

数
量
经
济
学
系
列
丛
书

计量经济学

张晓峒 著



清华大学出版社

数量经济学系列丛书

计量经济学

张晓峒 著



清华大学出版社
北京

内 容 简 介

这是一本面向经济类、管理类本科生和研究生的计量经济学教材,内容主要包括回归模型、时间序列 ARIMA 模型、单位根检验、误差修正模型和面板数据模型等。

本书在本领域内第一次增加蒙特卡洛模拟结果讨论统计量的分布特征,增强读者对统计量分布特征的理解。书中每一个知识点都用简练的语言介绍怎样用计量经济学软件 EViews 9 实现计算。书中所提供的案例基本上都是中国建模案例,为计量经济学理论与分析中国经济实际相结合提供切实范例。书中提供多种图(包括散点图、序列图、分布图等),辅助对所研究问题的理解。书中所用全部数据可免费下载。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

计量经济学/张晓峒著. —北京:清华大学出版社,2017
(数量经济学系列丛书)
ISBN 978-7-302-46583-6

I. ①计… II. ①张… III. ①计量经济学 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2017)第 030219 号

责任编辑:张 伟
封面设计:常雪影
责任校对:宋玉莲
责任印制:李红英

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62770175

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:29.5 字 数:646千字

版 次:2017年5月第1版 印 次:2017年5月第1次印刷

印 数:1~5000

定 价:49.00元

产品编号:056521-01

前 言

本书主要由回归模型、时间序列 ARIMA 模型、单位根检验和误差修正模型、面板数据模型等内容组成。读者对象是大专院校的本科生、硕士研究生以及从事经济、管理等领域研究的学者、工作者和教师。若讲授完本书全部内容再加上安排上机时间,大约需要 100 学时。

本书共分 14 章。其中第 1~9 章基本上属于经典计量经济学的内容。第 10 章介绍时间序列 ARIMA 模型,第 11 章和第 12 章介绍非平稳时间序列以及单位根检验。第 13 章介绍单方程误差修正模型。第 14 章介绍面板数据模型。第 11~14 章属于计量经济学中比较新的内容,系统介绍第 11~14 章内容的本科生教材以往并不多见。

本书具有如下一些特点。

用现代手段和视角分析经典计量经济学知识和非平稳相关统计量的分布特征。例如,对 OLS 回归估计量、异方差、自相关、多重共线性、动态模型回归系数估计量分布的讨论,既给出理论推导,又给出蒙特卡洛模拟结果,将会使读者更容易理解所学习的内容。蒙特卡洛模拟方法是分析统计量分布特征的重要方法,但是对于十多年前的国内来说,是不可能做到的事情,主要是受计算机运算速度的影响。现在,计算机运算速度早已今非昔比,所以是用蒙特卡洛模拟方法研究、讲授统计量分布特征的时候了。本书还给出了非平稳时间序列建模,虚假回归,虚假相关,单整、协整相关统计量分布特征的蒙特卡洛模拟结果。蒙特卡洛模拟方法有助于读者对统计量分布特征的理解。

把时间序列 ARIMA 模型引入计量经济学教材。从当前看,ARIMA 模型是计量经济学理论的重要组成部分。而 20 世纪的计量经济学教材则主要以介绍回归模型为主,很少或根本不涉及 ARIMA 模型部分。计量经济学理论发展到今天,如果不学习 ARIMA 模型,则单位根检验,单整、协整理论,组合(regARIMA)模型,误差修正模型等知识根本无法掌握。

在介绍计量经济模型的方式上坚持 3 个环节并举:介绍计量经济模型的理论知识,介绍与其相联系的典型案例分析,介绍与案例分析相对应的计量经济建模的 EViews 9 操作。如果读者自己再运用专用软件 EViews 和 STATA 进行练习,则一定会完美掌握本书所提供的计量经济学知识。

