高级计量经济学

[日]雨宫健（Takeshi Amemiya）著
朱保华　周亚虹　等译

汉译经济学文库
Translated Economics Library

ADVANCED ECONOMETRICS

上海财经大学出版社

汉译经济学文库

高级计量经济学

[日] 雨宫健(Takeshi Amemiya) 著
朱保华 周亚虹 等译

上海财经大学出版社

图书在版编目(CIP)数据

高级计量经济学/(日)雨宫健著;朱保华,周亚虹译. —上海:上海财经大学出版社,2010.2

(汉译经济学文库)

书名原文:Advanced Econometrics

ISBN 978-7-5642-0368-9/F·0368

Ⅰ.高… Ⅱ.①雨… ②朱… ③周… Ⅲ.计量经济学 Ⅳ.F224.0

中国版本图书馆 CIP 数据核字(2008)第 166916 号

□策　　划　黄　磊
□责任编辑　仲崇巍
□封面设计　钱宇辰

GAOJI JILIANG JINGJIXUE
高 级 计 量 经 济 学

[日] 雨宫健(Takeshi Amemiya) 著
朱保华　周亚虹　等译

上海财经大学出版社出版发行
(上海市武东路 321 号乙　邮编 200434)
网　　址:http://www.sufep.com
电子邮箱:webmaster@sufep.com
全国新华书店经销
上海市印刷七厂印刷
宝山葑村书刊装订厂装订
2010 年 2 月第 1 版　2010 年 2 月第 1 次印刷

787mm×1092mm　1/16　26.5 印张(插页:2)　515 千字
印数:0 001—4 000　　定价:49.00 元

图字：09-2006-836 号
Advanced Econometrics

译者序

在研究生层次的计量经济学教材中，两本日本学者撰写的教材非常有影响。一本是林文夫的《计量经济学》(Fumio Hayashi, *Econometrics*, Princeton University Press, 2000)；另一本就是雨宫健的《高级计量经济学》。从教材内容而言，两本计量经济学的教材都反映了计量经济学的最新进展。林文夫的《计量经济学》从一般矩法(Generalized Method of Moments)的视角介绍计量经济学的基本理论，从而成为流行于欧美国家知名大学的教材之一。雨宫健的《高级计量经济学》侧重介绍微观计量经济分析的定性反应模型，兼顾博士研究生的计量经济理论的教学需要，属于权威性的高级计量经济理论的著作之一。

《高级计量经济学》是雨宫健教授在长年担任 *Journal of Econometrics* 主编之后编写的研究生层次的计量经济学教材，融合了计量经济理论研究的方法和技巧，也是一本值得计量经济学的专业人员认真阅读的计量经济学著作。在计量经济学理论研究的学术论文中，《高级计量经济学》是一本被广泛引用的参考文献，迄今为止的累计引用数高达 3 200 次以上。本书着重讨论微观计量经济学涉及的各种理论问题，特别是在微观计量分析的定性模型的详细讨论中融入了作者的研究心得经验。本书从经典最小二乘法出发，结合拓展的各种回归分析方法，说明计量经济理论涉及的大样本理论，利用大样本理论讨论微观计量分析出现的极值统计量的性质及各种微观计量模型的统计推断问题。考虑到计量经济理论体系的完整性，本书也适当介绍了时间序列分析、一般最小二乘法、线性与非线性的联立方程模型，提供了计量经济分析常用的矩阵代数与统计分布函数的附录。为帮助学生进一步地理解消化正文的内容，各章配备大量练习题。由于本书作者设想读者已经具备中级计量经济学的知识，因而内容叙述严谨独特，简洁扼要的定理证明富有特色，有助于读者深入理解辨别各种容易混淆的概念。

