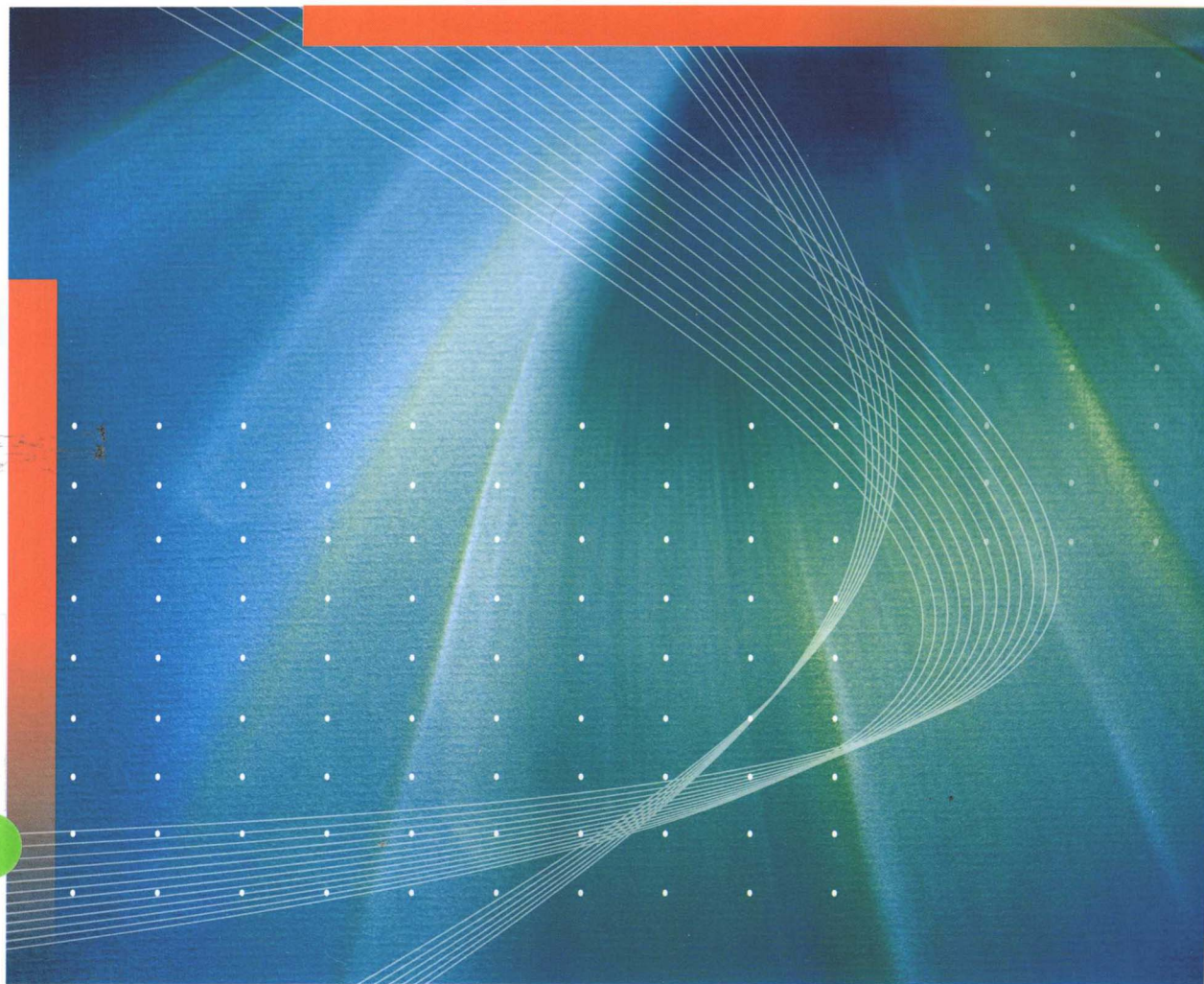


经济学、管理学类研究生教学用书

高级计量经济学 及Stata应用 (第二版)

陈强 编著



高等教育出版社

C14043926

F224.0-39

经济学、管理学类 研究生教学用书

12-2

高级计量经济学 及Stata应用 (第二版)

陈强 编著

GAOJI JILIANG JINGJIXUE JI STATA YINGYONG



北航

C1731327

高等教育出版社·北京

内容简介

本书较多地借鉴了现代计量经济学的最新发展,内容全面,除了介绍传统的横截面数据外,对面板数据(含长面板、动态面板、非线性面板)、时间序列(含VAR、单位根、协整)、自然实验、重复截面数据、GMM、自助法、蒙特卡罗法、分位数回归、门限回归、非参数估计、处理效应、空间计量、久期分析、贝叶斯估计等均做了较深入的分析。本书力图以生动的语言、较多的插图与经济意义来直观地解释计量方法,而又不失数学的严谨性。同时,结合目前欧美最为流行的Stata计量软件,及时地介绍相应的Stata命令与实例,为读者提供“一站式”服务。

