

北京市教委特色专业建设资助项目

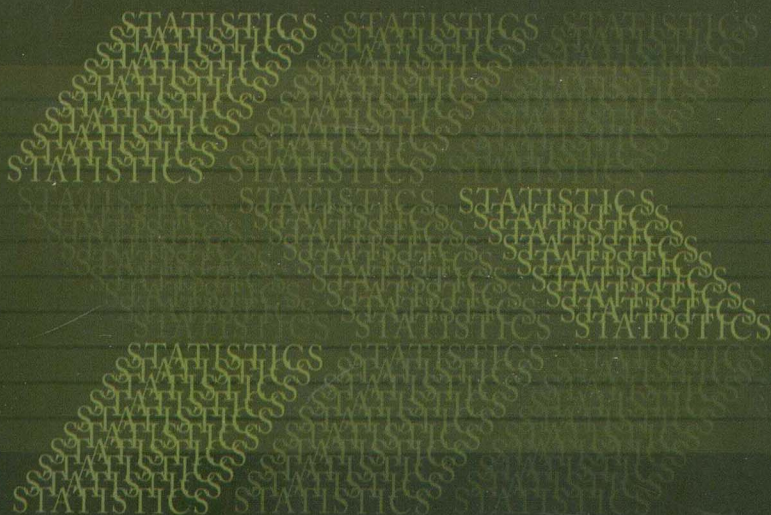


首都经济贸易大学统计学前沿文库

# 关于测量误差模型的 统计推断研究

GUANYU CELIANG WUCHA MOXING DE  
TONGJI TUIDUAN YANJIU

刘 强 / 著



首都经济贸易大学出版社

Capital University of Economics and Business Press



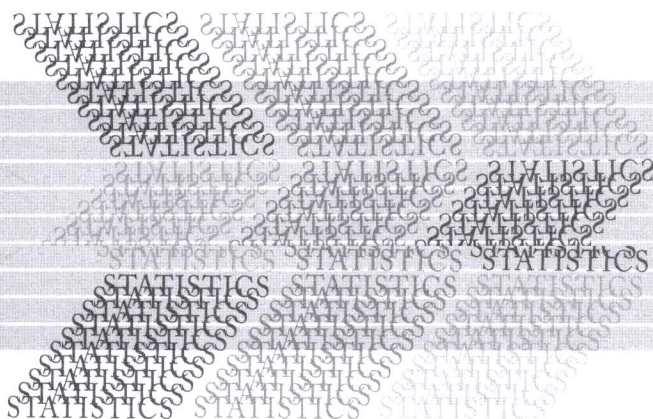
北京市教委特色专业建设资助项目

首都经济贸易大学统计学前沿文库

# 关于测量误差模型的 统计推断研究

GUANYU CELIANG WUCHA MOXING DE  
TONGJI TUIDUAN YANJIU

刘 强 / 著



首都经济贸易大学出版社  
Capital University of Economics and Business Press

· 北 京 ·

## 图书在版编目(CIP)数据

关于测量误差模型的统计推断研究/刘强著. —北京:首都经济贸易大学出版社, 2011. 6

(首都经济贸易大学统计学前沿文库)

ISBN 978 - 7 - 5638 - 1921 - 8

I. ①关… II. ①刘… III. ①测量误差—模型—统计推断—研究 IV. P207

中国版本图书馆 CIP 数据核字(2011)第 107361 号

### 关于测量误差模型的统计推断研究

刘 强 著

---

出版发行 首都经济贸易大学出版社

地 址 北京市朝阳区红庙(邮编 100026)

电 话 (010)65976483 65065761 65071505(传真)

网 址 <http://www.sjmcb.com>

E - mail [publish@cueb.edu.cn](mailto:publish@cueb.edu.cn)