本书不仅非常适合用作研究生层次的计量经济学教材，也可作为从事计量经济学的研究人员的专业手册使用。

尽管伍德里奇的《横截面与面板数据的经济计量分析》(Jeffrey M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002) 的写作风格也与本书相似，但本书的处理方法相对显得更简洁，特别是对微观计量分析的 Tobit 模型的分类方法已成为一种模型分类标准。由于本书的成书时间较早，微观计量分析涉及的面板数据等内容可能需要参阅萧政的《面板数据分析》(Cheng Hsiao, *Analysis of Panel Data*, 2^{nd}. ed, Cambridge University Press, 2003) 或阿雷拉诺的《面板数据计量经济学》(Manuel Arellano, *Panel Data Econometrics*, Oxford University Press, 2003) 等书籍。此外，对雨宫健的学术经历感兴趣的读者，可以参阅刊登于学术期刊《计量经济理论》2007 年第 1 期的访谈"The ET Interview: Takashi Amemiya"(*Econometric Theory*, Vol. 23, pp. 155～181.)。

本书的翻译过程如下：(姓氏笔画为序) 王侯友、刘俐含、许玲丽、张琦、苗作兴、梁娟参与了本书的初译工作。在翻译初稿的基础上，朱保华和周亚虹分别依据原文仔细校译修改了第 1 章至第 8 章和第 9 章至第 11 章的译稿，并相互通读全部译稿，对专业术语和文体进行了统一，改正了原书的印刷错误。限于译者的学识，译稿自然存在不足之处，敬请读者宽容和批评指正。

本书的翻译出版属于上海财经大学研究生精品课程建设"高级计量经济学"项目的组成部分，对上述资助表示感谢。本书的翻译出版得到了上海财经大学出版社的袁敏先生的关心，上海财经大学出版社的编辑为本书付出了辛勤的劳动，对此表示衷心感谢。

朱保华　周亚虹

2010 年 2 月

前　言

本书既是计量经济学家的专业参考书，也是研究生课程的教科书。根据本人在斯坦福大学的多年教学经验，作为教科书使用本书时，一学年的课程可以教完本书的内容。基于本书的各类课程的预修课包括一学年的微积分、一学期的矩阵分析和一学年的中级统计推断的课程。中级统计推断的教材清单参见第 1 章的注释。更理想的情况是除上述课程外，学生还修读过初级或中级的计量经济学课程。中级计量经济学的程度相当于 Johnston (1972)的教材水平。最后要求的计量经济学知识未必是必需的，但根据我过去的经验，大多数修读研究生层次高级计量经济学课程的经济学专业的学生具有初级或中级的计量经济学的知识。

本书的主要特色反映在以下几方面：细致地讨论经典最小二乘理论（第 1 章）和广义最小二乘理论（第 6 章）；严谨地说明大样本理论（第 3 章和第 4 章）；详细地讨论定性反应模型（第 9 章）、截取或断尾的回归模型（第 10 章）及马尔科夫链和持续期限模型（第 11 章）；介绍非线性联立方程组模型。

因为存在许多优秀的时间序列分析的教材，本书只介绍时间序列分析的基础知识（第 5 章和第 6 章的一部分）。时间序列分析的书籍清单参见第 5 章的起始部分。最近的计量经济应用分析已经大量使用本书最后三章讨论的模型，但还没有教科书像本书这样严谨地彻底讨论这些模型。部分教师可能希望使用时间序列分析的教材来弥补本书的时间序列分析的内容。

本书的线性联立方程组模型（第 7 章）的讨论也很简短。希望更全面地学习线性联立方程组模型的读者可以参考第 7 章提及的参考文献。鉴于非线性联立方程组模型还处于发展的初期阶段，因而大多数计量经济学教材并未全面地介绍非线性联立方程组模型，本书就选择使用更多的篇幅讨论非线性联立方程组模型。

本书的大部分及第 3 章和第 4 章的整体不仅使用了数学定理的证明模

式，还严密地导出了数学结论。然而，本书的目的不是基于数学的一般性来描述各种定理。由于本人希望本书是教科书而不是研究专著，即使在需要放松定理的前提与结论时，本书选择使用了相对容易理解和便于导出简单证明过程的各种假设条件。这样处理的目的是使得读者理解各种定理的基本结构，并能够根据自己的需要及能力拓展各种定理。通过各章的例题或练习题的形式提供了许多定理的应用例，以便更好地说明各种定理的本质内容。

