

分类号 O212.1  
密 级 公开  
学 号 151582



# 硕士学位论文

(学术型)

题目 部分线性模型和广义线性模型的惩罚经验似然

作 者 毛沥悦

指导教师 陈夏 副教授

一级学科名称 统计学

二级学科名称 统计学

提交日期 二〇一八年五月

同意提交学位论文

毛源悦

2018.5.31

## 学位论文原创性声明

本人声明所呈交的学位论文是我在导师的指导下进行研究工作所取得的研究成果。尽我所知，除文中已经注明引用的内容和致谢的地方外，本论文不包含其他个人或集体已经发表或撰写过的研究成果，也不包含本人或他人已申请学位或其他用途使用过的成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

本学位论文若有不实或者侵犯他人权利的，本人愿意承担一切相关的法律责任。

作者签名： 毛源悦 日期： 2018年 5月 31日

## 学位论文知识产权及使用授权声明书

本人在导师指导下所完成的学位论文及相关成果，知识产权归属陕西师范大学。本人完全了解陕西师范大学有关保存、使用学位论文的规定，允许本论文被查阅和借阅，学校有权保留学位论文并向国家有关部门或机构送交论文的纸质版和电子版，有权将本论文的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本论文。本人保证毕业离校后，发表本论文或使用本论文成果时署名单位仍为陕西师范大学。

保密论文解密后适用本声明。

作者签名： 毛源悦 日期： 2018年 5月 31日

## 摘要

部分线性模型由 Engle 等人在 1986 年首次提出, 之后有大量的研究与应用. 广义线性模型由 Nelder 和 Wedderburn 于 1972 年提出, 它是线性模型的重要推广. 广义线性模型在社会、经济、生物和医学领域有广泛的应用. 在实际问题中, 数据常带有测量误差. 另外, 随着科技的发展, 数据多以高维的形式出现. 因此, 关于带有测量误差的部分线性模型和高维广义线性模型的统计研究具有一定的实用价值.

经验似然是由 Owen 于 1988 年提出的一种非参数统计方法. 该方法有很多突出的优点, 比如置信域形状由数据决定、Bartlett 检验等. 经验似然方法掀开了统计推断方法的新篇章, 该方法已被应用到各种统计模型及不同领域中. 近年来, 高维数据成为一种常态, 使用有效的变量选择方法从高维数据中挖掘重要信息成为关注的焦点. 鉴于传统变量选择方法的特点, 研究者提出了可以同时进行变量选择与参数估计的惩罚函数法并且该方法得到了广泛的应用.

本文将惩罚经验似然方法应用到协变量都带有测量误差的部分线性模型和高维数据下广义线性模型中, 推广了该方法的应用领域. 通过最大化惩罚经验似然目标函数得到参数估计, 并从理论和数值模拟两方面研究了估计的渐近性质. 论文共分为三章, 主要内容如下:

第一章首先陈述了本论文的研究背景; 接着回顾了本文使用到的相关知识和理论, 详细介绍了经验似然, 变量选择方法; 最后列出了本文的主要工作.

第二章讨论了参数部分和非参数部分都带有测量误差时, 部分线性模型的惩罚经验似然估计. 通过逆卷积核估计方法构造辅助随机向量, 进而得到惩罚经验似然估计. 理论和数值模拟证明了惩罚经验似然估计具有相合性和 Oracle 性质. 与此同时, 考虑假设检验问题, 所构造的检验统计量的渐近分布为卡方分布.

第三章通过适当的辅助随机变量研究了自适应 Lasso 下高维广义线性模型的惩罚经验似然. 主要的结论有提出的方法具有 Oracle 性质以及在假设检验中构造的检验统计量的渐近分布为卡方分布. 随机模拟以及实例分析表明, 所提方法的效果是令人满意的.

**关键词:** 部分线性模型; 广义线性模型; 惩罚经验似然; 测量误差; 变量选择; 高维数据.

## Abstract

Partly linear model was first proposed by Engle in 1986, followed by a large number of studies and applications. Generalized linear model, which was proposed by Nelder and Wedderburn in 1972, is an important generalization of the linear model. This model is widely used in the fields of society, economy, biology and medicine. In many practical problems, we may encounter the situation that the covariates are measured with errors. In addition, with the development of science and technology, data often appear in the form of high dimensional. Therefore, the statistical study of the partial linear model with measurement error in all variables and high dimensional generalized linear model have practical value.

Empirical likelihood is a nonparametric statistical method proposed by Owen in 1988. It has a lot of outstanding advantages, for example, objective determination of the shape of the confidence region, Bartlett correctability, etc. As a new method, empirical likelihood has been applied to various statistical models and different fields. In recent years, data are usually in the form of high dimensional. Therefore, using effectively variable selection to excavate effective information from high dimensional data becomes the focus of attention. In view of the characteristics of traditional variable selection, statisticians proposed penalty function, which can select variables and estimate parameters simultaneously. This method has been widely used over this years.

In this paper, we apply penalized empirical likelihood to the partially linear model with measurement error in all variables and high dimensional generalized linear model. This applications extend the application field of this method . By maximizing penalized empirical likelihood objective function, we obtain parameter estimator. We study the asymptotic property of estimator in theory and numerical simulations. This thesis is divided into three chapters, the main contents are as follows:

The first chapter presents the background of the research and reviews the relevant knowledge and theory. Empirical likelihood and variable selection are introduced in detail. The main work of this paper is also listed.

In the second chapter, penalized empirical likelihood for partially linear model is discussed when both parameter and nonparametric have measurement errors. We con-

struct auxiliary random variable by the kernel estimation method, so the penalized empirical likelihood estimator is obtained. Theory and simulation studies prove our theory, that is, proposed estimator have consistency and oracle property. Finally, we consider the problem of hypothesis testing. The profiled penalized empirical likelihood ratio is asymptotically chi-square distribution.

In the third chapter, based on appropriate auxiliary random variable, we propose penalized empirical likelihood with adaptive lasso in high dimensional generalized linear models. Our main findings are that the extended penalized empirical likelihood has the oracle property and the asymptotic distribution of the test statistics constructed in the hypothesis test is chi-square distribution. Simulation studies and real data analysis indicate that the efficiency of the proposed penalized empirical likelihood estimator is encouraging.

**Keywords:** Partially linear models; Generalized linear models; Penalized empirical likelihood; Measurement errors; Variable selection; High-dimensional data.

# 目 录

摘要	I
Abstract	III
第1章 绪论	1
1.1 研究背景	1
1.2 主要研究方法	2
1.2.1 经验似然	2
1.2.2 变量选择	3
1.3 本文的主要工作	4
第2章 部分线性 EV 模型的惩罚经验似然	5
2.1 引言	5
2.2 主要方法和结果	6
2.3 计算	10
2.3.1 惩罚经验似然算法	10
2.3.2 调整参数的选择	11
2.4 数值模拟和实例分析	11
2.4.1 数值模拟	11
2.4.2 血浆 $\beta$ -胡萝卜素水平数据	15
2.5 定理证明	16
第3章 高维广义线性模型的惩罚经验似然	23
3.1 引言	23
3.2 方法和主要结果	24

3.3 数据模拟和实例分析 . . . . .	26
3.3.1 数据模拟 . . . . .	26
3.3.2 NMES 数据 . . . . .	29
3.4 定理证明 . . . . .	31
<b>总结与展望</b>	<b>39</b>
<b>参考文献</b>	<b>41</b>
<b>致谢</b>	<b>47</b>
<b>攻读硕士学位期间的科研成果</b>	<b>49</b>

## 第1章 绪论

### 1.1 研究背景

在分析数据时,需要在实际问题的背景上建立适当的统计模型.而统计模型只是对这些数据的近似拟合,所以统计学在不断地寻找更优的统计模型以求更准确的解释数据.

本文研究的第一个模型是部分线性模型 (partly linear models),它是半参数回归模型 (semi-parametric regression models) 的一种.由于线性模型和非线性模型在处理实际数据时缺少灵活性,所以 Engle 等人于 1986 年研究用电需求量与电价、收入以及气温等变量之间的关系时首次提出该模型<sup>[1]</sup>.部分线性模型自引入以来,受到了统计学界的极大关注.详细的信息可参见 Härdle 等人关于部分线性模型的专著<sup>[2]</sup>,以及 Ruppert 等人关于半参数回归的专著<sup>[3]</sup>.

第二个模型是广义线性模型 (generalized linear models).该模型于 1972 年由 Nelder 和 Wedderburn 引入.在经典广义线性模型 (响应变量具有指数族分布)<sup>[4]</sup>下,文献 [5] 建立了极大似然估计的大样本理论;Qian 和 Wu 研究了 Logistic 模型中参数极大似然估计的重对数律和模型选择问题<sup>[6]</sup>.自 1974 年 Wedderburn 提出拟似然函数以来<sup>[7]</sup>,广义线性模型对响应变量具有指数分布族的假设弱化为均值与方差函数的条件,这扩大了该模型的使用范围.广义线性模型被广泛的应用到教育,医学,经济等领域.1989 年 McCullagh 和 Nelder 出版了关于广义线性模型的专著<sup>[8]</sup>.

实际问题中,人们在工程、经济、物理、生物医学等领域得到的数据常常带有测量误差.例如,在获得性免疫缺陷综合症 (AIDS) 研究中,人类免疫缺陷病毒的血浆浓度 (HIV)-1 RNA 和 CD4+ 细胞数以及血压、尿钠氯水平、暴露于污染物水平等.在统计研究中通常把带有测量误差的模型称为 EV (errors-in-variables) 模型.研究者讨论了不同形式的 EV 模型,如线性 EV 模型<sup>[9-11]</sup>、非线性 EV 模型<sup>[12]</sup>、非参数 EV 模型<sup>[13-15]</sup>.对带有测量误差的数据进行统计分析是当今的一个热点研究问题.

随着互联网、云计算等科学技术的快速发展,全球数据呈爆炸式增长,特别是在生物、教育、工程等领域产生的实验数据维数甚至超过了样本大小,人类进入到大数据时代<sup>[16]</sup>.大数据常以高维数据形式呈现,如何有效地从数据中挖掘重要的信息成为人们关注的焦点.变量选择是解决该问题的有效方法,详细可参见 Bühlmann 和 van de Geer 关于高维数据的方法、理论、应用的专著<sup>[17]</sup>.



## 1.2 主要研究方法

### 1.2.1 经验似然

经验似然 (empirical likelihood, EL) 由 Owen 于 1988 年提出, 是一种非参数统计推断方法<sup>[18]</sup>. 因为不用预先给定数据所属的分布族, 所以该方法有很多突出的优点, 如无需估计渐近方差、数据确定置信域形状、保域性、无需构造轴统计量、Bartlett 检验等性质<sup>[19-22]</sup>. 该方法已被应用到各种领域, 例如总体均值的统计推断 [18]、估计方程 [20]、线性回归模型 [23]、广义线性模型 [24]、非参数回归 [25]、部分线性模型 [26]、半参数变系数部分线性模型 [27]. EL 方法得到广泛的应用并成为二十世纪来重要的统计推断方法.

下面我们简单介绍 Qin 和 Lawless 关于广义估计方程的经验似然<sup>[20]</sup>.

假设  $d$  维观测值  $x_1, x_2, \dots, x_n$  为独立同分布的样本, 分布函数  $F$  未知. 参数  $\theta \in \mathbb{R}^p$  与总体分布  $F$  有关. 假设  $\theta$  与  $F$  的关系可以用  $r \geq p$  个独立无偏估计方程来表示, 即满足  $E\{g_j(x, \theta)\} = 0$ , 其中  $j = 1, 2, \dots, r$ . 向量表达式为

$$E\{g(x, \theta)\} = 0,$$

其中  $g(x, \theta) = (g_1(x, \theta), \dots, g_r(x, \theta))^T$ . 文献 [20] 在此基础上定义对数经验似然比函数

$$\ell(\theta) = \max \left\{ \prod_{j=1}^n p_j : p_j \geq 0, \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j g(x_j, \theta) = 0 \right\}.$$

通过拉格朗日算法, 得到

$$p_j = \frac{1}{n} \frac{1}{1 + \lambda^T g(x_j, \theta)},$$

且

$$\ell(\theta) = \sum_{j=1}^n \log \{1 + \lambda^T g(x_j, \theta)\},$$

其中  $\lambda = \lambda(\theta)$  满足方程

$$\sum_{j=1}^n \frac{1}{n} \frac{g(x_j, \theta)}{1 + \lambda^T g(x_j, \theta)} = 0.$$

若  $\theta_0$  为参数真值,  $\tilde{\theta}$  是 EL 估计. 在适当的条件下, Qin 和 Lawless 有以下的结论<sup>[20]</sup>:

$$\sqrt{n}(\tilde{\theta} - \theta_0) \rightarrow N(0, V),$$

其中  $V = [E(\frac{\partial g}{\partial \theta})^T (E g g^T)^{-1} E(\frac{\partial g}{\partial \theta})]^{-1}$  且

$$W(\theta_0) = 2\ell(\theta_0) - 2\ell(\tilde{\theta}) \rightarrow \chi_p^2.$$

### 1.2.2 变量选择

变量选择在模型的构建中有重要的作用. 一般情况下, 变量越多越能拟合数据, 但多变量的引入可能会使得某些变量并不显著. 有效的变量选择可以达到仅保留显著变量的效果, 在降低模型复杂度的同时提高模型的预测能力.

传统的变量选择方法有逐步回归、全子集回归、岭回归和最优子集回归等. 这些方法存在着自身的不足, 如计算量大、模型选择不具有稳定性等. 随着高维数据的发展, 传统的方法显得力不从心. 近年来, 通过惩罚函数进行变量选择的方法越来越受到人们的关注. 该方法可同时进行变量选择和参数估计, 这大大提高了计算速度.

常见的惩罚函数有:  $L_q$  惩罚 (bridge penalty)<sup>[28]</sup>、Lasso 惩罚 (soft thresholding rule)<sup>[29]</sup>、Hard 惩罚 (hard thresholding rule)<sup>[30]</sup>、SCAD 惩罚 (smoothly clipped absolute deviation penalty)<sup>[31]</sup>、Alasso 惩罚 (adaptive lasso penalty)<sup>[32]</sup>. Fan 和 Li 指出好的惩罚函数应具有以下三条性质<sup>[31]</sup>, 一是无偏性: 为了避免模型的偏差, 当未知参数的真值很大时估计是无偏的; 二是稀疏性: 估计是一个自动将小系数压缩成零的门槛准则, 这样可以降低模型的复杂程度; 三是连续性: 为了避免模型预测时的不稳定性, 最小二乘估计是连续的. 在上面提到的五种惩罚函数中, Alasso、 $L_q$  和 SCAD 惩罚函数同时满足以上三条性质. 关于变量选择的研究进展可参见文献 [33].

结合经验似然和惩罚方法, Otsu 以及 Tang 和 Leng 提出使用惩罚经验似然 (penalized empirical likelihood, PEL) 来进行变量选择<sup>[34,35]</sup>. 正如文献 [35] 所述, PEL 方法既有非参数似然方法的效率和自适应性, 也有 EL 方法的无需估计方差以及由数据决定置信区间形状的性质 [35]. 因此, PEL 方法已被应用到各种统计模型及实际领域中<sup>[36-40]</sup>.

调整参数的选择也非常重要, 它直接决定所选模型的稀疏性. 好的调整参数产生的模型有好的预测性. 调整参数的选择方法有很多, 如贝叶斯信息准则 (BIC)<sup>[41]</sup>、广义交叉核实<sup>[42]</sup>、风险膨胀准则<sup>[43]</sup>、赤池信息准则等. 文献 [44] 证明了 BIC 准则在模型选择方面有相合性, Wang 等人在 2009 年将上述结论推广到参数个数发散的情形<sup>[45]</sup>. 本文使用 BIC 准则选择调整参数.

惩罚函数大多数是非凸的, 即在零点不可导, 这使得惩罚方法的计算存在一定的困难. 研究者提出了各种算法来解决非凸优化问题. Fan 和 Li 提出了局部二次近似算法 (local quadratic approximation, LQA)<sup>[31]</sup>. Hunter 和 Li 在 LQA 的基础上提出改进的 LQA<sup>[46]</sup>. 虽然改进的 LQA 可以克服 LQA 的缺点, 但是改进的 LQA 算法需要更多的计算时间并且对稀疏性有一定的影响. Efron 等人提出最小角回归算法<sup>[47]</sup>.

Zou 和 Li 提出了局部线性近似<sup>[48]</sup>. 其他非凸最优化算法可参见文献 [49, 50].

### 1.3 本文的主要工作

本文利用 PEL 方法, 研究了所有协变量均带有测量误差的部分线性模型和高维广义线性模型的变量选择与参数估计问题. 基于 Alasso 惩罚函数进行了理论的证明和数值的计算.

第二章研究了部分线性 EV 模型的惩罚经验似然估计. 使用逆卷积方法和纠偏方法处理测量误差, 并通过 PEL 方法构造目标函数. 证明了 PEL 估计具有相合性和 Oracle 性质. 即, 提出的 PEL 方法可以正确选出零元素, 并且非零元估计和真实模型有一样的渐近性质. 同时, 讨论了非零元素的假设检验问题, 构造的检验统计量具有渐近卡方分布. 探讨了 PEL 的计算方法, 采用 LQA 方法将非凸问题转化为凸优化问题. 最后, 在模拟中比较了五种不同的 PEL 估计以及两种非惩罚估计的效果, 结果表明 PEL 估计具有好的表现.

第三章研究了高维广义线性模型中的 PEL 问题. 基于三个方程的辅助统计量, 构造了同时估计参数  $\beta$  和  $\sigma^2$  的 PEL 函数并证明了 PEL 具有 Oracle 性质. 同时, 探讨了假设检验问题, 构造的检验统计量的渐近分布为卡方分布. 随机模拟和实例分析说明了所提 PEL 方法在参数估计与变量选择方面都有良好的表现. 值得注意的是, 无论是在参数估计还是变量选择方面, 三个方程下的 PEL 比两个方程下的 PEL 表现更好.