本书适合普通高等学校经济学、管理学类或社科类硕士生、博士生与研究人员使用。为便于读者学习高级计量经济学,本书在内容安排上,假设读者已经学过微积分、线性代数与概率统计,但不要求学过本科阶段的计量经济学(学过更好)。

图书在版编目(CIP)数据

高级计量经济学及Stata应用 / 陈强编著. -- 2版

. -- 北京 : 高等教育出版社, 2014. 4

ISBN 978-7-04-032983-4

I. ①高… II. ①陈… III. ①计量经济学—高等学校—教材②经济计量分析—应用软件—高等学校—教材
IV. ①F224.0

中国版本图书馆CIP数据核字(2014)第023987号

策划编辑 施春花
插图绘制 尹文军

责任编辑 施春花
责任校对 李大鹏

封面设计 赵阳
责任印制 张泽业

版式设计 马敬茹

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100120
印 刷 北京市四季青双青印刷厂
开 本 787 mm×1092 mm 1/16
印 张 42.5 插页 1
字 数 1110千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2010年10月第1版
2014年4月第2版
印 次 2014年4月第1次印刷
定 价 59.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 32983-00

作者简介



陈强，男，1971年出生，山东大学经济学院教授，泰岳经济研究中心副主任（主持工作）。分别于1992年、1995年获北京大学经济学学士、硕士学位，后留校任教。2007年获美国Northern Illinois University数学硕士与经济学博士学位。主要研究领域为发展经济学、计量经济学、经济史与制度经济学。已独立发表论文于*Economica*, *Journal of Comparative Economics*, 《经济学（季刊）》、《世界经济》等国内外期刊。曾获中国数量经济学会、中国制度经济学会优秀论文奖、山东省高等学校优秀科研成果论文一等奖。现为美国经济学会、英国皇家经济学会会员，*Applied Economics*, 《经济学（季刊）》与《产业经济评论》的匿名审稿人。2010年入选教育部新世纪优秀人才支持计划。



第一版

读者反馈

- 在图书馆比较了很多本书，最后决定买这本，讲解很清楚很好。
——京东读者
- 这是一本难得的优秀教材，对研究生而言是莫大的幸事，对我们学习经典的计量经济学理论和软件实现都起着不可估量的作用，感谢陈强老师。
——京东读者
- 这本书虽然面世时间不长，但是字里行间透着作者的严谨与对内容讲解的透彻。即使仅学习它的理论部分都让人受益匪浅。Stata部分讲得也很实用。
——卓越读者
- 慢慢品读，慢慢演习。先说说优点：
 1. 语言顺畅，版面充实。如果按照一般教材比较宽的边距，这本书会很厚；
 2. 讲解非常之贴心——时不时地告诉你一些小技巧，语言非常柔和，一改一般教材严苛死板的说辞语气；
 3. 慢慢发现。
——卓越读者
- 本书接轨现代计量经济学，语言生动，例题翔实，尤其适合Stata初学者和提高者。
——当当读者
- 国内对于深度专题探讨Stata的书很少，所以看到这本就果断买了，不过感觉涉及内容非常多，但是每章还是比较简洁，回去好好研究。
——当当读者
- 陈老师的书不错，强烈推荐，我出国留学就带了一本书，就是您这本书。
——人大经济论坛读者
- 这本书真的难得，陈老师是真用心写的，复杂理论通俗化，Stata操作很详细，现在第二遍阅读，有好多收获。强烈推荐，强烈推荐。
——人大经济论坛读者

第二版前言

《高级计量经济学及 Stata 应用》自 2010 年 10 月出版以来,受到广大读者热烈欢迎,在此表示特别感谢。随着时间的推移,尽管网上依旧好评如潮,但第一版的不足越发清晰。当代计量经济学博大精深且发展迅猛,厚爱本书的读者也提出了不少合理化建议,要求增加这样或那样的内容。为此,从 2012 年暑期即着手第二版写作,冬去春来,到第二版初稿完成时,竟又接近济南的盛夏了。

第二版是第一版的重大升级(a significant upgrade)。第二版新增了四章全新内容,即非线性面板(Nonlinear Panels)、处理效应(Treatment Effects)、空间计量经济学(Spatial Econometrics)与久期分析(Duration Analysis)。已有章节也得到不少充实与完善,不胜枚举,部分新增内容参见二级子目录。比如,原来的“离散被解释变量”那一章,由于增加了太多内容,不得不分为三章。另外,Stata 软件也更新到 Stata 12,功能更为强大。

