有一个概念值得我们注意——**因果关系**(causation)。在式(1-2)这一回归模型中,我们说城市劳动力参与率是应变量,城市失业率是自变量或解释变量。这两个变量之

1 我们甚至试图通过这些数据点做一条抛物线,但结果与线性假定并没有什么实质性的不同。

2 在统计语言中,随机误差项有两种英文表示: random error, stochastic error

间存在因果关系(即城市失业率是原因,城市劳动力参与率是结果)吗?换句话说,回归包含因果关系吗?并不一定。正如Kendall和Stuart所说,“统计关系,无论有多强,有多紧密,也绝不能建立起因果关系:因果关系的概念必须排除在统计学之外”。在本例中,是根据经济理论(比如受挫工人假说)在应变量和解释变量之间来建立起原因-结果这样的关系。如果不能建立起因果关系,不妨称之为断定的关系——即给定城市失业率,我们能够预测城市劳动力参与率。

1.3.5 经济计量模型参数的估计

利用表1-1给出的数据如何估计(estimates)式(1-2)中的参数 B_1 , B_2 呢?即如何确定这些参数(parameters)的具体数值呢?这是本书第二部分的重点,在那里,我们将用适当的方法,尤其是普通最小二乘法来计算。运用最小二乘法和这些数据,得到下面的结果:

$$\widehat{CLFPR} = 69.9355 - 0.6458CUNR \quad (1-3)$$

注意我们在CLFPR上加一符号^提醒大家:式(1-3)是式(1-2)的估计式。图1-1是根据真实数据估计得到的回归直线。

从式(1-3)可知, B_1 的估计值约为70, B_2 约为-0.64。因此,平均地,如果失业率上升一个单位(比如说一个百分点),则城市劳动力参与率将下降0.64个百分点;也就是说,当经济状况恶化时,劳动力参与率将平均净减少0.64个百分点,这或许表明了受挫工人效应占主导力量。我们讲“平均”是因为前面提到的误差项 u 可能会导致变量之间的关系有些出入,这一点可以从图1-1清楚地看到:真实数据点并未落在估计的回归线上。回归直线上的点与真实数据点的距离称为残差。简言之,估计的回归直线(式(1-3))给出了平均城市劳动力参与率与城市失业率之间的关系——每一单位城市失业率的变化所引起的城市劳动力参与率的平均改变量是多少。常数69.93即当城市失业率为零时城市劳动力参与率的平均值。也就是说,当充分就业时(即不存在失业),城市适龄工作人口的69.93%将参与就业。¹

1.3.6 检查模型的准确性:模型的假设检验

式(1-3)所描述模型的准确性如何呢?通常,人们在进入劳动力市场之前,会根据一些因素,比如说失业率的大小来考虑劳动市场的状况。例如,1982年(不景气的一年)城市失业率为9.7%,而1996年仅为5.4%。显然,失业率为5.4%,与失业率为9.7%时相比,人们更可能进入劳动力市场。但还有其他一些因素影响人们进入劳动力市场的决定。比如每小时工资或收入也是重要的决定变量。至少在短期内,工资越高越能吸引工人进入劳动力市场。其他因素也类似。为进一步说明其重要性,在表1-1中我们还给出了以美元为计量单位的真实平均小时工资(AHE82)的数据。现在加上AHE82这一影响因素,考虑下面这个模型:

$$CLFPR = B_1 + B_2CUNP + B_3AHE82 + u \quad (1-4)$$

式(1-4)是一个多元线性回归模型,而式(1-2)是一个简单的(双变量)线性回归模型。在双变量模型中,仅有一个解释变量,而在多元回归模型中有若干个(或称多元)解释变量。值得注意的是,在多元回归模型中,即式(1-4)中,同样包括误差项 u ,这是因为无论模型中有多少个解释变量,都不能完全解释应变量的行为。一个多元回归模型究竟需要包括多少个解释变量,须根据具体情况而定。当然,基本的经济理论通常会告诉我们哪些变量需要包括到模型之中,但是需要提醒注意的是,正如前面所提到的:回归并不意味着存在因果关系;一个或多个解释变量是否与应变变量存在因果关系,必须根据相关理论来判定。

如何估计式(1-4)中的参数呢?我们将在第5章和第6章讨论完双变量模型之后,在第7章中

¹ 然而,这仅是截据这一概念的机械的解释。我们将在第5章中阐述截据的概念。

给予详细说明。首先考虑双变量模型是因为它是多元回归模型的基础。在随后的第7章中将会看到,多元回归模型在许多方面是双变量模型的直接延伸。

用普通最小二乘法估计得到回归方程(式(1-4)):

$$\widehat{CLFPR} = 97.9 - 0.446 \text{ CUNP} - 3.86 \text{ AHE82} \quad (1-5)$$

这个结果很有意思,因为两个斜率系数均为负数。负的城市失业率表明失业率每增加1%,城市劳动力参与率将平均减少0.44%(假设平均小时工资为一常数)。这个结果又一次支持了受挫工人假说。另一方面,若城市失业率为一常数,则平均小时工资每增加一个百分点,城市劳动力参与率将平均减少3.86%。¹负的平均小时工资系数有经济意义吗?为什么不期望该系数为正(即小时工资越高,则劳动力市场的吸引力也就越高)呢?我们可通过微观经济两个孪生概念,收入效应和替代效应,来验证系数为负。²