## 第2章 部分线性 EV 模型的惩罚经验似然

### 2.1 引言

在工程学、经济、物理、流行病学领域中,由于测量误差的原因,通常不能获得变量的真实值.同时,线性和非线性模型在处理实际问题时不够灵活,而半参数回归模型可以解决这个问题.部分线性模型是半参数模型中的一种.本章将研究所有协变量带有测量误差时,部分线性模型的参数估计与变量选择问题.

部分线性模型有测量误差的统计推断,已有很多的研究,详细参见文献 [51–57].部分线性模型将函数分为参数和非参数部分,当只有参数部分带有测量误差时,文献 [58] 讨论了 EL 估计; Xu 等人做了缺失数据下的假设检验<sup>[59]</sup>.当只有非参数部分有测量误差时, Fan 和 Truong 研究了非参数函数的回归估计,并利用逆卷积 (deconvolution) 方法构造了一类新的核估计<sup>[15]</sup>; Chen 和 Cui 讨论了 EL 估计<sup>[60]</sup>.当参数和非参数部分同时有测量误差时, Koul 和 Song 设计了非参数元素是否具有参数形式的不相称检验<sup>[61]</sup>; Yan 和 Chen 做了经验似然推断<sup>[62]</sup>.

在变量选择问题中,惩罚函数方法成为当前较流行的模型选择方法.文献 [63, 64] 提出惩罚似然方法通过调整偏差方差权衡提高效率.在部分线性模型中, Ni 等人提出了利用双惩罚最小二乘方法进行自动模型选择<sup>[65]</sup>; Liang 和 Li 研究了当参数部分带有测量误差时变量选择问题,并指出直接用观察到的数据去代替未知的变量会使得估计不具有相合性<sup>[66]</sup>; Xie 和 Huang 基于 SCAD 惩罚函数,研究了高维部分线性模型的变量选择与参数估计<sup>[67]</sup>.

在已有的文献中,对部分线性模型进行变量选择时,通常不考虑测量误差或者误差只出现在参数部分或者非参数部分.当所有协变量带有测量误差时,变量选择是一项具有挑战性的工作.因此,本章研究所有协变量都带有测量误差时,部分线性模型的自适应 Lasso 惩罚经验似然估计问题.主要的结论是,提出的 PEL 估计具有相合性与 Oracle 性质.并且,构造的检验统计量的渐近分布为卡方分布.模拟研究以及实际数据分析说明了提出的 PEL 估计具有优良性质.需要注意的是,文章中的自适应 Lasso 惩罚函数可以换成其他形式,如: SCAD、桥惩罚等,只需在理论证明中稍作修改.

## 2.2 主要方法和结果

部分线性模型有如下的形式

$$Y = X^T \beta + g(T) + \varepsilon, \quad (2.2.1)$$

其中  $Y \in \mathbb{R}$  是响应变量,  $X \in \mathbb{R}^p$  是随机向量,  $\beta \in \mathbb{R}^p$  是未知参数向量,  $T$  是定义在  $[0, 1]$  上的随机变量,  $g(t)$  是定义在  $[0, 1]$  上未知的光滑函数并且随机误差  $\varepsilon$  满足  $E(\varepsilon|X, T) = 0$ .

假设变量  $X$  和  $T$  分别有测量误差  $u$  和  $v$ , 即

$$\begin{cases} Z = X + u, \\ W = T + v, \end{cases}$$

所以, 部分线性 EV 模型为:

$$\begin{cases} Y = X^T \beta + g(T) + \varepsilon, \\ Z = X + u, \\ W = T + v, \end{cases} \quad (2.2.2)$$

其中  $u, v$  和  $(X^T, T, \varepsilon)^T$  相互独立.

当非参数部分带有测量误差时, 为了模型识别, 假设  $v$  的分布已知<sup>[15]</sup>. 进一步, 假设

$$\begin{cases} E(u) = E(v) = 0, \quad \text{Cov}(u) = \Sigma_{uu}, \\ E(\varepsilon|X, T) = 0, \quad \text{Var}(\varepsilon|X, T) = \sigma_\varepsilon^2, \end{cases}$$

其中  $\sigma_\varepsilon^2$  未知,  $\Sigma_{uu} > 0$  已知<sup>[52, 55, 62]</sup>. 实际中, 若协方差矩阵  $\Sigma_{uu}$  未知, 则可利用部分重复的估计方法进行估计<sup>[52]</sup>.

设有样本  $(Z_j, W_j, Y_j), j = 1, \dots, n$ . 假设  $T$  的密度函数是  $f(t)$ ,  $f(t)$  的逆卷积核估计为

$$\hat{f}_n(t) = \frac{1}{nh} \sum_{i=1}^n K_n \left( \frac{t - W_i}{h} \right),$$

其中  $h$  是窗宽,  $K_n(x)$  是逆卷积核函数, 其定义形式为

$$K_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \frac{\phi_K(t)}{\phi_v(t/h)} dt,$$

其中  $\phi_K(t)$  是核函数  $K(t)$  的傅里叶变换,  $\phi_v(\cdot)$  是误差变量  $v$  的特征函数. 记

$$\omega_{ni}(t) = K_n \left( \frac{t - W_i}{h} \right) / \sum_j K_n \left( \frac{t - W_j}{h} \right) = \frac{1}{nh} K_n \left( \frac{t - W_i}{h} \right) / \hat{f}_n(t).$$

令  $g_1(t) = E(X|T=t) = E(Z|T=t)$ ,  $g_2(t) = E(Y|T=t)$ , 则定义

$$\hat{g}_{1n}(t) = \sum_{j=1}^n \omega_{nj}(t) Z_j$$

和

$$\hat{g}_{2n}(t) = \sum_{j=1}^n \omega_{nj}(t) Y_j$$

分别是  $g_1(t)$  和  $g_2(t)$  的估计. 因为  $g(t) = E(Y - X^T \beta | T=t) = E(Y - Z^T \beta | T=t)$ , 所以对于给定的  $\beta$ ,

$$g_n(t, \beta) = \sum_{j=1}^n \omega_{nj}(t) (Y_j - Z_j^T \beta) \quad (2.2.3)$$

是  $g(t)$  的估计. 在适当的条件下, 可证明对于某类测量误差的分布,  $\hat{g}_{1n}(t)$  和  $\hat{g}_{2n}(t)$  分别是  $g_1(t)$  和  $g_2(t)$  的相合估计<sup>[13,15]</sup>.

构造如下的辅助统计量

$$\eta_j(\beta) = \tilde{Z}_j(\tilde{Y}_j - \tilde{Z}_j^T \beta) + \Sigma_{uu} \beta,$$

其中  $\tilde{Y}_j = Y_j - \hat{g}_{2n}(W_j)$ ,  $\tilde{Z}_j = Z_j - \hat{g}_{1n}(W_j)$ . 如果  $\beta_0$  是  $\beta$  的真值, 那么  $E(\eta_j(\beta_0)) = 0$ . 根据 [18],  $\beta$  的对数经验似然比定义为

$$\ell(\beta) = -\max \left\{ \sum_{j=1}^n \log(np_j) : p_j \geq 0, \sum_{j=1}^n p_j = 1, \sum_{j=1}^n p_j \eta_j(\beta) = 0 \right\}.$$

由拉格朗日算法, 得到

$$p_j = \frac{1}{n} \frac{1}{1 + \lambda^T \eta_j(\beta)},$$

$\ell(\beta)$  可表示为

$$\ell(\beta) = \sum_{j=1}^n \log \{1 + \lambda^T \eta_j(\beta)\},$$

其中  $\lambda = \lambda(\beta)$  满足方程

$$\sum_{j=1}^n \frac{1}{n} \frac{\eta_j(\beta)}{1 + \lambda^T \eta_j(\beta)} = 0. \quad (2.2.4)$$

当  $\tilde{\beta}$  是  $\beta_0$  的  $\sqrt{n}$  相合估计时, 设权重向量  $\tilde{w} = 1/|\tilde{\beta}|^\gamma$  ( $\gamma > 0$ ). 自适应 Lasso 惩罚经验似然目标函数定义为

$$\ell_p(\beta) = \sum_{j=1}^n \log \{1 + \lambda^T \eta_j(\beta)\} + n\tau \sum_{i=1}^p \tilde{w}_i |\beta_i|, \quad (2.2.5)$$

其中  $\tilde{w}_i$  是  $\tilde{w}$  的第  $i$  个元素,  $\tau$  是调整参数. 设  $\Theta$  是参数空间, 定义 PEL 估计

$$\hat{\beta} = \arg \min_{\beta \in \Theta} \ell_p(\beta).$$

假设  $\mathcal{A} = \{i: \beta_{0i} \neq 0\}$  代表由参数真值  $\beta_0$  的非零元素组成的集合且  $|\mathcal{A}| = d$  代表  $\mathcal{A}$  的维度 (未知). 令  $\beta_0 = (\beta_{10}^T, \beta_{20}^T)^T$ ,  $\beta_{10} \in \mathbb{R}^d$ ,  $\beta_{20} \in \mathbb{R}^{p-d}$ . 不失一般性, 假设  $\beta_{10} \neq 0$  和  $\beta_{20} = 0$ , 其中,  $0$  是所有元素都为零的向量. 相应地, 令  $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$ , 其中  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  分别是  $\beta_{10}$ ,  $\beta_{20}$  的估计. 假设  $\Omega = E[(\varepsilon - u^T \beta_0)(X - E(X|T))]^{\otimes 2} + E[(uu^T - \Sigma_{uu})\beta_0]^{\otimes 2} + E(uu^T \varepsilon^2)$ , 其中  $A^{\otimes 2} = AA^T$ ,  $\Sigma = E[(X - E(X|T))(X - E(X|T))^T]$ ,  $V = \Sigma^{-1} \Omega \Sigma^{-1}$ . 对应的, 将  $V$  化分为矩阵块  $V_{ij}$  ( $i = 1, 2; j = 1, 2$ ), 其中  $V_{11} \in \mathbb{R}^{d \times d}$ . 在本文中, 取  $\tilde{w} = 1/|\tilde{\beta}_{EL}|^{-1}$ , 其中  $\tilde{\beta}_{EL}$  是  $\beta$  的 EL 估计<sup>[62]</sup>.

首先, 给出逆卷积方法的相关性定义<sup>[15]</sup>:

**定义 2.2.1** 对于正常数  $d_0, d_1, \varepsilon$ , 如果误差  $v$  的特征函数  $\phi_v(t)$  满足

$$d_0 \leq \liminf_{|t| \rightarrow \infty} |\phi_v(t)t^s| \leq \limsup_{|t| \rightarrow \infty} |\phi_v(t)t^s| \leq d_1,$$

那么称误差  $v$  的分布是  $s$  阶普通光滑的.

假设  $X_{ji}$  是  $X_j$  的第  $i$  个元素且  $h_i(t) = E(X_{ji}|T_j = t)$ ,  $1 \leq j \leq n$ ,  $1 \leq i \leq p$ . 下面是本章所需要的条件:

(C1)  $\sup_{0 \leq t \leq 1} E(\|X_1\|^3|T = t) < \infty$  并且  $\Sigma$  和  $\Omega$  的特征值有界.

(C2) 函数  $g(\cdot)$  和  $h_i(\cdot)$  ( $1 \leq i \leq p$ ) 一阶 Lipschitz 连续.

(C3)  $E(|\varepsilon|^3 + \|u\|^3 + |v|^3) < \infty$ .

(C4) 如果  $\beta_n^* = \min\{|\beta_{0j}|, j \in \mathcal{A}\}$ , 那么存在正数  $M$  满足  $\beta_n^* \geq M$ .

(C5) 当  $n \rightarrow \infty$ , 调整参数  $\tau$  满足  $\sqrt{n}\tau \rightarrow 0$  且  $n\tau \rightarrow \infty$ .

(C6)  $0 < \inf_{0 \leq t \leq 1} f(t) \leq \sup_{0 \leq t \leq 1} f(t) < +\infty$  且  $f(t)$  是  $k$  阶导数有界, 其中  $k$  是正整数.

(C7)  $K(t)$  是  $k$  阶核函数, 即,  $K(t) = K(-t)$ ,

$$\int_{-\infty}^{\infty} K(t)dt = 1, \quad \int_{-\infty}^{\infty} t^k K(t)dt \neq 0, \quad \int_{-\infty}^{\infty} t^j K(t)dt = 0$$

对  $1 \leq j \leq k-1$  和

$$\int_{-\infty}^{\infty} |x|^{s+1} \{\phi_K(x) + \phi'_K(x)\} dx < \infty, \quad \int_{-\infty}^{\infty} |x|^{s+1} \phi_K(x)^2 dx < \infty.$$

(C8) 误差变量  $v$  的分布是  $s$  阶普通光滑的且它的特征函数  $\phi_v(t)$  是非退化的. 取窗宽  $h = dn^{-1/(2k+2s+1)}$ ,  $d > 0$ . 对于某常数  $c \neq 0$ , 当  $t \rightarrow \infty$  时

$$t^s \phi_v(t) \rightarrow c, \quad t^{s+1} \phi_v'(t) = O(1).$$

注 2.2.1 条件 (C1)–(C8) 是相关文献中的一般性假定<sup>[36, 52, 54, 55, 68]</sup>. 条件 (C1)–(C3) 保证了 PEL 估计的存在性以及相合性. 条件 (C4)–(C5) 控制了惩罚函数对于非零元素的影响. 条件 (C6) 和 (C7) 保证回归方程以及  $T$  的密度函数的光滑性. 条件 (C8) 使得误差变量  $v$  的分布是普通光滑性的, 双指数分布和伽玛分布满足条件.<sup>[15]</sup>

注 2.2.2 文中的自适应 Lasso 惩罚可以换成其他的惩罚方式, 如 SCAD 和桥估计. 改变关于惩罚的条件以及证明中的相关部分, 文中结论仍然可以得到. 数值模拟将呈现不同惩罚函数下 PEL 估计量的效果.

接下来的定理给出了惩罚经验似然估计  $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$  的性质.

定理 2.2.1 在条件 (C1)–(C8) 下, 以趋近于 1 的概率, 存在  $\ell_p(\beta)$  的极小值点  $\hat{\beta}$  且满足  $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$ .

定理 2.2.2 假设条件 (C1)–(C8) 都成立. 当  $n \rightarrow \infty$  时, 有

(i) 稀疏性:  $P(\hat{\beta}_2 = 0) \rightarrow 1$ .

(ii) 渐近正态性:  $\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \rightarrow N(0, V_d)$ , 其中  $V_d = V_{11} - V_{12}V_{22}^{-1}V_{21}$ .

注 2.2.3 由定理 2.2.1 可知, PEL 估计具有  $\sqrt{n}$  相合性. 正如文献 [31] 所述, 定理 2.2.2 证明了提出的 PEL 估计具有 Oracle 性质. 即, PEL 在模型选择中有相合性并且与已知真实稀疏模型的 EL 估计有相同的渐近分布.

当  $\beta_2 = 0$  时, 讨论假设检验  $H_0: \beta_1 = \beta_{10}$ , 构造惩罚经验似然比检验统计量:

$$W(\beta_{10}) = 2\ell_p((\beta_{10}^T, 0)^T) - 2\ell_p(\hat{\beta}).$$

定理 2.2.3 假设条件 (C1)–(C8) 和原假设  $H_0$  成立, 当  $n \rightarrow \infty$  时, 有  $W(\beta_{10}) \rightarrow \chi_d^2$ .

根据定理 2.2.3, 可以构建置信区域. 其中,  $\beta_1$  的  $1 - \alpha$  水平渐近置信区域为

$$V(\alpha) = \{\beta_1 | W(\beta_1) \leq \chi_d^2(\alpha)\}, \quad (2.2.6)$$

其中  $\chi_d^2(\alpha)$  是自由度为  $d$  的卡方分布的  $(1 - \alpha)$  分位数.



## 2.3 计算

Tang 和 Leng 计算惩罚经验似然时, 在 LQA 的基础上使用牛顿迭代<sup>[35]</sup>, 即在  $\lambda$  和  $\beta$  之间迭代直到收敛. 这里采用 Ren 和 Zhang 在 2011 年的文章中使用的方法<sup>[69]</sup>, 直接使用迭代方法计算  $\hat{\beta}$ , 再使用 Owen 的方法计算相应的  $\lambda$ <sup>[21]</sup>, 这种计算方式提高了收敛的速度.

### 2.3.1 惩罚经验似然算法

当  $\beta_i \approx \beta_i^{(0)}$ , 设  $p_\tau(\beta_i) = \tau|\beta_i|/|\tilde{\beta}_i|$ .  $p_\tau(|\beta_i|)$  的局部二次近似为

$$p_\tau(|\beta_i^{(0)}|) + \frac{1}{2} \{p'_\tau(|\beta_i^{(0)}|)/|\beta_i^{(0)}|\} (\beta_i^2 - \beta_i^{(0)2}).$$

根据文献 [69], 记

$$Q_{1n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{[1 + \lambda^\top \eta_i(\beta)]},$$

$$Q_{2n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\eta'_i(\beta)}{[1 + \lambda^\top \eta_i(\beta)]} \lambda.$$

将  $Q_{1n}(\beta, \lambda)$  和  $Q_{2n}(\beta, \lambda)$  在  $(\beta, 0)$  处关于  $\beta$  和  $\lambda$  求导, 有

$$\frac{\partial Q_{1n}(\beta, 0)}{\partial \beta^\top} = \left[ \frac{\partial Q_{2n}(\beta, 0)}{\partial \lambda^\top} \right]^\top = \frac{1}{n} \sum_{i=1}^n \eta'_i(\beta),$$

$$\frac{\partial Q_{1n}(\beta, 0)}{\partial \lambda^\top} = -S_\beta, \quad \frac{\partial Q_{2n}(\beta, 0)}{\partial \beta^\top} = 0,$$

其中  $S_\beta = n^{-1} \sum_{i=1}^n \eta_i(\beta) \eta_i^\top(\beta)$ .

