

半参数模型的理论与应用

● 魏传华 著



科学出版社

半参数模型的理论与应用

魏传华 著

科学出版社

北 京

内 容 简 介

近三十年来,半参数模型在统计学和计量经济学领域里得到了广泛关注和深入研究.本书主要研究了部分线性变系数模型和半参数可加模型在测量误差或缺失数据等复杂情形下的估计和检验问题,此外,还研究了部分线性模型的检验问题和非参数协方差分析模型.

本书可供统计学和经济学等相关专业的科研人员和学生使用.

图书在版编目(CIP)数据

半参数模型的理论与应用/魏传华著. —北京:科学出版社, 2013.11

ISBN 978-7-03-038924-4

I. ①半… II. ①魏… III. ①半参数模型—研究 IV. ①O211.3

中国版本图书馆 CIP 数据核字 (2013) 第 248138 号

责任编辑:陈玉琢/责任校对:宣 慧

责任印制:赵德静/封面设计:王 浩

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

文林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2013 年 11 月第 一 版 开本: 720×1000 1/16

2013 年 11 月第一次印刷 印张: 10

字数: 200 000

定价: 56.00 元

(如有印装质量问题, 我社负责调换)

前 言

实际问题研究中, 我们往往需要探求变量之间的关系. 回归分析作为解决此类问题的一个有效统计分析方法而得到广泛的应用. 毫不夸张地说, 相关与回归分析(特别是线性回归分析) 是除了描述性统计方法之外应用最为广泛的统计分析方法. 同时, 回归分析也是数理统计领域一个内容极其丰富, 得到研究较多的分支.

近三十年来, 随着计算机技术的飞速发展, 越来越多的半参数模型被提出用以探求因变量和自变量之间蕴涵的复杂关系. 本书主要针对部分线性变系数模型和部分线性可加模型, 以及部分线性模型和非参数协方差分析模型这几类半参数模型的估计和检验问题做深入的探讨.

第 2 章 ~ 第 6 章研究了部分线性变系数模型的推断问题. 第 2 章首先提出了一种 profile Lagrange 乘子方法用以检验参数分量的线性约束条件, 其次构造了参数分量的 backfitting 估计. 第 3 章研究了异方差情形下的部分线性变系数模型, 讨论了参数分量的有效估计量的构造问题, 并提出了检验异方差的 profile 得分方法. 第 4 章针对一类部分线性变系数变量含误差模型构造了参数分量的约束估计, 并提出了针对约束条件的检验方法. 第 5 章讨论了当非参数部分的协变量不能精确观测时部分线性变系数模型的估计问题. 第 6 章讨论了部分线性变系数模型在因变量存在缺失和线性部分自变量不能精确观测时的估计问题. 第 7 章和第 8 章主要研究了部分线性可加模型. 第 7 章构造了模型的 profile 最小二乘估计及相应的约束估计, 同时提出了针对参数分量的广义似然比检验方法. 第 8 章则讨论了一类部分线性可加变量含误差模型的经验似然推断. 第 9 章主要讨论了几种利用部分线性模型检验线性关系的方法. 第 10 章则对非参数回归曲线的比较问题进行了深入研究.

本书内容大都已在国内外学术期刊上公开发表, 衷心感谢投稿过程中评审人的宝贵意见. 此外, 本书的第 2 章、第 3 章、第 9 章和第 10 章是作者博士学位论文中的内容. 衷心感谢导师吴喜之教授在我读博期间给予的指导和帮助以及我工作后一直给予的鼓励和支持. 感谢科学出版社陈玉琢编辑在本书出版过程中给予的帮助.

由于作者水平有限, 书中难免有不妥之处, 恳请同行和广大读者批评指正.

魏传华

2013 年 6 月

中央民族大学

目 录

前言

第 1 章	介绍	1
1.1	引言	1
1.2	常用非参数光滑方法和半参数模型介绍	3
1.2.1	权函数估计	4
1.2.2	样条方法	6
1.2.3	部分线性模型	7
1.2.4	变系数模型	8
1.2.5	可加模型	9
1.3	预备知识	9
1.3.1	精确方法	10
1.3.2	三阶矩 χ^2 逼近方法	10
1.3.3	F 分布逼近法	11
第 2 章	部分线性变系数模型的 profile Lagrange 乘子检验与 backfitting 估计	12
2.1	引言	12
2.2	profile Lagrange 乘子检验法	13
2.3	部分线性变系数模型的 backfitting 估计	18
2.4	数值模拟	21
2.5	定理的证明	22
第 3 章	异方差部分线性变系数模型的研究	28
3.1	引言	28
3.2	异方差情形下参数分量的有效估计	29
3.3	profile 得分检验统计量的构造及性质	36
3.4	关于一类部分线性模型异方差检验的一个注记	39
3.5	定理的证明	42
第 4 章	部分线性变系数变量含误差模型的约束估计与检验	47
4.1	引言	47
4.2	参数分量的约束估计	48
4.3	参数分量的检验	50

4.4	数值模拟	52
4.5	定理的证明	54
第 5 章	部分线性变系数测量误差模型的估计	57
5.1	引言	57
5.2	校正局部线性估计	58
5.3	估计的渐近性质	61
5.4	定理的证明	62
第 6 章	因变量缺失下部分线性变系数变量含误差模型的估计	66
6.1	引言	66
6.2	参数分量和非参数分量的几类估计	67
6.2.1	完整观测数据估计方法	67
6.2.2	插补估计	70
6.2.3	Surrogate 估计	71
6.3	因变量均值的估计	72
6.4	定理的证明	73
第 7 章	部分线性可加模型的估计与检验	80
7.1	引言	80
7.2	参数分量的 profile 最小二乘估计	82
7.3	参数分量的约束 profile 最小二乘估计	84
7.4	广义似然比检验	86
7.5	数值模拟	87
7.6	定理的证明	89
第 8 章	部分线性可加变量含误差模型的经验似然推断	93
8.1	介绍	93
8.2	参数分量的经验似然	95
8.3	数值模拟	97
8.4	定理的证明	98
第 9 章	利用部分线性模型检验线性回归关系	103
9.1	引言	103
9.2	广义似然比检验方法	106
9.3	基于导函数的非参数检验方法	113
9.4	部分线性模型的 profile 局部加权最小二乘估计	115
9.5	定理的证明	119
第 10 章	非参数协方差分析模型的研究	123
10.1	引言	123

10.2 两条回归曲线比较的虚拟变量法·····	124
10.3 光滑残差检验法·····	129
10.4 个体差异曲线检验法·····	131
参考文献·····	138
索引·····	150

第1章 介 绍

1.1 引 言

实际数据分析中, 大量问题涉及两组变量间的关系的研究. 回归分析作为解决此类问题的一个有效的统计方法而得到十分广泛的应用. “回归 (regression)”一词是由英国著名生物学家和统计学家 F.Galton 在研究人类遗传问题时提出的. 从早期的一元及多元线性回归模型, 到非线性回归模型, 以及后来为处理离散变量发展起来的广义线性模型, 回归分析一直是数理统计学研究的最热点之一. 多样化的回归模型及其拟合方法构成了数理统计学的一个内容极其丰富, 应用十分广泛的分支.

设因变量为 Y , 自变量为 $X_1, X_2, \dots, X_p$, 回归分析的主要内容就是探求因变量与自变量之间的关系, 即建立如下的模型

$$Y = f(X_1, X_2, \dots, X_p) + \varepsilon, \quad (1.1)$$

其中 ε 为误差项, $f(\cdot)$ 称为回归函数. 回归分析的主要任务包括基于 Y 和 $X_1, X_2, \dots, X_p$ 的观测值对回归函数 $f(X_1, X_2, \dots, X_p)$ 做估计, 选择出对因变量有显著影响的自变量等, 以建立 Y 与 $X_1, X_2, \dots, X_p$ 之间一个有效的关系, 从而基于该关系可以度量变量之间的相互影响程度以及用于预测.

回归模型的种类非常多. 根据回归函数 $f(\cdot)$ 的形式, 可以简单地分为线性回归模型和非线性回归模型. 线性回归模型一般可记为如下的形式

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \dots + X_p\beta_p + \varepsilon, \quad (1.2)$$

其中 $\beta_1, \beta_2, \dots, \beta_p$ 为未知参数. 该模型已经得到广泛的研究, 从参数的估计、自变量的选取、模型的诊断分析与统计推断等各个方面, 已建立起了一套完整的理论与方法体系并得到广泛的应用, 有关的内容可参考 Rao (1973), Ryan (1997) 和 Rao 等 (2008) 等. 非线性回归模型可记为如下形式

$$Y = g(X_1, X_2, \dots, X_p; \beta_1, \dots, \beta_p), \quad (1.3)$$

其中 $g(\cdot)$ 为某个关于未知参数 $\theta_1, \dots, \theta_p$ 非线性的已知函数. 此类回归模型无论从应用还是从理论都得到了较为充分的研究, 形成了较完整的体系, 可见 Bates 和 Watts (1988), Seber 和 Wild (1989) 等.

当因变量 Y 为 0-1 分布、Poisson 分布等离散分布时, 直接用最小二乘法拟合形式 (1.2) 的回归模型是不可行的. 为了解决这一问题, 20 世纪 70 年代由 J.A.Nelder 等提出并深入研究了广义线性模型, 这类模型将因变量的分布推广到了指数族分布. Logistic 回归模型和 Poisson 回归模型就是得到广泛应用的两类广义线性模型.