经 销 全国新华书店

照 排 首都经济贸易大学出版社激光照排服务部

印 刷 北京大华山印刷厂

开 本 787 毫米×960 毫米 1/16

字 数 140 千字

印 张 8

版 次 2011 年 6 月第 1 版第 1 次印刷

书 号 ISBN 978 - 7 - 5638 - 1921 - 8/P · 1

定 价 18.00 元

---

图书印装若有质量问题,本社负责调换

版权所有 侵权必究

## 出版总序

社会发展离不开数据,而数据必须使用统计方法来加以分析。自威廉·配第《政治算术》始,历史上几乎每一次对社会经济发展的深刻理解都是建立在统计分析方法变革的基础上的。正是这种变革所提供的各种数据分析工具加深了人们对社会经济本质的理解,使得人们的认识能够还原真实世界并与之无限接近。统计学数百年的发展历经两次方法上的“革命”:从最初不完整的全面调查方法到大样本统计推断,是统计方法的第一次革命;以大样本统计推断方法为基础,进一步发展出小样本统计推断方法,是统计方法的第二次革命。这两次革命都是施于用样本数据推断总体特征这一思想,而抽样误差的干扰导致统计方法日益复杂,使其应用受到限制。目前,以数据挖掘方法为代表的统计学的第三次革命即将到来。数据挖掘是在继承已有统计理论的基础上,与计算机技术紧密结合,充分发挥计算机运算速度快、存储量大的特点,将统计方法从抽样推断向海量数据分析推进,是统计学、计算机技术、仿真计算、机器学习、人工智能甚至哲学思想相融合的新学科,体现了科学发展“螺旋式上升”的哲学内涵。

统计学的发展过程不只是方法上的创新过程,更是一个统计学应用领域不断拓展的过程。从宏观经济计量、宏观经济统计分析,到涵盖消费、收入分配、投资、对外贸易等领域的宏观经济统计专题分析、博弈论等传统经济统计学以及国民经济核算等,再到如今在金融统计、财政统计、精算与风险管理、管理统计或商务统计(含企业微观统计、微观金融、微观核算、微观经济计量等)、市场调查、数据挖掘、质量控制与试验设计等社会经济生活方方面面的广泛应用,统计学的思想和方法无处不绽放出闪耀的光芒,为引领人类社会朝着客观、公平、公正的方向发展提供了必不可少的工具,同时也

为政策制定和决策提供了广泛而坚实的数据基础。

首都经济贸易大学统计学院是全国首家将统计学和数学融合在一起的统计学院,注重从方法的深入研究和应用领域的积极开拓两个方面进行发展,是实现统计学全面而迅速地与国际实质性接轨的先行者。近年来,该学院不断吸收国外统计前沿思想的精髓,引进高端人才,加强国际交流与合作,产生了一大批具有重要价值的前沿成果。本套丛书就是从中挑选出来的优秀成果之一,充分展示了首都经济贸易大学统计学院当前的研究方向和科研实力。

本套丛书的执笔人均为首都经济贸易大学统计学院处于教学科研一线岗位的具有博士学位的青年教师,他们思维活跃,在充分掌握当代数理统计理论与方法的基础上,对实际问题展开了深入而具体的研究,在翔实的数据基础上,对社会经济生活发展的多方面提出了谏言。本套丛书的研究范围具体涵盖了收入分配、宏观经济分析、金融计量、金融数学、数理统计、市场调查、数据挖掘等方向的前沿内容。从这些著作当中,可以清晰地看到他们深厚的统计理论功底与解决实际应用问题的能力,从而为推动统计学真正与国际接轨和良好发展打下了坚实的学术基础。

纪 宏

2009年11月

# 前 言

非参数、半参数回归模型是当今国内外统计学界和计量经济学界共同研究的一个热点课题。目前该类模型主要用于处理一些能够精确观测的数据,如独立数据、panel 数据、纵向数据等。事实上,目前众多统计学家和计量经济学家对非参数、半参数计量经济模型的研究在很大程度上也仅限于对可精确观测数据的处理。

然而经济变量能够精确观测的假定在实际问题中往往不能成立。例如在社会调研、经济管理等应用领域中,对兴趣变量进行观测时,经常会受到多种因素的影响,导致一些偏差,如抽样误差、记录误差等;另外,人们考察变量之间的关系时,常常只关心主要因素的影响,其他影响不大的因素之效应将体现于偏差中。文献中通常称这种观测数据带有误差的问题为“测量误差问题”,分析这些数据的统计模型通常称为“测量误差模型”或“EV (Errors-in-Variables) 模型”。如何利用非参数、半参数回归模型有效地处理带有测量误差的数据,进而如何在纵向数据、删失数据等复杂数据情形下讨论测量误差模型的大样本性质,是当今统计学界和计量经济学界亟待解决的热点课题之一。