以接近真值的值为初值  $\beta^{(0)}$ , 采用改进的牛顿迭代算法最优化 PEL 方程. 具体地, 定义  $\hat{\beta}^{(k)}$  是第  $k$  次迭代值,  $k = 0, 1, 2, \dots$ , 使用迭代式

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + [\Sigma_{(\tau, \hat{\beta}^{(k)})} - V(\hat{\beta}^{(k)})]^{-1} [Q(\hat{\beta}^{(k)}) - U_\tau(\hat{\beta}^{(k)})],$$

其中

$$V(\beta) = \frac{\partial Q_{2n}(\beta, 0)}{\partial \lambda^\top} \left[ \frac{\partial Q_{1n}(\beta, 0)}{\partial \lambda^\top} \right]^{-1} \frac{\partial Q_{1n}(\beta, 0)}{\partial \beta^\top},$$

$$Q(\beta) = \frac{\partial Q_{2n}(\beta, 0)}{\partial \lambda^\top} \left[ \frac{\partial Q_{1n}(\beta, 0)}{\partial \lambda^\top} \right]^{-1} Q_{1n}(\beta, 0).$$

因为非凸惩罚函数在零点不可导, 所以使用  $\Sigma_{(\tau, \beta)}$  的渐近式

$$\Sigma_{(\tau, \beta)} = \text{diag}\{p'_\tau(|\beta_1|)/|\beta_1|, \dots, p'_\tau(|\beta_p|)/|\beta_p|\}$$

和  $U_\tau(\beta) = \Sigma_{(\tau, \beta)} \beta$ . 若  $\max_{1 \leq j \leq p} |\hat{\beta}_j^{(k+1)} - \hat{\beta}_j^{(k)}| \leq 10^{-3}$ , 迭代终止.

### 2.3.2 调整参数的选择

对于给定的  $\tau$ , 其对应的估计为  $\hat{\beta}_\tau$ . BIC 准则为

$$\text{BIC}_\tau = \log(\hat{\sigma}^2) + df_\tau \log(n)/n,$$

其中  $df_\tau = \text{tr}\{[V(\hat{\beta}_\tau) - \Sigma_{(\tau, \hat{\beta}_\tau)}]^{-1}V(\hat{\beta}_\tau)\}$ ,  $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \{[Y_i - Z_i^T \hat{\beta}_\tau - g_n(W_i, \hat{\beta}_\tau)]^2 - \hat{\beta}_\tau^T \Sigma_{uu} \hat{\beta}_\tau\}$ .

## 2.4 数值模拟和实例分析

本小节将给出三个随机模拟和一个实例数据来说明所提 PEL 的性质. 本文都使用 BIC 准则选择惩罚函数的调整参数, 采用局部二次近似以及牛顿迭代算法. 在随机模拟中, 比较惩罚方法和非惩罚方法以及不同惩罚函数在模型误差和模型复杂性的表现.

### 2.4.1 数值模拟

从模型 (2.2.1) 中产生  $n = 100,200$  个随机数据, 每次实验重复 1000 次. 通过计算误差:  $\|\hat{\beta} - \beta_0\|_2 = \{(\hat{\beta} - \beta_0)^T(\hat{\beta} - \beta_0)\}^{1/2}$  的中位数 (ME) 来评价估计的精确度. T 代表零系数正确估计为零的平均数, F 代表非零系数错误估计为零的平均数. 这里比较 ALasso、SCAD、 $L_q(0 < q < 1)$ 、Lasso 和 Hard 惩罚经验似然估计以及 EL、OEL (oracle empirical likelihood) 的效果, OEL 指在模型正确稀疏性已知的情况下进行的经验似然估计. 由文献 [70] 知,  $L_q(0 \leq q < 1/2)$  的压缩表现能力与  $L_{1/2}$  相当,  $L_q(1/2 < q < 1)$  的压缩能力弱于  $L_{1/2}$ . 所以,  $L_{1/2}$  可替代  $L_q(0 < q < 1)$ . 我们也比较了七种估计量在忽略测量误差 (直接用  $Z_i$  和  $W_i$  替代  $X_i$  和  $T_i$ ) 时的效果.

设测量误差  $v_i$  具有双指数分布

$$f(v) = (\sqrt{2}\delta_0)^{-1} e^{-\sqrt{2}|v|/\delta_0} \text{ for } \delta_0^2 = (3/7)\text{Var}(T).$$

取  $K(x)$  为高斯核函数,

$$K(x) = (\sqrt{2\pi})^{-1} e^{-(x^2/2)}.$$

则

$$K_n(x) = (\sqrt{2\pi})^{-1} e^{-(x^2/2)} \{1 - \delta_0^2(x^2 - 1)/2h^2\}.$$

**例 2.4.1** 取  $\beta = (1.5, 0.75, 1.25, 1, 0, 0, 0, 0)$ , 变量  $X_i \sim N(0, I_8)$ ,  $T_i \sim U(0, 1)$ . 误差  $\varepsilon_i \sim U(-0.2\sqrt{3}, 0.2\sqrt{3})$ ,  $u_{ij} \sim U(-0.15\sqrt{3}, 0.15\sqrt{3}), i = 1, \dots, n, j = 1, \dots, 8$ . 取  $g(t) = t_+^3(1-t)_+^3$ .

表 2.1 例 2.4.1 中考虑测量误差时 ME 误差 ( $\times 10^{-1}$ ) 和模型选择结果.

		OEL	EL	Alasso	SCAD	$L_{1/2}$	Lasso	Hard
$n = 100$	ME	0.053	0.140	0.072	0.071	0.074	0.111	0.077
	T	4	-	3.715	3.528	3.665	3.22	3.48
	F	0	-	0	0	0	0	0
$n = 200$	ME	0.028	0.063	0.030	0.034	0.032	0.073	0.034
	T	4	-	3.925	3.908	3.943	3.767	3.794
	F	0	-	0	0	0	0	0

表 2.2 例 2.4.1 中不考虑测量误差时 ME 误差 ( $\times 10^{-1}$ ) 和模型选择结果.

		OEL	EL	Alasso	SCAD	$L_{1/2}$	Lasso	Hard
$n = 100$	ME	0.076	0.148	0.091	0.101	0.095	0.208	0.095
	T	4	-	3.667	3.486	3.618	3.187	3.367
	F	0	-	0	0	0	0.003	0
$n = 200$	ME	0.048	0.084	0.059	0.058	0.055	0.164	0.056
	T	4	-	3.924	3.895	3.926	3.720	3.722
	F	0	-	0	0	0	0	0

在考虑测量误差时 ME 误差和模型选择结果在表 2.1 中给出. 由表 2.1 得出七种估计的 ME 误差随着  $n$  的增大而减小. 随着  $n$  的递增, 正确估计零的平均数 (T) 接近 4. 所有的惩罚估计都可以降低模型的复杂度, 其中 Lasso 表现得最差. 具有 Oracle 性质的估计 (Alasso、SCAD、 $L_{1/2}$ ) 比不具有 Oracle 性质的估计 (Lasso) 表现好. Alasso、SCAD、 $L_{1/2}$  的效果接近 OEL. 表 2.2 给出不考虑测量误差时估计的 ME 误差以及模型选择能力的结果. 对比表 2.1 与表 2.2, 不难发现, 无论是估计误差还是变量选择, 估计在考虑了测量误差时都比不考虑测量误差时表现好.

值得注意的是, 最小的 ME 误差和最好的模型选择能力 (最大化的正确估计零和最小化的错误估计零) 并不能同时达到.

**例 2.4.2** 取  $\beta = (3, 0, 1.5, 0, 0, 0, 2, 0, 0, 0, 1, 0)$ ,  $X_i \sim N(0, I_{12})$ .  $T_i = \Phi\{(X_{i1} + \sqrt{3}\xi_i)/2\}$ ,  $\xi_i \sim N(0, 1)$  其中  $\Phi(\cdot)$  是标准正态分布函数. 所以,  $T_i \sim U(0, 1)$  且与变量  $X_i$  有关. 误差  $\varepsilon_i = \sin(t_i)$  且  $t_i \sim N(0, 1)$ . 取  $g(t) = (1/2)^3 \sin(2\pi t^3)$ .  $X_i$  带有测量误差  $u_i$ ,  $u_i$  的元素服从均值为 0, 方差为  $c0.2^2$  的正态分布, 并且元素间的相关系数为 0.2. 考虑两种情况  $c = 1, 2$ , 它们分别代表不同程度的污染水平. 模拟中将考虑不同的  $c$  对模型估

计的影响.

表 2.3 例 2.4.2 中 ME 误差 ( $\times 10^{-1}$ ) 结果.

		OEL	EL	Alasso	SCAD	$L_{1/2}$	Lasso	Hard
考虑测量误差								
$n = 100$	$c = 1$	0.512	1.962	0.585	0.617	0.575	1.043	0.954
	$c = 2$	0.941	3.759	1.127	1.161	0.970	1.660	1.702
$n = 200$	$c = 1$	0.262	0.915	0.271	0.286	0.258	0.573	0.364
	$c = 2$	0.468	1.682	0.517	0.510	0.460	0.875	0.719
不考虑测量误差								
$n = 100$	$c = 1$	0.993	2.347	1.169	1.088	1.075	2.317	1.593
	$c = 2$	0.737	1.439	0.855	0.777	0.785	1.806	0.975
$n = 200$	$c = 1$	2.607	4.829	3.010	2.997	2.629	4.940	4.012
	$c = 2$	2.215	3.592	2.432	2.291	2.275	4.238	3.038

表 2.3 列出了参数估计误差的结果. 从表 2.3 中可以得到和例 2.4.1 一样的结论. 无论考虑测量误差与否, 惩罚估计可以降低相应的 EL 估计的模型误差. 然而, Lasso 在忽略测量误差时并不满足这样的结论. 从表中数据, 可以看出测量误差的水平对结果有影响. 随着  $c$  的增加, 所有估计在考虑测量误差和不考虑测量误差的 ME 误差都增加. 同样地, 在考虑测量误差时估计方法的 ME 误差比不考虑测量误差时小.

表 2.4 给出了模型选择的结果. 从表 2.4 可以看出,  $T$  随着  $c$  的增加而减小. 当  $c$  增加时, 测量误差对非零元素有更多的干扰, 所有更难正确估计零. 同样地, 考虑误差时, 变量选择的效果更好.

最后, 考虑 PEL 检验问题.

例 2.4.3 取  $\beta_0 = (1, 0.7, 0.5, 0, 0, 0)$ , 变量  $X_i \sim N(0, \Sigma_x)$ ,  $\Sigma_x = (\sigma_{ij})$  其中  $\sigma_{ij} = 0.5^{|i-j|}$ ,  $T_i \sim U(0, 1)$ . 误差  $\varepsilon_i \sim U(-0.2\sqrt{3}, 0.2\sqrt{3})$ ,  $u_i \sim N(0, \sigma^2 I_6)$ ,  $\sigma^2 = 0.25^2, 0.3^2, 0.35^2$  和  $0.4^2$  代表不同水平的测量误差. 非参数方程  $g(t) = t_+^3(1-t)_+^3$ .

设  $\beta_1 = \beta_{10} + \Delta/\sqrt{n}$ ,  $\Delta = -0.2, -0.1, 0, 0.1, 0.2$ . 表 2.5 给出实际检验水平和功效即拒绝原假设  $H_0: \beta_1 = \beta_{10}$  的经验频率, 显著性水平为 0.05. 当  $\Delta \neq 0$  时, 备择假设存在密集的转变. 随着样本  $n$  的增加, 实际检验水平接近 0.05, 功效接近 1. 当  $\Delta$  偏离  $\Delta = 0$  越多, 功效越接近 1. 另外, 测量误差的水平影响检验的效果. 误差越小, 实际检验水平和功效的结果越好. 各种误差水平下的检验都随着  $n$  的增加, 实际检验水

表 2.4 例 2.4.2 中模型选择的结果.

		$n=100$					$n=200$				
		ALasso	SCAD	$L_{1/2}$	Lasso	Hard	ALasso	SCAD	$L_{1/2}$	Lasso	Hard
考虑测量误差											
$c=1$	T	7.693	7.806	7.894	6.211	6.515	7.958	7.984	7.979	7.412	7.191
	F	0	0	0	0	0	0	0	0	0	0
$c=2$	T	7.640	7.741	7.822	5.981	6.486	7.949	7.941	7.948	7.041	6.798
	F	0	0	0	0	0	0	0	0	0	0
不考虑测量误差											
$c=1$	T	7.588	7.585	7.160	5.849	5.717	7.915	7.910	7.385	6.906	6.522
	F	0	0	0	0	0	0	0	0	0	0
$c=2$	T	7.493	6.987	7.096	4.601	4.620	7.754	7.451	7.279	5.632	5.231
	F	0	0	0	0	0	0	0	0	0	0

表 2.5 例 2.4.3 在 0.05 水平下拒绝  $H_0$  的经验频率.

		$\Delta$					$\Delta$				
$n$	$\sigma^2$	-0.2	-0.1	0	0.1	0.2	-0.2	-0.1	0	0.1	0.2
考虑测量误差											
100	$0.25^2$	91.7	68.5	5.3	68.9	92.4	93.3	73.2	25.2	85.0	97.6
	$0.3^2$	89.8	67.7	9.0	70.6	92.2	96.0	80.5	47.0	90.9	99.0
	$0.35^2$	87.8	65.7	13.7	67.3	91.1	97.1	90.0	70.7	94.6	99.2
	$0.4^2$	89.6	71.9	21.5	72.2	91.2	98.4	95.2	86.5	97.6	99.8
200	$0.25^2$	99.1	87.1	5.1	88.2	98.9	99.6	94.3	49.1	98.2	100
	$0.3^2$	97.0	84.3	7.8	85.6	98.9	99.9	98.3	77.4	99.3	99.9
	$0.35^2$	97.7	80.4	12.8	82.5	98.1	99.8	98.7	94.3	100	100
	$0.4^2$	95.2	79.3	17.3	82.0	96.4	100	99.8	99.3	100	100
300	$0.25^2$	99.7	94.8	5.0	95.9	99.8	100	98.7	67.8	99.4	100
	$0.3^2$	99.7	92.9	6.2	93.3	99.9	99.9	98.9	91.8	99.7	100
	$0.35^2$	99.6	89.2	10.6	91.6	99.4	100	100	99.3	100	100
	$0.4^2$	98.2	88.2	14.8	88.5	99.0	100	100	99.9	100	100

平和功效分别趋近于 0.05 和 1. 结果表明, 提出的检验统计量表现令人满意.

值得注意的是, 在忽略测量误差时不能维持实际检验水平的一致性. 所以, 相应检验的功效也就没有任何的意义. 也就是说, 当变量带有测量误差时, 忽略测量误差会导致假设检验失去意义.

### 2.4.2 血浆 $\beta$ -胡萝卜素水平数据

这里利用文献 [71] 的血浆  $\beta$ -胡萝卜素水平数据来说明所提 PEL 的运用. 关于此数据更多的分析可以参见文献 [72–74].

考虑血浆  $\beta$ -胡萝卜素水平 (BetaPlasma) 与下面变量之间的关系: 年龄 (Age)、克托莱指数 (Quetelet)、维他命的使用 (Vitamin)、每天消耗的卡路里 (Calories)、酒精饮料的数量 (Alcohol)、胆固醇摄入 (Cholesterol) 和饮食  $\beta$ -胡萝卜素 (BetaDiet), 模型拟合

$$\text{BetaPlasma} = \text{Age} * \beta_1 + \text{Quetelet} * \beta_2 + \text{Vitamin} * \beta_3 + \text{Calories} * \beta_4 + \text{Alcohol} * \beta_5 + \text{Cholesterol} * \beta_6 + g(\text{BetaDiet}) + \varepsilon.$$

$\beta$ -胡萝卜素水平数据通常认为有测量误差. 基于此, 表 2.6 给出 Speckman [75], EL 以及 PEL 估计及其标准误差结果. 从表 2.6 知道 PEL 方法有较小的模型误差并且降低了模型的复杂度. 从 PEL 估计结果可以看出 Calories 和 Alcohol 是不显著的变量. BetaPlasma 与 Age 正相关, 与 Quetelet, Vitamin 和 Cholesterol 负相关. 图 2.1 是  $g(\cdot)$  的估计曲线, 此曲线表示标准化数据 BetaDiet 与标准化数据 BetaPlasma 之间的关系.

表 2.6 血浆  $\beta$ -胡萝卜素水平数据的参数估计与标准误差结果.

Variable	PEL	EL	Speckman
Age	0.0546(0.0111)	0.0737(0.0129)	0.0750(0.0125)
Quetelet	-0.1989(0.0087)	-0.2005(0.0085)	-0.2022(0.0086)
Vitamin	-0.1947(0.0111)	-0.1994(0.0112)	-0.2021(0.0115)
Calories	0	0.0389(0.0206)	0.0391(0.0195)
Alcohol	0	-0.0113(0.0168)	-0.0064(0.0141)
Cholesterol	-0.1143(0.0084)	-0.1378(0.0130)	-0.1375(0.0123)

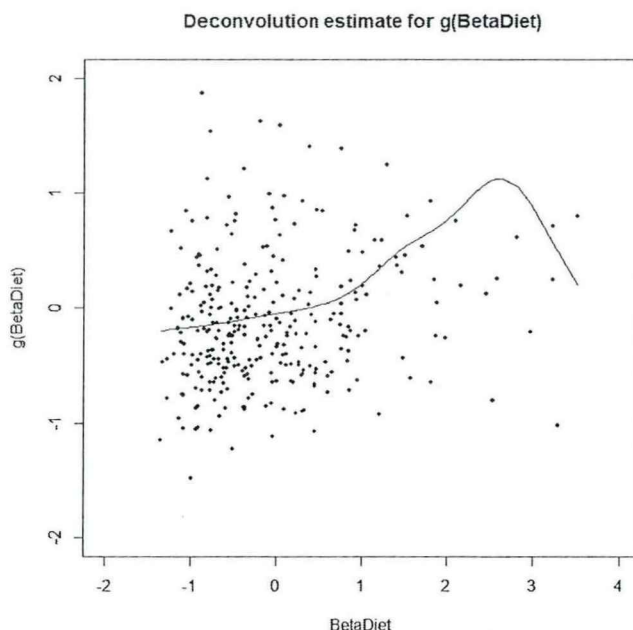


图 2.1 血浆  $\beta$ -胡萝卜素水平数据的  $g(\cdot)$  的估计曲线. 点代表部分残差  $r_i = Y_i - Z_i^T \hat{\beta}$ ,  $\hat{\beta}$  是 PEL 估计量. 实曲线代表  $g(\cdot)$  的估计.