线性回归模型和非线性回归模型以及广义线性模型都属于参数模型, 即它们共同的特点是因变量和自变量之间的函数关系是指定的特定形式. 正如梅长林 (2000) 所说, 参数回归模型有很多优点. 首先, 模型中的参数一般具有某种物理意义, 从而有利于对分析结果的解释. 事实上, 在许多回归分析的问题中, 人们对参数的兴趣胜过回归函数本身. 在经济学中应用最为广泛的 C-D 生产函数模型可以转化为如下形式的线性回归模型

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \varepsilon,$$

其中 Y, K, L 分别是产出、资本投入和劳动投入. 模型系数 α 和 β 在经济学中的含义分别是资本和劳动的产出弹性. 其次, 参数回归模型易于统计分析, 以线性回归模型为例, 显然对于回归函数的研究就转化为对几个参数 $\beta_1, \dots, \beta_p$ 的研究上. 最后, 当我们所设定的参数模型形式与实际数据比较吻合时, 显然对整个回归函数的估计在统计上具有更高的有效性, 分析结果更具有针对性和说服力.

实际数据分析中, 因变量与自变量之间到底是什么样的关系? 我们需要从实际数据以及问题专业背景出发进行探索, 而不是为了推断上的简便而将这种关系简单地假定为线性或者其他具体的形式. 在大多数实际问题中, 人们对回归函数的有关信息知之甚少, 通常对其具体形式的假定具有很大的主观性. 设定不当的参数回归模型有可能导致对回归曲线 (或曲面) 的不符合实际的甚至是完全错误的推断. 此外, 对于越来越多的复杂数据, 用某一特定形式的参数回归模型去分析显然是不能够刻画其复杂的变化规律. 从上面的分析可以看出, 参数回归模型在应用上有一定的局限性, 如果当分析者对参数模型形式的信息了解甚少时, 利用这类模型去分析实际问题有可能会冒较大的风险, 甚至得出不符合实际的结论.

面对上面提到的问题, 那什么样的建模方式才能更好地探索出因变量和自变量之间蕴涵的复杂关系呢? 一个合理的做法是充分依靠数据本身而不是主观假定某种具体的模型形式, “让数据自己说话”, 即在模型 (1.1) 中, 对函数 $f(\cdot)$ 的具体形式不作任何假定或只作一些简单的光滑性要求, 依靠观测数据寻求 $f(\cdot)$ 的特征. 这时, 模型 (1.1) 称为非参数回归模型. 因此, 非参数回归模型尤其适合于当对回归函数的本质特征了解较少时的回归分析问题. 近三十年来借助于计算机计算能力的快速提高, 非参数回归模型 (特别是当自变量为一维的情况) 得到了广泛的研究, 已经有多种光滑方法提出用以估计未知回归函数.

虽然从理论上讲, 各种非参数光滑方法都可以推广到多个回归变量的情形, 但

是随着回归变量个数的增加, 这些方法都面临着所谓的“维数祸根”(curse of dimensionality) 问题. 这是因为非参数函数估计方法本质上讲都是局部估计或局部光滑, 要想使 $m(x)$ 在 x 点得到比较充分的估计, 必须使得 x 的邻域包含有足够多的数据. 但当 x 为多维数据时, 这个条件不易满足, x 的维数越高, 包含一定数量观测点所对应的邻域将会更大以致失去局部特征 (详细讨论可参见 Hastie 和 Tibshirani(1990)), 也就是说要使得光滑方法在高维情况下仍能适用, 则需要的观测点数据会大到难以实现的地步. 为了克服上面提到的“维数祸根”问题, 近年来统计学家与计量经济学家提出了多种多元非参数建模方法.

1.2 常用非参数光滑方法和半参数模型介绍

非参数光滑方法大致可以分为三类, 第一类权函数估计主要是基于局部加权的思想利用周围的点拟合某点处回归函数的值, 常见的最近邻估计、核估计以及后来提出的局部多项式估计都属于该类方法. 第二类最小二乘方法主要是利用参数空间来逼近 (近似) 无穷维参数空间, 基于不同的逼近思想构造参数空间的基函数, 将未知函数 (无穷维参数) 的估计问题转为 (有限个) 未知参数的估计, 从而可以通过最小二乘方法得到未知参数的估计. 该类方法主要包括 Fourier 级数估计、小波方法、多项式 (或分块多项式) 方法以及样条估计. 文献中也常称基于这类方法的估计为 Sieve 估计. 第三类光滑样条估计是一种惩罚最小二乘方法. 关于核估计 (包括局部多项式估计) 方法的详细内容可参考 Muller (1988), Nadaraya (1989), Hardle (1990,1991), Wand 和 Jones (1995), Fan 和 Gijbels (1996), Simonoff (1996), Bowman 和 Azzalini (1997), Hart (1997), Loader (1999), Pagan 和 Ullah (1999) 与 Fox (2000). 关于各类样条方法的详细内容可参考 Eubank (1988), Wahba (1990), Green 和 Silverman (1994), Dierckx (1995), Gu (2002), Ruppert, Wand 和 Carroll (2003) 和 Hansen 等 (2003). 关于经典级数估计的详细内容可参考 Thompson 和 Tapia(1990), Tarter 和 Lock(1993) 和 Efromovich (1999). 关于小波方法的详细内容可参考 Ogden (1996), Louis, Maass 和 Rieder (1997), Hardle 等 (1998), Schimernk (2000) 和 Walter 和 Shen (2001). 下面将简单介绍几类常用的光滑方法以及几种常见的多元非 (半) 参数模型.

考虑如下的一元非参数回归模型

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1.4)$$

其中 Y_i 是被解释变量观测值, x_i 为解释变量观测值, x 既可以是随机变量也可以是固定设计变量, 其主要目的是基于观测数据 $\{Y_i, x_i\}_{i=1}^n$ 得到未知回归函数 $m(x)$ 的估计值 $\hat{m}_n(x)$.

1.2.1 权函数估计

假设要估计 x_0 对应的 $m(x_0)$, 一个直观的想法就是利用周围的点, 其中距离 x_0 越近的对 $m(x_0)$ 的估计所起的作用越大, 距离 x_0 越远的点对 $m(x_0)$ 的估计所起的作用越小. 而这种作用大小的度量主要用权函数来表达. 如果令 $\{y_i, x_i\}$ 在估计 $m(x_0)$ 中对应的权函数为 $W_{ni}(x_0)$, 则有 $m(x_0)$ 的加权估计为

$$\hat{m}(x_0) = \sum_{i=1}^n W_{ni}(x_0) Y_i.$$

显然常用的三点平均、分段平均光滑以及最近邻估计都属于简单的权函数估计方法. 下面介绍应用非常广泛的核估计方法.

核估计方法

设 $K(\cdot)$ 是一个实值函数, 通常为对称的概率密度函数并称为核函数, h 是用来控制局部邻域大小的非负实数称为窗宽 (bandwidth) 或者光滑参数 (smoothing parameter). 令 $K_h(\cdot) = K(\cdot/h)/h$, $m(x_0)$ 的 Nadaraya-Watson 核估计为

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n K_h(x_i - x_0) Y_i}{\sum_{i=1}^n K_h(x_i - x_0)},$$

通常使用的核函数包括 Gauss 核函数

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$$

和对称 β 族函数

$$K(t) = \frac{1}{\beta\left(\frac{1}{2}, \gamma+1\right)} (1-t^2)_+^{\gamma},$$

其中 “+” 表示函数 $(1-t^2)^\gamma$ 的正部, $\beta(\cdot, \cdot)$ 为 β 函数. 当 $\gamma = 0, 1, 2, 3$ 时, 分别对应于均匀 (uniform) 核函数、Epanechnikov 核函数、二次 (biweight) 核函数和立方权 (triweight) 函数.

上面估计中随机分母给估计性质的研究带来一定的困难, 为了克服这一缺点, Gasser 和 Muller 提出了一种新的估计方法. 令 $x_0 = -\infty, x_{n+1} = +\infty$, 并假定 $x_0 < x_1 \leq x_2 \leq \cdots \leq x_n < x_{n+1}$, 定义新的估计为

$$\hat{m}(x_0) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(u - x_0) du Y_i,$$

其中 $s_i = \frac{x_i + x_{i+1}}{2}$.

虽然核估计在一定的条件下具有相合性与渐近正态性, 但该估计存在边界效应, 即估计在边界处收敛于实际函数的速度慢于在内点处的收敛速度. 而且 Nadaraya-Watson 估计的偏差通常较大, 尤其是在回归函数的导数值较大的地方, 而 Gasser-Muller 估计的方差通常较大.

局部多项式估计方法

注意到上面的核估计可以认为是一种加权最小二乘估计, 即对于 $m(x_0)$ 通过使得

$$\sum_{i=1}^n \{Y_i - m(x_0)\}^2 K_h(x_i - x_0)$$

达到最小来估计. 为了克服核估计的缺点, 局部多项式估计方法得到了深入的研究.

设回归函数 $m(x)$ 有 $p+1$ 阶连续导数, 则由 Taylor 展式可知, 对 $m(x)$ 定义域内的 x_0 , 其邻域内的函数值 $m(x)$ 有

$$m(x) \approx \sum_{j=0}^p \frac{m^{(j)}(x_0)}{j!} (x - x_0)^j \doteq \sum_{j=0}^p \beta_j(x_0) (x - x_0)^j,$$

其中 $m^{(j)}(\cdot)$ 表示 $m(\cdot)$ 的第 j 阶导数. 利用局部加权多项式回归方法估计, 对于 $\beta_j(x_0)$, 使得

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j(x_0) (x - x_0)^j \right\}^2 K_h(x_i - x_0)$$

达到最小予以估计. 记 $\beta_j(x_0)$ 的估计为 $\hat{\beta}_j(x_0)$, 则得回归函数及其各阶导数在 x_0 处的估计为

$$\hat{m}^{(v)}(x_0) = v! \hat{\beta}_j(x_0), \quad v = 0, 1, \dots, p,$$

特别是当上面 $p = 1$ 时, 即将 $m(x)$ 线性展开, 称对应的估计为局部线性估计.