关于非参数、半参数回归模型的具体问题可以参见柴根象与洪圣岩的专著《半参数回归模型》(1995),Härdle, Liang 和 Gao 的专著《Partially Linear Models》(2000)以及叶阿忠的专著《非参数计量经济学》(2003);关于测量误差模型的一些基本结果被总结在 Fuller 的专著《Measurement Error Model》(1987)和 Carroll 等的专著《Measurement Error in Nonlinear Models》(2006)。

本书研究的模型主要包括:线性 EV 模型、半参数 EV 模型以及单指标 EV 模型等统计模型。研究的主要目的是:研究 EV 模型中兴趣参数以及兴

趣函数估计的大样本问题,更进一步,基于各种复杂数据(如纵向数据、删失数据、缺失数据等等),研究此类模型中兴趣参数及兴趣函数估计的大样本问题,如估计量的渐近正态性、相合性及其收敛速度等统计性质。

本书的结构安排如下:

第1章是绪论,主要介绍本书涉及的一些基本模型、基本方法和常见数据。第2~6章主要介绍作者近几年对测量误差模型估计理论与估计方法的一些研究成果。具体而言,第2章主要讨论了半参数EV模型的小波估计问题;第3章讨论了纵向数据下半参数EV模型的估计问题;第4章讨论了响应变量随机删失情形下的半参数EV模型的经验似然估计问题;第5章讨论了缺失数据下线性EV模型的估计问题;第6章讨论了核实数据下带有一般测量误差的单指标EV模型的估计问题。

本书是国家社科基金项目“非参数、半参数计量经济中的带有测量误差数据的统计推断研究”(批准号:10CTJ001)和北京市教委社科计划项目(批准号SM201110038014)的研究成果之一,也是北京市市属高等学校人才强教计划资助项目——中青年骨干人才培养计划(批准号:PHR201008214)和北京市教委统计学特色专业建设项目的阶段性研究成果之一。

该课题的研究得到了我的导师北京工业大学薛留根教授和西安交通大学吴可法教授的悉心指导,在此书出版之际谨向他们表示诚挚的感谢!

在本书的撰写过程中,首都经济贸易大学统计学院纪宏教授、马立平教授、刘黎明教授、刘娟教授、沈大庆教授、郭文英教授给予了极大的支持和热心的帮助;中国科学院数学与系统科学研究院王启华研究员,首都师范大学崔恒建教授和北京理工大学徐兴忠教授,北京工业大学张忠占教授和程维虎教授曾对本书的部分内容提出了很好的修改意见,在此表示诚挚的感谢!另外,还要特别感谢首都经济贸易大学出版社的薛晓红编辑对本书的最终出版所作出的努力。

由于作者水平有限,不妥之处在所难免,恳请同行和读者批评指正。

刘强

2011年5月于首都经济贸易大学

E-mail:cuebliuqiang@163.com

# 符号表

|                                 |                                                                                      |
|---------------------------------|--------------------------------------------------------------------------------------|
| $\triangleq$                    | “定义为”或“记为”                                                                           |
| $T$                             | 向量或矩阵的转置                                                                             |
| $\xrightarrow{L}$               | 依分布收敛                                                                                |
| $\xrightarrow{P}$               | 依概率收敛                                                                                |
| $a. s.$                         | 强收敛(依概率 1 收敛)                                                                        |
| $y = O(1)$                      | $y$ 是有界变量                                                                            |
| $y = o(1)$                      | $y$ 是无穷小量                                                                            |
| $\xi_n = o_p(\eta_n)$           | 对任一 $\varepsilon > 0$ , 有 $P\{\ \xi_n\  \geq \varepsilon \ \eta_n\ \} \rightarrow 0$ |
| $\xi_n = o_p(1)$                | $\xi_n$ 依概率收敛到 0                                                                     |
| $O_p(1)$                        | 随机有界                                                                                 |
| $K(\cdot)$                      | 核函数                                                                                  |
| $h$ 或 $h_n$                     | 窗宽                                                                                   |
| $N(\mu, \Sigma)$                | 均值为 $\mu$ , 协方差阵为 $\Sigma$ 的正态分布                                                     |
| $\chi_p^2$                      | 自由度为 $p$ 的卡方分布                                                                       |
| i. i. d.                        | 独立同分布                                                                                |
| $\ \cdot\ $                     | Euclidean 范数                                                                         |
| $A^{\otimes 2}$                 | $AA^T$                                                                               |
| $\text{tr}(A)$                  | 方阵 $A$ 的迹                                                                            |
| $\text{mineig}(A)$              | 矩阵 $A$ 的最小特征值                                                                        |
| $\text{diag}(a_1, \cdots, a_n)$ | 由元素 $a_1, \cdots, a_n$ 组成的对角阵                                                        |
| $c$                             | 正的常数, 在不同地方可以表示不同的值                                                                  |