## 2.5 定理证明

引理 2.5.1 给出了辅助随机向量在参数真值  $\beta_0$  处的渐近性质. 引理 2.5.2 保证了  $\ell(\beta)$  的渐近展开. 记  $D_n = \{\beta : \|\beta - \beta_0\| \leq Cn^{-1/2}\}$ ,  $C > 0$ . 本文证明中, 采用矩阵  $A$  的 Frobenius 范数, 即  $\|A\| = \{\text{tr}(A^T A)\}^{1/2}$ .

引理 2.5.1<sup>[62]</sup> 在条件 (C1-C3) 和 (C6-C8) 下, 有

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\beta_0) \xrightarrow{d} N(0, \Omega),$$

$$\frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) \eta_i(\beta_0)^T \xrightarrow{p} \Omega,$$

$$\max_i \|\eta_i(\beta_0)\| = o_p(n^{1/2}).$$

引理 2.5.2 在条件 (C1)-(C8) 下, 对于  $\beta \in D_n$  有  $\|\lambda_\beta\| = O_p(n^{-1/2})$ .

证明 对于  $\beta \in D_n$ , 设  $\lambda = \rho u$  其中  $\|u\| = 1$ . 根据文献 [21], 有

$$\begin{aligned} \rho u^T S_\beta u &< \rho u^T \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta) \eta_i(\beta)^T}{1 + \lambda^T \eta_i(\beta)} u \cdot \max_i \{1 + \lambda^T \eta_i(\beta)\} \\ &= \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta) \max_i \{1 + \lambda^T \eta_i(\beta)\} \\ &< \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta) \right| \{1 + \rho \max_i \|\eta_i(\beta)\|\}. \end{aligned}$$

所以

$$\rho \left\{ u^T S_\beta u - \max_i \|\eta_i(\beta)\| \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta) \right| \right\} \leq \frac{1}{n} \left| \sum_{i=1}^n u^T \eta_i(\beta) \right|.$$

由引理 2.5.1, 有

$$\left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta_0) \right| = O_p(n^{-1/2}).$$

根据条件 (C4), 得到  $\max_i \|\partial \eta_i(\beta_0) / \partial \beta^T (\beta - \beta_0)\| = o_p(1)$ . 所以

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta) \right| &= \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta_0) + \frac{\partial \eta_i(\beta_0)}{\partial \beta^T} (\beta - \beta_0) \right| \\ &\leq \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta_0) \right| + \max_i \left\| \frac{\partial \eta_i(\beta_0)}{\partial \beta^T} (\beta - \beta_0) \right\| \\ &= O_p(n^{-1/2}). \end{aligned}$$

易知

$$\max_i \|\eta_i(\beta)\| \leq \max_i \|\eta_i(\beta_0)\| + \max_i \|\partial \eta_i(\beta_0) / \partial \beta^T (\beta - \beta_0)\|,$$

这结合引理 2.5.1, 知

$$\max_i \|\eta_i(\beta)\| \left| \frac{1}{n} \sum_{i=1}^n u^T \eta_i(\beta) \right| = o_p(1).$$

因此

$$u^T S_\beta u = u^T S_{\beta_0} u + o_p(1) = O_p(1),$$

引理 2.5.2 的证明完成.

**定理 2.2.1 的证明** 只需证明对于任意给定的  $\varepsilon > 0$  都存在足够大的常数  $C$  使得

$$P \left\{ \inf_{\beta \in \partial D_n} \ell_p(\beta) > \ell_p(\beta_0) \right\} \geq 1 - \varepsilon, \quad (2.5.1)$$

其中  $\partial D_n$  是  $D_n$  的边界. 这就指出, 以概率至少  $1 - \varepsilon$ , 在  $\{\beta_0 + n^{-1/2}u : \|u\| \leq C\}$  内存在局部最小. 也就是说, 存在  $\ell_p(\beta)$  的局部最小值点  $\hat{\beta}$  且满足  $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$ .



由 (2.2.3) 知

$$Q_{1n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{[1 + \lambda^T \eta_i(\beta)]} = 0.$$

由引理 2.5.2 可知  $\lambda^T \eta_i(\beta)$  和  $\|\lambda\|$  是随机任意小. 对  $Q_{1n}(\beta, \lambda)$  泰勒展开, 就有

$$0 = \bar{\eta}(\beta) - S_\beta \lambda + r_n, \quad (2.5.2)$$

其中  $\bar{\eta}(\beta) = n^{-1} \sum \eta_i(\beta)$ ,  $r_n = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) [\lambda^T \eta_i(\beta)]^2 (1 + \xi_i)^{-3}$  是剩余项并且  $|\xi_i| \leq |\lambda^T \eta_i(\beta)|$ . 反解 (2.5.2), 得到  $\lambda = S_\beta^{-1} [\bar{\eta}(\beta) + r_n]$ . 由  $\ell(\beta) = \ell(\beta, \lambda)$  的泰勒展开式得到

$$2\ell(\beta, \lambda) = 2 \sum_{i=1}^n \lambda^T \eta_i(\beta) - \sum_{i=1}^n [\lambda^T \eta_i(\beta)]^2 + \frac{2}{3} \sum_{i=1}^n \frac{[\lambda^T \eta_i(\beta)]^3}{(1 + \xi_i)^3},$$

其中  $|\xi_i| \leq |\lambda^T \eta_i(\beta)|$ . 将  $\lambda$  带入到  $\ell(\beta)$  中, 就有

$$2\ell(\beta) = n\bar{\eta}(\beta)^T S_\beta^{-1} \bar{\eta}(\beta) - nr_n^T S_\beta^{-1} r_n + \frac{2}{3} \sum_{i=1}^n \frac{[\lambda^T \eta_i(\beta)]^3}{(1 + \xi_i)^3}. \quad (2.5.3)$$

将 (2.5.3) 分解为

$$2\ell(\beta) = T_0 + T_1 + T_2,$$

其中

$$T_0 = n\bar{\eta}(\beta_0)^T S_{\beta_0}^{-1} \bar{\eta}(\beta_0),$$

$$T_1 = n\{\bar{\eta}(\beta) - \bar{\eta}(\beta_0)\}^T S_\beta^{-1} \{\bar{\eta}(\beta) - \bar{\eta}(\beta_0)\}$$

以及

$$\begin{aligned} T_2 = & n\{\bar{\eta}(\beta_0)\}^T [S_\beta^{-1} - S_{\beta_0}^{-1}] \bar{\eta}(\beta_0) + 2\bar{\eta}(\beta_0)^T S_\beta^{-1} \{\bar{\eta}(\beta) - \bar{\eta}(\beta_0)\} \\ & - nr_n^T S_\beta^{-1} r_n + \frac{2}{3} \sum_{i=1}^n [\lambda^T \eta_i(\beta)]^3 (1 + \xi_i)^{-3}. \end{aligned}$$

因为

$$\frac{1}{n} \sum_{i=1}^n [\eta_i(\beta) - \eta_i(\beta_0)] = \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta^*)}{\partial \beta^T} (\beta - \beta_0),$$

其中  $\beta^* \in D_n$ . 由条件 (C1)-(C3), 得到

$$\begin{aligned} T_1 & > n \|\bar{\eta}(\beta) - \bar{\eta}(\beta_0)\|_{r_p^{-1}(\Omega)}^2 = n \left\| \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta_0)}{\partial \beta^T} \right\|^2 \|\beta - \beta_0\|_{r_p^{-1}(\Omega)}^2 \\ & = C^2 O_p(1). \end{aligned}$$

因为  $T_2/T_1 = O_p(1)$  且  $2\ell(\beta_0) - T_0 \rightarrow 0$ , 所以当  $C$  足够大时,  $T_1$  控制了  $2\ell(\beta) - 2\ell(\beta_0)$  的符号.

由泰勒展开, 有

$$\frac{1}{|\tilde{\beta}_i|} = \frac{1}{|\beta_{i0}|} - \frac{\text{sgn}(\beta_{i0})}{|\beta_{i0}|^2}(\tilde{\beta}_i - \beta_{i0}) + o_p(|\tilde{\beta}_i - \beta_{i0}|) = \frac{1}{|\beta_{i0}|} + O_p(n^{-1/2}).$$

由条件 (C5)-(C6), 当  $n$  足够大时,

$$n\tau \sum_{i=1}^d \frac{n^{-1/2}|e_i|}{|\tilde{\beta}_i|} < \frac{n^{1/2}\tau dC}{M} + \tau d O_p(1) = o_p(1).$$

因为

$$\begin{aligned} \ell_p(\beta) - \ell_p(\beta_0) &= [\ell(\beta) - \ell(\beta_0)] + n\tau \sum_{i=1}^p (|\beta_i| - |\beta_{i0}|) / |\tilde{\beta}_i| \\ &> [\ell(\beta) - \ell(\beta_0)] - n\tau \sum_{i=1}^d (n^{-1/2}|e_i|) / |\tilde{\beta}_i|, \end{aligned}$$

其中第一项是  $C^2$  阶, 第二项被第一项控制. 所以对于充分大的  $C$ , (2.5.1) 式成立, 证明完成.

**定理 2.2.2 的证明** 由引理 2.5.2 可知, 对于  $\beta \in D_n$ , 有  $\lambda^\top \eta_i(\beta) = o_p(1)$ . 由泰勒展开, 有

$$\begin{aligned} \frac{1}{n} \frac{\partial \ell_p(\beta)}{\partial \beta_j} &= n^{-1} \sum_{i=1}^n \lambda^\top \frac{\partial \eta_i(\beta)}{\partial \beta_j} (1 + o_p(1)) + \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} \\ &= \lambda^\top \frac{1}{n} \sum_{i=1}^n \left[ \frac{\partial \eta_i(\beta_0)}{\partial \beta_j} + \frac{\partial^2 \eta_i(\beta_0)}{\partial \beta_j \partial \beta^\top} (\beta - \beta_0) \right] (1 + o_p(1)) + \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} \\ &= \lambda^\top \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta_0)}{\partial \beta_j} + \lambda^\top \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \eta_i(\beta_0)}{\partial \beta_j \partial \beta^\top} (\beta - \beta_0) + \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} + o_p(1) \\ &= I_j + II_j + III_j + \text{small order terms.} \end{aligned}$$

在条件 (C1-C3) 下, 有

$$\max_{j \notin \mathcal{A}} I_j = \max_{j \notin \mathcal{A}} \lambda^\top \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta_0)}{\partial \beta_j} = O_p(n^{-1/2}).$$

因为  $\partial^2 \eta_i(\beta_0) / \partial \beta_j \partial \beta^\top = 0$ , 所以  $II_j = 0$ . 由条件 (C5), 知

$$\max_{j \notin \mathcal{A}} III_j = \max_{j \notin \mathcal{A}} \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} = O_p(\tau/n^{-1/2}) \text{sgn}(\beta_j)$$

因为  $\max_{j \notin \mathcal{A}} III_j$  一致性的控制了  $\partial \ell_p(\beta) / \partial \beta_j$  的符号, 所以, 对于任意的  $j \notin \mathcal{A}$ , 以概率趋近于 1, 有

$$\frac{\partial \ell_p(\beta)}{\partial \beta_j} > 0, \quad \beta_j \in (0, cn^{-1/2}e_j); \quad \frac{\partial \ell_p(\beta)}{\partial \beta_j} < 0, \quad \beta_j \in (-cn^{-1/2}e_j, 0).$$

对于任意的  $j \notin \mathcal{A}$  有  $\hat{\beta}_j = 0$ , 这证明了第一部分.

接下来证明第二部分估计的有效性. 由第一部分的结论和 PEL 定义可知,  $\hat{\beta}$  是满足  $H_2\beta_0 = 0$  时 (2.2.4) 的最小值. 根据文献 [20], 由拉格朗日算法得到估计量等价于最小化下面的目标函数

$$\tilde{\ell}(\beta, \lambda, \nu) = n^{-1} \sum_{i=1}^n \log\{1 + \lambda^T \eta_i(\beta)\} + \sum_{j=1}^p p_\tau(|\beta_j|) + \nu^T H_2 \beta,$$

其中  $\nu \in \mathbb{R}^{p-d}$  是拉格朗日乘子向量.

记  $b(\beta) = (\tau \text{sgn}(\beta_1)/|\hat{\beta}_1|, \dots, \tau \text{sgn}(\beta_p)/|\hat{\beta}_p|)^T$ ,  $(\hat{\beta}, \hat{\lambda}, \hat{\nu})$  满足  $Q_{1n}(\hat{\beta}, \hat{\lambda}) = 0$ ,  $Q_{2n}(\hat{\beta}, \hat{\lambda}) + b(\hat{\beta}) + H_2^T \hat{\nu} = 0$  和  $Q_{3n}(\hat{\beta}, \hat{\lambda}) = H_2 \hat{\beta} = 0$ . 由引理 2.5.1 和文献 [20], 得到  $\|\hat{\lambda}\| = \|\hat{\nu}\| = O_p(n^{-1/2})$ . 将  $Q_{1n}(\hat{\beta}, \hat{\lambda}) = 0$ ,  $Q_{2n}(\hat{\beta}, \hat{\lambda}) + b(\hat{\beta}) + H_2^T \hat{\nu} = 0$  和  $Q_{3n}(\hat{\beta}, \hat{\lambda}) = 0$  在  $(\beta_0, 0, 0)$  处展开, 得到

$$\begin{pmatrix} -Q_{1n}(\beta_0, 0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\Omega & -\Sigma & 0 \\ -\Sigma^T & 0 & H_2^T \\ 0 & H_2 & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \\ \hat{\nu} \end{pmatrix} + R_n^{(1)} + R_n^{(2)} + R_n^{(3)} + o_p(n^{-1/2}), \quad (2.5.4)$$

其中

$$R_n^{(1)} = \{0, b(\beta_0)^T, 0\}^T,$$

$$R_n^{(2)} = \left\{ \left\{ \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) \eta_i^T(\beta_0) - \Omega \right\} \hat{\lambda} \right\}^T + \left\{ \left[ \frac{1}{n} \sum_{i=1}^n \partial \eta_i(\beta_0) / \partial \beta^T + \Sigma \right] (\hat{\beta} - \beta_0) \right\}^T, 0, 0 \right\}^T,$$

$$R_n^{(3)} = \{0, \left\{ \left[ \frac{1}{n} \sum_{i=1}^n \partial \eta_i(\beta_0)^T / \partial \beta + \Sigma^T \right] \hat{\lambda} \right\}^T, 0\}^T.$$

由条件 (C5) 得到  $\|R_n^{(1)}\| = o_p(n^{-1/2})$ . 类似于文献 [52], 有

$$\frac{1}{n} \sum_{i=1}^n \partial \eta_i(\beta_0) / \partial \beta^T \xrightarrow{p} -\Sigma.$$

由引理 2.5.1, 得到  $\|R_n^{(2)}\| = \|R_n^{(3)}\| = o_p(n^{-1/2})$ .

定义  $K_{11} = -\Omega$ ,  $K_{12} = (-\Sigma, 0)$ ,  $K_{21} = K_{12}^T$  和

$$K_{22} = \begin{pmatrix} 0 & H_2^T \\ H_2 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}.$$

反解 (2.5.4) 得到

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \\ \hat{\nu} \end{pmatrix} = K^{-1} \left\{ \begin{pmatrix} -\tilde{Q}_{1n}(\beta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} + R_n \right\}.$$

由矩阵块的逆运算得到

$$K^{-1} = \begin{pmatrix} K_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -K_{11}^{-1}K_{12} \\ I \end{pmatrix} A^{-1} \begin{pmatrix} -K_{21}K_{11}^{-1} & I \end{pmatrix},$$

其中  $A = K_{22} - K_{21}K_{11}^{-1}K_{12}$ . 所以

$$\hat{\beta} - \beta_0 = \{V - VH_2^T(H_2VH_2^T)^{-1}H_2V\}\Sigma^T\Omega^{-1}Q_{1n}(\beta_0, 0) + o_p(n^{-1/2}).$$

非零元素  $\beta_1$  的估计  $\hat{\beta}_1$  可以表示为

$$\hat{\beta}_1 - \beta_{10} = \{H_1V - H_1VH_2^T(H_2VH_2^T)^{-1}H_2V\}\Sigma^T\Omega^{-1}Q_{1n}(\theta_0, 0) + o_p(n^{-1/2}).$$

所以,

$$\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \rightarrow N(0, V_d).$$

证明完成.