下面对 Nadaraya-Watson 估计、Gasser-Muller 估计以及局部线性估计这三种估计的偏差以及方差做一对比, 内容见下表. 其中记 Bias 和 Var 分别为 $\hat{m}(x_0)$ 的

条件偏差与条件方差. 记 $b_n = \frac{h^2}{2} \int u^2 K(u) du$, $V_n = \frac{\sigma^2(x_0)}{nhf(x_0)} \int K^2(u) du$.

方法	Bias	Var
Nadaraya-Watson 估计	$\left(m''(x_0) + \frac{2m'(x_0)f'(x_0)}{f(x_0)} \right) b_n$	V_n
Gasser-Muller 估计	$m''(x_0)b_n$	$1.5V_n$
局部线性估计	$m''(x_0)b_n$	V_n

从上表可以看到, 局部线性估计与 Gasser-Muller 估计相比, 二者偏差一样, 前者的方差小于后者的方差. 而局部线性估计与 Nadaraya-Watson 估计相比, 二者方差一样, 前者的偏差小于后者的偏差, 而 Nadaraya-Watson 估计与 Gasser-Muller 估计相比, 前者的偏差大于后者的偏差, 但是方差小于后者的方差. 所以局部线性估计是优于 Gasser-Muller 估计与 Nadaraya-Watson 估计的.

一般来讲局部多项式方法具有以下优点.

(i) 局部多项式估计有相对小的偏差和方差. 与 Nadaraya-Watson 估计以及 Gasser-Muller 估计对应的比较见上表.

(ii) 局部多项式方法适用于各种设计. 它适用于随机设计 (random designs)、固定设计 (fixed designs)、均匀设计 (uniform designs)、分组设计 (clustered designs) 等.

(iii) 局部多项式估计没有边界效应. 局部多项式估计在边界点的估计偏差与内点的估计偏差阶数一样, 不需要在边界点处用特殊的权函数来减少边界效应.

(iv) 局部多项式估计有很好的极小极大效率 (minimax efficiency).

1.2.2 样条方法

样条方法可以看成多项式回归的一种推广. 利用分段不同阶数的多项式拟合数据, 使得两个多项式函数在连接点 (knots) 处可以允许有不连续的导数, 这样便使得估计的回归函数比整体拟合一个固定阶数的多项式更具灵活性. 通常使用的是三次样条, 即利用有二阶连续导数的分段多项式函数逼近未知的回归函数. 具体的构造思想如下.

设 $t_1, t_2, \dots, t_J$ 为固定节点, 满足 $-\infty < t_1 < t_2 < \dots < t_J < +\infty$, 即这些节点将实直线划分为如下的区间 $(-\infty, t_1], [t_1, t_2], \dots, [t_{J-1}, t_J], [t_J, +\infty)$. 三次样条函数 $s(x)$ (这里以三次样条为例, 其余阶数的情形类似) 有连续二阶导数, 并且在上面每个小区间里都是 (不超过) 三次多项式. 所有的三次样条函数形成一个 $J+4$ 维的线性空间, 这个空间常用的三次样条基有如下两个:

(1) 幂基 (power basis): $1, x, x^2, x^3, (x - t_j)_+^3 (j = 1, 2, \dots, J)$;

(2) B 样条基 (详细的定义以及相关性质可参见 Boor (1978) 和 Schumaker (1981)).

选定定的三次样条基为 $B_1, \dots, B_{J+4}$, 则三次样条函数为

$$s(x) = \sum_{j=1}^{J+4} \theta_j B_j,$$

上面的未知参数 θ_j 可通过极小化

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=1}^{J+4} \theta_j B_j(x_i) \right\}^2$$

予以估计. 记 θ_j 的估计为 $\hat{\theta}_j, j = 1, \dots, J+4$, 从而可以定义 $m(x)$ 的估计为

$$\hat{m}(x) = \sum_{j=1}^{J+4} \hat{\theta}_j B_j(x).$$

其中的 J 称为光滑参数, 其作用相当于前面介绍的局部估计中的窗宽 h . m 次 B 样条函数的基由节点唯一确定, 若选择均匀节点或其他固定方式确定节点的位置, 则基由节点个数唯一确定. 节点个数 k 在拟合数据和回归函数估计的光滑程度之间起平衡作用, 随着节点数目的增加, B 样条估计的方差越大, 偏差越小, 此时该估计更加充分地拟合数据; 随着节点数目的减少, B 样条估计的方差越小, 偏差越大, 估计出的回归函数越光滑. 因此节点控制回归函数的光滑程度.

光滑样条方法是自动选取节点的方法. 如果单纯考虑最小二乘估计, 即求函数 m , 使得

$$\sum_{i=1}^n \{Y_i - m(x_i)\}^2$$

达到最小, 显然符合条件的函数 m 为将所有观测点依次连接起来的折线, 这样的估计是没有任何意义的. 为此, 将对 m 附加一定的限制条件, 比如, 要求 m 具有二阶连续导数 (其他阶类似), 那么现在要极小化

$$\sum_{i=1}^n \{Y_i - m(x_i)\}^2 + \lambda \int (m''(x))^2 dx$$

来估计 m . 上面的解是区间 $[x_{(1)}, x_{(n)}]$ 上的唯一的一个三次样条, 且对于任一 x , 其解是 $Y_1, Y_2, \dots, Y_n$ 的线性函数. 其中 λ 称为光滑参数.

回归 B 样条与惩罚样条和光滑样条都是用样条函数来估计未知的非参数回归函数, 主要区别在于惩罚样条先保守地选择较多数目的节点 (光滑样条甚至选择每一个数据点为节点), 然后再通过惩罚系数防止估计过分地拟合数据, 调节估计函数的光滑程度. 而回归 B 样条估计直接通过节点来平衡拟合数据和估计函数的光滑程度, 所选择的光滑参数与惩罚样条不同, 因此一般回归 B 样条估计所需要的节点个数较少, 从而待估的参数也较少.

1.2.3 部分线性模型

为了克服线性模型的局限性及非参数回归模型可能带来的“维数祸根”问题, 一个努力的方向是 Engle 等 (1986) 提出的部分线性回归模型, 也称为半参数回归

模型. 其一般形式是

$$Y = X_1\beta_1 + X_2\beta_2 + \cdots + X_m\beta_m + f(T) + \varepsilon, \quad (1.5)$$

其中 $(X_1, \dots, X_p, T)$ 是自变量, 而 Y 为因变量, ε 为模型误差. 可以看出模型 (1.5) 有两部分组成, 线性部分 $X_1\beta_1 + X_2\beta_2 + \cdots + X_m\beta_m$ 称为参数分量, 非线性部分 $g(t)$ 称为非参数分量. 线性部分中的未知参数 β_j 与未知函数 $f(\cdot)$ 的估计及其性质是研究该模型的一个主要任务.

模型 (1.5) 自从 1986 年 Engle 等用于讨论关于气温与供电及电力价格的关系提出以来, 受到了统计学家与计量经济学家的广泛关注, 基本上也可以说是得到研究最多的一类多元非 (半) 参数模型. Heckman (1986), Speckman (1988), Chen (1988), Donaldand 和 Newey (1994), Yatchew (1997) 分别基于不同的光滑思想研究了模型的估计问题; Severini 和 Staniswalis (1994) 研究了广义部分线性模型, 即因变量为离散数据的情况; Liang 等 (1999) 研究了变量含误差 (error in variable) 部分线性模型; 而 Wang 等 (2004) 以及 Liang 等 (2004) 基于该模型分别研究了因变量与自变量存在缺失的情况; Zhang 等 (1998) 利用该建模方法研究了纵向数据 (longitudinal data), Li 和 Ullah (1998) 研究了 Panel data. 关于该模型的详细论述包括一些实际数据的分析可参见柴根象和洪圣岩 (1995), Hardle, Liang 和 Gao (2000), Yatchew (2003), Hardle 等 (2004).

1.2.4 变系数模型

变系数模型的建模思想早已存在, 但直到 Hastie 和 Tibshirani (1993) 的文献发表才引起了人们的关注. 其形式如下

$$Y = \sum_{j=1}^p \beta_j(U) X_j + \varepsilon, \quad (1.6)$$

其中 $(U, X_1, \dots, X_p)$ 是自变量, Y 为因变量, $\beta_j(\cdot)$ 为未知光滑函数, ε 为模型误差.

对于模型中 $\beta_j(\cdot)$ 的估计, Hastie 和 Tibshirani (1993) 提出了利用局部加权核方法以及光滑样条方法拟合模型; Fan 和 Zhang (1999) 基于局部多项式方法提出了两步估计; Eubank 等 (2004) 深入地讨论了光滑样条估计. 由于其良好的解释能力, 该模型在很多领域得到了应用. Cai, Fan 和 Li (2000) 研究了广义变系数模型; Fan, Zhang 和 Zhang (2001) 提出了广义似然比检验, 研究了针对该模型的假设检验问题; Chen 和 Tsay (1993) 以及 Cai, Fan 和 Yao (2000) 将该模型用到了非线性时间序列数据上; Fotheringham, Brundom 和 Charlton (2002) 将该模型用到了空间数据分析上, 用以揭示空间数据的非平稳特征, 并将该模型称之为地理加权回归 (geographically weighted regression) 模型. 在纵向数据的处理上, 该模型可以很好

地揭示各协变量对因变量影响的变化趋势,从而得到了人们的广泛研究,针对纵向数据变系数模型的估计、核方法 (Wu 和 Chiang(2000))、局部线性方法 (Hoover et al., 1998)、基函数法 (Huang et al., 2002)、光滑样条 (Chiang et al., 2001)、多项式样条方法 (Huang et al., 2004) 都已经得到了研究. 关于变系数模型的介绍可参考王启华, 史宁中和耿直 (2010).