# 目 录

|                            |           |
|----------------------------|-----------|
| <b>1 绪论</b>                | <b>1</b>  |
| 1.1 半参数回归模型                | 2         |
| 1.2 单指标模型                  | 6         |
| 1.3 测量误差模型(EV 模型)          | 8         |
| 1.4 数据集                    | 12        |
| 1.5 小波估计                   | 16        |
| 1.6 经验似然                   | 17        |
| 1.7 本书的内容及结构               | 17        |
| <b>2 半参数 EV 模型的小波估计</b>    | <b>20</b> |
| 2.1 固定设计下半参数 EV 模型的小波估计    | 20        |
| 2.2 随机设计下半参数 EV 模型的小波估计    | 31        |
| <b>3 纵向数据下半参数 EV 模型的估计</b> | <b>35</b> |
| 3.1 固定设计下半参数 EV 模型的估计      | 36        |
| 3.2 随机设计下半参数 EV 模型的估计      | 50        |
| <b>4 删失数据下半参数 EV 模型的估计</b> | <b>61</b> |
| 4.1 方法与主要结果                | 62        |
| 4.2 数值模拟                   | 66        |
| 4.3 定理的证明                  | 68        |

5 缺失数据下线性 EV 模型的估计 ..... 77

5.1 引言 ..... 77

5.2 方法与主要结果 ..... 78

5.3 数值模拟 ..... 81

5.4 定理的简要证明 ..... 83

6 核实数据下单指标 EV 模型的估计 ..... 87

6.1 方法与主要结果 ..... 87

6.2 定理的证明 ..... 92

参考文献 ..... 105

# 1 绪 论

在科学实验、工农业生产、经济分析以及社会调研等领域中,利用所获数据分析变量之间的关系时,往往假定自变量不带有测量误差,换言之,自变量可精确观测。然而在实际生活中,事实并非如此。在对兴趣变量进行观测时,往往会受到多种因素的影响(如抽样误差、仪器误差、记录误差等),导致一些观测偏差;另一方面,人们考察变量之间的关系时,往往只关心主要因素对兴趣变量的影响,诸多影响不大的因素的效应将反映于兴趣变量取值的偏差中。比如研究身高和体重的关系,在测量身高和体重的时候往往不是很精确,得到的数据经常带有误差;研究某种药物对某种病症的影响,在将病症量化时也不可能做到精确;调查研究影响工资收入的因素时,某些被调查的人可能因为种种原因,不愿意透露真实的工资数额,此时调查得到的数据也带有测量误差,等等。在文献中这种观测变量带有误差的问题通常称为“测量误差问题”,分析这些数据的统计模型通常称为“测量误差模型”或“EV(Errors - in - Variables)模型”。EV 模型是处理变量间数据关系的一种有力工具,目前,无论在理论上还是在实际应用中,该模型都有了长足的发展。

在解决实际问题时,实际工作者和学者们又提出了多种回归模型。例如,线性回归模型,参数回归模型,非参数回归模型,半参数回归模型(包括部分线性模型),单指标模型以及半参数变系数模型等。利用这些模型在处理实际问题时,常常又会遇到各种各样的复杂数据,如删失数据、纵向数据、缺失数据以及时间序列数据等。因此,研究各种复杂数据下各类 EV 模型(如线性 EV 模型,半参数 EV 模型,非线性半参数 EV 模型,单指标 EV 模型等)具有较高的实用价值,目前此类问题已经成为统计学界和计量经济学界研究的热点课题之一。