**定理 2.2.3 的证明** 由定理 2.2.2 的结论, 可知, 以概率趋近于 1 有  $\hat{\beta} = (\hat{\beta}_1^T, 0)$ . 根据  $\hat{\beta}_1$  得到对应的  $\hat{\lambda}_1$ . 假设  $\hat{\lambda} = (\hat{\lambda}_1^T, 0)$ ,  $S_{21}^T = S_{12}$  是  $\Sigma$  左上角  $d \times d$  个元素构成的子矩阵,  $S_{11}$  是  $\Omega$  左上角  $d \times d$  个元素构成的子矩阵,  $W = S_{21}S_{11}^{-1}S_{12}$ . 所以,

$$\begin{aligned} \ell_p(\hat{\beta}) &= \sum_{i=1}^n \log(1 + \hat{\lambda}^T \eta_i(\hat{\beta})) + n \sum_{j=1}^p p_\tau(|\hat{\beta}_j|) \\ &= \frac{n}{2} Q_{1n}^{(1)T}(\beta_{10}, 0) S_{11}^{-1} Q_{1n}^{(1)}(\beta_{10}, 0) - \frac{n}{2} \{[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau]^T [\Sigma_\tau + W]^{-1} \\ &\quad \times [S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau] + n \sum_{j=1}^d p_\tau(|\beta_{j0}|) + o_p(1)\}, \end{aligned}$$

其中  $Q_{1n}^{(1)}(\beta_{10}, 0)$  是  $Q_{1n}(\beta_{10}, 0)$  前  $d$  个分量构成的子向量.

$$b_\tau = (p'_\tau(|\beta_{10}|)\text{sgn}(\beta_{10}), \dots, p'_\tau(|\beta_{d0}|)\text{sgn}(\beta_{d0}))^T,$$

$$\Sigma_\tau = \text{diag}\{p''_\tau(|\beta_{10}|)\text{sgn}(\beta_{10}), \dots, p''_\tau(|\beta_{d0}|)\text{sgn}(\beta_{d0})\}.$$

类似地,

$$\begin{aligned} \ell_p((\beta_{10}^T, 0)^T) &= \sum_{i=1}^n \log(1 + (\lambda_{10}^T, 0)^T \eta_i(\beta_0)) + n \sum_{j=1}^d p_\tau(|\beta_{j0}|) \\ &= \frac{n}{2} Q_{1n}^{(1)T}(\beta_{10}, 0) S_{11}^{-1} Q_{1n}^{(1)}(\beta_{10}, 0) + n \sum_{j=1}^d p_\tau(|\beta_{j0}|) + o_p(1). \end{aligned}$$

因此,

$$\begin{aligned} W(\beta_{10}) &= n\{[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau]^T [\Sigma_\tau + W]^{-1}[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau] + o_p(1)\} \\ &= \{\sqrt{n}[\Sigma_\tau + W]^{-1/2}[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau]\}^T \\ &\quad \times \sqrt{n}[\Sigma_\tau + W]^{-1/2}[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau] + o_p(1). \end{aligned}$$

因为  $P(b_\tau = 0) \rightarrow 1$ , 所以

$$\sqrt{n}[\Sigma_\tau + W]^{-1/2}[S_{21}S_{11}^{-1}Q_{1n}^{(1)}(\beta_{10}, 0) + b_\tau] \rightarrow N(0, I).$$

故

$$W(\beta_{10}) \rightarrow \chi_d^2,$$

证明完成.

## 第3章 高维广义线性模型的惩罚经验似然

### 3.1 引言

在上章中,我们研究了低维情况下部分线性模型的参数估计与变量选择问题.随着科技的发展,在实际应用中,如计量经济学,生物信息学,生物遗传学以及航空航天方面等领域,数据常以高维的形式出现,这使得高维的相关研究有必要性.在本章中,我们讨论高维情况下广义线性模型的参数估计与变量选择问题.

广义线性模型是在经典线性模型中发展起来,它不要求因变量具有正态分布.广义线性模型对连续数据和离散数据都可进行统计分析,这使得该模型在社会,经济,生物医学等领域中有广泛的应用.早期广义线性模型被用来分析人口增长等社会现象,1972年Nelder和Wedderburn正式提出经典广义线性模型<sup>[4]</sup>,其中响应变量具有指数族分布.自Wedderburn提出拟似然方法以来,广义线性模型只需要期望和方差形式<sup>[7]</sup>.

1994年,Kolaczyk研究了标准广义线性模型的经验似然<sup>[24]</sup>.文献[24]的优点是构造自然形状和方向的置信区间,而不是参数估计.为了充分利用方差结构中的信息,2003年Chen和Cui在文献[24]的基础上推广了广义线性模型的经验似然方法.Yan和Chen在固定和自适应情况下讨论了广义线性模型的经验似然<sup>[76]</sup>.近年来,高维数据研究得到了广泛的关注.文献[77]证明了当参数的维数变化时,经验似然方法仍然有效.

Tang和Leng在2010年将PEL方法被应用到高维数据分析中<sup>[35]</sup>.此后,统计学家将该方法应用到各种模型中,但在广义线性模型中的相关研究却很少.同时,大多数研究者通常将散布参数 $\sigma^2$ 设成已知或者根本不考虑,如文献[78]比较缺失数据下广义线性模型的四种估计方法;Liang等人讨论了高维广义线性模型的贝叶斯子集建模[79].当散布参数 $\sigma^2$ 未知时,相关的研究却很少.在经济学和社会学中,数据常常是超散布或零很多的计数模型.这使得在广义线性模型中考虑散布参数很有必要.

因此,基于以上问题的考虑,本章提出基于三个估计方程的高维广义线性模型的PEL,同时估计参数 $\beta$ 和散布参数 $\sigma^2$ .主要研究结果是所提出的PEL估计具有Oracle性质以及构造的检验统计量有渐近卡方分布.模拟实验和实例分析表明提出的PEL有较好的效果.

### 3.2 方法和主要结果

假设响应变量  $Y$  与随机变量  $X$  具有如下的关系:

$$E(Y|X) = \mu(\beta) = G(X^T \beta) \quad , \quad \text{Var}(Y|X) = \sigma^2 V\{G(X^T \beta)\}. \quad (3.2.1)$$

其中  $\beta \in \mathbb{R}^p$  是未知向量,  $G(\cdot)$  是光滑联系函数的逆,  $V(\cdot)$  是方差函数,  $\sigma^2$  是散布参数. 假设  $\mu(\beta) = G(X_i^T \beta)$ ,  $\beta$  的拟似然函数定义为<sup>[7]</sup>

$$Q\{y; \mu(\beta)\} = \int_y^{\mu(\beta)} \frac{y-u}{\sigma(u)} du. \quad (3.2.2)$$

假设  $(X_1, Y_1), \dots, (X_n, Y_n)$  是来自模型 (3.2.1) 的独立同分布样本, 且  $\mu_i(\beta) = G(X_i^T \beta)$ . 对式子 (3.2.2) 关于  $\beta$  求导, 得到拟得分函数为

$$\frac{\partial Q(Y_i, \mu_i(\beta))}{\partial \beta} = \frac{\{Y_i - \mu_i(\beta)\} G'(X_i^T \beta) X_i}{V\{\mu_i(\beta)\}}.$$

从模型的方差结构出发可得到

$$E\{[Y_i - G(X_i^T \beta)]^2\} = \sigma^2 V\{G(X_i^T \beta)\}.$$

记

$$Z_i^{(1)}(\sigma^2, \beta) = \frac{[Y_i - G(X_i^T \beta)]^2}{\sigma^4 V\{G(X_i^T \beta)\}} - \frac{1}{\sigma^2},$$

$$Z_i^{(2)}(\beta) = \frac{[Y_i - G(X_i^T \beta)] G'(X_i^T \beta) X_i}{V\{G(X_i^T \beta)\}}.$$

当  $V' = V'(G(X_i^T \beta)) \neq 0$ , 方差结构中包含更多的有用信息. 为了充分利用这些信息, Chen 和 Cui 引入了如下的辅助随机向量<sup>[80]</sup>

$$Z_i^{(3)}(\sigma^2, \beta) = \left( \frac{[Y_i - G(X_i^T \beta)]^2}{\sigma^4 V\{G(X_i^T \beta)\}} - \frac{1}{\sigma^2} \right) \omega(X_i^T \beta, X_i),$$

其中  $\omega$  是  $r$  维权重向量 ( $1 \leq r \leq p$ ), 论文给出了最优的  $\omega$ . 本章采用

$$Z_i(\sigma^2, \beta) = (Z_i^{(1)T}(\sigma^2, \beta), Z_i^{(2)T}(\beta), Z_i^{(3)T}(\sigma^2, \beta))^T$$

作为辅助随机向量. 由数值计算可知, 引入的限制条件降低了参数估计误差.

如果  $(\sigma_0^2, \beta_0)$  是参数真值, 那么  $E(Z_i(\sigma_0^2, \beta_0)) = 0$ . 由文献 [18] 知, 参数  $(\sigma^2, \beta)$  的对数经验似然比函数定义为

$$\ell(\sigma^2, \beta) = -\max \left\{ \sum_{i=1}^n \log(np_i) : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i Z_i(\sigma^2, \beta) = 0 \right\}.$$

由拉格朗日乘数法, 得到

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T Z_i(\sigma^2, \beta)}$$

且  $\ell(\sigma^2, \beta)$  可表示为

$$\ell(\sigma^2, \beta) = \sum_{i=1}^n \log\{1 + \lambda^T Z_i(\sigma^2, \beta)\},$$

其中  $\lambda = \lambda(\sigma^2, \beta)$  满足

$$\sum_{i=1}^n \frac{1}{n} \frac{Z_i(\sigma^2, \beta)}{1 + \lambda^T Z_i(\sigma^2, \beta)} = 0. \quad (3.2.3)$$

假设  $\tilde{\beta}$  是  $\beta_0$  的  $\sqrt{n}$  相合估计, 权重函数  $\tilde{w} = 1/|\tilde{\beta}|^\gamma$  ( $\gamma > 0$ ). 结合自适应 Lasso 惩罚函数, 参数  $(\sigma^2, \beta)$  的 PEL 估计定义为如下目标函数

$$\ell_p(\sigma^2, \beta) = \sum_{i=1}^n \log\{1 + \lambda^T Z_i(\sigma^2, \beta)\} + n\tau \sum_{j=1}^p \tilde{w}_j |\beta_j| \quad (3.2.4)$$

的最小值点  $(\hat{\sigma}^2, \hat{\beta})$ , 其中  $\tilde{w}_j$  是  $\tilde{w}$  的第  $j$  个元素.

设  $\mathcal{A} = \{j : \beta_{0j} \neq 0\}$ ,  $\mathcal{A}$  的基数记为  $|\mathcal{A}| = d$ .  $d$  未知且随着  $n \rightarrow \infty$  而递增. 令  $\theta_0 = (\sigma_0^2, \beta_{10}^T, \beta_{20}^T)^T$ ,  $\beta_{10} \in \mathbb{R}^d$ ,  $\beta_{20} \in \mathbb{R}^{p-d}$ . 不失一般性, 假设  $\beta_{10} \neq 0$ ,  $\beta_{20} = 0$ , 其中  $0$  是所有元素都为零的向量. 类似地, 令  $\hat{\theta} = (\hat{\sigma}^2, \hat{\beta}_1, \hat{\beta}_2)^T = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$ ,  $\hat{\theta}_1 = (\hat{\sigma}^2, \hat{\beta}_1^T)^T$ ,  $\hat{\theta}_2 = \hat{\beta}_2$ .  $\hat{\sigma}^2$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  分别是  $\sigma_0^2$ ,  $\beta_{10}$ ,  $\beta_{20}$  的估计. 假设  $\Omega = E(\partial Z_i(\theta_0)/\partial \theta^T)$ ,  $\Sigma = E(Z_i(\theta_0)Z_i^T(\theta_0))$  以及  $V^{-1} = \Omega^T \Sigma^{-1} \Omega$ . 相应地, 将  $V$  划分为分块矩阵  $V_{ij}$  ( $i = 1, 2$ ;  $j = 1, 2$ ), 其中  $V_{11} \in \mathbb{R}^{(d+1) \times (d+1)}$ . 下面是本章需要的条件.

(C1) 假设  $\varepsilon_i = Y_i - G(X_i^T \beta)$ ,  $\{\varepsilon_i\}_{i=1}^n$  独立同分布且与  $X_i$  独立. 当  $\alpha > 4$  时,  $E\|X_i\|^{3\alpha} < +\infty$ ,  $E\|\varepsilon_i\|^{2\alpha} < +\infty$  且  $E\|\partial w_i / \partial \theta^T\|^\alpha < +\infty$ .

(C2)  $\Sigma$  的特征值满足  $0 < C_1 \leq \gamma_1(\Sigma) \leq \dots \leq \gamma_{p+r+1}(\Sigma) \leq C_2$ , 其中  $C_2 > C_1 > 0$ .

(C3)  $G(\cdot)$  三阶连续可导且  $V(\cdot)$  二阶连续可导.

(C4) 当  $n \rightarrow \infty$  时, 有  $p \rightarrow \infty$  且  $p^2/n \rightarrow 0$ .

(C5) 假设  $\beta_n^* = \min\{|\beta_{0j}|, j \in \mathcal{A}\}$ , 存在正常数  $M$  使得  $\beta_n^* \geq M$ .

(C6) 当  $n \rightarrow \infty$  时, 调整参数  $\tau$  满足  $\tau \rightarrow 0$ ,  $\tau/a_n \rightarrow \infty$  且  $\sqrt{nd}\tau \rightarrow 0$ .

**注 3.2.1** 条件 (C1)–(C3) 保证了 PEL 估计量的存在性以及相合性. 当条件 (C4) 满足时,  $p$  和  $d$  都可以变化. 条件 (C5)–(C6) 控制了惩罚函数对非零元素的影响.

首先, 定理 3.2.1 给出 PEL 估计  $\hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)^T$  的性质.



**定理 3.2.1** 在条件 (C1) – (C6) 下, 当  $n \rightarrow \infty$  时,

(i) 稀疏性:  $P(\hat{\theta}_2 = 0) \rightarrow 1$ ;

(ii) 渐近正态性:  $n^{1/2}W_n B^{-1/2}(\hat{\theta}_1 - \theta_{10})^T \rightarrow N(0, G)$ , 其中  $W_n \in \mathbb{R}^{q \times (d+1)}$  满足  $W_n W_n^T \rightarrow G$ ,  $G \in \mathbb{R}^{q \times q}$  ( $q$  固定),  $B = V_{11} - V_{12}V_{22}^{-1}V_{21}$ .

**注 3.2.2** 正如文献 [81] 所述, 定理 3.2.1 证明了 PEL 估计量的 Oracle 性质. 即, PEL 估计量在模型选择方面具有相合性并且与 OEL 估计具有相同的效率.

接下来, 考虑假设检验问题和构造  $\theta$  的置信区域. 检验  $H_0 : L_n \theta_0 = 0$  对  $H_1 : L_n \theta_0 \neq 0$ , 其中  $L_n \in \mathbb{R}^{q \times (p+1)}$  且满足  $L_n L_n^T = I_q$  ( $q$  固定),  $I_q$  是  $q$  维单位矩阵. 轮廓惩罚经验似然比检验统计量定义为

$$\tilde{\ell}(L_n) = -2 \left\{ \ell_p(\hat{\theta}) - \min_{\theta: L_n \theta = 0} \ell_p(\theta) \right\}.$$

定理 3.2.2 给出惩罚经验似然比检验统计量的渐近分布.

**定理 3.2.2** 若定理 3.2.1 的条件和原假设  $H_0$  成立, 当  $n \rightarrow \infty$  时, 有  $\tilde{\ell}(L_n) \rightarrow \chi_q^2$ .

因此,  $L_n \theta$  的  $(1 - \alpha)$  水平置信区域为

$$V_\alpha = \left[ v : -2 \left\{ \ell_p(\theta) - \min_{\theta: L_n \theta = v} \ell_p(\theta) \right\} \leq \chi_{q, 1-\alpha}^2 \right], \quad (3.2.5)$$

其中  $\chi_{q, 1-\alpha}^2$  是  $\chi_q^2$  分布的  $(1 - \alpha)$  分位数. 因此, 由定理 3.2.2 得到, 当  $n \rightarrow \infty$  时, 有  $P(L_n \theta_0 \in V_\alpha) \rightarrow 1 - \alpha$ .

### 3.3 数据模拟和实例分析

本节采用上章相同的算法, 并使用 BIC 准则选择合适的调整参数.

#### 3.3.1 数据模拟

在模拟中  $n$  和  $p$  同时变化, 每次实验重复 1000 次. 比较基于三种不同估计方程, 即:  $Z_i^{(2)}$ ,  $(Z_i^{(1)}, Z_i^{(2)})$  和  $(Z_i^{(1)}, Z_i^{(2)}, Z_i^{(3)})$  以及三种方法 OEL、EL 和 PEL 估计的效果. 计算 ME 中位数:  $\|\hat{\theta} - \theta_0\|_2 = \{(\hat{\theta} - \theta_0)^T(\hat{\theta} - \theta_0)\}^{1/2}$ ,  $\|\hat{\beta} - \beta_0\|_2 = \{(\hat{\beta} - \beta_0)^T(\hat{\beta} - \beta_0)\}^{1/2}$  和  $\|\hat{\sigma}^2 - \sigma_0^2\|_2 = |\hat{\sigma}^2 - \sigma_0^2|$ . T 和 F 的定义和上章相同. 估计量的小脚标 1, 2, 3 分别对应于估计方程  $Z_i^{(2)}$ ,  $(Z_i^{(1)}, Z_i^{(2)})$ ,  $(Z_i^{(1)}, Z_i^{(2)}, Z_i^{(3)})$ .  $D$ ,  $D_1$  和  $D_2$  分别代表  $\theta$ ,  $\beta$  和  $\sigma^2$  的 ME 误差.

考虑如下的广义线性模型

$$Y_i = G(X_i^T \beta) + \sigma V^{1/2} \{G(X_i^T \beta)\} \varepsilon_i.$$

例 3.3.1 取  $G(t) = t^2$ ,  $V(t) = t$ ,  $\varepsilon_i$  是  $[-\sqrt{3}, \sqrt{3}]$  上的均匀分布.  $\sigma = 1.2$ , 参数  $\beta = (3, 1.5, 0, 0, 2, 0, \dots)^T$ .  $X_{ij} \sim U[0, 1]$ ,  $i = 1 \dots n$ ,  $j = 1 \dots p$ , 其中  $X_{ij}$  是  $X_i$  的第  $j$  个分量.