虽然本书是计量经济学方法的教科书，但本书介绍了说明如何使用理论研究结论的经验分析数值例。本书的最后三章特别显著地采用了上述处理方法。

许多学者为本书的写作和修改做出了贡献，由于篇幅关系难以列出全部姓名。特别感谢 Trevor Breusch、Hidehiko Ichimura、Tom MaCurdy、Jim Powell 和 Gene Savin 对教材全部初稿的有益评论。感谢 Carl Christ、Art Goldberger、Cheng Hsiao、Roger Koenker、Tony Lancaster、Chuck Manski 和 Hal White 对本书部分初稿的有益评论。同时感谢 Colin Cameron、Tom Downes、Harry Paarsch、Aaron Han 和 Choon Moon 校对了本书，还感谢 Colin Cameron、Tom Downes 和 Harry Paarsch 订正了英语错误。此外，Tom Downes 和 Choon Moon 帮助整理了本书的索引。Dzung Pham 帮助打印了本书初稿的几个修正版本，尽管属于加班工作，但 Dzung Pham 的不懈耐心和良好素养值得赞赏。David Criswell、Cathy Shimizu 和 Bach-Hong Tran 也帮助打印了本书。非常感谢自然科学基金的赞助研究产生了本书介绍的许多定理结论。最后，感谢 *Journal of Economic Literature* 的编辑允许本书的第 9 章使用本人论文"Qualitative Response Models: A Survey"(*Journal of Economic Literature*, 19: 1483-1536, 1981)的部分内容，还感谢 North-Holland 出版公司允许本书的第 10 章使用本人论文"Tobit Models: A Survey"(*Journal of Econometrics*, 24: 3-61, 1984)的修订版。

目　录
CONTENTS

译者序 / *1*

前言 / *1*

1 经典最小二乘估计理论 / *1*

1.1　线性回归模型　/ *1*
1.2　最小二乘理论　/ *4*
1.3　正态分布假设的模型 1　/ *10*
1.4　带线性约束的模型 1　/ *17*
1.5　检验线性假设　/ *22*
1.6　预测　/ *32*

2 回归分析的最新发展 / *37*

2.1　选择回归变量　/ *37*
2.2　岭回归和 Stein 估计量　/ *45*
2.3　稳健性回归　/ *57*

3 大样本理论 / *66*

3.1　随机变量和分布函数　/ *66*

3.2 不同收敛方式 / 69
3.3 大数定律和中心极限定理 / 73
3.4 limE、AE 和 plim 的关系 / 76
3.5 最小二乘估计量的相合性和渐近正态性 / 78

4 极值估计量的渐近性质 / 85

4.1 一般结论 / 85
4.2 极大似然估计量 / 92
4.3 非线性最小二乘估计量 / 103
4.4 迭代方法 / 111
4.5 渐近检验及相关问题 / 114
4.6 最小绝对偏差估计量 / 120

5 时间序列分析 / 129

5.1 简介 / 129
5.2 自回归模型 / 131
5.3 残差为移动平均的自回归模型 / 138
5.4 自回归模型的最小二乘估计量和极大似然估计量的渐近性质 / 140
5.5 预测 / 144
5.6 分布滞后模型 / 145

6 广义最小二乘理论 / 148

6.1 已知协方差矩阵 / 148
6.2 协方差矩阵未知 / 152
6.3 系列相关 / 152
6.4 似不相关回归模型 / 161
6.5 异方差性 / 162
6.6 误差成分模型 / 170
6.7 随机系数模型 / 179

7 线性联立方程模型　/ 187

7.1　模型和识别　/ 187
7.2　完全信息极大似然估计量　/ 190
7.3　有限信息模型　/ 192
7.4　三阶段最小二乘估计量　/ 198
7.5　前沿性论题　/ 199