曾国藩云:带勇以能打仗为要。类似地,论文以创新为要,而教材则以易懂为要。为此,第二版秉承第一版的写作风格,在深入浅出、通俗易懂方面痛下功夫。时而感叹何必自苦,顶着海归的科研压力,却将宝贵时间投入此无底洞式的教材写作;但一想到我之费时费力可使得数以千计的学子省时省力,心下也便释然了。

自从 2012 年“千人计划学者”波士顿学院肖志杰教授在山东大学执教以来,给我教益良多;山东大学经济学院王永副教授、韩青博士对第二版新增内容进行了仔细的校对,在此一并感谢(当然文责自负)。第二版的修订得到山东大学研究生精品课程项目的资助,以及高等教育出版社编辑一如既往的支持,在此表示衷心感谢。第二版的错漏与不足之处,依然恳请读者指正。本书用到的所有数据集均可在我的个人网页(www.econ.sdu.edu.cn/tree/Faculty.php)下载,或通过 qiang2chen2@126.com 联系索取。正是您的一贯支持,给予我不断前行的动力。

陈 强

2013 年 10 月

第一版前言

本书是在山东大学经济学院硕士生、博士生《高级计量经济学》教案的基础上编著而成,适合高等学校经济管理类或社科类研究生与研究人员使用。

本书的主要特色如下:

(1) 接轨现代计量经济学。本书较多地借鉴了 Angrist and Pischke (2008), Baum (2006), Cameron and Trivedi (2005, 2010), Greene (2012), Hamilton (1994), Hayashi (2000), Hsiao (2003), Kennedy (2003), Poirier (1995), Verbeek (2004), Wooldridge (2010), 其中尤以 Hayashi (2000) 对本书的影响最深。

(2) 内容全面。除了介绍传统的横截面数据外,本书对面板数据(含长面板、动态面板)、时间序列(含 VAR、单位根、协整)、自然实验、重复截面数据、GMM、蒙特卡罗法、自助法、分位数回归、门限回归、非参数估计、贝叶斯估计等方法均进行了较深入的介绍。

(3) 计量理论与软件操作相结合。学习计量的学生,既需要了解计量原理,也需要知道如何在电脑上实现。为此,本书提供了“一站式”服务,在讲解每个估计方法后,随即介绍相应的 Stata 电脑操作及实例(Stata 为目前欧美最为流行的计量软件)。

(4) 本书力图以生动的语言、较多的插图与经济意义来直观地解释计量方法,而不仅仅是从数学推导到数学推导;另一方面,又不失数学的严谨性(部分证明放在附录)。

(5) 先修课不包括本科水平的计量经济学。在中国的国情下,不少经济类研究生并未学过本科阶段的计量经济学。因此,本书在内容安排上,假设读者已经学过微积分、线性代数与概率统计,但不要求学过本科阶段的计量经济学(当然,如果学过更好)。

学习计量经济学不是一件容易的事(我也经历过,以后还要经历),但回报却很丰厚(其为实证研究不可或缺之工具),可以说是“高投入、高产出”。对于许多初学者而言,或许计量经济学难就难在使用了较多的数学^①。但数学只是一种语言,而任何数学符号原则上都可以“翻译”为汉语。事实上,看似复杂的数学公式后面,常常有着非常直观的道理。因此,只要渐渐地掌握数学这门语言,学会看数学符号背后的含义,学习计量也就不难了。

套用一個參禪的故事,学习计量大致可以分为三个境界。第一境界是“见山是山,见水是水”,第二境界是“见山不是山,见水不是水”,第三境界是“见山又是山,见水又是水”。在第一阶段,以为计量就是作最小二乘回归而已,自然不在话下。在第二阶段,开始体会到计量的精妙之处,心中时时产生疑问。在第三阶段,通过考前复习及实践应用,对计量的理论与方法逐渐融会贯通,进而内化为熟练掌握的工具。其中,尤以第二阶段最为漫长。“取法乎上,仅得其中”。貌似难懂之处,其实正是取得进步的地方。这是一个“痛并快乐着”的过程,时常伴有顿悟之喜悦。我曾为学生们写了一首打油诗,收录在此,以博一笑。