表 3.1 给出  $\hat{\theta}$ ,  $\hat{\beta}$  和  $\hat{\sigma}^2$  的估计误差以及模型选择的结果. 从表 3.1 可以得出, 三种方法估计效果随着估计方程的增多而变好. PEL 的估计误差比 EL 小, 并且 PEL 的效果接近 OEL. 当  $n$  较大时, 三种估计方程下的 PEL 效果都非常接近 OEL. 随着  $n$  的增加, 正确估计零的个数接近  $p - 3$ . 所以, 所提 PEL 方法的模型选择能力是令人满意的, 这也验证了定理 3.2.1 的结论. 同时, 可以看出估计方程越多, 正确估计零的效果越好.

表 3.1 例 3.3.1 中估计误差以及模型选择结果.

$n$	$p$	$L_2$	$OEL_1$	$EL_1$	$PEL_1$	$OEL_2$	$EL_2$	$PEL_2$	$OEL_3$	$EL_3$	$PEL_3$
100	10	$D$	-	-	-	0.343	1.385	0.530	0.338	1.327	0.377
		$D_1$	0.169	1.083	0.244	0.166	1.046	0.215	0.157	1.061	0.174
		$D_2$	-	-	-	0.093	0.232	0.211	0.089	0.151	0.126
		T	-	-	6.801	-	-	6.845	-	-	6.896
		F	-	-	0	-	-	0	-	-	0
200	15	$D$	-	-	-	0.165	0.964	0.257	0.157	0.912	0.180
		$D_1$	0.081	0.795	0.103	0.081	0.780	0.101	0.077	0.777	0.082
		$D_2$	-	-	-	0.041	0.114	0.101	0.040	0.058	0.056
		T	-	-	11.906	-	-	11.906	-	-	11.948
		F	-	-	0	-	-	0	-	-	0
400	20	$D$	-	-	-	0.079	0.600	0.117	0.070	0.562	0.076
		$D_1$	0.040	0.529	0.049	0.039	0.520	0.048	0.036	0.509	0.036
		$D_2$	-	-	-	0.019	0.049	0.045	0.017	0.023	0.021
		T	-	-	16.926	-	-	16.939	-	-	16.979
		F	-	-	0	-	-	0	-	-	0

例 3.3.2 设  $G(t) = e^t$ ,  $V(t) = t^2$ ,  $\varepsilon_i \sim U[-\sqrt{3}, \sqrt{3}]$ ,  $\beta = (1, 2, \dots)^T$ ,  $X_{ij} \sim U[0, 2]$  以及  $\sigma = 0.5$ .

模拟结果在表 3.2 给出. 关于估计方法和估计方程个数对误差的影响可以得出与例 3.3.1 相同的结论.

表 3.2 例 3.3.2 中估计误差以及模型选择结果.

$n$	$p$	$L_2$	$OEL_1$	$EL_1$	$PEL_1$	$OEL_2$	$EL_2$	$PEL_2$	$OEL_3$	$EL_3$	$PEL_3$
100	10	$D$	-	-	-	0.060	0.768	0.082	0.038	0.571	0.055
		$D_1$	0.049	0.756	0.058	0.049	0.753	0.061	0.0275	0.551	0.036
		$T$	-	-	7.505	-	-	7.691	-	-	7.836
		$F$	-	-	0	-	-	0	-	-	0
200	15	$D$	-	-	-	0.030	0.597	0.042	0.018	0.364	0.025
		$D_1$	0.025	0.579	0.028	0.025	0.591	0.031	0.014	0.357	0.016
		$T$	-	-	12.564	-	-	12.703	-	-	12.899
		$F$	-	-	0	-	-	0	-	-	0
400	20	$D$	-	-	-	0.014	0.375	0.019	0.009	0.206	0.011
		$D_1$	0.013	0.384	0.014	0.012	0.372	0.014	0.007	0.203	0.007
		$T$	-	-	17.716	-	-	17.871	-	-	17.951
		$F$	-	-	0	-	-	0	-	-	0

例 3.3.3 取  $G(t) = \sin(t)$ ,  $V(t) = t^2$ ,  $\varepsilon_i \sim N(0, 1)$ ,  $\beta = (1, 1, 0, 0, 1, 0, \dots)^T$  以及  $\sigma = 0.5$ .  $X_i$  和例 3.3.1 中有相同的分布.

表 3.3 例 3.3.3 中 0.05 水平下拒绝原假设  $H_0$  的经验频率.

$p$	$n$	0.8	0.9	1.0(size)	1.1	1.2
6	50	18.6	15	8	15.9	23.7
10	100	40.5	27.3	6.7	31.7	40.1
15	200	76.4	61.6	6.2	67.3	83
20	400	93.7	86	5.9	91.4	96.4
26	800	99.3	95.8	6	97.1	99.3

假设 (3.2.5) 中  $L_n = (0, 1, 0, \dots, 0)$ , 这就构建了关于  $\theta$  的第二个元素  $\theta_2$  的  $1 - \alpha$  水平的置信区域. 表 3.3 给出了在  $\alpha = 0.05$  下不同的  $\theta_2$  值,  $\theta_2 \notin V_\alpha$  的经验频率. 在真值  $\theta_2 = 1$  时, 拒绝原假设的经验频率接近  $\alpha = 0.05$ , 所以随机模拟维持了实际检验水平并且证实了定理 3.2.2. 当  $\theta_2$  和真实值的距离越大, 拒绝的频率越高. 尤其是,

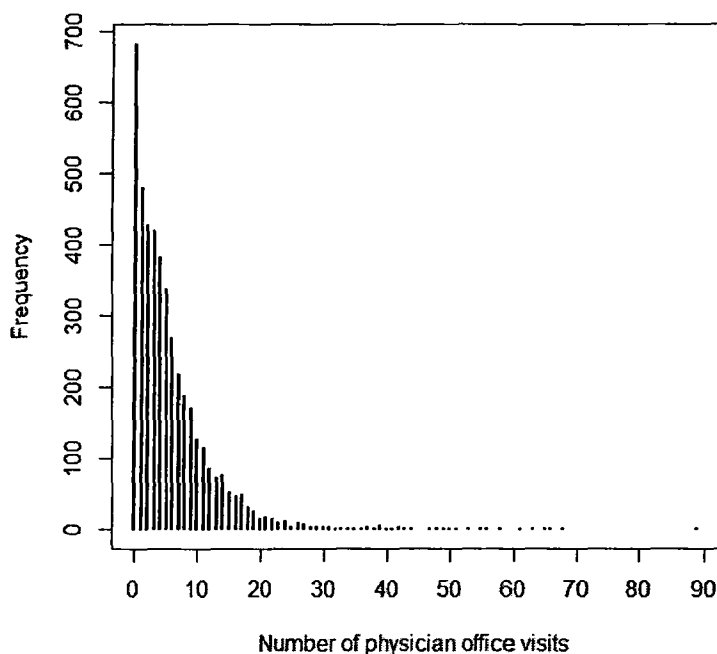


图 3.1 拜访医生办公室次数的频率图.

当  $n$  较大并且  $\theta_2$  和真实值的距离为 0.2 时, 拒绝的频率接近 1. 提出的检验统计量在假设检验中有很好的功效.

### 3.3.2 NMES 数据

本节使用美国医疗支出调查 (NMES) 数据来解释提出的 PEL 方法的有效性. 该数据研究了在 1987 到 1988 年之间, 美国人如何使用和支付医疗服务. 研究对象是年龄大于 66 岁的公民, 共有 4406 个观察者. 这个数据也被研究者用来分析医疗服务的需要<sup>[82,83]</sup>. 建立医生办公室的拜访次数 (OFP) 与下面变量之间的关系: EXCLHLTH (=1, 自我感知身体状况非常好)、POORHLTH (=1, 自我感知身体状况非常不好)、NUMCHRON (慢性疾病的种数)、ADLDIFF (=1, 有日常活动被限制的状况)、NOREAST (=1, 生活在美国的东北部)、MIDWEST (=1, 生活在美国的中西部)、WEST (=1, 住在美国的西部)、AGE (年龄除以10)、BLACK (=1, 非洲裔美国人)、MALE (=1, 男人)、MARRIED (=1, 已婚)、SCHOOL (受教育的年数)、FAMINC (家庭收入除以10000)、EMPLOYED (=1, 在职)、PRIVINS (=1, 有私人的医疗保险)、MEDICAID (=1, 有医疗补助计划). 因变量 OFP 的均值为 5.77, 方差为 45.69, 原始数据呈现出高度的分散性, 散布参数在该模型中有存在的意义. 同时, OFP 变量

中包含很多的零, 图 3.1 是变量 OFP 的频率直方图. 根据文献 [83], 考虑如下的模型:

$$E(Y|X) = G(X^T \beta) \quad , \quad \text{Var}(Y|X) = \sigma^2 G(X^T \beta),$$

其中  $G(t) = e^t$ . 在该模型下, 计算拟泊松似然估计、最优子集估计以及 PEL 估计. 结果在表 3.4 中给出, 其中 PEL 的误差应除以 10. PEL 估计是最简的模型并且接近最优子集方法.

表 3.4 NMES 数据的参数估计与标准误差.

Variable	Likelihood	PEL	Best Subset
OFP	-	-	-
INTERCEPT	1.296(0.228)***	1.277(0.013)	1.229(0.082)
EXCLHLTH	-0.386(0.079)***	-0.388(0.010)	-0.378(0.030)
POORHLTH	0.287(0.047)***	0.284(0.012)	0.286(0.018)
NUMCHRON	0.163(0.011)***	0.184(0.002)	0.164(0.004)
ADLDIFF	0.093(0.043)*	0.081(0.009)	0.094(0.016)
NOREAST	0.107(0.046)*	0.110(0.009)	0.113(0.016)
MIDWEST	-0.010(0.043)	0	0
WEST	0.123(0.047)**	0.121(0.007)	0.129(0.016)
AGE	-0.055(0.028).	-0.078(0.003)	-0.051(0.010)
BLACK	-0.065(0.058)	0	0
MALE	-0.071(0.037).	-0.065(0.009)	-0.084(0.013)
MARRIED	-0.040(0.038)	0	0
SCHOOL	0.025(0.005)***	0.047(0.001)	0.026(0.001)
FAMINC	-0.002(0.006)	0	0
EMPLOYED	0.052(0.057)	0	0
PRIVINS	0.321(0.051)***	0.317(0.001)	0.325(0.019)
MEDICAID	0.288(0.065)***	0.276(0.003)	0.286(0.024)
$\sigma^2$	6.842	6.839	

从结果可知, 慢性疾病种数 (NUMCHRON) 以及自我感知身体状况在模型中是重要的变量. 学历越高的人, 拜访医生办公室的次数越多. 这可能是因为高学历使得他们对待医疗服务更明智. 由于私人保险基本涵盖了理疗、检查等而医疗补助计划却没有的原因, 所以拥有私人医疗保险的人拜访医生办公室的次数越多. 是否拥有

医疗补助计划是拜访医生次数的重要因素,这可能是出于价格的因素<sup>[82]</sup>.一些身体上的缺陷可以通过治疗和康复来改善或推迟,所以该变量对结果有一些影响.男人比女人拜访医生办公室的次数少,一个有趣的事实是,男人出现病症时通常要等很长的时间才去看医生.另外,年龄以及居民居住的地方也是显著的变量.

### 3.4 定理证明

为了方便书写,记  $I_{p+1} = (H_1^T, H_2^T)$ , 其中  $H_1 \in \mathbb{R}^{(d+1) \times (p+1)}$ ,  $H_2 \in \mathbb{R}^{(p-d) \times (p+1)}$ .

**引理 3.4.1** 令  $a_n = (p/n)^{1/2}$ ,  $D_n = \{\theta : \|\theta - \theta_0\| \leq ca_n\}$ ,  $c > 0$ . 如果条件 (C1)-(C4) 成立, 则对于  $\theta \in D_n$ , 有  $\|\lambda_\theta\| = O_p(a_n)$ .

**证明** 当  $\theta \in D_n$ , 记  $\lambda = \rho u$ , 其中  $\|u\| = 1$  是单位向量. 根据文献 [21], 有

$$\rho \left\{ u^T S_\theta u - \max_i \|Z_i(\theta)\| \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta) \right| \right\} \leq \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta) \right|,$$

其中  $S_\theta = n^{-1} \sum_{i=1}^n Z_i(\theta) Z_i^T(\theta)$ .

由于

$$\left\| \frac{1}{n} \sum_{i=1}^n Z_i(\theta_0) \right\| = O_p((p/n)^{1/2}),$$

那么

$$\frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta_0) \right| = O_p((p/n)^{1/2}).$$

由于

$$S_{\theta_0} = \frac{1}{n} \sum_{i=1}^n Z_i(\theta_0) Z_i^T(\theta_0) \xrightarrow{p} \Sigma,$$

所以  $u^T S_{\theta_0} u = O_p(1)$ . 因为,

$$\max_i \|Z_i(\theta_0)\| \leq \max_i \|Z_i^{(1)}(\theta_0)\| + \max_i \|Z_i^{(2)}(\theta_0)\| + \max_i \|Z_i^{(3)}(\theta_0)\|,$$

结合条件 (C1) 有  $\max_i \|Z_i(\theta_0)\| = O_p(n^{1/\alpha})$ . 所以

$$\max_i \|Z_i(\theta_0)\| \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta_0) \right| = O_p(n^{1/\alpha}) O_p((p/n)^{1/2}) = o_p(1),$$

并且  $\|\lambda_{\theta_0}\| = O_p((p/n)^{1/2})$ . 由泰勒展开式, 有

$$\begin{aligned} \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta) \right| &= \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta_0) + \frac{\partial Z_i(\theta_0)}{\partial \theta} (\theta - \theta_0) + o_p(1) \right| \\ &\leq \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta_0) \right| + \max_i \left\| \frac{\partial Z_i(\theta_0)}{\partial \theta} (\theta - \theta_0) \right\| + o_p(1). \end{aligned}$$

由条件 (C1), 有  $\max_i \|\partial Z_i(\theta_0)/\partial \theta^T(\theta - \theta_0)\| = o_p(1)$ . 因此

$$\frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta) \right| = O_p((p/n)^{1/2}).$$

易知

$$\max_i \|Z_i(\theta)\| \leq \max_i \|Z_i(\theta_0)\| + \max_i \|\partial Z_i(\theta_0)/\partial \theta^T(\theta - \theta_0)\|,$$

所以

$$\max_i \|Z_i(\theta)\| \frac{1}{n} \left| \sum_{i=1}^n u^T Z_i(\theta) \right| = o_p(1).$$

由于  $u^T S_\theta u = u^T S_{\theta_0} u + o_p(1) = O_p(1)$ , 引理 3.4.1 得证.

**引理 3.4.2** 当  $n \rightarrow \infty$  时, 以概率趋近于 1 成立,  $\ell_p(\theta)$  在  $D_n$  里有最小值.

证明当  $\theta \in D_n$ , 由 EL 定义, 有

$$Q_{1n}(\theta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{Z_i(\theta)}{[1 + \lambda^T Z_i(\theta)]} = 0.$$

由引理 3.4.1, 当  $\theta \in D_n$  时,  $\lambda^T Z_i(\theta)$  和  $\|\lambda\|$  随机小. 对  $Q_{1n}(\theta, \lambda)$  使用泰勒展开, 有

$$0 = \bar{Z}(\theta) - S_\theta \lambda + r_n,$$

其中  $\bar{Z}(\theta) = \frac{1}{n} \sum_{i=1}^n Z_i(\theta)$ ,  $r_n = \frac{1}{n} \sum_{i=1}^n Z_i(\theta) [\lambda^T Z_i(\theta)]^2 (1 + \xi_i)^{-3}$  是余项且  $|\xi_i| \leq |\lambda^T Z_i(\theta)|$ .