凡是样本容量不大的数据在书中相应例子和案例的位置都已经给出,而书中全部例子和案例的样本数据则以 EViews 和 STATA 数据文件的形式在清华大学出版社官方网站(<http://www.tup.com.cn>)和书中的二维码上给出。读者可免费下载,也可以直接向作者索取。所有数据文件的编号都与书中例子和案例的编号相对应。用 li 表示例子的编号,case 表示案例的编号。比如,li 13-3 表示第 13 章第 3 个例子;case 15-1 表示第 15 章第 1 个案例。STATA 文件名与 EViews 数据文件的命名方法相同。



EViews 数据文件



STATA 数据文件

书中所提供的案例基本上都是中国建模案例。因为读者对中国的国情最为熟悉,用中国案例分析建模过程和估计结果,读者最容易理解,同时也为计量经济学理论与分析中国经济实际相结合提供切实的范例。

书中在表达模型时,所有的变量都用英文字母,所有模型的参数都用希腊字母表示。

书的最后提供三个附录。附录 A 给出推断统计学与矩阵知识的简要介绍,供读者查阅。附录 B 给出 15 个假设检验用表。附录 C 给出 EViews 9 使用简介,有助于读者对计量经济学软件 EViews 9 的运用与掌握。书中每一个知识点都给出 EViews 9 操作的简要说明。

非平稳序列相关统计量极限分布的推导过程本书未给出,作者认为已超出了本科生和硕士研究生的知识范围,如果读者感兴趣,可以参考更高层次的计量经济学著作。

本书第 4 章和第 5 章初稿由赵娜博士撰写,其余部分则均由张晓峒撰写。张晓峒为全书最终定稿。

本书是作者在多年教学讲稿基础之上撰写而成的。徐鹏博士、何永涛博士、郭小稚博士、涂晓枫博士生、刘笑时博士生等参与了本书的案例数据收集工作,以及计算机操作方法的整理等大量工作,博士生梁方参与了第 11 章编程工作,在此表示感谢。

本书在撰写过程中得到清华大学出版社的支持,在此向清华大学出版社表示感谢。本书在出版过程中策划编辑张伟付出很多,在此一并表示感谢。

书中难免存在不足和错误,还请读者不吝赐教、指正。

张晓峒

zhangnk710@126.com

2017 年 1 月 10 日

目 录

第 1 章 一元线性回归模型	1
1.1 计量经济学简介与建模步骤	1
1.2 模型的建立及其假定条件	3
1.2.1 建立模型的意义	3
1.2.2 一元线性回归模型的定义	3
1.2.3 一元线性回归模型的经济含义与特征	3
1.2.4 模型的假定条件	4
1.3 一元线性回归模型的参数估计	5
1.3.1 估计方法初探	5
1.3.2 最小二乘估计法原理	6
1.3.3 最小二乘估计的计算	6
1.4 $y_t, \hat{\beta}_1$ 和 $\hat{\beta}_0$ 的分布	7
1.4.1 y_t 的分布	7
1.4.2 $\hat{\beta}_1$ 的分布	8
1.4.3 $\hat{\beta}_0$ 的分布	8
1.5 σ^2 的估计	10
1.6 最小二乘估计量的统计性质	10
1.6.1 线性特性	10
1.6.2 无偏性	11
1.6.3 最小方差性	11
1.6.4 渐近无偏性	13
1.6.5 一致性	13
1.7 最小二乘回归函数的性质	13
1.8 拟合优度的测量	14
1.9 回归系数的显著性检验	16
1.10 回归系数的置信区间	17
1.11 单方程回归模型的预测	17
1.11.1 单个 y_{T+1} 的点预测	18
1.11.2 单个 y_{T+1} 的区间预测	18
1.11.3 $E(y_{T+1})$ 的区间预测	20
1.12 相关分析	21