^① 对于从理科转学经济学的同学来说,可能面临另一问题,即如何更快地建立经济学的直觉,加深对经济意义的理解。

计量啊计量

辛苦读研学计量,推来导去费思量。
只因成绩盼优良,折腾数据叫爹娘。
爱恨交加为那般,实证研究是桥梁。
此情可待成佳酿,奈何当下心已凉。

在本书出版之际,特别要感激以下曾教授过我统计学或计量经济学的授业恩师们(以时间先后为序):范培华、胡健颖、靳云汇、陈良焜(北京大学);Dale Poirier (University of California, Irvine);Susan Porter-Hudak, Nader Ebrahimi, Mohsen Pourahmadi (Northern Illinois University)。没有他们的谆谆教诲,本书是绝不可能完成的。

山东大学经济学院的领导与同事们对本书的写作给予了大力支持与鼓励,另外,2008级与2009级硕士与博士生在听课过程中提出了很多好建议,在此一并感谢。山东大学经济学院王永老师,博士生韩青,硕士生李欢、李晶、林兴兰、戚传萍、吴振华、张甜等参与了校对,陈丽云同学协助制作了部分插图,在此表示衷心感谢(当然,文责自负)。最后,要特别感谢高等教育出版社的于明编辑、边晓娜编辑及其同仁们,为本书的撰写提出了许多宝贵意见,并付出了辛勤的劳动。

正如苏格拉底所说,学习是学生自我发现的过程,而教师不过是助产婆,但愿这本教科书能起到这个作用。

当然,由于本人知识有限,对于本书中的错误与不足之处,恳请各位老师与同学及时指出,以便在本书的网站上公布勘误表,并在未来的版本中更新。邮箱为 qiang2chen2@126.com。

陈 强

2010年2月于济南



北航

C1731327

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010) 58581897 58582371 58581879

反盗版举报传真 (010) 82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第 1 章 绪论	1	5.4 统计量的大样本性质	52
1.1 什么是计量经济学	1	5.5 渐近分布的推导	53
1.2 经济数据的特点与类型	2	5.6 随机过程的性质	53
第 2 章 概率统计回顾	3	5.7 大样本 OLS 的假定	57
2.1 概率与条件概率	3	5.8 OLS 的大样本性质	58
2.2 分布与条件分布	4	5.9 线性假设的大样本检验	60
2.3 随机变量的数字特征	5	5.10 大样本 OLS 的 Stata 命令及 实例	61
2.4 迭代期望定律	8	习题	63
2.5 随机变量无关的三个层次概念	9	附录	63
2.6 常用连续型统计分布	9	第 6 章 最大似然估计法	66
2.7 统计推断的思想	11	6.1 最大似然估计法的定义	66
习题	12	6.2 线性回归模型的最大似然估计	68
附录	12	6.3 最大似然估计的数值解	69
第 3 章 小样本 OLS	13	6.4 信息矩阵与无偏估计的最小 方差	70
3.1 古典线性回归模型的假定	13	6.5 最大似然法的大样本性质	71
3.2 OLS 的代数推导	14	6.6 最大似然估计量的渐近协方差 矩阵	74
3.3 OLS 的几何解释	17	6.7 三类渐近等价的统计检验	75
3.4 拟合优度	17	6.8 准最大似然估计法	78
3.5 OLS 的小样本性质	18	6.9 对正态分布假设的检验	80
3.6 对单个系数的 t 检验	20	6.10 最大似然估计法的 Stata 命令及 实例	80
3.7 对线性假设的 F 检验	23	习题	84
3.8 F 统计量的似然比原理表达式	25	附录	84
3.9 分块回归与偏回归 (选读)	26	第 7 章 异方差与 GLS	87
3.10 预测	27	7.1 异方差的后果	87
习题	28	7.2 异方差的例子	87
附录	29	7.3 异方差的检验	88
第 4 章 Stata 简介	30	7.4 异方差的处理	90
4.1 为什么使用 Stata	30	7.5 处理异方差的 Stata 命令及实例	93
4.2 Stata 的窗口	30	7.6 Stata 命令的批处理	96
4.3 Stata 操作实例	31	习题	98
4.4 Stata 命令库的更新	46	附录	98
4.5 进一步学习 Stata 的资源	47	第 8 章 自相关	100
习题	48	8.1 自相关的后果	100
第 5 章 大样本 OLS	49		
5.1 为何需要大样本理论	49		
5.2 随机收敛	49		
5.3 大数定律与中心极限定理	51		