我们选择哪一个模型呢?式(1-3)还是式(1-5)呢?既然式(1-5)包含(1-3),而且增加了一个分析变量(收入),所以我们可能选择式(1-5)。毕竟,式(1-2)暗含地假定了除失业率以外其余变量均为常数。但是,我们的分析在哪才是尽头呢?例如,劳动力参与率可能还依赖于家庭财富,六岁以下孩子的个数(这一点对于已婚妇女考虑进入劳动市场特别重要),孩子日托的便利程度,宗教信仰,福利事业的好坏,事业保险等等。即使这些变量的数据都可得到,我们也不会把他们都包括到模型中来,因为建模的目的不是包纳现实中的所有因素,而仅仅是一些显著因素。如果我们试图在回归模型中包括每一个可以想像到的变量,那么这个模型将会极为庞大以至于没有任何实际用处。最终选择的模型应该是对现实的合理的复制。在第14章中我们将进一步讨论这个问题,并且探讨如何建立和发展模型。

1.3.7 检验来自模型的假设

模型最终确定之后,我们进行**假设检验**(hypothesis testing)。即验证估计的模型是否有经济含义,以及用模型估计的结果是否与经济理论相符。例如,受挫工人假说假设劳动力参与与失业率之间负相关。这个假说与结果相符吗?我们统计的结果与假说相一致,因为估计得到的城市失业率系数为负。

然而,假设检验或许更复杂。在这个例子中,假设得知在先前的研究中,城市失业率的系数约为-1,那么得到的结果还会与假设一致吗?如果以式(1-3)这个模型为基础,我们可能得到一个结果,但是如果以式(1-5)模型为基础,则可能得到另一个结果。怎样解决这个问题呢?我们会在适当的章节中利用一些必要的工具来解决诸如此类的问题,但是需要提醒注意的是:根据某一特定的假说所得到的结果将依赖于最终所选择的模型。

还有一点,在回归分析中,我们不仅对模型参数的估计感兴趣,而且对检验来自于某个经济理论(或先验经验)的假设感兴趣。

1.3.8 运用模型进行预测

经过上述多个阶段,很自然地提出这样一个问题:我们用估计的模型干什么呢?比如说式(1-5)所表示的模型。一般地,我们用模型进行**预测**(prediction, forecasting)。举个例子,假设现在有1997年的城市失业率和平均小时工资的数据,分别是5.2和1.2。将其带入式(1-5),得到1997年城市劳动力参与率的预测值为49.26%。即,如果1997年的失业率为5.2%,真实

1 式(1-5)中的城市失业率系数和平均小时工资系数称为偏回归系数,我们将在第7章中讨论偏回归系数的确切含义。

2 可查阅任意一本微观经济学的标准教科书。一种直接判断结果的方法是:假设夫妻双方都参加工作,则在不影响家庭收入的前提下某一方收入的大量提高,将会促使另一方撤出劳动市场。

小时工资为12美元,则该年的城市劳动力参与率约为49%。当然,当得到1997年的城市劳动力参与率的真实值后,可与估计值做一比较,两者之间的差距代表了预测误差,从本质上说,我们希望预测误差尽可能小,这是否总是可能呢?我们将在第6章和第7章中回答这个问题。

总结一下经济计量分析的步骤:

序号	步骤	例子
1	理论的陈述	增加或受挫工人假说
2	收集数据	表1-1
3	理论的数学模型	$CLFPR = B_1 + B_2 CUNR$
4	理论的经济计量模型	$CLFPR = B_1 + B_2 CUNR + u$
5	参数估计	$CLFPR = 69.9 - 0.646CUNR$
6	检查模型的准确性	$CLFPR = 97.9 - 0.446CUNR - 3.86AHE82$
7	假设检验	$B_2 < 0$ 或 $B_2 > 0$
8	预测	给定CUNR和AHE82值,预测CLFPR的值

虽然我们仅用劳动经济学中的一个例子来阐述经济计量学的方法论,但需指出的是我们可用同样的步骤来分析任何领域中不同变量之间的定量关系。事实上,回归分析已用于政治、国际关系、心理学、社会学、气象学和其他许多领域。

1.4 全书结构

以上我们粗略地介绍了经济计量学的特性及研究范围。本书共分四个部分。

第一部分包括第2章、第3章和第4章。主要是为那些淡忘了统计知识的读者介绍概率和统计的基础知识。本章假定读者知道一些统计学的入门知识。

第二部分向读者介绍经济计量学的基本分析工具,称之为古典的线性回归模型(CLRM)。读者必须对古典线性回归模型有一个完整的理解,这样才能进行经济和商业领域的研究。

第三部分介绍了回归分析在实践中的运用,并讨论了当古典回归模型的假设不满足时须解决各类问题。

第四部分讨论了联立方程模型(第15章)。

在联立方程模型中,我们所关心的是方程中的应变量以及这些变量之间的关系。一个熟悉的例子是微观经济学中所提到的需求和供给函数。因为均衡数量和均衡价格是由需求曲线和供给曲线的交点决定的,因此我们必须同时考虑需求函数和供给函数。我们将在第15章中讨论均衡价格和均衡数量是如何确定的。

通览全书,初学者需要时刻注意,大多数论题的讨论是直接的,不牵涉数学定理和推论等。¹学生需记住:经济计量学入门课程就像已学过的统计入门课程一样,经济计量学主要讨论的也是估计和假设检验,所不同的或者说更有意思、更有用的是被估计或检验的参数并不仅仅是均值和方差,而且还有变量之间的关系,这也正是经济学和其他社会科学所关心的。

最后一点值得提出的是,通过运用一些价格便宜的计算机软件包,会使初学者易于掌握经济计量学这门课程。读者将会在本书的学习中遇到这样的软件。一旦熟悉了一两个标准的软件,你将会意识到学习经济计量学很有意思而且会对经济学有更好的理解。

¹ 一些定理和推论可参见本书作者著《经济计量学基础》(Basic Econometrics, 3d ed., McGraw-Hill, New York, 1995.)