反解上式得到  $\lambda = S_\theta^{-1} [\bar{Z}(\theta) + r_n]$ . 对  $\ell(\theta)$  使用泰勒展开, 有

$$2\ell(\theta, \lambda) = 2 \sum_{i=1}^n \lambda^T Z_i(\theta) - \sum_{i=1}^n [\lambda^T Z_i(\theta)]^2 + \frac{2}{3} \sum_{i=1}^n \frac{[\lambda^T Z_i(\theta)]^3}{(1 + \xi_i)^3},$$

其中  $|\xi_i| \leq |\lambda^T Z_i(\theta)|$ . 将  $\lambda$  带入到  $\ell(\theta)$ , 得到

$$2\ell(\theta) = n \bar{Z}(\theta)^T S_\theta^{-1} \bar{Z}(\theta) - n r_n^T S_\theta^{-1} r_n + \frac{2}{3} \sum_{i=1}^n \frac{[\lambda^T Z_i(\theta)]^3}{(1 + \xi_i)^3}.$$

将分解式子得到

$$2\ell(\theta) = T_0 + T_1 + T_2,$$

其中

$$T_0 = n \bar{Z}(\theta_0)^T S_{\theta_0}^{-1} \bar{Z}(\theta_0),$$

$$T_1 = n \{\bar{Z}(\theta) - \bar{Z}(\theta_0)\}^T S_\theta^{-1} \{\bar{Z}(\theta) - \bar{Z}(\theta_0)\},$$

$$T_2 = n \{\bar{Z}(\theta_0)\}^T [S_\theta^{-1} - S_{\theta_0}^{-1}] \bar{Z}(\theta_0) + 2 \bar{Z}(\theta_0)^T S_\theta^{-1} [\bar{Z}(\theta) - \bar{Z}(\theta_0)]$$

$$- n r_n^T S_\theta^{-1} r_n + \frac{2}{3} \sum_{i=1}^n [\lambda^T Z_i(\theta)]^3 (1 + \xi_i)^{-3}.$$

当  $n \rightarrow \infty$  时,

$$\frac{1}{n} \sum_{i=1}^n [Z_i(\theta) - Z_i(\theta_0)] \xrightarrow{p} \Omega(\theta - \theta_0).$$

由于  $\|\Omega\| < +\infty$ , 所以

$$\begin{aligned} T_1 &> n \|\bar{Z}(\theta) - \bar{Z}(\theta_0)\|_{r_{p+r+1}}^{-2}(\Sigma) = n \|\Omega\|^2 \|\theta - \theta_0\|_{r_{p+r+1}}^{-2}(\Sigma) \\ &= O_p(na_n^2). \end{aligned}$$

由  $T_2/T_1 \rightarrow 0$  且  $2\ell(\theta_0) - T_0 \rightarrow 0$ ,  $2\ell(\theta) - 2\ell(\theta_0)$  被正项控制住. 因为  $na_n^2 \rightarrow \infty$ , 所以对于任意给定  $C > 0$ ,

$$P\{2\ell(\theta) - 2\ell(\theta_0) > C\} \rightarrow 1.$$

假设  $\beta = \beta_0 + a_n e$ ,  $\|e\| \leq C$  且  $M = \min_{j \notin \mathcal{A}} |\beta_{j0}|$ ,

$$\begin{aligned} \ell_p(\theta) - \ell_p(\theta_0) &= \ell(\theta) - \ell(\theta_0) + n\tau \sum_{i=1}^p (|\beta_i| - |\beta_{i0}|) / |\tilde{\beta}_i| \\ &> \ell(\theta) - \ell(\theta_0) - n\tau \sum_{i=1}^d (a_n |e_i|) / |\tilde{\beta}_i|. \end{aligned}$$

由泰勒展式,

$$\frac{1}{|\tilde{\beta}_i|} = \frac{1}{|\beta_{i0}|} - \frac{\text{sgn}(\beta_{i0})}{|\beta_{i0}|^2} (\tilde{\beta}_i - \beta_{i0}) + o_p(|\tilde{\beta}_i - \beta_{i0}|) = \frac{1}{|\beta_{i0}|} + O_p(a_n).$$

在条件 (C5) 下, 对于充分大的  $n$ ,

$$n\tau \sum_{i=1}^d \frac{a_n |e_i|}{|\tilde{\beta}_i|} < \frac{n\tau d a_n C}{M} + n\tau d O_p(a_n^2) = o_p(1).$$

所以, 对于  $\theta \in D_n$  有  $\ell_p(\theta) > \ell_p(\theta_0)$ , 引理 3.4.2 得证.

定理 3.2.1 的证明 由引理 3.4.1 可知, 当  $\theta \in D_n$  时,  $\lambda^T Z_i(\theta) = o_p(1)$ . 所以

$$\begin{aligned} \frac{1}{n} \frac{\partial \ell_p(\theta)}{\partial \beta_j} &= n^{-1} \sum_{i=1}^n \lambda^T \frac{\partial Z_i(\theta)}{\partial \beta_j} (1 + o_p(1)) + \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} \\ &= \lambda^T \frac{1}{n} \sum_{i=1}^n \left[ \frac{\partial Z_i(\theta_0)}{\partial \beta_j} + \frac{\partial^2 Z_i(\theta_0)}{\partial \beta_j \partial \theta} (\theta - \theta_0) \right] (1 + o_p(1)) + \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} \\ &= I_j + II_j + III_j + \text{small order terms.} \end{aligned}$$

由于

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial Z_i(\theta_0)}{\partial \beta_j} \xrightarrow{p} \Omega_{j+1},$$

其中  $\Omega_{j+1}$  是  $\Omega$  的第  $(j+1)$  列元素, 所以

$$\max_{j \notin \mathcal{A}} I_j = \max_{j \notin \mathcal{A}} \lambda^T \frac{1}{n} \sum_{i=1}^n \frac{\partial Z_i(\theta_0)}{\partial \beta_j} = O_p(a_n).$$



在条件 (C1) 下,

$$\max_{j \notin \mathcal{A}} II_j = \max_{j \notin \mathcal{A}} \lambda^T \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 Z_i(\theta_0)}{\partial \beta_j \partial \theta} (\theta - \theta_0) = O_p(a_n^2) = O_p(a_n).$$

在条件 (C6) 下,

$$\max_{j \notin \mathcal{A}} III_j = \max_{j \notin \mathcal{A}} \tau \frac{\text{sgn}(\beta_j)}{|\tilde{\beta}_j|} = O_p(\tau/a_n) \text{sgn}(\beta_j).$$

$\max_{j \notin \mathcal{A}} III_j$  一致性的控制了  $\partial \ell_p(\theta)/\partial \beta_j$  的符号. 所以, 对于所有的  $j \notin \mathcal{A}$ , 以概率趋近于 1, 有

$$\frac{\partial \ell_p(\theta)}{\partial \beta_j} > 0, \quad \beta_j \in (0, ca_n e_j); \quad \frac{\partial \ell_p(\theta)}{\partial \beta_j} < 0, \quad \beta_j \in (-ca_n e_j, 0).$$

由文献 [81] 可知, 对于所有的  $j \notin \mathcal{A}$ , 有  $\hat{\beta}_j = 0$ , 这证明了第一部分.

接下来证明第二部分估计的有效性. 由第一部分的结论以及 PEL 的定义, PEL 估计量  $\hat{\theta}$  是满足  $H_2 \theta = 0$  下式子 (3.2.4) 的最小值. 根据文献 [20], 由拉格朗日算法, 估计量是最小化下面的目标函数

$$\tilde{\ell}(\theta, \lambda, \nu) = n^{-1} \sum_{i=1}^n \log\{1 + \lambda^T Z_i(\theta)\} + \sum_{j=1}^p p_\tau(|\beta_j|) + \nu^T H_2 \theta,$$

其中  $\nu \in \mathbb{R}^{p-d}$  是朗格朗日算子.

记

$$\tilde{Q}_{1n}(\theta, \lambda, \nu) = \frac{1}{n} \sum_{i=1}^n \frac{Z_i(\theta)}{[1 + \lambda^T Z_i(\theta)]},$$

$$\tilde{Q}_{2n}(\theta, \lambda, \nu) = \left( \frac{1}{n} \left\{ \sum_{i=1}^n \frac{[\partial Z_i(\theta)^T / \partial \sigma^2] \lambda}{[1 + \lambda^T Z_i(\theta)]} \right\}^T, \frac{1}{n} \left\{ \sum_{i=1}^n \frac{[\partial Z_i(\theta)^T / \partial \beta] \lambda}{[1 + \lambda^T Z_i(\theta)]} + b(\beta) \right\}^T \right)^T + H_2^T \nu,$$

以及

$$\tilde{Q}_{3n}(\theta, \lambda, \nu) = H_2 \theta,$$

其中  $b(\beta) = (\tau \text{sgn}(\beta_1)/|\tilde{\beta}_1|, \dots, \tau \text{sgn}(\beta_p)/|\tilde{\beta}_p|)^T$ .

估计量  $(\hat{\theta}, \hat{\lambda}, \hat{\nu})$  满足  $\tilde{Q}_{jn}(\hat{\theta}, \hat{\lambda}, \hat{\nu}) = 0 (j = 1, 2, 3)$ . 由引理 3.4.1 得到  $\|\hat{\lambda}\| = O_p(a_n)$  随机小. 类似于文献 [20], 由  $\tilde{Q}_{2n}(\hat{\theta}, \hat{\lambda}, \hat{\nu}) = 0$  得到  $\|\hat{\nu}\| = O_p(a_n)$ . 所以, 将  $\tilde{Q}_{jn}$  在  $(\theta_0, 0, 0)$  处展开 ( $j = 1, 2, 3$ ). 由展开式直接得到

$$\begin{pmatrix} -\tilde{Q}_{1n}(\theta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\Sigma & \Omega & 0 \\ \Omega^T & 0 & H_2^T \\ 0 & H_2 & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\theta} - \theta_0 \\ \hat{\nu} \end{pmatrix} + R_n^{(1)} + R_n^{(2)} + R_n^{(3)} + R_n^{(4)}, \quad (3.4.1)$$

其中  $R_n^{(1)} = (R_{1n}^{(1)T}, R_{2n}^{(1)T}, 0)^T$ ,  $R_{1n}^{(1)} \in \mathbb{R}^{r+p+1}$ ,  $R_{2n}^{(1)} \in \mathbb{R}^{p+1}$ .  $R_{jn}^{(1)}$  ( $j = 1, 2$ ) 的第  $k$  个元素为

$$R_{jn,k}^{(1)} = \frac{1}{2}(\hat{\eta} - \eta_0)^T \frac{\partial^2 \tilde{Q}_{jn,k}(\eta^*)}{\partial \eta \partial \eta^T} (\hat{\eta} - \eta_0),$$

$\eta = (\theta^T, \lambda^T)^T$ ,  $\eta^* = (\theta^{*T}, \lambda^{*T})^T$  满足  $\|\theta^* - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$  且  $\|\lambda^*\| \leq \|\hat{\lambda}\|$ , 所以  $\|R_n^{(1)}\| = o_p(n^{-1/2})$ . 另外三项分别为  $R_n^{(2)} = \{0, \tilde{b}(\beta)^T, 0\}^T$ , 其中  $\tilde{b}(\beta_0) = (0, b(\beta_0)^T)^T \in \mathbb{R}^{p+1}$ ,  $R_n^{(3)} = \{[\Sigma - \frac{1}{n} \sum_{i=1}^n Z_i(\theta_0) Z_i^T(\theta_0)] \hat{\lambda}\}^T + \{(\frac{1}{n} \sum_{i=1}^n \partial Z_i(\theta_0) / \partial \theta^T - \Omega)(\hat{\theta} - \theta_0)\}^T, 0, 0\}^T$  以及  $R_n^{(4)} = \{0, \{(\frac{1}{n} \sum_{i=1}^n \partial Z_i(\theta_0)^T / \partial \theta - \Omega^T) \hat{\lambda}\}^T, 0\}^T$ . 由条件 (C6), 有  $\|R_n^{(2)}\| = o_p(n^{-1/2})$  且  $\|R_n^{(3)}\| = \|R_n^{(4)}\| = o_p(n^{-1/2})$ . 因此, 有  $\|R_n\| = o_p(n^{-1/2})$ .

定义  $K_{11} = -\Sigma$ ,  $K_{12} = (\Omega, 0)$  且  $K_{21} = K_{12}^T$ , 其中

$$K_{22} = \begin{pmatrix} 0 & H_2^T \\ H_2 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}.$$

假设  $\vartheta = (\theta^T, v^T)^T$ , 反解 (3.4.1), 得到

$$\begin{pmatrix} \hat{\lambda} - 0 \\ \hat{\vartheta} - \vartheta_0 \end{pmatrix} = K^{-1} \left\{ \begin{pmatrix} -\tilde{Q}_{1n}(\theta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} + R_n \right\}.$$

由分块矩阵的逆运算, 得到

$$K^{-1} = \begin{pmatrix} K_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} -K_{11}^{-1} K_{12} \\ I \end{pmatrix} A^{-1} \begin{pmatrix} -K_{21} K_{11}^{-1} & I \end{pmatrix},$$

其中  $A = K_{22} - K_{21} K_{11}^{-1} K_{12}$ . 所以  $\hat{\vartheta} - \vartheta_0 = A^{-1} K_{21} K_{11}^{-1} \tilde{Q}_{1n}(\theta_0, 0, 0) + o_p(n^{-1/2})$ . 其中逆矩阵

$$A^{-1} = \begin{pmatrix} V - V H_2^T (H_2 V H_2^T)^{-1} H_2^T V & V H_2^T (H_2 V H_2^T)^{-1} \\ (H_2 V H_2^T)^{-1} H_2 V & -(H_2 V H_2^T)^{-1} \end{pmatrix}.$$

所以

$$\hat{\theta} - \theta_0 = -\{V - V H_2^T (H_2 V H_2^T)^{-1} H_2 V\} \Omega^T \Sigma^{-1} \tilde{Q}_{1n}(\theta_0, 0, 0) + R_{1n},$$

其中  $R_{1n}$  是  $K^{-1} R_n$  的对应元素并且  $\|R_{1n}\| = o_p(n^{-1/2})$ . 显然, 对于非零元素  $\theta_1$  的估计  $\hat{\theta}_1$  有

$$\hat{\theta}_1 - \theta_{10} = -\{H_1 V - H_1 V H_2^T (H_2 V H_2^T)^{-1} H_2 V\} \Omega^T \Sigma^{-1} \tilde{Q}_{1n}(\theta_0, 0, 0) + o_p(n^{-1/2}).$$

假设  $Z_{ni} = -W_n B^{-1/2} \{H_1 V - H_1 V H_2^T (H_2 V H_2^T)^{-1} H_2 V\} \Omega^T \Sigma^{-1} Z_i(\theta_0)$ ,  $B = V_{11} - V_{12} V_{22}^{-1} V_{21}$ ,  $Y_{ni} = n^{-1/2} Z_{ni}$ . 接下来, 证明 Lindeberg-Feller 条件成立. 易知,  $E(Z_{ni}) = 0$ ,  $\text{Var}(Z_{ni}) =$

$W_n W_n^T$ ,  $E(\|Y_{ni}\|^4) = O(1/n^2)$ . 对  $\forall \varepsilon > 0$

$$P(\|Y_{ni}\| > \varepsilon) \leq \frac{1}{n\varepsilon^2} E(\|Z_{ni}\|^2) = \frac{\text{trVar}(Z_{ni})}{n\varepsilon^2} = O(1/n).$$

因此,

$$\sum_{i=1}^n E(\|Y_{ni}\|^2) I(\|Y_{ni}\| > \varepsilon) \leq n \{E(\|Y_{n1}\|^4)\}^{1/2} \{P(\|Y_{n1}\| > \varepsilon)\}^{1/2} \xrightarrow{p} 0.$$

由于

$$\sum_{i=1}^n \text{Var}(Y_{ni}) = W_n W_n^T \xrightarrow{p} G,$$

所以

$$n^{1/2} W_n B^{-1/2} (\hat{\theta}_1 - \theta_{10}) \longrightarrow N(0, G).$$

最后, 由于  $\|n^{1/2} W_n B^{-1/2} R_{1n}\|^2 = o_p(1)$ , 证明完成.

定理 3.2.2 的证明首先给出  $\ell(\hat{\theta})$  的渐近展开式, 其中  $\hat{\theta}$  是  $\ell(\theta)$  的最小值. 由引理 3.4.1 得到  $\max_i |\hat{\lambda}^T Z_i(\hat{\theta})| = o_p(1)$ . 令  $h_i = \hat{\lambda}^T Z_i(\hat{\theta})$ , 由泰勒展开式得到

$$\ell(\hat{\theta}) = \sum_{i=1}^n h_i - \sum_{i=1}^n \frac{h_i^2}{2} + \sum_{i=1}^n \frac{h_i^3}{\{3(1 + \xi_i)^3\}},$$

其中  $|\xi_i| < |\hat{\lambda}^T Z_i(\hat{\theta})|$ . 在引理 3.4.2 中, 对于  $\theta \in D_n$  有  $\lambda = S_\theta^{-1} [\bar{Z}(\theta) + r_n]$ , 其中  $r_n = n^{-1} \sum_{i=1}^n \{Z_i(\theta) [\lambda^T Z_i(\theta)]^2 (1 + \xi_i)^{-3}\}$  且  $|\xi_i| \leq |\lambda^T Z_i(\theta)|$ . 将  $\hat{\lambda}$  带入到  $h_i$  得到

$$2\ell(\hat{\theta}) = n \bar{Z}(\theta_0)^T \Sigma^{-1} \Omega V H_2^T (H_2 V H_2^T)^{-1} H_2 V \Omega^T \Sigma^{-1} \bar{Z}(\theta_0) + o_p(1). \quad (3.4.2)$$

由于  $L_n L_n^T = I_q$ , 所以存在  $\tilde{H}$  满足  $\tilde{H}_2 \theta = 0$  且  $\tilde{H}_2 \tilde{H}_2^T = I_{p-d+q}$ . 重复定理 3.2.1 的证明, 在原假设下,  $\theta$  可以通过最小化下式得到

$$\tilde{\ell}_p(\theta, \lambda, \nu) = \sum_{i=1}^n \log\{1 + \lambda^T Z_i(\theta)\} + n \sum_{j=1}^p p_\tau(|\beta_j|) + \nu^T \tilde{H}_2 \theta. \quad (3.4.3)$$

记 (3.4.3) 的最小值为  $(\check{\theta}, \check{\lambda}, \check{\nu})$ , 由定理 3.2.1 的第一部分可知, 以概率趋近于 1 有  $\check{\theta}_2 = 0$ . 类似于式子 (3.4.2), 有

$$2\ell(\theta)|_{L_n \theta = 0} = n \bar{Z}(\theta_0)^T \Sigma^{-1} \Omega V \tilde{H}_2^T (\tilde{H}_2 V \tilde{H}_2^T)^{-1} \tilde{H}_2 V \Omega^T \Sigma^{-1} \bar{Z}(\theta_0) + o_p(1) \quad (3.4.4)$$

由方程 (3.4.2) 和 (3.4.4), 得到

$$\tilde{\ell}(L_n) = n \bar{Z}(\theta_0)^T \Sigma^{-1/2} \{P_1 - P_2\} \Sigma^{-1/2} \bar{Z}(\theta_0) + o_p(1),$$

其中

$$P_1 = \Sigma^{-1/2} \Omega V \tilde{H}_2^T (\tilde{H}_2 V \tilde{H}_2^T)^{-1} \tilde{H}_2 V \Omega^T \Sigma^{-1/2},$$

$$P_2 = \Sigma^{-1/2} \Omega V H_2^T (H_2 V H_2^T)^{-1} H_2 V \Omega^T \Sigma^{-1/2}.$$

因为  $P_1 - P_2$  是秩为  $q$  的幂等矩阵, 所以  $P_1 - P_2 = \Xi_n^T \Xi_n$ , 其中  $\Xi_n$  是  $q \times (p+r+1)$  的矩阵且  $\Xi_n \Xi_n^T = I_q$ . 进一步, 容易得到

$$n^{1/2} \Xi_n \Sigma^{-1/2} \bar{Z}(\theta_0) \longrightarrow N(0, I_q).$$

所以

$$n \bar{Z}(\theta_0)^T \Sigma^{-1/2} (P_1 - P_2) \Sigma^{-1/2} \bar{Z}(\theta_0)^T \longrightarrow \chi_q^2,$$

定理 3.2.2 得证.