8 非线性联立方程模型　/ 202

8.1　单方程估计　/ 202
8.2　方程组估计　/ 211
8.3　假设检验、预测和计算　/ 216

9 定性反应模型　/ 221

9.1　简介　/ 221
9.2　单维二分变量模型　/ 222
9.3　多分模型　/ 237
9.4　多元模型　/ 256
9.5　基于选择的抽样　/ 262
9.6　任意分布法　/ 279
9.7　面板数据 QR 模型　/ 286

10 Tobit 模型　/ 295

10.1　简介　/ 295
10.2　标准 Tobit 模型(第一类 Tobit 模型)　/ 296
10.3　实证事例　/ 298
10.4　标准假设的估计量的性质　/ 299
10.5　非标准假设的 Tobit 极大似然估计量的性质　/ 310
10.6　广义 Tobit 模型　/ 314
10.7　第二类 Tobit 模型:$P(y_1<0)\cdot P(y_1>0,y_2)$　/ 316
10.8　第三类 Tobit 模型:$P(y_1<0)\cdot P(y_1,y_2)$　/ 320
10.9　第四类 Tobit 模型　/ 325
10.10　第五类 Tobit 模型:$P(y_1<0,y_3)\cdot P(y_1>0,y_2)$　/ 328

11 马尔科夫链和持续期限模型 / 339

11.1 马尔科夫链模型 / 339
11.2 持续期限模型 / 357

附录 1 矩阵分析的常用定理 / 378

附录 2 分布理论 / 382

参考文献 / 384

1 经典最小二乘估计理论

本章介绍经典线性回归模型统计推断的基本结论。在经典线性模型中，假定回归变量与误差项独立，误差项无序列相关且同方差。经典线性模型是研究的起点，随后章节讨论的模型都是对经典线性模型的修正。

1.1 线性回归模型

本节解释为什么要研究线性回归模型及如何设定线性模型。我们将定义贯穿本章讨论的模型 1 作为研究的起点。

1.1.1 *介绍*

考虑 K 个随机变量$(y_t, x_{2t}, x_{3t}, \cdots, x_{Kt})$，$t=1,2,\cdots,T$ 组成的序列。定义 T 维向量 $\mathbf{y}=(y_1, y_2, \cdots, y_T)'$、$K-1$ 维向量 $\mathbf{x}_t^* =(x_{2t}, x_{3t}, \cdots, x_{Kt})'$和一个$(K-1)\times T$ 维向量 $\mathbf{x}^* =(\mathbf{x}_1^{*\prime}, \mathbf{x}_2^{*\prime}, \cdots, \mathbf{x}_T^{*\prime})'$。假定它们的联合密度函数为 $f(\mathbf{y}, \mathbf{x}^*, \boldsymbol{\theta})$，其中 $\boldsymbol{\theta}$ 为未知参数向量。我们考虑根据观测向量 $\mathbf{y}$ 和 $\mathbf{x}^*$ 对参数向量 $\boldsymbol{\theta}$ 进行统计推断的方法。

在计量经济学的研究中，常常关注在给定一个随机变量集的条件下另一个随机变量集的条件分布，例如给定收入下消费的条件分布、给定价格下需求量的条件分布。假定我们的目标是得到已知 $\mathbf{x}^*$ 时 $\mathbf{y}$ 的条件分布。我们将联合密度表示成条件密度和边际密度的乘积：

$$f(\mathbf{y}, \mathbf{x}^*, \boldsymbol{\theta}) = f(\mathbf{y} | \mathbf{x}^*, \boldsymbol{\theta}_1) f(\mathbf{x}^*, \boldsymbol{\theta}_2) \qquad (1.1.1)$$