1.2.5 可加模型

为了使得模型更易解释且不受“维数祸根”问题的影响, Hastie 和 Tibshirani (1986) 提出了可加模型并给出了相应的拟合方法. 该模型为

$$Y = \alpha + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) + \varepsilon, \quad (1.7)$$

其中 $X_1, \dots, X_p$ 为自变量, Y 为因变量, $f_j(\cdot)$ 为任意的一元函数且只与相应的回归变量有关.

对于未知函数 $f_j(\cdot)$, 主要有两种估计方法: 第一种为 Friedman 和 Stuetzle (1981) 提出的 Back-Fitting 算法, 该算法被 Hastie 和 Tibshirani (1986) 用到可加模型以及广义可加模型上, 在 R(S-Plus) 上并有专门的程序包; 第二种方法为 Linton 和 Nielsen (1995), Linton (1997) 提出的边界积分法 (marginal intergation method). Hastie 和 Tibshirani (1990) 是关于该模型早期研究的一个系统总结, 关于最近的研究内容可参考 Hardle 等 (2004).

1.3 预备知识

在非参数与半参数模型的检验中, 经常遇到如下形式的检验统计量

$$T = \frac{\boldsymbol{\varepsilon}^T \mathbf{P}_1 \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^T \mathbf{P}_2 \boldsymbol{\varepsilon}}, \quad (1.8)$$

如果令 T 的观测值为 t , 则 T 的检验 p 值为

$$p_0 = P_{H_0}(T > t) = P_{H_0}\{\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon} > 0\}, \quad (1.9)$$

其中 $\mathbf{W} = \mathbf{P}_1 - t\mathbf{P}_2$, 即在 H_0 之下, 事件 $\{T > t\}$ 发生的概率. 对于给定的显著水平 α , 若 $p_0 > \alpha$, 则接受 H_0 ; 若 $p_0 \leq \alpha$, 则拒绝 H_0 . 如果模型误差服从正态分布, 显然上面的检验统计量为正态变量二次型之比的形式. 从而可以利用正态变量二次型分布的有关结果来求得检验 p 值. 为了以后章节的应用, 下面列出一种精确算法和两种逼近算法. 详细内容可参考梅长林 (2000).

1.3.1 精确方法

基于 Imhof (1961) 关于正态变量二次型分布的一个结果, 我们有下面的结论.

结论 1.1 设 n 维随机变量 $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)^T \sim N(\mathbf{0}, \mathbf{I})$, $\mathbf{A}$ 为实对称矩阵, 其互不相同的非零特征值为 $\lambda_1, \lambda_2, \dots, \lambda_m$, 相应的重数分别为 $h_1, h_2, \dots, h_m$, 则对于二次型 $Q = \boldsymbol{\xi}^T \mathbf{A} \boldsymbol{\xi}$, 有

$$P(Q > 0) = \frac{1}{2} + \frac{1}{\pi} \int_0^{+\infty} \frac{\sin[\theta(u)]}{u\rho(u)} du, \quad (1.10)$$

其中

$$\theta(u) = \frac{1}{2} \sum_{k=1}^m [h_k \tan^{-1}(\lambda_k u)], \quad (1.11)$$

$$\rho(u) = \prod_{k=1}^m (1 + \lambda_k^2 u^2)^{h_k/4}, \quad (1.12)$$

进一步, 若令

$$T_U = \left(\pi H U^H \prod_{k=1}^m |\lambda_k|^{h_k/2} \right)^{-1}, \quad (1.13)$$

则有 $|t_U| < T_U$, 其中 $U > 0$ 为任意常数, $H = \frac{1}{2} \sum_{k=1}^m h_k$.

实际应用中, 补充式 (1.10) 中被积函数在原点的定义为 $\lim_{t \rightarrow 0} \frac{\sin[\theta(t)]}{t\rho(t)} = \frac{1}{2} \sum_{k=1}^m h_k \lambda_k$, 对于任一 $U > 0$, 可利用数值方法计算积分 $\int_0^U \frac{\sin[\theta(t)]}{t\rho(t)} dt$, 并可由式 (1.13) 估计其余项. 但用数值方法计算 $\int_0^U \frac{\sin[\theta(t)]}{t\rho(t)} dt$ 所造成的误差取决于具体的数值积分方法.

显然可以直接利用该推论计算式 (1.9), 将推论中的 $\mathbf{A}$ 换作 $\mathbf{W}$. 值得注意的是由于 T 关于 ε 的常数倍具有不变性, 所以假定模型 $\varepsilon \sim N(\mathbf{0}, \mathbf{I})$ 不影响最后结果.

1.3.2 三阶矩 χ^2 逼近方法

前面给出的是精确的计算公式, 计算量非常大. 下面给出计算 p 值的三阶矩逼近方法, 该方法可以大大降低计算量. 三阶矩 χ^2 逼近最早由 Pearson (1959) 用于逼近非中心 χ^2 分布, Imhof (1961) 将其应用于逼近正态变量二次型的分布.

结论 1.2 设式 (1.9) 中的误差项 ε 为正态分布 $N(0, \mathbf{I}_n)$, 且 $\text{tr}(\mathbf{M})^3 \neq 0$. 若应用三阶矩 χ^2 逼近求检验 p 值 p_0 , 则有

(i) 当 $\text{tr}(\mathbf{M}^3) > 0$ 时,

$$p_0 = P_{H_0}(R > r) \approx P(\chi_d^2 > d - h), \quad (1.14)$$

(ii) 当 $\text{tr}(\mathbf{M}^3) < 0$ 时,

$$p_0 = P_{H_0}(R > r) \approx P(\chi_d^2 < d - h), \quad (1.15)$$

其中 χ_d^2 是自由度为 d 的 χ^2 变量, 且

$$\begin{cases} d = \frac{\{\text{tr}(\mathbf{M}^2)\}^3}{\{\text{tr}(\mathbf{M}^3)\}^2}, \\ h = \frac{\text{tr}(\mathbf{M}^2) \text{tr}(\mathbf{M})}{\text{tr}(\mathbf{M}^3)}. \end{cases} \quad (1.16)$$

1.3.3 F 分布逼近法

前面给出了计算统计量检验 p 值的精确方法和三阶矩 χ^2 逼近方法, 下面给出一种更简单的方法, 我们称之为 F 分布逼近法, 即用 F 分布逼近统计量 T 的分布, 其主要思想是分别用某适当倍数和自由度的 χ^2 分布逼近 T 中的分子与分母中的正态变量二次型的分布, 然后再用相应自由度的 F 分布逼近 T 的分布, 该方法也是 Cleveland 和 Devlin (1988) 等在讨论非参数回归模型中的假设检验问题中使用的.

结论 1.3 设式 (1.9) 中的误差项 ε 为正态分布 $N(0, \mathbf{I}_n)$, 若应用 F 分布逼近方法求检验 p 值 p_0 , 则有

$$p = P(T > t) \approx P\left(F(r_1, r_2) > \frac{\text{tr}\mathbf{A}_2}{\text{tr}\mathbf{A}_1}t\right),$$

$$\text{其中 } r_1 = \frac{[\text{tr}(\mathbf{A}_1)]^2}{\text{tr}(\mathbf{A}_1^2)}, r_2 = \frac{[\text{tr}(\mathbf{A}_2)]^2}{\text{tr}(\mathbf{A}_2^2)}.$$

第2章 部分线性变系数模型的 profile Lagrange 乘子检验与 backfitting 估计

部分线性变系数模型作为变系数模型与部分线性模型的推广, 是一种应用非常广泛的数据分析模型. 类似于其他半参数模型, 模型中参数分量 (常值系数) 的估计和检验往往是我们感兴趣的问题. 本章首先研究的是针对常值系数的假设检验问题, 在传统的 Lagrange 乘子检验基础上提出了一种新的 profile Lagrange 乘子检验统计量, 并证明了该统计量在原假设下的渐近分布为卡方分布, 从而将 Lagrange 乘子检验方法推广到半参数模型上. 本章的另外一部分是从可加模型的角度给出了部分线性变系数模型的 backfitting 估计, 并研究了常值系数估计的渐近性质. 本章的主要内容来自于文献魏传华和吴喜之 (2008b, 2008c).

2.1 引言

为了更好地探求变量之间蕴涵的复杂关系, 近三十年来借助于计算机强大计算能力而发展起来的非参数回归方法得到了人们的广泛关注. 非参数回归方法克服了参数回归方法主观假定函数形式的严重缺陷, 充分依靠数据本身探求变量之间的关系. 但是非参数回归方法在处理高维数据时会遇到“维数祸根”问题, 为了克服这一局限性, 多种多元非参数和半参数模型相继提出, 比如可加模型 (Breiman and Friedman, 1985; Hastie and Tibshirani, 1990)、变系数模型 (Hastie and Tibshirani, 1993; Fan and Zhang, 1999)、部分线性模型 (Green and Silverman, 1994; Hardle et al., 2000)、多指标模型 (Hardle and Stoker, 1989) 以及这些模型的混合形式 (Carroll et al. (1997)). 其中的变系数模型在最近得到了人们的重视. 该模型的结构如下

$$Y = \sum_{j=1}^p \beta_j(U) X_j + \varepsilon, \quad (2.1)$$

其中 $(U, X_1, \dots, X_p)$ 是自变量, Y 为因变量, ε 为模型误差.