8.2	自相关的例子	101	11.7	双变量 Probit 模型(选读)	187
8.3	自相关的检验	101	11.8	部分可观测的双变量 Probit 模型(选读)	189
8.4	自相关的处理	103	习题		190
8.5	处理自相关的 Stata 命令及实例	108	第 12 章 多值选择模型	192	
习题		115	12.1	多项 Logit 与多项 Probit	192
第 9 章 模型设定与数据问题	116		12.2	条件 Logit 模型	193
9.1	遗漏变量	116	12.3	混合 Logit 模型	193
9.2	无关变量	117	12.4	嵌套 Logit	205
9.3	建模策略:“由小到大”还是“由大到小”	118	习题		208
9.4	解释变量个数的选择	118	第 13 章 排序与计数模型	209	
9.5	对函数形式的检验	120	13.1	排序模型	209
9.6	多重共线性	123	13.2	泊松回归	211
9.7	极端数据	124	13.3	负二项回归	213
9.8	虚拟变量	126	13.4	零膨胀泊松回归与负二项回归	215
9.9	经济结构变动的检验	127	13.5	计数模型的 Stata 实例	215
9.10	缺失数据与线性插值	132	习题		222
9.11	变量单位的选择	133	第 14 章 受限被解释变量	223	
习题		133	14.1	断尾回归	223
附录		133	14.2	零断尾泊松回归与负二项回归	226
第 10 章 工具变量, 2SLS 与 GMM	135		14.3	随机前沿模型(选读)	228
10.1	解释变量与扰动项相关的例子	135	14.4	偶然断尾与样本选择	235
10.2	工具变量法作为一种矩估计	138	14.5	归并回归	238
10.3	二阶段最小二乘法	140	14.6	归并数据的两部分模型	243
10.4	有关工具变量的检验	141	14.7	含内生解释变量的 Tobit 模型(选读)	246
10.5	GMM 的假定	146	习题		248
10.6	GMM 的推导	147	附录		248
10.7	GMM 的大样本性质	148	第 15 章 短面板	250	
10.8	如何获得工具变量	151	15.1	面板数据的特点	250
10.9	MLE 也是 GMM	152	15.2	面板数据的估计策略	251
10.10	工具变量法的 Stata 命令及实例	153	15.3	混合回归	252
习题		167	15.4	个体固定效应模型	252
附录		167	15.5	时间固定效应	253
第 11 章 二值选择模型	169		15.6	一阶差分法	254
11.1	离散被解释变量的例子	169	15.7	随机效应模型	254
11.2	二值选择模型	169	15.8	组间估计量	255
11.3	二值选择模型的微观基础	177	15.9	拟合优度的度量	255
11.4	二值选择模型中的异方差问题	178	15.10	非平衡面板	256
11.5	稀有事件偏差(选读)	179	15.11	究竟该用固定效应还是随机效应模型	257
11.6	含内生变量的 Probit 模型(选读)	183	15.12	个体时间趋势	257
			15.13	短面板的 Stata 命令及实例	258

习题	271	19.1 蒙特卡罗法的思想与用途	346
第 16 章 长面板与动态面板	272	19.2 蒙特卡罗法实例:模拟中心极限定理	347
16.1 长面板的估计策略	272	19.3 蒙特卡罗法实例:服从卡方分布的扰动项	348
16.2 面板校正标准误	272	19.4 蒙特卡罗积分	349
16.3 仅解决组内自相关的 FGLS	274	19.5 最大模拟似然法与模拟矩估计	350
16.4 全面 FGLS	278	19.6 自助法的思想与用途	351
16.5 组间异方差的检验	279	19.7 自助法的分类	352
16.6 组内自相关的检验	280	19.8 使用自助法估计标准误	352
16.7 组间同期相关的检验	281	19.9 使用自助法进行区间估计	353
16.8 变系数模型	283	19.10 使用自助法进行假设检验	353
16.9 面板工具变量法	287	19.11 自助法的一致性(选读)	354
16.10 豪斯曼-泰勒估计量(选读)	288	19.12 异方差情况下的自助法	354
16.11 动态面板	289	19.13 面板数据与时间序列的自助法	355
16.12 动态面板的 Stata 命令及实例	291	19.14 自助法的 Stata 命令	355
16.13 偏差校正 LSDV 法	300	19.15 使用自助法进行稳健的豪斯曼检验	356
16.14 重复截面数据与组群分析	301	习题	358
习题	302	附录	358
第 17 章 非线性面板	303	第 20 章 平稳时间序列	361
17.1 面板二值选择模型	303	20.1 时间序列的数字特征	361
17.2 面板二值选择模型的随机效应估计	304	20.2 自回归模型	362
17.3 面板二值选择模型的固定效应估计	305	20.3 移动平均模型	364
17.4 面板二值选择模型的 Stata 实例	307	20.4 ARMA	364
17.5 面板泊松回归	313	20.5 自回归分布滞后模型	365
17.6 面板负二项回归	314	20.6 ARMA 模型的 Stata 命令及实例	366
17.7 面板计数模型的 Stata 实例	315	20.7 误差修正模型	371
17.8 面板 Tobit	325	20.8 MA(∞)与滞后算子	372
17.9 面板随机前沿模型	327	20.9 向量自回归过程	375
习题	332	20.10 VAR 的脉冲响应函数	377
第 18 章 随机实验与自然实验	334	20.11 预测误差的方差分解	380
18.1 实验数据	334	20.12 格兰杰因果检验	381
18.2 理想的随机实验	335	20.13 面板格兰杰因果检验	381
18.3 引入更多的解释变量	335	20.14 VAR 的 Stata 命令及实例	381
18.4 随机实验执行过程中可能出现的问题	336	20.15 季节调整	399
18.5 自然实验	337	习题	407
18.6 双重差分法	339	第 21 章 单位根与协整	409
18.7 三重差分法	343	21.1 非平稳序列	409
18.8 观测数据的处理效应	344	21.2 ARMA 的平稳性	410
习题	345	21.3 VAR 的平稳性	411
第 19 章 蒙特卡罗法与自助法	346	21.4 单位根所带来的问题	411