1.12.1	相关的定义与分类	21
1.12.2	相关系数	21
1.12.3	线性相关系数的局限性	24
1.12.4	简单相关系数的检验	24
1.13	回归系数 $\hat{\beta}_1$ 与相关系数 r 的关系	25
1.14	案例分析	25
第2章	多元线性回归模型	32
2.1	多元线性回归模型及其假定条件	32
2.1.1	模型的建立	32
2.1.2	模型的假定条件	33
2.2	最小二乘法	34
2.3	最小二乘估计量的特性	35
2.3.1	线性特性	35
2.3.2	无偏特性	35
2.3.3	最小方差性	35
2.3.4	渐近无偏性	36
2.3.5	一致性	36
2.4	残差的方差	37
2.5	\mathbf{Y} 与最小二乘估计量 $\hat{\beta}$ 的分布	38
2.6	多重可决系数(多重确定系数)	38
2.6.1	总平方和、回归平方和与残差平方和	38
2.6.2	多重确定系数 R^2	39
2.6.3	调整的多重确定系数 \bar{R}^2	39
2.7	F 检验	40
2.8	t 检验和回归系数的置信区间	41
2.9	预测	43
2.9.1	y_{T+1} 的点预测	43
2.9.2	单个 y_{T+1} 的置信区间预测	43
2.9.3	$E(y_{T+1})$ 的置信区间预测	43
2.9.4	预测的评价指标	44
2.10	多元线性回归计算举例	45
2.11	偏相关与复相关	51
2.11.1	偏相关	51
2.11.2	复相关	54
2.12	案例分析	55
2.13	实际建模过程中应该注意的若干问题	57

第 3 章 可线性化的非线性回归模型	63
3.1 可线性化的 7 种非线性函数	63
3.1.1 幂函数模型	63
3.1.2 指数函数模型	66
3.1.3 对数函数模型	69
3.1.4 双曲线函数模型	71
3.1.5 多项式函数模型	74
3.1.6 生长曲线函数模型	77
3.1.7 龚伯斯曲线函数模型	81
3.2 可线性化的非线性模型综合案例	82
3.3 可线性化的非线性模型一览表	90
第 4 章 特殊解释变量	92
4.1 虚拟变量	92
4.1.1 测量截距移动	92
4.1.2 测量斜率变化	98
4.2 工具变量	101
4.2.1 工具变量在一元线性回归模型中的应用	102
4.2.2 工具变量在多元线性回归模型中的应用	105
4.3 滞后变量	106
4.3.1 分布滞后模型	106
4.3.2 自回归模型	108
4.4 随机解释变量	108
第 5 章 异方差	110
5.1 同方差假定	110
5.2 异方差的表现与来源	111
5.3 模型存在异方差的后果	113
5.4 异方差检验	115
5.4.1 定性分析异方差	115
5.4.2 戈德菲尔德-匡特检验	115
5.4.3 怀特检验	116
5.4.4 戈列瑟检验	117
5.4.5 ARCH(自回归条件异方差)检验	118
5.5 克服异方差的方法	118
5.5.1 用解释变量或解释变量的算术根除原回归式克服异方差	118
5.5.2 用戈列瑟检验式克服异方差	119
5.5.3 通过对数据取自然对数消除异方差	120
5.5.4 克服异方差的矩阵描述	120

5.6	案例分析	122
第6章	自相关	129
6.1	非自相关假定	129
6.2	自相关的来源与后果	132
6.3	自相关检验	135
6.3.1	图示法	135
6.3.2	DW 检验法	136
6.3.3	LM 检验(亦称 BG 检验)法	138
6.3.4	回归检验法	138
6.4	自相关的解决方法	139
6.5	克服自相关的矩阵描述	140
6.6	自相关系数的估计	142
6.7	案例分析	143
第7章	多重共线性	150
7.1	非多重共线性假定	150
7.2	多重共线性的来源	151
7.3	多重共线性的后果	152
7.3.1	完全多重共线性对参数估计的影响	152
7.3.2	近似共线性对参数估计的影响	153
7.3.3	多重共线性后果的矩阵描述与蒙特卡洛模拟	155
7.4	多重共线性的检测	156
7.5	多重共线性的解决方法	159
7.5.1	直接合并解释变量	159
7.5.2	利用已知信息合并解释变量	160
7.5.3	增加样本容量或重新抽取样本	160
7.5.4	合并截面数据与时间序列数据	161
7.5.5	剔除引起多重共线性的变量	162
7.6	案例分析	164
7.7	多重共线性与解释变量的不正确剔除	169
7.8	违反模型假定条件的其他几种情形	170
7.8.1	被解释变量存在测量误差	170
7.8.2	被解释变量、解释变量同时存在测量误差	170
7.8.3	随机解释变量	171
7.8.4	模型的设定误差	171
第8章	联立方程模型	174
8.1	联立方程模型的概念	174
8.2	联立方程模型的分类	174