回归分析可以定义为关于 $\boldsymbol{\theta}_1$ 的统计推断。为此，若 $\boldsymbol{\theta}_1$ 与 $\boldsymbol{\theta}_2$ 之间无联系，可以不考虑 $f(\mathbf{x}^*, \boldsymbol{\theta}_2)$。向量 $\mathbf{y}$ 称为因变量或内生变量，向量 $\mathbf{x}^*$ 称为自变量或外生变量。

在回归分析中，我们通常仅需要估计条件分布的一阶矩和二阶矩，而非整个参数向量 $\boldsymbol{\theta}_1$（在特定情况下一阶矩和二阶矩可以完全确定 $\boldsymbol{\theta}_1$ 的特征分

布)。因此,回归分析可定义成条件期望 $E(\mathbf{y}|\mathbf{x}^*)$ 和条件方差协方差矩阵 $V(\mathbf{y}|\mathbf{x}^*)$ 的统计推断。通常,各阶矩是 $\mathbf{x}^*$ 的非线性函数。然而,本章将考虑一种特殊情况,即 $E(\mathbf{y}_t|\mathbf{x}^*)$ 等于 $E(\mathbf{y}_t|\mathbf{x}_t^*)$,且 $E(\mathbf{y}_t|\mathbf{x}_t^*)$ 是 $\mathbf{x}_t^*$ 线性函数,$V(\mathbf{y}|\mathbf{x}^*)$ 是单位矩阵常数倍。该模型被称为经典(或标准)线性回归模型或同方差(方差不变)线性回归模型,因为在第 1 章学习该模型,我们简单称之为模型 1。

1.1.2 *模型 1*

记 $\mathbf{x}_t=(1,\mathbf{x}_t^{*\prime})'$,我们可以定义模型 1 如下:假定

$$y_t=\mathbf{x}_t'\boldsymbol{\beta}+u_t \quad t=1,2,\cdots,T \tag{1.1.2}$$

其中,y_t 为观测到的随机变量,$\boldsymbol{\beta}$ 为 K 维未知参数,$\mathbf{x}_t$ 为满足 $\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'$ 非奇异的 K 维已知常数,未观测到的随机变量 u_t 满足以下条件:对所有的 t,$Eu_t=0,Vu_t=\sigma^2$(另一个未知参数);对所有 $t\neq s,Eu_tu_s=0$。

注意,我们假设 $\mathbf{x}^*$ 为已知的常数向量,相当于我们仅关心估计给定 $\mathbf{x}^*$ 时 $\mathbf{y}$ 的条件分布。模型 1 的最重要假设是 $E(y_t|\mathbf{x}_t^*)$ 是线性函数,我们将在下面一小节来讨论该假设的含义。我们也做了同方差的假定(对所有的 t,$Vu_t=\sigma^2$),以及无序列相关的假定(对所有的 $t\neq s,Eu_tu_s=0$),之所以做出这些假定,并非是我们相信大多数应用问题满足这些条件,而是因为这些假定可成为一个很好的研究起点。后续章节将去掉这些假定。

有时,我们对模型 1 增加假设来获得一些特殊的结论。特别是有时假设 $\{u_t\}$ 是序列独立的,或假设 u_t 服从正态分布。一般而言,独立性的假设强于无序列相关的假设,尽管在正态分布下这两个概念是一致的。其他增加的假设将在引入模型 1 时再作介绍。

1.1.3 *线性假设的含义*

假定随机变量 y_t 和 $\mathbf{x}_t^*$ 存在有限的二阶矩,相应的方差协方差矩阵定义如下:

$$V\begin{bmatrix} y_t \\ \mathbf{x}_t^* \end{bmatrix}=\begin{bmatrix} \sigma_1^2 & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

我们总能写出:

$$y_t=\beta_0+\mathbf{x}_t^{*\prime}\boldsymbol{\beta}_1+v_t \tag{1.1.3}$$

其中,$\boldsymbol{\beta}_1=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$,$\beta_0=Ey_t-\boldsymbol{\sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}E\mathbf{x}_t^*$,$Ev_t=0$,$Vv_t=\sigma_1^2-\boldsymbol{\sigma}_{12}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}$,$E\mathbf{x}_t^*v_t=\mathbf{0}$。认识到(1.1.3)式并不意味着模型 1 包含的特定假设条件是重要的。因为 $E(v_t|\mathbf{x}_t^*)$ 一般不为零,并不能从(1.1.3)式导出 $E(y_t|\mathbf{x}_t^*)$ 的线