本书研究的模型主要包括:线性 EV 模型、半参数 EV 模型以及单指标 EV 模型等统计模型;研究的主要目的是:研究 EV 模型中未知参数和未知函数的估计问题,更进一步,基于各种复杂数据(如纵向数据、删失数据、缺失数据等等),研究此类模型中兴趣参数及兴趣函数估计的大样本问题,如估计量的渐近正态性、相合性及其收敛速度等统计性质。

下面分别对半参数回归模型、单指标模型、EV 模型、不同的数据类型、小波估计方法、经验似然方法以及它们的研究现状进行简单的回顾。

### 1.1 半参数回归模型

根据回归函数的形式,回归模型可以分为参数型、非参数型、半参数型三种形式。回顾回归分析发展的历史,在 20 世纪 70 年代以前,研究重点在于参数回归,尤其是线性回归,目前这个方面仍在向纵深方向发展。自 20 世纪 70 年代以来,非参数回归模型的研究日渐兴起,吸引了众多统计学者的关注,尤其是到了 20 世纪 90 年代,非参数模型广泛应用于计量经济学的研究领域,出现了大量的优秀成果。

参数回归模型对回归函数提供了大量的额外信息(通常由经验和历史资料提供),因而当假设模型成立时,其推断有较高的精度,例如熟知的线性模型的最小二乘估计,其方差有  $O(1/n)$  的阶。但当参数假定与实际相背离时,基于假定模型所做的统计推断的表现可以很差,这种情况(也包括参数回归可能作出的正态误差与实际相背离的情况)促使人们寻找别的出路,非参数回归模型的出现有效地解决了这个问题。

非参数回归模型的特点是:回归函数的形式是任意的,变量的分布也限制很少,因而具有较大的实用性。虽然非参数模型有如上优点,然而从实际应用的角度来讲,仍有一定的局限性。首先它忽略了各个解释变量对因变量作用的差别,在实际问题未提供任何信息时,非参数模型会明显降低其解释能力;其次,若影响  $Y$  的因素可以分为两个部分,例如  $x_1, x_2, \dots, x_n$  和  $t_1, t_2, \dots, t_m$ ,若根据经验或历史资料可以认为因素  $x_1, x_2, \dots, x_n$  是主要的,而  $t_1, t_2, \dots, t_m$  是某种干扰因素,它同  $Y$  的关系完全未知,这时无无论用参数回归或者是非参数回归,效果都不会太好。为弥补非参数回归的不足,一个努力的

方向就是 Engle 等<sup>[1]</sup>在研究气象条件对电力需求这一实际问题时提出了部分线性回归模型。部分线性回归模型由两个部分组成,线性部分称为参数分量,非线性部分称为非参数分量。显然其中参数分量中的回归参数  $\beta$  是有限维,而非参数分量涉及无限维的未知参数(详见文献[2])。该模型自提出以来一直是统计学界和计量经济学界讨论的一个热点课题。作为部分线性模型的一般推广,半参数回归模型可以表示如下:

$$Y_i = f(X_i, \beta) + g(T_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.1.1)$$

式中,  $f(\cdot)$  为已知的联系函数,  $g(\cdot)$  为未知的非参数函数,  $\{(X_i^T, T_i), i = 1, 2, \dots, n\}$  为协变量,  $\beta$  为  $p$  维未知参数,  $\varepsilon_i$  为随机误差。当  $f(X_i, \beta) = X_i^T \beta$  时,模型(1.1.1)变为

$$Y_i = X_i^T \beta + g(T_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.1.2)$$

这就是在文献中频繁出现的部分线性回归模型。

半参数回归模型(包括部分线性回归模型)的优点在于:它既含有参数分量,又含有非参数分量,可以概括和描述众多实际问题,因而自模型提出以来就引起统计学界和计量经济学界的广泛重视。从实践意义上讲,半参数回归模型可以说是更接近于真实,更能充分利用数据信息的一种提法。在处理方法上,这种模型融合了常用的参数方法以及各种非参数方法(例如,核估计、局部多项式估计、近邻估计、样条估计、小波估计等),但又不是两类方法的简单叠加,因而其复杂性和难度超过了单一性质的回归模型。可以说,半参数回归模型的核心在于它既集中了主要部分(参数分量)的信息,又兼顾了次要因素(非参数部分)的信息,因而具有较强的解释能力。随着半参数回归模型在理论和方法上的日益成熟,它必将具有广阔的应用前景。