8.2.1	结构模型	174
8.2.2	简化型模型	177
8.2.3	递归模型	179
8.3	联立方程模型的识别	179
8.3.1	识别概念	179
8.3.2	结构模型的识别方法	181
8.4	联立方程模型的估计方法	183
8.4.1	递归模型的估计方法	183
8.4.2	简化型模型的估计方法	184
8.4.3	结构模型的估计方法	184
8.5	联立方程模型举例	185
第 9 章	模型诊断常用统计量与检验	193
9.1	检验模型中全部解释变量都无解释作用的 F 统计量	193
9.2	检验单个回归系数显著性的 t 统计量	194
9.3	检验回归系数线性约束条件是否成立的 F 统计量	195
9.4	似然比统计量	199
9.5	沃尔德(Wald)统计量	201
9.6	拉格朗日乘子统计量	205
9.7	赤池、施瓦茨和汉南-奎因统计量	209
9.8	检验正态分布性的 JB 统计量	213
9.9	格兰杰因果性检验	215
9.10	邹突变点检验	218
9.11	回归系数稳定性的邹检验	226
9.12	递归分析	229
第 10 章	时间序列 ARIMA 模型	235
10.1	随机过程与时间序列定义	235
10.2	ARIMA 模型的分类	239
10.2.1	自回归模型	239
10.2.2	移动平均模型	244
10.2.3	自回归移动平均模型	246
10.2.4	单整自回归移动平均模型	247
10.3	伍尔德(Wold)分解定理	249
10.3.1	伍尔德分解定理	249
10.3.2	随机过程期望与漂移项的关系	250
10.4	自相关函数及其估计	252
10.4.1	自相关函数	252
10.4.2	自回归过程的自相关函数	253

10.4.3	移动平均过程的自相关函数	257
10.4.4	ARMA 过程的自相关函数	258
10.4.5	自相关函数的估计(相关图)	259
10.5	偏自相关函数及其估计	261
10.5.1	偏自相关函数定义	261
10.5.2	偏自相关函数的计算	262
10.5.3	AR、MA、ARMA 过程偏自相关函数特征	263
10.5.4	偏自相关函数的估计	264
10.5.5	ARIMA 过程自相关函数和偏自相关函数特征总结	265
10.6	ARIMA 模型的建立、估计过程与预测	266
10.6.1	模型的识别	266
10.6.2	模型参数的估计	268
10.6.3	模型的诊断与检验	273
10.6.4	ARIMA 模型预测	274
10.7	ARIMA 模型建模案例	277
10.8	季节时间序列 ARIMA 模型	283
10.8.1	季节时间序列模型定义	284
10.8.2	季节随机过程的自相关函数和偏自相关函数	287
10.8.3	季节 ARIMA 模型的识别、拟合、检验与预测	301
10.8.4	季节 ARIMA 模型建模案例	302
10.9	回归与 ARMA 组合模型(regARIMA 模型)	312
10.9.1	回归与 ARMA 组合模型定义	312
10.9.2	回归与 ARMA 组合模型案例分析	313
第 11 章	虚假回归	319
11.1	问题的提出	319
11.2	单整性定义	320
11.3	单整序列的统计特征	321
11.4	虚假回归	324
第 12 章	单位根检验	332
12.1	4 种典型的非平稳过程	332
12.1.1	随机游走过程	332
12.1.2	随机趋势过程	333
12.1.3	趋势平稳过程	334
12.1.4	趋势非平稳过程	335
12.2	$DF, T(\hat{\beta}-1)$ 统计量的分布特征	336
12.2.1	DF 统计量的分布特征	336
12.2.2	AR(p) 含单位根过程的 DF 统计量分布特征	342