性假设。

我们称(1.1.3)式中$\beta_0+\mathbf{x}_t^{*\prime}\boldsymbol{\beta}_1$为给定$\mathbf{x}_t^*$时$y_t$的最优线性预测算子，因为可以证明$\beta_0$和$\boldsymbol{\beta}_1$是最小化$E(y_t-b_0-\mathbf{x}_t^{*\prime}\mathbf{b}_1)^2$的$b_0$和$\mathbf{b}_1$的估计量。条件期望$E(y_t|\mathbf{x}_t^*)$为已知$\mathbf{x}_t^*$时$y_t$的最优预测算子，其理由是对任意函数$g$都有$E[y_t-E(y_t|\mathbf{x}_t^*)]^2\leqq E[y_t-g(\mathbf{x}_t^*)]^2$。

读者可能会问为什么我们用(1.1.2)式而非(1.1.3)式进行研究，因为(1.1.3)式太一般化了，我们很难得到有意义的结论。例如，尽管可以用确定β_0和$\boldsymbol{\beta}_1$的特征的相应样本矩代替y_t和$\mathbf{x}_t^*$的总体矩定义β_0和$\boldsymbol{\beta}_1$的自然估计量(事实上相应的估计量与最小二乘估计量相一致)，但若无法确定更多的$\mathbf{x}_t^*$和v_t之间的相互联系，就无法得到估计量的均值。

$E(y_t|\mathbf{x}_t^*)$的线性约束性如何？如果y_t和$\mathbf{x}_t^*$服从联合正态分布或y_t和$\mathbf{x}_t^*$都是伯努利(Bernoulli)分布[1]，该约束就成立。但是在很多重要的分布下，线性约束都不成立。然而，线性假设并不像初看起来那样限制，因为可以通过多种方式变换原始自变量变形得到$\mathbf{x}_t^*$。例如，若商品供给的条件均值y_t是价格p_t的二次函数，就可以令$\mathbf{x}_t^*=(p_t,p_t^2)'$，从而使$E(y_t|\mathbf{x}_t^*)$满足线性要求。

1.1.4 *矩阵记号*

为方便以下的分析，我们将(1.1.2)式用矩阵形式来表示：

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u} \tag{1.1.4}$$

其中，$\mathbf{y}=(y_1,y_2,\cdots,y_T)'$，$\mathbf{u}=(u_1,u_2,\cdots,u_T)'$，$\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_T)'$。换言之，$\mathbf{X}$为$T\times K$矩阵，第$t$行元素为$\mathbf{x}_t'$。矩阵$\mathbf{X}$可以记为：

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{TK} \end{bmatrix}.$$

如果我们关心$\mathbf{X}$的列向量，可记为$\mathbf{X}=[\mathbf{x}_{(1)},\mathbf{x}_{(2)},\cdots,\mathbf{x}_{(K)}]$，其中$\mathbf{x}_{(i)}$为$T$维向量。在没有混淆$\mathbf{x}_{(i)}$和$\mathbf{x}_t$的危险时，可以将括号省略简记为$\mathbf{x}_i$。采用矩阵表示方法时，关于$\mathbf{X}$和$\mathbf{u}$的假设表示成：$\mathbf{X}'\mathbf{X}$为非奇异阵，这也等价于$\text{rank}(\mathbf{X})=K$，如果$T\geqq K$；$E(\mathbf{u})=0$；$\mathbf{E}\mathbf{u}\mathbf{u}'=\sigma^2\mathbf{I}_T$，其中$\mathbf{I}_T$为$T\times T$阶单位矩阵。(可从上下文推断出单位矩阵的阶数时，可将$\mathbf{I}_T$简记为$\mathbf{I}$。)