21.5	单位根检验与平稳性检验	414	25.1	非线性最小二乘法	503
21.6	单位根检验的 Stata 实例	418	25.2	非线性回归的 Stata 命令及 实例	504
21.7	面板单位根检验	422	25.3	门限回归	505
21.8	协整的思想与初步检验	432	25.4	面板数据的门限回归	507
21.9	Beveridge-Nelson 分解公式	433	25.5	门限回归的计算机操作 习题	508
21.10	协整的定义与最大似然估计	434	第 26 章 分位数回归	509	
21.11	协整分析的 Stata 实例	437	26.1	为什么需要分位数回归	509
习题		445	26.2	总体分位数	509
附录		445	26.3	样本分位数	510
第 22 章 自回归条件异方差模型	447		26.4	分位数回归的估计方法	512
22.1	条件异方差模型的例子	447	26.5	分位数回归的 Stata 命令及 实例	513
22.2	ARCH 模型的性质	448	习题		517
22.3	ARCH 模型的 MLE 估计	449	第 27 章 非参数与半参数估计	518	
22.4	GARCH 模型	450	27.1	为什么需要非参数与半参数 估计	518
22.5	何时使用 ARCH 或 GARCH 模型	451	27.2	对密度函数的非参数估计	518
22.6	ARCH 与 GARCH 模型的扩展	451	27.3	核密度估计的性质	520
22.7	ARCH 与 GARCH 的 Stata 命令及 实例	453	27.4	最优带宽	521
22.8	多维 GARCH 模型(选读)	460	27.5	多元密度函数的核估计	523
习题		467	27.6	非参数核回归	523
第 23 章 似不相关回归	468		27.7	多元核回归	525
23.1	单一方程估计与系统估计	468	27.8	k 近邻回归	525
23.2	似不相关回归的假定	468	27.9	局部线性回归	526
23.3	SUR 的 FGLS 估计	470	27.10	非参数估计的 Stata 命令及 实例	526
23.4	SUR 的假设检验	471	27.11	半参数估计	530
23.5	似不相关回归的 Stata 命令及 实例	471	习题		533
23.6	变系数面板数据的 SUR 估计	475	附录		533
习题		478	第 28 章 处理效应	537	
附录		479	28.1	处理效应与选择难题	537
第 24 章 联立方程模型	482		28.2	通过随机分组解决选择难题	539
24.1	联立方程模型的结构式与 简化式	482	28.3	依可测变量选择	539
24.2	联立方程模型的识别	483	28.4	匹配估计量的思想	540
24.3	单一方程估计法	486	28.5	倾向得分匹配	542
24.4	三阶段最小二乘法	487	28.6	倾向得分匹配的 Stata 实例	545
24.5	三阶段最小二乘法的 Stata 实例	489	28.7	偏差校正匹配估计量	555
24.6	结构 VAR	493	28.8	双重差分倾向得分匹配	557
24.7	SVAR 的 Stata 实例	496	28.9	断点回归的思想	559
习题		502	28.10	精确断点回归	561
第 25 章 非线性回归与门限回归	503		28.11	模糊断点回归	563
			28.12	断点回归的 Stata 实例	565