12.2.3	误差项为 ARMA 形式的 $I(1)$ 过程 DF 分布特征	343
12.2.4	DF 检验式中 $t(\hat{\alpha})$, $t(\hat{\gamma})$ 和 F 统计量的分布特征	344
12.2.5	$T(\hat{\beta}-1)$ 统计量的分布特征	346
12.2.6	趋势过程中 t 统计量的分布特征	347
12.3	单位根检验	347
12.3.1	单位根检验原理	347
12.3.2	单位根检验步骤	349
12.4	单位根检验的 EViews 9 操作	351
12.5	单位根检验案例分析	352
12.6	结构突变序列的单位根检验	358
第 13 章	单方程误差修正模型	369
13.1	均衡概念	369
13.2	误差修正模型	370
13.2.1	自回归分布滞后模型	370
13.2.2	误差修正模型定义	374
13.3	协整定义	377
13.4	协整检验	379
13.4.1	以残差为基础的协整检验法	379
13.4.2	协整系数的分布滞后模型估计法	385
13.5	格兰杰定理	388
13.5.1	多项式矩阵	389
13.5.2	格兰杰(Granger)定理	390
13.5.3	举例验证格兰杰定理	391
13.6	建立单方程误差修正模型的 EG 两步法	394
13.6.1	EG 两步法	394
13.6.2	单方程误差修正模型案例分析	395
第 14 章	面板数据模型	399
14.1	面板数据的定义	399
14.2	面板数据模型分类	403
14.2.1	混合模型	403
14.2.2	固定效应模型	404
14.2.3	随机效应模型	407
14.3	面板数据模型估计方法	408
14.3.1	混合最小二乘估计	408
14.3.2	组内估计	409
14.3.3	最小二乘虚拟变量估计法	410
14.3.4	一阶差分估计	410

14.3.5	可行 GLS 估计法(随机效应估计法)	411
14.3.6	面板数据模型拟合优度的测量	413
14.4	面板数据模型的设定与检验	414
14.4.1	F 检验	414
14.4.2	H 检验	415
14.4.3	Wald 检验	417
14.4.4	F 检验和 LR 检验	417
14.5	面板数据建模案例分析	418
14.6	面板数据建模的 EViews 9 操作	435
14.6.1	Pool(混合)数据工作文件的建立,模型的估计、检验与预测	435
14.6.2	面板数据 panel 型工作文件的建立,模型估计与检验	441
附录 A	随机变量、概率极限、矩阵代数知识简介	445
附录 B	统计分布表	446
附表 1	相关系数临界值表	446
附表 2	标准正态分布函数表	447
附表 3	t 分布百分位数表	448
附表 4	χ^2 分布百分位数表	449
附表 5	F 分布百分位数表	450
附表 6	DW 检验临界值表($\alpha=0.05$)	452
附表 7	DF 分布百分位数表	453
附表 8	$t_{(\hat{\alpha})}$ 检验临界值表($\Delta y_t = \alpha + \rho y_{t-1} + u_t$ 中检验 $\alpha=0$)	453
附表 9	F 检验临界值表($\Delta y_t = \alpha + \rho y_{t-1} + u_t$ 中检验 $\alpha=\rho=0$)	454
附表 10	$t_{(\hat{\alpha})}$ 检验临界值表($\Delta y_t = \alpha + \gamma t + \rho y_{t-1} + u_t$ 中检验 $\alpha=0$)	454
附表 11	$t_{(\hat{\gamma})}$ 检验临界值表($\Delta y_t = \alpha + \gamma t + \rho y_{t-1} + u_t$ 中检验 $\gamma=0$)	454
附表 12	F 检验临界值表($\Delta y_t = \alpha + \gamma t + \rho y_{t-1} + u_t$ 中检验 $\gamma=\rho=0$)	455
附表 13	$T(\hat{\beta}-1)$ 分布百分位数表	455
附表 14	EG 和 AEG 协整检验临界值表	456
附表 15	协整检验临界值表	457
附录 C	EViews 9 使用简介	458
参考文献		459