对于半参数回归模型的研究,目前仍集中在大样本性质上。研究的基本问题大致有如下几点:

(1) 未知回归参数及误差方差的相合估计。主要讨论参数的估计在什么条件下是强相合或弱相合的。

(2) 回归参数及误差方差估计的渐近性质。这主要包括:估计的收敛速度、渐近分布等。关于渐近分布是研究在什么条件下,所构造的估计是渐近正态或  $\chi^2$  分布的,这也是目前文献研究的热点。

(3) 回归参数的渐近置信域。要想获得回归参数估计的渐近置信域,通常需要考虑回归参数估计的渐近分布以及渐近方差的相合估计。

(4) 非参数回归函数  $g(\cdot)$  的估计及其收敛速度和渐近分布。通过非参数方法构造估计,例如可以利用核估计、近邻估计、局部多项式估计以及小波估计等方法构造其估计量,讨论的问题往往是估计的相合性和收敛速度。目前,部分文献已经将重点放到了估计量的渐近分布上。

(5) 回归参数估计的稳健性。稳健性从直观意义上是说,当假设模型与真实模型相符时估计量具有良好的性质;当假设模型与真实模型稍许偏离时,其性能所受影响较小;当假设模型与真实模型发生严重偏离时,其性能也还过得去。如在线性回归情形中已经证明  $M$  估计是一种稳健估计。

对于部分线性回归模型(1.1.2),当设计点列  $\{(X_i^T, T_i), i=1, \dots, n\}$  独立同分布时,文献[3-7]先后讨论了当  $g(\cdot)$  的估计分别取样条估计、核估计和最近邻估计时参数  $\beta$  的加权最小二乘估计的渐近正态性及强弱收敛速度等;Hamilton 和 Truong<sup>[8]</sup> 采用局部线性回归构造了参数和非参数分量的估计,并证明了估计量的渐近正态性;Mammen 和 Van de Geer<sup>[9]</sup> 应用经验过程理论构造了惩罚拟似然估计,并给出了该估计的渐近性质。

Ma 等<sup>[10]</sup> 研究了异方差的部分线性回归模型,对加权估计方程进行了修正,即

$$\frac{1}{n} \sum_{i=1}^n \omega(X_i, T_i) \{Y_i - X_i^T \beta - \hat{g}(T_i, \beta)\} \left[ X_i - \frac{\hat{E}\{\omega(X, T)X | T_i\}}{\hat{E}\{\omega(X, T) | T_i\}} \right] = 0 \quad (1.1.3)$$

式中,  $\omega(X_i, T_i)$  表示权函数,通常使用方差的逆(Inverse-to-Variation)权,  $\hat{g}(T_i, \beta)$  表示  $g(\cdot)$  的非参数估计,  $\hat{E}\{\omega(X, T)X | T_i\}$  和  $\hat{E}\{\omega(X, T) | T_i\}$  分别表示  $E\{\omega(X, T)X | T_i\}$  和  $E\{\omega(X, T) | T_i\}$  的非参数相合估计。

研究发现:当  $\hat{g}$  被错误指定时,式(1.1.3)的左边仍可收敛到 0 时,甚至错误地取  $\hat{g} \equiv 0$  时,解方程(1.1.3)仍可得到  $\beta$  的相合估计  $\hat{\beta}$ ,该估计具有渐近正态性,并且是半参数有效估计。

Severini 等<sup>[11]</sup> 和 Härdle 等<sup>[12]</sup> 研究了部分线性回归模型的推广形式,即广义部分线性模型

$$E(Y_i | X_i, T_i) = H\{X_i^T \beta + g(T_i)\}, \quad i=1, \dots, n \quad (1.1.4)$$

式中,  $H(\cdot)$  为一个已知函数, 称为联系函数。

为了估计  $\beta$  和  $g(\cdot)$ , Severini 等<sup>[11]</sup> 引进了拟似然估计方法, 该方法有类似于似然函数的性质, 但仅需指定  $Y$  的二阶矩而不是完全分布。基于 Severini 等人的方法, Härdle 等<sup>[12]</sup> 考虑了  $g(\cdot)$  的线性检验问题, 他们对检验问题的研究补充了 Severini 等人的工作。