28.13 处理效应模型	570	30.12 不可观测的异质性	613
习题	574	30.13 其他久期分析模型	614
第 29 章 空间计量经济学	575	30.14 久期分析的 Stata 命令及 实例	615
29.1 地理学第一定律	575	习题	630
29.2 空间权重矩阵	575	第 31 章 贝叶斯估计简介	631
29.3 空间自相关	578	31.1 贝叶斯估计的思想	631
29.4 空间自回归模型	583	31.2 贝叶斯定理	631
29.5 空间杜宾模型	586	31.3 贝叶斯估计的一个例子	632
29.6 空间误差模型	586	31.4 基于后验分布的统计推断	634
29.7 一般的空间计量模型	589	31.5 先验分布的选择	635
29.8 含内生解释变量的 SARAR 模型	592	31.6 多元回归的贝叶斯分析	637
29.9 空间面板模型	593	31.7 马尔可夫链蒙特卡罗法	639
29.10 空间计量方法的局限性	598	习题	640
第 30 章 久期分析	599	第 32 章 如何做规范的实证研究	641
30.1 久期数据的处理方法	599	32.1 计量理论与现实数据	641
30.2 风险函数	599	32.2 实证研究的主要步骤	642
30.3 久期数据的归并问题	601	32.3 实证论文的结构	644
30.4 描述性分析	602	32.4 计量实践的十诫	645
30.5 久期模型的最大似然估计	603	32.5 结束语	646
30.6 比例风险模型	604	习题	646
30.7 加速失效时间模型	606	附录:常用数据来源	647
30.8 Cox 模型	607	参考书目	649
30.9 比例风险模型的设定检验	610	数学符号	664
30.10 分层 Cox 模型	611	英文缩写	666
30.11 随时间而变的解释变量	612		

第1章 绪论

1.1 什么是计量经济学

顾名思义,“计量经济学”(Econometrics,也译为“经济计量学”)就是运用概率统计的方法对经济变量之间的(因果)关系进行定量分析的科学。之所以把“因果”两个字加括号,是因为计量经济学常常不足以确定经济变量之间的因果关系(由于实验数据的缺乏)^①,另一方面,大多数实证分析的目的恰恰正是要确定变量之间的因果关系(即是否 X 导致 Y),而非仅仅是相关关系^②。因此,在学习与应用计量经济学的过程中,很有必要时时以“因果关系”作为思考的框架与指引。

比如,你看到街上人们带伞,于是预测今天要下雨。这是一种相关关系。然而,“人们带伞”并不是造成“下雨”的原因。因此,计量分析必须建立在经济理论的基础之上。然而,即使有理论基础,因果关系依然不好分辨。首先,可能存在“逆向因果关系”(reverse causality)。比如,FDI(外商直接投资)能促进经济增长,但也可能是FDI被吸引到增长潜力高的国家或地区。其次,也可能是被遗漏的第三个变量(Z)对这两个变量(X, Y)同时产生了作用,参见图1.1。

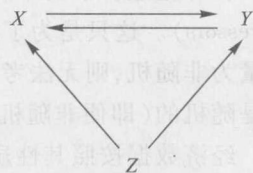


图1.1 可能的因果关系

作为一个例子,考虑决定教育投资回报率(returns to schooling)的因素:

$$\ln W_i = \alpha + \beta S_i + \varepsilon_i \quad (1.1)$$

其中, $\ln W$ (工资收入的自然对数)为“被解释变量”(dependent variable), S (教育年限)为“解释变量”(explanatory variable 或 regressor)、“自变量”(independent variable)或“协变量”(covariate), ε 为“随机扰动项”(stochastic disturbance)或“误差项”(error term),而下标 i 表示第 i 个观测值(即个体 i)。

如果用数据估计这个简单的一元回归,其结果一般会显示,对数工资收入与受教育年限显著正相关,而且教育投资回报率 β 还挺高。然而,一个人的工资收入也与能力有关,但能力不能直接观测,而能力高的人通常选择接受更多教育。因此,在这个简单的回归中,教育的高回报率其实包含了对能力的回报。

另外,影响工资收入的因素还可能包括工作经验、毕业学校、人种、性别、外貌等。因此,需要引入更多的“控制变量”(control variables),也就是多元回归的方法,才能较准确地估计我们“感兴趣的参数”(parameters of interest),即本例中的教育投资回报率 β 。然而,现实中总有某些相关

① 计量经济学家 Guy Orcutt 曾说过,“做计量经济学就像试图通过播收音机来研究电的规律”(Doing econometrics is like trying to learn the laws of electricity by playing the radio),可见其难度。