第 1 章 一元线性回归模型

本章在 1.1 节给出计量经济学简介与建模步骤,第 1 章介绍一元线性回归模型。内容包括模型的建立及其假定条件、一元线性回归模型的系数估计、最小二乘(OLS)估计方法、回归系数估计量的分布、最小二乘估计量的统计性质、最小二乘回归方程的性质、拟合优度的测量、回归系数的显著性检验、回归系数的置信区间、模型的预测、案例分析等。

1.1 计量经济学简介与建模步骤

“计量经济学”,国内也称“经济计量学”对应的英文词都是 econometrics。

计量经济学是指用定量与定性相结合的方法研究经济活动规律及其应用的科学。它是经济学与统计学、数学相结合的交叉学科。

“计量经济学”作为一个专有名词是 1926 年由挪威经济学家弗里希(R. Frisch)提出的。随后 1930 年成立了国际计量经济学学会,1933 年创办了《计量经济学》杂志。至今已有 80 多年的历史。

我国 1980 年正式引进计量经济学。标志是中国社会科学院邀请美国以诺贝尔经济学奖获得者、美国宾夕法尼亚大学克莱因教授为首的 7 位计量经济学家开办“计量经济学讲习班”。之后计量经济学在中国得到迅速发展。1998 年教育部高等学校经济学学科教学指导委员会正式将计量经济学列为高等学校经济学门类各专业本科生的 8 门必修课之一。

计量经济学以 20 世纪 70 年代为界,之前的研究成果多属于经典计量经济学范畴,之后的研究成果多属于非经典计量经济学范畴。随着时间的推移,计量经济学逐渐渗透到经济学各个领域形成了新的计量经济学分支,如金融计量经济学、时间序列计量经济学、空间计量经济学等。也有人按研究对象把计量经济学分为宏观计量经济学和微观计量经济学。

计量经济模型,即研究经济、人文问题所建立的定量分析模型。其中使用时间最久的是回归模型,如果以高斯(C. F. Gauss)提出最小二乘估计方法为标志,则已经有 208 年的历史了。但真正建立起一套完整的设定、估计、推断、检验体系是在 20 世纪 30 年代。20 世纪 40 年代以前建立的基本上是单方程回归模型。40 年代以后随着计算机的发展,以及人们着眼于对宏观经济的研究,开始建立联立方程模型。20 世纪 70 年代初,Box-Jenkins 提出研究时间序列的 ARIMA 模型。20 世纪 70 年代以后随着计算机以及计算机专用软件的逐步普及,各种模型被提出,研究成果呈爆炸式增长。

计量经济学的研究内容与目的主要有以下两个。

(1) 定量描述与分析经济活动。包括描述宏观、微观经济问题,寻找和验证经济规

律、建立计量经济模型。通过计量模型得到回归系数(边际系数、弹性系数、技术系数、比率、速率等)的可靠估计值,从而为分析经济问题、制定相关经济政策、实施宏观经济调控提供依据。

例如,1962—2001年中国储蓄存款总额(Y ,亿元)与国内生产总值(GDP,亿元)关系的估计结果如下:

$$\begin{aligned} \ln Y_t = & -8.7350 + 1.7443 \ln \text{GDP}_t + 1.1840 \text{AR}(1) - 0.3511 \text{AR}(2) \\ & (-13.6) \quad (25.2) \quad (7.8) \quad (-2.3) \\ R^2 = & 0.998, DW = 1.64, T = 40, (1962-2001) \end{aligned}$$

通过回归系数估计值可知,1962—2001年中国储蓄存款总额(Y)与GDP的关系是,GDP每增加1%,中国储蓄存款总额平均增加1.7443%。中国储蓄存款总额的增长速度远大于GDP的增长速度。这为了解、掌控中国储蓄存款总额,制定相关政策提供重要依据。