对于变系数模型 (2.1), 实际问题研究中最普遍的情况是一部分自变量对因变量的影响是随变量 U 而改变的, 但另外一部分自变量对因变量的影响是常值, 此时变系数模型转化为如下的部分线性变系数模型, 其形式可记为

$$Y = \boldsymbol{\alpha}^T(U) \mathbf{X} + \boldsymbol{\beta}^T \mathbf{Z} + \varepsilon, \quad (2.2)$$

其中 $(U, \mathbf{X}^T, \mathbf{Z}^T)$ 是自变量, 不失一般性, 下面的叙述中假定 U 为一维变量, Y 为因变量, ε 为模型误差, 有 $E(\varepsilon) = 0$ 和 $\text{Var}(\varepsilon) = \sigma^2$, 常值系数 $\beta = (\beta_1, \beta_2, \dots, \beta_q)^T$ 为 q 维未知待估参数, $\alpha(\cdot) = (\alpha_1(\cdot), \alpha_1(\cdot), \dots, \alpha_p(\cdot))^T$ 为一列未知函数. 一般称 $\alpha^T(U)\mathbf{X}$ 为模型的非参数部分, 而称 $\beta^T\mathbf{Z}$ 为模型的线性部分. 显然, 当 $\mathbf{Z} = \mathbf{0}$ 时, 该模型即为形式 (2.1) 的变系数模型. 当 $p = 1, \mathbf{X} = \mathbf{1}$ 时, 模型即为已经得到了充分研究的部分线性模型.

Zhang 等 (2002) 基于局部多项式估计最早研究了模型 (2.2), Zhou 和 You (2004) 基于小波方法研究了该模型, Xia 等 (2004) 基于局部线性方法对该模型提出了一种新的有效估计, Ahmad 等 (2005) 提出了级数估计, Fan 和 Huang (2005) 对参数分量提出了 profile 最小二乘估计方法, 并基于广义似然比方法研究了模型的检验问题.

正如其他的半参数模型, 模型 (2.2) 中常值系数 β 往往是研究的重点, 本章首先要研究的是如下的假设检验问题

$$H_0 : \mathbf{A}\beta = \mathbf{b}, \quad (2.3)$$

其中 $\mathbf{A}$ 为 $k \times q$ 的已知矩阵, 且 $\text{rank}(\mathbf{A}) = k$, $\mathbf{b}$ 为 $k \times 1$ 已知向量. 同线性模型一样, 模型 (2.2) 中的很多问题都可以归结到形如 (2.3) 的假设检验问题, Fan 和 Huang (2005) 基于广义似然比方法也讨论了这种类型的假设检验问题. 2.2 节将给出模型在原假设条件下的约束 profile 最小二乘估计, 并且基于约束估计构造了 profile Lagrange 乘子 (PLM) 检验统计量.

部分线性变系数模型从结构上可以看成一种特殊的具有两部分的可加模型, 2.3 节将基于 Backfitting 方法给出了模型 (2.2) 的估计, 并研究了常值系数估计的渐近性质.

2.2 profile Lagrange 乘子检验法

最近有多种方法提出用以估计模型 (2.2), 下面将采用 Fan 和 Huang(2005) 提出的基于局部线性光滑的 profile 最小二乘估计方法. 为了叙述的连贯性, 下面首先对该估计方法做一简单介绍.

假设有对模型 (2.2) 的 n 组观测值为 $\{u_k, x_{k1}, \dots, x_{kp}, z_{k1}, \dots, z_{kq}, y_k\}$, $k = 1, 2, \dots, n$. 首先假定模型 (2.2) 中 $(\beta_1, \beta_2, \dots, \beta_q)^T$ 已知, 则模型转化为如下形式的变系数模型

$$y_i^* = \alpha_1(u_i)x_{i1} + \dots + \alpha_p(u_i)x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.4)$$

其中 $y_i^* = y_i - (z_{i1}\beta_1 + \cdots + z_{iq}\beta_q)$. 针对变系数模型 (2.4), 利用基于局部线性光滑的局部加权最小二乘法来估计未知系数函数. 给定 u_0 邻域内的一点 u , 对 $\alpha_j(u)$ 利用 Taylor 展开式有

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha'_j(u_0)(u - u_0), \quad j = 1, 2, \dots, p, \quad (2.5)$$

从而可通过针对 $a_j(u_0), a'_j(u_0)$ 的极小化

$$\sum_{i=1}^n \left[y_i^* - \sum_{j=1}^p \{ \alpha_j(u_0) + \alpha'_j(u_0)(u - u_0) \} x_{ij} \right]^2 K_h(u_i - u_0) \quad (2.6)$$

来估计 $(\hat{\alpha}_j(u_0), \hat{\alpha}'_j(u_0), j = 1, 2, \dots, p)$, 其中 $K_h(\cdot) = K(\cdot/h)/h$, K 是核函数, h 是窗宽.

为了叙述方便, 采用下面的矩阵形式, 记

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_n^T \end{bmatrix} = \begin{bmatrix} z_{11} & \cdots & z_{1q} \\ z_{21} & \cdots & z_{2q} \\ \vdots & & \vdots \\ z_{n1} & \cdots & z_{nq} \end{bmatrix}, \\ \mathbf{M} &= \begin{bmatrix} \alpha_1(u_1)x_{11} + \cdots + \alpha_p(u_1)x_{1p} \\ \alpha_1(u_2)x_{21} + \cdots + \alpha_p(u_2)x_{2p} \\ \vdots \\ \alpha_1(u_n)x_{n1} + \cdots + \alpha_p(u_n)x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}^T(u_1)\mathbf{x}_1 \\ \boldsymbol{\alpha}^T(u_2)\mathbf{x}_2 \\ \vdots \\ \boldsymbol{\alpha}^T(u_n)\mathbf{x}_n \end{bmatrix}, \\ \mathbf{D}_{u_0} &= \begin{bmatrix} \mathbf{x}_1^T & \mathbf{x}_1^T \frac{u_1 - u_0}{h} \\ \mathbf{x}_2^T & \mathbf{x}_2^T \frac{u_2 - u_0}{h} \\ \vdots & \vdots \\ \mathbf{x}_n^T & \mathbf{x}_n^T \frac{u_n - u_0}{h} \end{bmatrix}, \end{aligned}$$

以及 $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$; $\mathbf{W}_{u_0} = \text{diag}(K_h(u_1 - u_0), K_h(u_2 - u_0), \dots, K_h(u_n - u_0))$. 从而模型 (2.4) 可写为如下的矩阵形式

$$\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{M} + \boldsymbol{\varepsilon}. \quad (2.7)$$

基于问题 (2.6), 由广义最小二乘法可得

$$[\hat{\alpha}_1(u_0), \dots, \hat{\alpha}_p(u_0), h\hat{\alpha}'_1(u_0), \dots, h\hat{\alpha}'_p(u_0)]^T$$

$$=(D_{u_0}^T W_{u_0} D_{u_0})^{-1} D_{u_0}^T W_{u_0} (Y - Z\beta), \quad (2.8)$$

则 M 的初次估计为

$$\widetilde{M} = S(Y - Z\beta),$$

其中

$$S = \begin{bmatrix} (x_1^T \ 0) \{D_{u_1}^T W_{u_1} D_{u_1}\}^{-1} D_{u_1}^T W_{u_1} \\ (x_2^T \ 0) \{D_{u_2}^T W_{u_2} D_{u_2}\}^{-1} D_{u_2}^T W_{u_2} \\ \vdots \\ (x_n^T \ 0) \{D_{u_n}^T W_{u_n} D_{u_n}\}^{-1} D_{u_n}^T W_{u_n} \end{bmatrix}.$$

将 M 的初次估计代入式 (2.7), 整理可得如下的线性回归模型

$$(I - S)Y = (I - S)Z\beta + \varepsilon. \quad (2.9)$$

利用最小二乘法估计上面的模型, 得 β 的 profile 最小二乘估计为

$$\hat{\beta} = [Z^T(I - S)^T(I - S)Z]^{-1} Z^T(I - S)^T(I - S)Y, \quad (2.10)$$

得 M 的最终估计为

$$\hat{M} = S(Y - Z\hat{\beta}).$$

从而 Y 的拟合值为

$$\hat{Y} = \hat{M} + Z\hat{\beta} = LY, \quad (2.11)$$

其中

$$L = S + (I - S)Z[Z^T(I - S)^T(I - S)Z]^{-1} Z^T(I - S)^T(I - S).$$

上面拟合过程中, 窗宽 h 可由下述计算量较小的交叉证实法见 (Hastie 和 Tibshirani(1990) 第 3 章) 来确定. 设 $h > 0$ 为任一给定的参数值, 为明确拟合值与 h 的关系, 记

$$\hat{Y}(h) = (\hat{y}_1(h), \hat{y}_2(h), \dots, \hat{y}_n(h))^T = L(h)Y,$$

其中 $L(h) = (L_{ij}(h))$ 如式 (2.11) 所示. 令

$$CV(h) = \sum_{i=1}^n \left\{ \frac{y_i - \hat{y}_i(h)}{1 - L_{ii}(h)} \right\}^2 = \sum_{i=1}^n \frac{\hat{\varepsilon}_i^2(h)}{[1 - L_{ii}(h)]^2}, \quad (2.12)$$

其中 $\hat{\varepsilon}_i(h) = y_i - \hat{y}_i(h)$ ($i = 1, 2, \dots, n$) 为利用所有观测拟合模型 (2.2) 所得残差. 由于 L 有明确的表达式, 因此其主对角线上的各元素 $L_{ii}(h)$ 是容易求得的. 广义交叉证实法即选择 h_0 使得 $CV(h)$ 达到最小.