## 总结与展望

本文利用 PEL 方法研究了部分线性 EV 模型和高维广义线性模型的参数估计与变量选择问题. 基于该方法, 主要有以下的结论:

对于部分线性 EV 模型, 证明了提出的方法有相合性和 Oracle 性质并且检验统计量收敛到卡方分布. 虽然在理论部分只证明了自适应 Lasso 这一种特殊的惩罚函数, 但在其他惩罚函数下, 如 SCAD 和桥估计等, PEL 仍有相同的性质. 从模拟结果可以看到, SCAD 以及  $L_q$  估计也有很好的表现. 本文结果拓展了测量误差的存在范围, 推广了模型理论及其应用.

对于高维广义线性模型的情形, 提出的方法同样具有 Oracle 性质并且检验统计量收敛到卡方分布. 本文采用的三个估计方程提高了方差函数所包含信息的利用度, 它较两个方程的估计有小的模型误差和好的变量选择能力. 这说明所提方法是有效的. 另外, 本文同时对  $\beta$  和散布参数  $\sigma^2$  进行了估计, 拓宽了现有文献中少有研究散布参数的情形.

本文仍然存在需要进一步研究和改进的地方. 如, 第二章是在固定维数下进行研究. 但在实际问题中, 数据常常以高维的形式出现. 这就使得对高维下带有测量误差的部分线性模型进行研究成为必要. 甚至  $p \gg n$  的情形<sup>[64,84]</sup>, 如生物统计中出现的微列阵基因表达数据. 同样地, 在第三章中, 广义线性模型中也可能有测量误差. 这些都是以后进一步研究的方向.



## 参考文献

- [1] Engle R. F, Granger C. W. J, Rice J, et al. Semiparametric estimates of the relation between weather and electricity sales[J]. *J. Amer. Statist. Assoc.*, 1986, 81: 310–320.
- [2] Härdle W, Liang H, Gao J. T. Partially Linear Models[M]. Heidelberg: Springer Physica-Verlag, 2000.
- [3] Ruppert D, Wand M. P, Carroll R. J. Semiparametric Regression[M]. Cambridge: Cambridge University Press, 2003.
- [4] Nelder J. A, Wedderburn R. W. M. Generalized linear models[J]. *J. R. Statist. Soc. A*, 1972, 135: 370-384.
- [5] Fahrmeir L, Kaufmann H. Correction: Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models[J]. *Ann. Statist.*, 1985, 13(1): 1643-1643.
- [6] Qian G, Wu Y. H. Strong limit theorems on model selection in generalized linear regression with binomial responses[J]. *Statist. Sinica*, 2006,16: 1335-1365.
- [7] Wedderburn R. W. M. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method[J]. *Biometrika*, 1974, 61: 439-447.
- [8] McCullagh P, Nelder J. A. Generalized Linear Models[M]. New York: Chapman and Hall, 1989.
- [9] Cui H. J, Chen S. X. Empirical likelihood confidence regions for parameters in error-in-variables models[J]. *J. Multivar. Anal.*, 2002, 84: 101-115.
- [10] Wang Q, Rao J. N. Empirical likelihood-based inference in linear errors-in-covariables models with validation data[J]. *Biometrika*, 2002, 89(2): 345-358.
- [11] Fuller W. A. Measurement error models[M]. 2nd ed. New York: John Wiley, 2006.
- [12] Carroll R. J, Ruppert D, Stefanski L. A, et al. Measurement error in nonlinear models[M]. 2nd ed. London: Chapman and Hall, 2006.
- [13] Stefanski L. A, Carroll R. J. Deconvolving kernel density estimators[J]. *Statistics*, 1990, 21(2): 169-184.
- [14] Fan J. On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems[J]. *Ann. Statist.*, 1991, 19(3): 1257-1272.



- [15] Fan J, Truong Y. K. Nonparametric Regression with Errors in Variables[J]. *Ann. Statist.*, 1993, 21(4): 1900-1925.
- [16] Fan J, Han F, Liu H, et al. Challenges of Big Data Analysis.[J]. *National Science Review*, 2014, 1(2): 293-314.
- [17] Bühlmann P, van de Geer S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*[M]. Heidelberg: Springer, 2011.
- [18] Owen A. B. Empirical likelihood ratio confidence intervals for a single function[J]. *Biometrika*, 1988, 75: 237-249.
- [19] Diccio T. J, Hall P, Romano J. P. Empirical likelihood is Bartlett-correctable[J]. *Ann. Statist.*, 1991, 19: 1053-1061.
- [20] Qin J, Lawless J. Empirical likelihood and generalized estimating equations[J]. *Ann. Statist.*, 1994, 22: 300-325.
- [21] Owen A. B. *Empirical Likelihood*[M]. New York: Chapman and Hall-CRC, 2001.
- [22] Chen S. X, Cui H. J. On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika*, 2006, 16: 1101-1115.
- [23] Chen S. X. Empirical Likelihood Confidence Intervals for Linear Regression Coefficients[J]. *J. Multivariate Anal.*, 1994, 49(1): 24-40.
- [24] Kolaczyk E. D. Empirical likelihood for generalized linear model[J]. *Statist. Sinica*, 1994, 4: 199-218.
- [25] Chen S. X. Empirical likelihood confidence intervals for nonparametric density estimation[J]. *Biometrika*, 1996, 83(2): 329-341.
- [26] Shi J, Lau T. Empirical Likelihood for Partially Linear Models[J]. *J. Multivar Anal.*, 2000, 72(1): 132-148.
- [27] You J, Zhou Y. Empirical likelihood for semiparametric varying-coefficient partially linear regression models[J]. *Statist. Probab. Lett.*, 2006, 76(4): 412-422.
- [28] Frank L. E, Friedman J. H. A Statistical View of Some Chemometrics Regression Tools[J]. *Technometrics*, 1993, 35(2): 109-135.
- [29] Tibshirani R. Regression shrinkage and selection via the lasso[J]. *J. R. Statist. Soc. Ser. B*, 1996, 58: 267-288.
- [30] Antoniadis A. Wavelets in statistics: a review[J]. *J. I. Statist. Assoc.*, 1997, 6: 97-144.
- [31] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. *J. Amer. Statist. Assoc.*, 2001, 96: 1348-1360.

- [32] Zou H. The Adaptive Lasso and its oracle properties[J]. *J. Amer. Statist. Assoc.*, 2006, 101: 1418-1429.
- [33] Fan J, Lv J. A selective overview of variable selection in high dimensional feature space[J]. *Statist. Sinica*, 2010, 20: 101 - 148.
- [34] Otsu T. Penalized empirical likelihood estimation of semiparametric models[J]. *J. Multivar. Anal.*, 2007, 98: 1923-1954.
- [35] Tang C. Y, Leng C. L. Penalized high-dimensional empirical likelihood[J]. *Biometrika*, 2010, 97: 905-920.
- [36] Leng C. L, Tang C. Y. Penalized empirical likelihood and growing dimensional general estimating equations[J]. *Biometrika*, 2012, 99: 703-716.
- [37] Lahiri S. N, Mukhopadhyay S. A penalized empirical likelihood method in high dimensions[J]. *Ann. Statist.*, 2013, 40: 2511-2540.
- [38] Fan G, Liang H, Shen Y. Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors[J]. *J. Multivar. Anal.*, 2016, 147: 183-201
- [39] Fang J, Liu W, Lu X. Penalized empirical likelihood for semiparametric models with a diverging number of parameters[J]. *J. Statist. Plan. Inference*, 2017, 186: 42-57.
- [40] Wang S. S, Xiang L. M. Penalized empirical likelihood inference for sparse additive hazards regression with a diverging number of covariates[J]. *Stat. Comput.*, 2017, 27: 1347-1364.
- [41] Schwarz G. Estimating the Dimension of a Model[J]. *Ann. Statist.*, 1978, 6(2): 461-464.
- [42] Konishi S, Kitagawa G. Generalised information criterion in model selection[J]. *Biometrika*, 1996, 83:875-890.
- [43] Foster D. P, George E. I. The Risk Inflation Criterion for Multiple Regression[J]. *Ann. Statist.*, 1994, 22(4): 1947-1975.
- [44] Wang H, Li R, Tsai C, et al. Tuning parameter selectors for the smoothly clipped absolute deviation method[J]. *Biometrika*, 2007, 94(3): 553-568.
- [45] Wang H, Li B, Leng C, et al. Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters[J]. *J. R. Statist. Soc. B*, 2009, 71(3): 671-683.
- [46] Härdle D. R, Li R. Variable selection using MM algorithms[J]. *Ann. Statist.*, 2005, 33: 1617-1642.
- [47] Efron B, Hastie T, Johnstone I. M, et al. Least angle regression[J]. *Ann. Statist.*, 2004, 32(2): 407-499.

- [48] Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models[J]. *Ann. Statist.*, 2008, 36(4): 1509-1533.
- [49] James G. M, Radchenko P, Lv J, et al. DASSO: connections between the Dantzig selector and lasso[J]. *J. R. Statist. Soc. B*, 2009, 71(1): 127-142.
- [50] Zhang C. Nearly unbiased variable selection under minimax concave penalty[J]. *Ann. Statist.*, 2010, 38(2): 894-942.
- [51] Cui H. J, Li R. C. On parameter estimation for semi-linear errors-in-variables models[J]. *J. Multivar. Anal.*, 1998, 64: 1-24.
- [52] Liang H, Härdle W, Carroll R. J. Estimation in a semiparametric partially linear errors-in-variables model[J]. *Ann. Statist.*, 1999, 27: 1519-1535.
- [53] He X, Liang H. Quantile regression estimates for a class of linear and partially linear errors-in-variables models[J]. *Statist. Sinica*, 2000, 10: 129-140.
- [54] Liang H. Asymptotic normality of parametric part in partially linear models with measurement error in the nonparametric part[J]. *J. Statist. Plann. Inference*, 2000, 86: 51-62.
- [55] Zhu L. X, Cui H. J. A semi-parametric regression model with errors in variables[J]. *Scand. J. Stat.*, 2003, 30: 429-442.
- [56] Liang H, Wang S, Carroll R. J. Partially linear models with missing response variables and error-prone covariates[J]. *Biometrika*, 2007, 94: 185-198.
- [57] Jin J, Tong X. W. Corrected-loss estimation for error-in-variable partially linear model[J]. *Sci. China. Math.*, 2015, 58: 1101-1114.
- [58] Cui H. J, Kong E. F. Empirical likelihood confidence region for parameters in semi-linear errors-in-variables models[J]. *Scand. J. Statist.*, 2006, 33: 153-168.
- [59] Xu H. X, Fan G. L, Chen Z. L. Hypothesis tests in partial linear errors-in-variables models with missing response[J]. *Statist. Probab. Lett.*, 2017, 126: 219-229.
- [60] Chen X, Cui H. J. Empirical likelihood inference for parameters in a partially linear errors-in-variables model[J]. *Statistics*, 2012, 46: 745-757.
- [61] Koul H. L, Song W. Model checking in partial linear regression models with Berkson measurement errors[J]. *Statist. Sinica*, 2010, 20: 1551-1579.
- [62] Yan L, Chen X. Empirical likelihood for partly linear models with errors in all variables[J]. *J. Multivar. Anal.*, 2014, 130: 275-288.
- [63] Candès E. J, Tao T. The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ [J]. *Ann. Statist.*, 2007, 35(6): 2313-2351.

- [64] Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space[J]. *J. R. Statist. S. B*, 2008, 70(5): 849-911.
- [65] Ni X, Zhang H. H, Zhang D, et al. Automatic model selection for partially linear models[J]. *J. Multivar. Anal*, 2009, 100(9): 2100-2111.
- [66] Liang H, Li R. Variable selection for partially linear models with measurement errors[J]. *J. Amer. Statist. Assoc*, 2009, 104: 234-248.
- [67] Xie H, Huang J. SCAD-penalized regression in high-dimensional partially linear models[J]. *Ann. Statist*, 2009, 37(2): 673-696.
- [68] Zhang H, Lu W. Adaptive Lasso for Cox's proportional hazards model[J]. *Biometrika*, 2007, 94: 691-703.
- [69] Ren Y. W, Zhang X. S. Variable selection using penalized empirical likelihood[J]. *Sci. China. Math*, 2011, 54: 1829-1845.
- [70] Xu Z. B, Zhang H, Wang Y, et al.  $L_{1/2}$  regularization[J]. *Sci. China. Inf. Sci*, 2010, 53: 1159-1169.
- [71] Nierenberg D. W, Stukel T. A, Baron J. A, et al. Determinants of plasma levels of beta-carotene and retinol[J]. *Am. J. Epidemiol*, 1989, 130: 511-521.
- [72] Kai B, Li R, Zou H. New efficient estimation and variable selection methods semiparametric varying-coefficient partially linear models[J]. *Ann. Statist*, 2011, 39: 305-332.
- [73] Guo J, Tang M, Tian M, et al. Variable selection in high-dimensional partially linear additive models for composite quantile regression[J]. *Comput. Statist. Data. Anal*, 2013, 65: 56-67.
- [74] He H, Tang W, Zuo G. Statistical inference in the partial linear models with the double smoothing local linear regression method[J]. *J. Statist. Plann. Inference*, 2014, 146: 102-112.
- [75] Müller M. Estimation and testing in generalized partial linear models—a comparative study[J]. *Statist. Comput*, 2001, 11: 299-309.
- [76] Yan L, Chen X. Empirical likelihood for generalized linear models with fixed and adaptive designs[J]. *Statistics*, 2015, 49(5): 978-988.
- [77] Chen S. X, Peng L, Qin Y, et al. Effects of data dimension on empirical likelihood[J]. *Biometrika*, 2009, 96(3): 711-722.
- [78] Ibrahim J. G, Chen M, Lipsitz S R, et al. Missing-Data Methods for Generalized Linear Models: A Comparative Review[J]. *J. Amer. Statist. Assoc*, 2005, 100(469): 332-346

- [79] Liang F, Song Q, Yu K, et al. Bayesian Subset Modeling for High-Dimensional Generalized Linear Models[J]. J. Amer. Statist. Assoc, 2013, 108(502): 589-606.
- [80] Chen S. X, Cui H. J. An extended empirical likelihood for generalized linear models. Statist. Sinica, 2003, 13: 69-81.
- [81] Fan J, Peng H. Nonconcave penalized likelihood with a diverging number of parameters[J]. Ann. Statist, 2004, 32(3): 928-961.
- [82] Deb P, Trivedi P. K. Demand for medical care by the elderly: a finite mixture approach[J]. J. Appl. Econ, 1997, 12(3): 313-336.
- [83] Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R[J]. J. Statist. Sof, 2008, 27(1): 1-25.
- [84] Candès E. J, Tao T. The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ [J]. Ann. Statist, 2005, 35(6): 2313-2351.

## 致 谢

论文是在陕西师范大学数学与信息科学学院陈夏老师的指导下完成的. 在论文的选题, 撰写, 修改过程中, 老师给了耐心细致的指导.

首先, 特别感谢我的硕士研究生导师陈夏老师. 在科研上, 陈夏老师有扎实的专业知识, 严谨的治学态度, 一丝不苟的科研精神. 他督促我们要以踏实稳重的态度对待知识, 以严密谨慎的心态对待科研. 陈老师是我科研路上的指引者, 是学习的榜样! 在生活上, 陈老师关爱学生, 让异地求学的我们感到非常的温馨. 同时, 陈老师严于律己, 宽于待人的态度给了我很多生活的感触. 非常感谢这位博学宽容谦虚的导师, 是他的帮助让我在研究生期间能够顺利的完成学业, 并且他的以身作则是我以后工作学习的榜样. 在此, 学生对您表示由衷的感谢!

其次, 我要感谢研究生期间的学科老师: 唐三一老师, 宁丽娟老师, 王霞老师, 闫莉老师等. 是老师们的无私奉献给我们呈现了知识的广阔, 是老师们的耐心指导帮我们拨开了迷雾. 在此, 诚挚的说一声: 老师们辛苦了!

再次, 感谢我的家人朋友对我的关心, 支持以及鼓励. 感谢我的同门师姐, 崔小花学姐以及崔艳学姐, 感谢我的同窗好友谭晓燕, 何军伟, 白璐. 当我学习生活上遇到困难时, 是你们毫不吝啬的伸出援助之手, 让我研究生生活增添了几分色彩!

最后, 向参与论文评阅, 答辩的老师以及专家致以最诚挚的感谢!

毛沥悦

2018年5月



## 攻读硕士学位期间的科研成果

### 一、完成的论文

- [1] 毛沥悦, 陈夏, 闫莉. 部分线性 EV 模型的惩罚经验似然[J]. 陕西师范大学学报(自然科学版), 2018, 接收.
- [2] Chen X, Mao L.Y. Penalized empirical likelihood for high-dimensional generalized linear models [J]. Statistica Sinica, 2017, under review.
- [3] Mao L.Y, Chen X. Penalized empirical likelihood for partially linear errors-in-variables models[J]. Metrika, 2018, under review.