(2) 做经济预测。这是计量经济学利用模型所要解决的最重要任务,但也是最困难的任务。计量经济学的发展史就是谋求对经济变量做出更精确预测的发展史。

要想得到经济变量之间的精确关系,就必须使用计量经济学的建模研究方法。建立计量经济模型一般分为4个步骤。

(1) 确定研究对象和影响因素的测量变量。比如,研究中国经济,使用GDP还是GNP(国民生产总值)做测量变量要首先确定下来。不容易度量的对象要找合理的替代变量。比如,商品需求量常用销售量代替。注意:研究对象必须是可量化的、可观测的。

(2) 收集数据。收集数据分为直接收集和间接收集两种。

直接收集数据即亲自做调查。调查分普查和抽样调查两种。普查即对每一个观测对象做调查。抽样调查方法分多种,有简单随机抽样、分层抽样、整群抽样、系统抽样等。抽样调查时,还可把这些方法结合在一起使用。如何使调查的数据最大限度地反映总体特征,应该是研究者关注的内容。理论上属于统计学领域研究的内容。

间接收集数据即从各种统计年鉴、网站、数据库等处引用数据。引用数据时要时刻注意,引用的数据是否与自己想要得到的数据定义相符。比如,想要得到农业劳动人口数据,但引用的是统计年鉴上的农村人口数据,这其实是两个概念。其后果相当于给变量的测量引入测量误差。

无论亲自收集数据还是引用数据都要注意不要出现错误。数据出现错误,相当于给变量的测量引入误差,将直接导致对模型参数的估计出现偏倚。在估计模型前,一定要检查数据,避免存在错误。

(3) 仔细观察,分析变量的时间序列图和散点图。一定要养成这样的习惯。通过对变量的观察可以为建立计量经济模型提供许多有用的信息。

(4) 根据经济理论和对问题的调查研究与深入了解,设定计量经济模型的具体形式,估计模型,对估计结果进行诊断与检验,最终确定模型估计结果。分析回归系数,解释其经济含义,对研究对象进行预测,等等。

以往在计量经济学著作中常强调在经济理论的基础上建立计量经济模型。建立计量

经济模型的目的也只是验证经济理论。实际上这只是问题的一个方面。建立计量经济模型的目的还有发现经济理论的一面。社会发展是无止境的,人类对经济发展规律的认识与探索也是无止境的。人们对经济活动规律的研究绝不只是验证,一定还包括发现。

以经济学中的恩格尔定律为例,德国统计学家恩斯特·恩格尔在1857年利用埃朵·杜皮扬(Edouard Ducpetiaux)收集的198个比利时家庭的收入与食物支出数据,采用一元线性回归的方法,发现了著名的恩格尔定律。虽然那时还没有计量经济学这个名称,但是,这毕竟是运用计量经济学的方法发现了一个经济理论,恩格尔定律。

计量经济学研究的主要内容是上述建模步骤的第(3)步和第(4)步。实际上,主要是第(4)步。本书主要围绕第(4)步内容展开。

1.2 模型的建立及其假定条件

1.2.1 建立模型的意义

在经济领域,一个变量的变化常常受其他多个经济变量的影响。为描述这些变量之间的关系,研究这些变量之间的变化规律,通常要建立计量经济模型,研究模型回归系数,进而利用计量经济模型进行预测。比如只有一个重要变量 x_t 影响变量 y_t 变化,且它们之间的关系是线性的,则建立的应是一元线性回归模型。

1.2.2 一元线性回归模型的定义

一元线性回归模型表示如下:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (1-1)$$

式(1-1)表示变量 y_t 和 x_t 之间的真实关系。其中 y_t 称作被解释变量(相依变量、因变量), x_t 称作解释变量(独立变量、自变量、回归因子), u_t 称作随机误差项(随机扰动项), β_0 称作常数项(截距项), β_1 称作回归系数。通常, β_0 和 β_1 又统称为模型的回归系数。