本章的余下部分不再用$\boldsymbol{\beta}'=(\beta_0,\boldsymbol{\beta}_1')$的分割形式，而是将$\boldsymbol{\beta}$的元素记为$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_K)'$。同样，我们不必假定$\mathbf{x}_{(1)}$是元素均为1的向量，尽管通

---

〔1〕 这个结论仅在$\mathbf{x}_t^*$为标量时正确。如果y、x_1、x_2是存在两种取值可能的(dichotomous)随机变量，存在$E(y|x_1,x_2)=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1x_2$，且可适当地定义$\beta$。

常是那样假设。我们的大多结论都基于 $\mathbf{X}$ 是常数矩阵的简单假定，而不对 $\mathbf{X}$ 限定特殊取值。

1.2 最小二乘理论

本节我们将定义模型 1 的参数 $\boldsymbol{\beta}$ 的最小二乘估计量，并将证明它是最优线性无偏估计量（best linear unbiased estimator，BLUE）。我们还将讨论方差 σ^2 的估计。

1.2.1 $\boldsymbol{\beta}$ 和 σ^2 的最小二乘估计量的定义

模型 1 的回归参数 $\boldsymbol{\beta}$ 的最小二乘估计量（least squares，LS）$\hat{\boldsymbol{\beta}}$ 定义为最小化残差平方和的 $\boldsymbol{\beta}$ 值：[2]

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned} \tag{1.2.1}$$

令 $S(\boldsymbol{\beta})$ 对 $\boldsymbol{\beta}$ 的一阶导数为零，可得到：

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y}+2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=\mathbf{0} \tag{1.2.2}$$

其中，$\partial S/\partial \beta_i$ 为 K 维向量 $\partial S/\partial \boldsymbol{\beta}$ 的第 i 个元素，β_i 为向量 $\boldsymbol{\beta}$ 的第 i 个元素。从(1.2.2)式求解 $\boldsymbol{\beta}$ 得到：

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{1.2.3}$$

显然，$S(\boldsymbol{\beta})$ 在 $\hat{\boldsymbol{\beta}}$ 处达到全局最小。

现在，考虑 $K=2$、$\mathbf{x}_t'=(1,x_{2t})$ 的特殊情况。利用平面上的点来表示 (y_t,x_{2t}) 的 T 个观测值，在几何上最小二乘估计是在 y 轴方向上实现观察值和 $\mathbf{x}_t'$ 对应线之间离差平方和最小化的投影线的截距和斜率。如果在其他不同方向上最小化离差平方和，就将得到不同的估计结果。

知道了最小二乘估计量 $\hat{\boldsymbol{\beta}}$ 时，我们定义：

$$\hat{\mathbf{u}} = \mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}} \tag{1.2.4}$$

为最小二乘残差向量。我们可以用 $\hat{\mathbf{u}}$ 来估计 σ^2：

$$\hat{\sigma}^2 = T^{-1}\hat{\mathbf{u}}'\hat{\mathbf{u}} \tag{1.2.5}$$

尽管此处的最小二乘估计不像估计回归参数估计那样引人注目，但 $\hat{\sigma}^2$ 仍称为

[2] (1.2.1)式中的 $\boldsymbol{\beta}$ 表示方程 S 的定义域，所以严格而言，它应该与参数 $\boldsymbol{\beta}$ 加以区分，参数 $\boldsymbol{\beta}$ 是未知的，并可以理解成一个固定值。因此，为了严格区分，我们应该使用两个不同的符号；但是，为了避免使符号复杂化，我们使用了同一个符号，读者必须从上下文中判断这个符号的具体含义。

σ^2的最小二乘估计量。

从(1.2.4)式可得到：

$$\mathbf{y}=\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\mathbf{u}}=\mathbf{P}\mathbf{y}+\mathbf{M}\mathbf{y} \tag{1.2.6}$$