约束 profile 最小二乘估计

为了构造检验统计量, 还需要在附加假设条件 (2.3) 的情况下估计模型 (2.2). 由上面的估计过程可以看到 β 的估计是基于线性模型 (2.9) 利用最小二乘估计得到的. 下面给出模型的约束 profile 最小二乘估计方法.

基于模型 (2.9) 以及约束条件 (2.3), 构造辅助函数

$$F(\beta, \lambda) = [(I - S)Y - (I - S)Z\beta]^T[(I - S)Y - (I - S)Z\beta] + 2\lambda^T(A\beta - b), \quad (2.13)$$

其中 λ 为 k 维 Lagrange 乘子. 对函数 $F(\beta, \lambda)$ 分别对 β, λ 求导, 并令偏导数等于 0, 有

$$\begin{cases} \frac{\partial F(\beta, \lambda)}{\partial \beta} = -2Z^T(I - S)^T(I - S)Y + 2Z^T(I - S)^T(I - S)Z\beta + 2A^T\lambda = 0, \\ \frac{\partial F(\beta, \lambda)}{\partial \lambda} = A\beta - b = 0. \end{cases} \quad (2.14)$$

整理可得 β 和 λ 的解为

$$\begin{cases} \hat{\beta}_r = \hat{\beta} - (\tilde{Z}^T \tilde{Z})^{-1} A^T [A(\tilde{Z}^T \tilde{Z})^{-1} A^T]^{-1} (A\hat{\beta} - b), \\ \hat{\lambda} = [A(\tilde{Z}^T \tilde{Z})^{-1} A^T]^{-1} (A\hat{\beta} - b), \end{cases} \quad (2.15)$$

其中 $\tilde{Z} = (I - S)Z$, $\hat{\beta}$ 为 β 的无约束 profile 最小二乘估计 (见式 (2.10)). 从而可得到变系数部分 M 的约束估计为

$$\hat{M}_r = S(Y - Z\hat{\beta}_r).$$

最后得 Y 在约束条件下的拟合值为

$$\hat{Y}_r = \hat{M}_r + Z\hat{\beta}_r. \quad (2.16)$$

类似于线性回归模型的约束最小二乘估计, 下面要证明 $\hat{\beta}_r$ 确实是线性约束条件 $A\beta = b$ 下 β 的估计, 只需证明如下两点.

(a) $\hat{\beta}_r$ 满足约束条件, 即 $A\hat{\beta}_r = b$.

(b) 对一切满足 $A\beta = b$ 的 β , 都有 $\|\bar{Y} - \bar{Z}\beta\|^2 \geq \|\bar{Y} - \bar{Z}\hat{\beta}_r\|^2$.

对于 (a), 由式 (2.15) 即得

$$A\hat{\beta}_r = A\hat{\beta} - (A\hat{\beta} - b) = b.$$

对于 (b), 证明方法与线性回归模型约束估计的证明完全相同, 在此省略.

profile Lagrange 乘子检验统计量

针对检验问题 (2.3), Fan 和 Huang (2005) 构造了 profile 广义似然比检验统计量和 Wald 检验统计量, 并给出了这两类统计量在原假设和备择假设下的渐近分布. 下面基于上一部分给出的约束估计, 提出 profile Lagrange 乘子检验统计量并研究其渐近性质.

类似于线性回归模型中 Lagrange 乘子检验统计量的构造, 下面基于 Lagrange 乘子的估计构造检验统计量. 首先假设问题 $H_0: \mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ 等价于针对 Lagrange 乘子的检验 $H_0^*: \boldsymbol{\lambda} = \mathbf{0}$. 由上一部分知 Lagrange 乘子 $\boldsymbol{\lambda}$ 的估计为

$$\hat{\boldsymbol{\lambda}} = \left[\mathbf{A}(\overline{\mathbf{Z}}^T \overline{\mathbf{Z}})^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{b}). \quad (2.17)$$

由 $\hat{\boldsymbol{\beta}}$ 的渐近性质 $\hat{\boldsymbol{\beta}} \rightarrow N(\boldsymbol{\beta}, \sigma^2(\overline{\mathbf{Z}}^T \overline{\mathbf{Z}})^{-1})$, 得

$$\hat{\boldsymbol{\lambda}} \rightarrow N\left(\mathbf{A}\boldsymbol{\beta} - \mathbf{b}, \sigma^2 \left[\mathbf{A}(\overline{\mathbf{Z}}^T \overline{\mathbf{Z}})^{-1} \mathbf{A}^T \right]^{-1}\right).$$

基于上面的结果, 针对假设检验 H_0^* 构造如下检验统计量

$$T_{\text{PLM}} = \frac{\hat{\boldsymbol{\lambda}}^T \left[\mathbf{A}(\overline{\mathbf{Z}}^T \overline{\mathbf{Z}})^{-1} \mathbf{A}^T \right] \hat{\boldsymbol{\lambda}}}{\sigma^2}, \quad (2.18)$$

由于其中的误差方差 σ^2 往往是未知的, 构造其在约束条件下的估计为

$$\hat{\sigma}_r^2 = \frac{(\overline{\mathbf{Y}} - \overline{\mathbf{Z}}\hat{\boldsymbol{\beta}}_r)^T (\overline{\mathbf{Y}} - \overline{\mathbf{Z}}\hat{\boldsymbol{\beta}}_r)}{n}, \quad (2.19)$$

其中 $\overline{\mathbf{Y}} = (\mathbf{I} - \mathbf{S})\mathbf{Y}$. 可以证明 $\hat{\sigma}_r^2$ 在原假设成立时是 σ^2 的相合估计, 所以具体计算中可用 $\hat{\sigma}_r^2$ 代替检验统计量中的 σ^2 .

参数回归模型的分析中, 对于类似 (2.3) 形式的线性假设检验问题, 广义似然比检验方法、Wald 检验方法和 Lagrange 乘子检验方法对应的检验统计量在原假设成立的情况下都趋于 χ^2 分布, 即有 Wilks 现象. 对于针对半参数模型 (2.2) 的检验问题 (2.3), 由于讨厌函数 $\boldsymbol{\alpha}(\cdot)$ 是完全非参数化的, 即为无穷维参数, 所以一个自然的问题就是该检验问题对应的统计量在原假设成立的情况下的渐近分布是否为 χ^2 分布? Fan 和 Huang (2005) 证明了 profile 广义似然比检验统计量和 Wald 检验统计量是存在 Wilks 现象的. 下面给出关于前面所构造的 profile Lagrange 乘子检验统计量 T_{PLM} 的回答.

定理 2.1 如果原假设 H_0 成立以及满足 2.5 节的条件 (A.1)~ 条件 (A.6), T_{PLM} 的渐近分布为 χ_k^2 . 其中 χ_k^2 表示自由度为 k 的 χ^2 分布.

该定理表明 T_{PLM} 的渐近零分布和参数 $\boldsymbol{\beta}, \sigma^2$, 以及 U 的设计密度 $f(\cdot)$, 未知系数函数 $\boldsymbol{\alpha}(\cdot)$ 是独立的, 是服从于自由度为 k 的 χ^2 分布. Fan 和 Huang (2005) 提

出的 profile 广义似然比以及 Wald 统计量也具有这些优良的性质, 从而本书的研究也证明了这三个基本的检验方法在处理半参数问题上的渐近等价性, 从而把它们的研究推广到了一个新的领域. 基于上面的结论, 可以利用渐近分布或者模拟求出检验问题 (2.3) 的临界值.

针对原假设 (2.4), 考虑相对应的局部备择假设为

$$H_a : \mathbf{A}\boldsymbol{\beta} = \mathbf{b} + n^{-1/2}\boldsymbol{\delta}, \quad (2.20)$$

其中 $\boldsymbol{\delta}$ 为一有限常数列.

定理 2.2 在备择假设 H_a 情况下, 以及满足式 (2.5) 中的条件 (A.1)~条件 (A.6), T_{PLM} 渐近分布为自由度为 k , 非中心参数为 λ 的 χ^2 分布. 其中

$$\lambda = \boldsymbol{\delta}^T (\sigma^2 \mathbf{A} \boldsymbol{\Sigma}^{-1} \mathbf{A}^T)^{-1} \boldsymbol{\delta} = n(\mathbf{A}\boldsymbol{\beta} - \mathbf{b})^T (\sigma^2 \mathbf{A} \boldsymbol{\Sigma}^{-1} \mathbf{A}^T)^{-1} (\mathbf{A}\boldsymbol{\beta} - \mathbf{b}), \quad (2.21)$$

其中 $\boldsymbol{\Sigma} = E(\mathbf{Z}\mathbf{Z}^T) - E[E(\mathbf{Z}\mathbf{X}^T|U)E(\mathbf{X}\mathbf{X}^T|U)^{-1}E(\mathbf{X}\mathbf{Z}^T|U)]$.

值得一提的是我们所提出的 profile Lagrange 乘子检验也能很容易地推广到其他类型的半参数模型上, Zhang 等 (2011) 将该方法推广到模型 (2.2) 中自变量 $\mathbf{Z}$ 不能精确观测的情形, Feng 和 Xue (2013) 利用该方法研究了模型 (2.2) 中自变量 $\mathbf{X}$ 不能精确观测的情形. 利用该检验方法研究单指标模型等复杂形式的半参数模型是值得研究的问题.