在模型(1-1)中, x_t 是影响 y_t 变化的重要解释变量。回归系数 β_0 和 β_1 具体描述这种关系。 β_0 和 β_1 通常是未知的,需要估计。如果 x_t 和 y_t 是截面数据, t 表示序数;如果 x_t 和 y_t 是时间序列数据, t 表示时间序数。 u_t 则包括除 x_t 以外的影响 y_t 变化的众多微小因素。 u_t 的变化是不可控的。

上述模型可以分为两部分:① $\beta_0 + \beta_1 x_t$ 是非随机部分;② u_t 是随机部分。

1.2.3 一元线性回归模型的经济含义与特征

这种模型可以赋予各种实际意义,如支出与收入的关系;商品价格与供给量的关系;基本建设投资与国内生产总值的关系;林区木材采伐量与其剩余物的关系;脉搏与血压的关系;身高与体重的关系;等等。

以研究家庭支出与收入的关系为例。假设家庭支出与收入呈线性函数关系。实际上,数据来自各个不同家庭,来自各个不同收入水平,从而使收入以外的影响支出变化的

其他因素维持不变是不可能的。随机误差项 u_t 中包括了家庭人口数、消费习惯、不同地域的物价水平、家庭的额外收入等因素。所以在研究经济问题中“控制其他因素不变”是不可能的。因此,即便 y_t 与 x_t 呈完全线性关系,由 y_t 与 x_t 数据得到的观测点也不在一条直线上(不呈线性函数关系),而是散布在一条直线周围,这些观测点服从回归关系,见图 1-1,其中直线 $E(y_t)=\beta_0+\beta_1x_t$ 称作真实的回归直线[$E(y_t)$ 是 y_t 的期望,在 1.4 节将进一步介绍],描述 y_t 与 x_t 的真实关系。式(1-1)中的 u_t ,即观测点到 $E(y_t)$ 的垂直距离,表示由于众多随机因素的影响使一个具体观测点偏离回归直线的幅度。

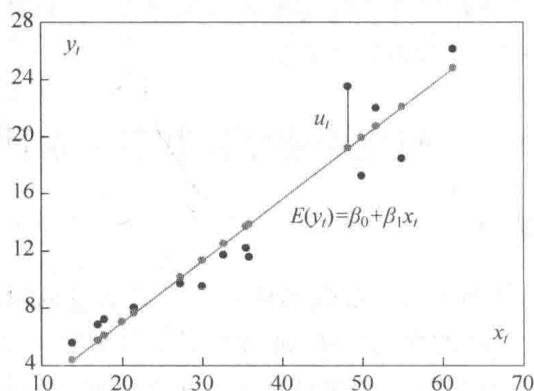


图 1-1 真实的回归直线

一般来说,回归模型的随机误差项中包括如下几项内容。

(1) 未在模型中专门列出的影响 y_t 变化的非重要解释变量。如上例中家庭人口数、消费习惯、物价水平差异等因素的影响都包括在随机误差项中。

(2) 人的随机行为。经济活动都是人参与的。人的经济行为的变化也会对随机误差项产生影响。

(3) 数学模型形式欠妥。对于同一组观测值,若拟合的数学模型形式不同,则相应的随机误差项的值也不同。显然当模型形式欠妥时,会直接对随机误差项的值造成影响。

(4) 归并误差。模型中被解释变量的值常常是归并而成的。当归并不合理时,会产生归并误差。比如由不同种类粮食合并构成的粮食产量的不合理归并会带来归并误差。

(5) 测量误差等。当对被解释变量的测量存在误差时,这种误差将包括在随机误差项中。

1.2.4 模型的假定条件

在对回归函数进行估计之前应该对回归模型的随机误差项 u_t 和解释变量做出如下假定。随后介绍的估计回归系数的最小二乘法是以如下假定条件为基础的。

(1) $u_t, (t=1, 2, \dots, T)$, 是 T 个随机变量, u_t 的取值服从概率分布。

(2) $E(u_t)=0, (t=1, 2, \dots, T)$ 。上式表示 u_t 在 t 的每一个点的期望都为零。在模型中如果能保证 u_t 中所包含的都是影响 y_t 的微小因素,那么在众多微小因素的作用下,