其中，$\mathbf{P}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$，$\mathbf{M}=\mathbf{I}-\mathbf{P}$。因为$\hat{\mathbf{u}}$与 $\mathbf{X}$ 正交(即$\hat{\mathbf{u}}'\mathbf{X}=\mathbf{0}$)，最小二乘估计可以解释为将 $\mathbf{y}$ 分解为两个正交的部分：一部分为 $\mathbf{X}$ 列向量的线性组合，另一部分是与 $\mathbf{X}$ 正交的部分。我们可以称 $\mathbf{P}\mathbf{y}$ 为 $\mathbf{y}$ 在 $\mathbf{X}$ 列向量张成的空间上的投影，称 $\mathbf{M}\mathbf{y}$ 为 $\mathbf{y}$ 在与 $\mathbf{X}$ 垂直的空间上的投影。附录 1 的定理 14 给出了投影矩阵 $\mathbf{P}$、$\mathbf{M}$ 的性质。在 $\mathbf{y}$ 和 $\mathbf{X}$ 为二维向量(即 $K=1,T=2$)的简单情况下，分解(1.2.6)式可以用图 1.1 表示，其中垂直轴和水平轴分布表示第一个和第二个观测值，射线表示向量。

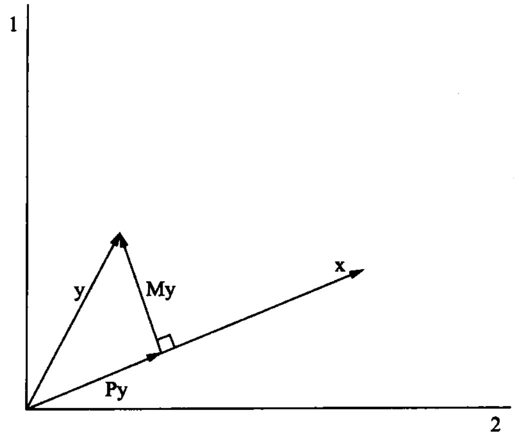

图 1.1　y 的正交分解

从(1.2.6)式可得到：

$$\mathbf{y}'\mathbf{y}=\mathbf{y}'\mathbf{P}\mathbf{y}+\mathbf{y}'\mathbf{M}\mathbf{y} \tag{1.2.7}$$

可用 $\mathbf{y}'\mathbf{P}\mathbf{y}/\mathbf{y}'\mathbf{y}$ 度量 $\mathbf{y}$ 对 $\mathbf{X}$ 回归的拟合优度，有时称拟合优度为 R^2。不过，更常见的方法是将 R^2定义为 $\mathbf{y}$ 和 $\mathbf{P}\mathbf{y}$ 样本相关度的平方：

$$R^2=\frac{(\mathbf{y}'\mathbf{L}\mathbf{P}\mathbf{y})^2}{\mathbf{y}'\mathbf{L}\mathbf{y}\cdot\mathbf{y}'\mathbf{P}\mathbf{L}\mathbf{P}\mathbf{y}} \tag{1.2.8}$$

其中，$\mathbf{L}=\mathbf{I}_T-T^{-1}\mathbf{1}\mathbf{1}'$，$\mathbf{1}$ 为元素为 1 的 T 维向量。若我们假设 $\mathbf{X}$ 的一个列向量为 $\mathbf{1}$(通常该假设成立)，则 $\mathbf{LP}=\mathbf{PL}$。因此，可将(1.2.8)式改写为：

$$R^2=\frac{\mathbf{y}'\mathbf{L}\mathbf{P}\mathbf{L}\mathbf{y}}{\mathbf{y}'\mathbf{L}\mathbf{y}}=1-\frac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{L}\mathbf{y}} \tag{1.2.9}$$

于是，R^2可以解释为 $\mathbf{y}$ 的均值离差对 $\mathbf{X}$ 列向量的均值离差进行回归的拟合优度的测度[2.1.4 节给出 Theil(1961)提出的修正 R^2]。