我国许多统计学者在部分线性模型的研究中也做了大量的工作。Liang<sup>[13]</sup> 系统地研究了多种场合下  $\beta$  的渐近有效估计的构造。Shi<sup>[14]</sup> 利用分块多项式逼近方法得到了  $\beta$  和  $g(\cdot)$  的稳健 M 估计  $\hat{\beta}$  和  $\hat{g}(\cdot)$ , 在一定条件下证明了  $\hat{\beta}$  具有渐近正态性。

值得注意的是, 小波作为一种技术在 20 世纪 90 年代已经应用到密度估计和非参数回归估计中来。受此启发, 柴根象、钱伟民等学者将小波技术推广到半参数回归模型中, 利用小波估计给出了未知参数的估计, 讨论了估计的相合性、收敛速度、渐近正态性等大样本性质, 取得较好的效果, 参见文献 [15-17] 等。

文献 [3-14] 大都是在  $(X_i^T, T_i)$  随机设计情形下进行讨论的, 文献 [15-17] 是在  $X_i$  为随机设计而  $T_i$  为固定设计时进行讨论的, 当  $(X_i^T, T_i)$  均为固定设计时的研究成果相对较少。我们知道, 固定设计情形并不是随机设计情形的特例, 随机设计情形下的结果往往也不能简单地推广到固定设计情形, 因为二者的处理方法和假设条件有一定的区别。对于固定设计情形, 胡舒合<sup>[18]</sup> 和高集体等<sup>[19-21]</sup> 分别研究了未知参数  $\beta$  的最小二乘估计和加权最小二乘估计的强相合性、渐近正态性、收敛速度、Berry-Esseen 界限以及重对数律等方面的大样本性质。姜玉英等<sup>[22]</sup> 利用小波估计法研究了半参数回归模型中未知参数估计的弱收敛速度和强相合性。王启华<sup>[23-24]</sup> 在截断样本下研究了  $\beta$  和  $g(\cdot)$  的估计的强相合性,  $p(p \geq 2)$  阶平均相合性和渐近正态性。陈明华<sup>[25-26]</sup> 讨论了  $\beta$  和  $g(\cdot)$  的估计的强相合性,  $p(p \geq 2)$  阶平均相合性和收敛速度。

柴根象和孙平<sup>[27]</sup> 基于部分线性模型的可加性, 提出了二阶段估计方法, 得到了  $\beta$  和  $g(\cdot)$  的核权函数形式的估计量  $\hat{\beta}$  和  $\hat{g}(\cdot)$ ; 在  $X_i$  为固定设计而  $T_i$  为随机设计的情形下, 得到了  $\hat{\beta}$  的渐近正态性, 证明了  $\hat{g}(\cdot)$  的强相合性

和一致强相合性,其一致强收敛速度可达到非参数回归函数估计的一致最优强收敛速度 $(n^{-1} \log n)^{\frac{1}{3}}$ 。

薛留根<sup>[28-29]</sup>将随机加权方法应用于部分线性模型的研究中,证明了用随机加权统计量的分布逼近原估计量误差的分布的强有效性,并给出了估计量的最优收敛速度和随机加权逼近速度。薛留根<sup>[30-31]</sup>研究了部分线性模型中误差方差估计之分布的 Berry - Esseen 界和非一致收敛速度。

Xue 和 Liu<sup>[32]</sup>讨论了半参数回归模型中小波估计的 Bootstrap 逼近问题。刘强和薛留根<sup>[33]</sup>在误差序列为  $\psi$ -混合或  $\varphi$ -混合下,通过小波估计法给出了半参数回归模型中未知参数  $\beta$  和未知函数  $g(\cdot)$  的估计。在较弱的条件下得到  $\hat{\beta}$  的强相合性以及  $\hat{g}(\cdot)$  的一致强相合性、一致  $r$  阶矩相合性。