2.3 部分线性变系数模型的 backfitting 估计

Backfitting 方法是由 Friedman 和 Stuetzle (1981) 提出用以研究投影寻踪 (projection pursuit) 回归模型的, 后来被 Hastie 和 Tibshirani (1986) 用来拟合可加模型 (additive models), 从而引起了人们的重视. 近年来, 一些文献利用 backfitting 方法研究了其他类型的半参数模型. 对于部分线性模型, Speckmen (1988) 研究了基于核方法的 backfitting 估计. 而 Liang (2006) 研究了基于局部线性方法的 backfitting 估计, Hu 等 (2004) 则基于纵向数据 (longitudinal) 研究了 backfitting 估计, 并且他们都将 backfitting 估计与 profile 最小二乘估计做了比较. 对于部分线性可加模型, Opsomer 和 Ruppert (1999) 研究了线性部分未知参数的 backfitting 估计的渐近性质, 并提供了一种窗宽选择方法.

对于模型 (2.2), 可将其视为由线性部分 $\boldsymbol{\beta}^T \mathbf{Z}$ 与变系数部分 $\boldsymbol{\alpha}^T(U)\mathbf{X}$ 组成的可加模型, 从而可以利用 backfitting 方法进行估计. 值得注意的是, 如果 U 表示数据 $(Y, \mathbf{X}, \mathbf{Z})$ 的 (空间) 观测位置, 那么模型 (2.2) 即为 Fotheringham 等 (2002) 以及魏传华和梅长林 (2005, 2006) 研究的半参数空间变系数回归模型 (混合地理加权回

归模型). 魏传华和梅长林 (2006) 结合局部加权回归技术 (基于核方法的局部加权最小二乘估计) 与最小二乘估计给出了模型的 backfitting 估计, 但是该文献只是通过数值实验考察了常值系数 backfitting 估计的优良性, 并没有给出估计的大样本性质. 下面将给出模型 (2.2) 的 backfitting 估计以及常值系数估计的渐近性质, 其中对变系数部分的拟合将基于局部线性方法.

如果模型 (2.2) 中 $(\beta_1, \beta_2, \dots, \beta_q)^T$ 已知 (给定), 则模型 (2.2) 转化为如下形式的变系数模型

$$y_i^* = \alpha_1(u_i)x_{i1} + \dots + \alpha_p(u_i)x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

其中 $y_i^* = y_i - (z_{i1}\beta_1 + \dots + z_{iq}\beta_q)$. 针对上面的变系数模型, 利用基于局部线性光滑的局部加权最小二乘法来估计未知系数函数. 同 2.2 节中的 profile 最小二乘估计, 可得到变系数部分 $\mathbf{M}$ 的拟合值为

$$\hat{\mathbf{M}} = \mathbf{S}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}), \quad (2.22)$$

其中 $\mathbf{S}$ 同的定义同 2.2 节.

假设模型 (2.2) 中的变系数部分已知, 则模型变为如下的线性回归模型

$$y_i - (\alpha_1(u_i)x_{i1} + \dots + \alpha_p(u_i)x_{ip}) = z_{i1}\beta_1 + \dots + z_{iq}\beta_q + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

从而得 $\boldsymbol{\beta}$ 的最小二乘估计为

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_q)^T = [\mathbf{Z}^T \mathbf{Z}]^{-1} \mathbf{Z}^T (\mathbf{Y} - \mathbf{M}).$$

从而线性部分拟合值为

$$\hat{\mathbf{f}} = (z_1^T \hat{\boldsymbol{\beta}}, z_2^T \hat{\boldsymbol{\beta}}, \dots, z_n^T \hat{\boldsymbol{\beta}})^T = \mathbf{Z} \hat{\boldsymbol{\beta}} = \mathbf{Z} [\mathbf{Z}^T \mathbf{Z}]^{-1} \mathbf{Z}^T (\mathbf{Y} - \mathbf{M}). \quad (2.23)$$

根据 backfitting 原理 (详见 Hastie 和 Tibshirani(1990)¹¹⁸⁻¹²⁰), 由式 (2.22) 和式 (2.23) 得到下面的估计方程

$$\begin{cases} \mathbf{M} = \mathbf{S}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}), \\ \mathbf{Z}\boldsymbol{\beta} = \mathbf{Z}[\mathbf{Z}^T \mathbf{Z}]^{-1} \mathbf{Z}^T (\mathbf{Y} - \mathbf{M}). \end{cases} \quad (2.24)$$

将第一式代入第二式, 并假定矩阵 $\mathbf{Z}^T (\mathbf{I} - \mathbf{S}) \mathbf{Z}$ 可逆, 则可求得

$$\begin{cases} \hat{\boldsymbol{\beta}}_{\text{BF}} = [\mathbf{Z}^T (\mathbf{I} - \mathbf{S}) \mathbf{Z}]^{-1} \mathbf{Z}^T (\mathbf{I} - \mathbf{S}) \mathbf{Y}, \\ \hat{\mathbf{M}}_{\text{BF}} = \mathbf{S}(\mathbf{Y} - \mathbf{Z} \hat{\boldsymbol{\beta}}). \end{cases} \quad (2.25)$$

从而 $\mathbf{Y}$ 的拟合值为

$$\hat{\mathbf{Y}} = \hat{\mathbf{M}}_{\text{BF}} + \mathbf{Z} \hat{\boldsymbol{\beta}}_{\text{BF}} = \mathbf{L}_{\text{BF}} \mathbf{Y},$$

其中

$$L_{\text{BF}} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\mathbf{Z}[\mathbf{Z}^T(\mathbf{I} - \mathbf{S})\mathbf{Z}]^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{S}).$$

容易看出参数分量 β 与系数函数 $\alpha(\cdot)$ 的 backfitting 估计与 2.2 节中对应的 profile 最小二乘估计在表达上非常类似. 如果矩阵 $(\mathbf{I} - \mathbf{S})$ 对称幂等, 则这两种估计方法是相同的. 但是模型的特点决定了矩阵 $(\mathbf{I} - \mathbf{S})$ 一般不是对称幂等的, 所以这两种估计方法是不同的, 同时也决定了这两种估计在很多方面的差别. 对于部分可加模型以及部分线性模型, 前面提到的 Speckman (1988), Opsomer 和 Ruppert (1999) 和 Liang (2006) 都分别针对常值系数的两种估计做了渐近性质方面的比较.

下面是关于常值系数 β 的 Basck-Fitting 估计的性质, 证明过程中需要的一些常规条件将在附录中给出, 另外记 $\mu_i = \int \mu^i K(\mu) d\mu, \nu_i = \int \mu^i K^2(\mu) d\mu$, 并令 $\Gamma(U) = E(\mathbf{X}\mathbf{X}^T|U)$, $\Phi(U) = E(\mathbf{X}\mathbf{Z}^T|U)$.

定理 2.3 对于常值系数 β 的 backfitting 估计有

$$E(\hat{\beta}_{\text{BF}} - \beta | \mathbf{X}, \mathbf{Z}, U) = [\mathbf{Z}^T(\mathbf{I} - \mathbf{S})\mathbf{Z}]^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{S})\mathbf{M},$$

$$\text{Var}(\hat{\beta}_{\text{BF}} | \mathbf{X}, \mathbf{Z}, U) = \sigma^2 [\mathbf{Z}^T(\mathbf{I} - \mathbf{S})\mathbf{Z}]^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^T\mathbf{Z}[\mathbf{Z}^T(\mathbf{I} - \mathbf{S})\mathbf{Z}]^{-1}.$$

如果满足 2.5 节的条件 (A.1)~条件 (A.6), 则其渐近偏与渐近方差分别为

$$E(\hat{\beta}_{\text{BF}} - \beta | \mathbf{X}, \mathbf{Z}, U) = -\frac{h^2}{2}\mu_2(K)E[\mathbf{z}_i\mathbf{x}_i^T\alpha''(u_i)]\boldsymbol{\Xi}^{-1} + o_p(h^2),$$

$$\text{Var}(\hat{\beta}_{\text{BF}} | \mathbf{X}, \mathbf{Z}, U) = \frac{\sigma^2}{n}\boldsymbol{\Xi}^{-1} + o_p\left(\frac{1}{n}\right),$$

其中 $\boldsymbol{\Xi} = E(\mathbf{Z}\mathbf{Z}^T) - E\left[E(\mathbf{Z}\mathbf{X}^T|U)E(\mathbf{X}\mathbf{X}^T|U)^{-1}E(\mathbf{X}\mathbf{Z}^T|U)\right]$.

上面的结论表明如果选择一般的窗宽 $h \sim n^{1/5}$, 那么常值系数 β 的估计便不是 $\sqrt{n}$ 相合的. 与此不同的是, β 的 profile 最小二乘估计在该情况下是 $\sqrt{n}$ 相合的, 这表明要想使得 $\hat{\beta}_{\text{BF}}$ 达到 $\sqrt{n}$ 相合, 拟合不足 (undersoothing) 是不可避免的, 而对于 profile 最小二乘估计, 利用最优窗宽 $h \sim n^{-1/5}$ 即可满足 $\sqrt{n}$ 相合. Spenkman (1988), Liang (2006) 与 Opsomer 和 Ruppert (1999) 对于部分线性模型与部分线性可加模型也得到了同样的结论. 同样地, 如果对窗宽进行调整, 可得到下面的结果.