② 如果使用计量经济学做预测,则只需要相关关系,不必顾及因果关系。

的变量无法观测,即存在“遗漏变量”(omitted variables),而这些遗漏变量统统被纳入随机扰动项 ε_i 中了。

随机扰动项 ε_i 中还可能包含哪些因素呢? 如果真实模型(true model)为

$$\ln W_i = \alpha + \beta S_i + \gamma S_i^2 + \varepsilon_i \quad (1.2)$$

那么 γS_i^2 也被纳入扰动项中了(可以视为广义的遗漏变量)。如果变量测量得不准确,则测量误差也被放入扰动项中了。总之,扰动项就像是一个“垃圾桶”,所有你不想要、无法把握的东西都往里面扔。另一方面,我们又希望扰动项有很好的性质。在很多情况下,这是自相矛盾的。西方有个谚语“The devil is in the details”,意即“魔鬼就在细节中”^①。套用到计量经济学上来,或许可以说“The devil is in the error term”,意即魔鬼就在扰动项中。计量经济学的很多玄妙之处就在于扰动项。如果真正理解了扰动项,也就加深了对计量经济学的理解。

1.2 经济数据的特点与类型

由于在经济学中通常无法像自然科学那样做“控制实验”(controlled experiment),故经济数据一般不是“实验数据”(experimental data)^②,而是自然发生的“观测数据”(observational data)。由于个人行为的随机性,所有经济变量原则上都是随机变量^③。

在计量经济学的本科课程中,为了简单起见,有时假设解释变量是非随机的、固定的(fixed regressors)。这只是为了教学法上的方便,却给更深入的理论探讨带来了不便。比如,如果解释变量为非随机,则无法考虑其与扰动项的相关性。因此,在这本研究生水平的教材中,所有变量都是随机的(即便非随机的常数,也可以视为退化的随机变量)。

经济数据按照其性质,可大致分成以下三种类型。

(1) 横截面数据(cross-sectional data,简称截面数据):指的是多个经济个体的变量在同一时点上的取值。比如,2012年中国各省的GDP。

(2) 时间序列数据(time series data):指的是某个经济个体的变量在不同时点上的取值。比如,在1978—2012年山东省每年的GDP。

(3) 面板数据(panel data):指的是多个经济个体的变量在不同时点上的取值。比如,在1978—2012年中国各省每年的GDP。

本书介绍的计量经济理论将包括以上三种数据类型,并使用国际上最为流行的Stata计量软件。在此之前,我们首先要对概率统计进行简短的回顾,并引入一些新概念(比如,均值独立、迭代期望定律)。

① 比如,对于一份冠冕堂皇的合同,可能让你以后吃尽苦头的正是那些合同附录中的小字部分。

② 第18章将讨论随机实验与自然实验数据。

③ 你能举出哪些经济数据(变量)不是随机变量吗?

第2章 概率统计回顾

2.1 概率与条件概率

1. 概率

假如街上有个老太太问你：“什么是概率？”那么你会怎么回答呢？若回答“事情发生的可能性”，老太太可能反问你：“说‘可能性’不就行了，为什么又造了一个新词‘概率’？”也许她会问你一个更具体的问题：“天气预报说明天70%概率下雨。这是啥意思？”也许你想说，这表明“明天70%的时间会下雨”，但更好的答案则是：“如果有100天的天气预报都报了70%的概率明天降雨，则大约有70天会下雨。”

总之，可以将“概率”理解为在大量重复实验下，事件发生的频率趋向的某个稳定值。记事件“下雨”为 A ，其发生的“概率”(probability)为 $P(A)$ ^①。

2. 条件概率

例 已知明天会出太阳，下雨的概率有多大？

记事件“出太阳”为 B ，则在出太阳的前提条件下降雨的“条件概率”(conditional probability)为

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)} \quad (2.1)$$

其中，“ \cap ”表示事件的交集(intersection)，故 $P(A \cap B)$ 为“太阳雨”的概率，参见图2.1。条件概率是计量经济学的重要概念之一。

例 股市崩盘的可能性为无条件概率；而在已知经济陷入严重衰退的情况下，股市崩盘的可能性则为条件概率。

3. 独立事件

如果条件概率等于无条件概率，即 $P(A|B) = P(A)$ ，即 B 是否发生不影响 A 的发生，则称 A, B 为相互独立的随机事件。此时， $P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$ ，故

$$P(A \cap B) = P(A)P(B) \quad (2.2)$$

也可以将此式作为独立事件的定义。

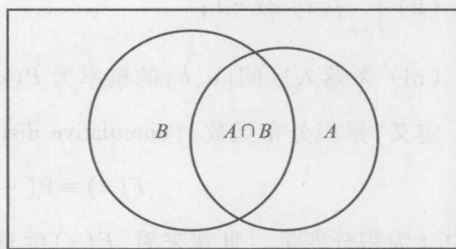


图2.1 条件概率示意图

^① 这里不讨论概率的严格公理化定义。