关于非参数、半参数回归模型的具体问题可以参见柴根象与洪圣岩的专著《半参数回归模型》<sup>[2]</sup>, Härdle, Liang 和 Gao 的专著《Partially Linear Models》<sup>[34]</sup>以及叶阿忠的专著《非参数计量经济学》<sup>[35]</sup>。文献[2]着重研究大样本领域的一些传统问题,包括估计的相合性、渐近分布、收敛速度和渐近效率等;文献[34]涉及若干较新题材,包括 Bootstrap、截尾数据、非线性和非参数时间序列等;文献[35]主要讨论非参数、半参数模型在计量经济学中的应用及计算机模拟程序实现。

## 1.2 单指标模型

单指标模型作为一种广义的回归模型,是 20 世纪 80 年代中后期发展起来的一种重要的统计模型。单指标回归模型的形式如下:

$$Y_i = g(X_i^T \beta) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.2.1)$$

式中,  $g(\cdot)$  为一元未知函数,  $\beta$  为  $p$  维未知参数,  $X_i$  为  $p$  维解释向量,  $\varepsilon_i$  为随机误差。在生物工程、网络工程等领域的数据处理中,经常遇到所谓的“维数灾祸”问题,由于单指标模型通过对  $X^T \beta$  而不是  $X$  回归,有效地避免了上述问题,因而该模型在实际问题中具有广泛的应用前景。

考虑到模型的可识别性,须要求  $\|\beta\| = 1$ , 并且  $\beta$  的第一个非 0 元素大于 0。Ichimura<sup>[36]</sup>和 Horowitz<sup>[37]</sup>研究了单指标模型的可识别性, Manski<sup>[38]</sup>讨论了二元响应模型的可识别性。

关于  $\beta$  的估计,已有不少的文献讨论。概括而言,根据是否需要求解非线性最优化问题,可以把  $\beta$  的估计分为两大类:M 估计(M - Estimate)和直接估计(Direct Estimate)。其中  $\beta$  的 M 估计包括半参数最小二乘估计(Semiparametric Least Squares Estimate)和半参数极大似然估计(Semiparametric Maximum Likelihood Estimate)。

Ichimura<sup>[36]</sup>和 Härdle 等<sup>[39]</sup>详细地研究了半参数最小二乘估计方法。该方法采用了参数回归模型中最小二乘估计的思想。在未知参数给定的情况下,用去掉一点的 NW 估计方法估计未知函数  $g(\cdot)$ ,所得的估计是未知参数  $\beta$  的一个函数,然后用非线性最小二乘的思想极小化残差,可得未知参数  $\beta$  的估计。由于该方法需要解决复杂的非线性最优化问题,因此计算过程较为复杂。

半参数极大似然估计沿用了参数极大似然估计的思想。在单指标模型中, $X$  和  $Y$  的联合分布以及给定  $X$  的情况下  $Y$  的条件密度都既依赖于参数  $\beta$  又依赖于回归函数  $g(\cdot)$ 。Delecroix 等<sup>[40]</sup>证明了  $\beta$  的半参数极大似然估计是渐近有效的,保持了参数极大似然估计最重要的性质。

虽然 M 估计具有有效性、渐近正态性和自动窗宽选取等许多优点,但该方法的最大缺点是它需要解决复杂的最优化问题。尽管直接估计方法的理论性质不如 M 估计,但由于其能够提供估计量的解析形式,因此,在单指标模型的估计中直接估计方法具有一定的吸引力。关于直接估计方法可参考 Stoker<sup>[41]</sup>和 Härdle 等<sup>[42]</sup>提出的平均导数方法以及 Powell 等<sup>[43]</sup>描述的密度加权平均导数估计方法,其中密度加权平均导数估计方法不需要迭代并且容易计算,还适用于所有  $X$  的分量都是连续型随机变量的情形。Horowitz 等<sup>[44]</sup>和 Hristache 等<sup>[45]</sup>把这种方法进一步推广和修正,推广到  $X$  含有离散变量的情形,并且举例说明了单指标模型的应用。

最近,Xue 和 Zhu<sup>[46]</sup>利用经验似然方法讨论了单指标模型(1.2.1)中未知参数的估计问题,给出了未知参数  $\beta$  的两种经验似然比统计量,在适当条件下,证明了所构造的统计量的分布渐近于  $\chi^2$  分布。

Kong 和 Xia<sup>[47]</sup>考虑了单指标模型(1.2.1)中的变量选择问题,指出在线性回归模型或非参数模型中常用的去掉  $m$  个个体的交叉核实法在单指标模