推论 2.1 如果 $h \sim n^{r/5}$, 其中 $-1 < r < -1/4$, 则有

$$\sqrt{n}(\hat{\beta}_{\text{BF}} - \beta) \rightarrow N(0, \sigma^2 \boldsymbol{\Sigma}^{-1}).$$

由推论可知, 如果正确的选择窗宽, backfitting 估计一样可以达到半参数效率界限.

2.4 数值模拟

下面将通过数值模拟考察前面所提出的 backfitting 估计的有效性. 下面的模拟实验中核函数选取的是 Epanechnikov 核 $K(u) = 0.75(1 - u^2)_+$.

考虑如下的部分线性变系数模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2(u_i)x_{i2} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (2.26)$$

令 $\beta_0 = 3, \beta_1 = 2, \beta_2(u_i) = \sin(2\pi u_i)$, 自变量 x_{1i} 是独立产生的服从 $U(0, 1)$ 分布的随机数, 而 x_{2i} 是服从 $N(2, 1)$ 的随机数, 模型误差 ε_i 是来自于 $N(0, \sigma^2)$ 的随机数. 为了考察样本量以及误差方差对估计的影响, 下面分别考察样本量 $n = 100, 200$ 和误差方差 $\sigma^2 = 0.2^2, 0.5^2, 0.8^2$ 几种情况. 每种情况下模拟的重复次数为 500 次, 其中窗宽利用 Cross-Validation 方法进行选择. 下面将基于 500 次对应的均值 (mean) 与标准差 (SD) 来考察估计的有效性, 具体的结果见表 2.1. 从上面的结果可以看出如下结论.

(1) 随着 n 的增大, 即观测点的增多, 虽然估计的精度 (估计值逼近精确值的程度) 提高不大, 但其稳定性有明显的提高.

(2) 随着 σ 的增大, 即噪声方差变大, 对模型的干扰增强, 估计的精度变化不大, 但稳定性有明显的降低.

表 2.1 常值系数的 backfitting 估计的均值与标准差

		$\beta_0 = 3$			$\beta_1 = 2$		
		$\sigma = 0.2$	$\sigma = 0.5$	$\sigma = 0.8$	$\sigma = 0.2$	$\sigma = 0.5$	$\sigma = 0.8$
$n = 100$	均值	2.9868	2.7473	3.6684	2.1762	2.1598	1.6129
	SD	0.0558	0.1432	0.1948	0.0673	0.1779	0.2858
$n = 200$	均值	3.0026	3.2567	3.0708	2.2049	1.6836	1.9189
	SD	0.0427	0.1023	0.1574	0.0466	0.1223	0.1964

下面考察光滑参数对常值系数的 backfitting 估计的影响, 在 $n = 100, \sigma^2 = 0.2^2$ (下面若不做特殊说明, 样本量与误差方差取值与此相同) 的情况下考察光滑参数 h 分别等于 0.1, 0.3, 0.5, 0.7, 0.9 和 1.1 六种情形. 模拟结果见表 2.2. 模拟结果表示窗宽的变化对常值系数的 backfitting 估计的精度以及稳定性影响不大, 这也与关于部分线性模型等半参数模型的研究结果相一致.

表 2.2 不同窗宽下各常值系数 backfitting 估计的均值与标准差

h	0.1	0.3	0.5	0.7	0.9	1.1
Mean(β_0)	3.0152	2.9610	2.9738	2.9495	2.9719	2.999
SD(β_0)	0.0572	0.0628	0.0584	0.0620	0.0584	0.0536
Mean(β_1)	2.0047	2.1071	2.1668	2.1891	2.1806	2.1732
SD(β_1)	0.0048	0.0720	0.0664	0.0699	0.0708	0.0668

接下来将基于估计均方误差 $MSE(\hat{\beta}) = \frac{1}{500} \sum_{j=1}^{500} (\hat{\beta}_j - \beta)^2$ 来比较常值系数的

backfitting 估计 (简称 BF 估计) 与 profile 最小二乘估计 (简称 PL 估计) 的有效性, 模拟结果见表 2.3. 模拟结果表明 backfitting 估计与 profile 最小二乘估计在对应的几种窗宽下相差不大. 这也与我们的结果相一致.

表 2.3 不同窗宽下各常值系数 BF 估计与 PL 估计的估计均方误差

h	0.1	0.3	0.5	0.7	0.9	1.1
$MSE_{PL}(\hat{\beta}_0)$	0.0032	0.0039	0.0059	0.0069	0.0071	0.0088
$MSE_{BF}(\hat{\beta}_0)$	0.0035	0.0054	0.0040	0.0063	0.0042	0.0028
$MSE_{PL}(\hat{\beta}_1)$	0.0048	0.0083	0.0171	0.0250	0.0253	0.0217
$MSE_{BF}(\hat{\beta}_1)$	0.0047	0.0166	0.0322	0.0406	0.0376	0.0344

2.5 定理的证明

下面将给出定理 2.1 至定理 2.3 的证明. 在给出具体证明之前, 首先列出一些条件.

(A.1) 随机变量 U 具有有界支撑 Ω , 其密度函数 $f(\cdot)$ 在其支撑上满足 Lipschitz 连续, 且不为 0.

(A.2) 对于任一 $U \in \Omega$, 矩阵 $E(\mathbf{X}\mathbf{X}^T|U)_{k \times K}$ 为非奇异, $E(\mathbf{X}\mathbf{X}^T|U)_{k \times K}$, $E(\mathbf{X}\mathbf{X}^T|U)_{k \times K}^{-1}$ 和 $E(\mathbf{X}\mathbf{Z}^T|\mathbf{U})_{k \times K}$ 都是 Lipschitz 连续的.

(A.3) 存在 $s > 2$, 使得 $E\|\mathbf{X}\|^{2s} < \infty$ 和 $E\|\mathbf{Z}\|^{2s} < \infty$, 对于 $\varepsilon < 2 - s^{-1}$, 使得 $n^{2\varepsilon-1}h \rightarrow \infty$.

(A.4) $\alpha(\cdot)$ 二阶连续可导.

(A.5) 函数 $K(\cdot)$ 为对称密度函数, 具有紧支撑.

(A.6) $nh^8 \rightarrow 0$ 和 $nh^2/(\log n)^2 \rightarrow \infty$.

上述条件是 Fan 和 Huang (2005) 证明定理所附加的条件, 这些条件只是为了证明的方便, 并非最弱条件.

另外为了定理的证明, 给出如下引理.

引理 2.1 令 $(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ 为独立同分布 (iid) 的随机序列, 其中

$\mathbf{Y}_i, i = 1, 2, \dots, n$ 为一元随机变量, 进一步假定 $E|y|^s < \infty$ 与 $\sup_x \int |y|^s f(x, y) dy < \infty$, 其中 f 表示 $(\mathbf{X}, \mathbf{Y})$ 的联合密度. 令 K 为一有界正函数, 并有有界支撑且满足 Lipschitz 条件, 则有

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n [K_h(\mathbf{X}_i - x) \mathbf{Y}_i - E(K_h(\mathbf{X}_i - x) \mathbf{Y}_i)] \right| = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right), \quad (2.27)$$

其中对于 $\varepsilon < 1 - s^{-1}$, 有 $n^{2\varepsilon-1}h \rightarrow \infty$.

引理 2.1 由 Mack 和 Silverman(1982) 即可证得. 下面记 $c_n = \left\{ \frac{\log(1/h)}{nh} \right\}^{1/2}$

引理 2.2 如果满足上面的条件 (A.1)~条件 (A.6), 那么 β 无约束下的 Profile 最小二乘估计是渐近正态的, 即

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow N(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}^{-1}).$$

引理 2.3 如果满足上面的条件 (A.1)~条件 (A.6), 有

$$n^{-1} \bar{\mathbf{Z}}^T \bar{\mathbf{Z}} \rightarrow \boldsymbol{\Sigma}.$$

引理 2.4 如果满足上面的条件 (A.1)~条件 (A.6), 有

$$\frac{1}{n} \bar{\mathbf{Z}}^T (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^T \bar{\mathbf{Z}} \rightarrow \boldsymbol{\Sigma}.$$

注 引理 2.2 至引理 2.4 分别为 Fan 和 Huang (2005) 的定理 4.1, 引理 7.2 与引理 7.3.

定理 2.1 和 2.2 的证明 由引理 2.2 和 2.3 可得, 原假设成立的情况下有

$$\mathbf{A} \hat{\beta} - \mathbf{b} \rightarrow N(\mathbf{0}, \hat{\sigma}_r^2 \mathbf{A}(\bar{\mathbf{Z}}^T \bar{\mathbf{Z}})^{-1} \mathbf{A}^T),$$

从而有

$$\hat{\lambda} \rightarrow N(\mathbf{0}, \hat{\sigma}_r^2 [\mathbf{A}(\bar{\mathbf{Z}}^T \bar{\mathbf{Z}})^{-1} \mathbf{A}^T]^{-1}),$$

则有正态分布的性质有

$$\frac{\hat{\lambda}^T [\mathbf{A}(\bar{\mathbf{Z}}^T \bar{\mathbf{Z}})^{-1} \mathbf{A}^T] \hat{\lambda}}{\hat{\sigma}_r^2} \rightarrow \chi_k^2.$$

备择假设下, 则有

$$\mathbf{A} \hat{\beta} - \mathbf{b} \rightarrow N(\mathbf{A} \beta - \mathbf{b}, \hat{\sigma}_r^2 \mathbf{A}(\bar{\mathbf{Z}}^T \bar{\mathbf{Z}})^{-1} \mathbf{A}^T),$$