

分类号: O221.6, F224.9

学校代码: 10363

密 级: 公开

学 号: 2170810103



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于高维半参数回归模型的惩罚
经验似然估计及其应用研究

论文作者 吕升日

指导教师 何帮强

学科(专业) 应用数学

研究方向 经济统计

论文提交日期: 2020 年 6 月 20 日

分类号: (O221.6, F224.9)

密 级: 公开

单位代码: 10363

学 号: 2170810103

题 目 **基于高维半参数回归模型的惩罚经验似然估计
及其应用研究**

英文并列

题 目 **PENALIZED EMPIRICAL LIKELIHOOD AND ITS
APPLICATION FOR HIGH-DIMENSIONAL
SEMI-PARAMETRIC REGRESSION MODELS**

学生姓名: 吕升日

指导教师: 何帮强

专 业: 应用数学

研究方向: 经济统计

论文答辩日期: 2020年06月20日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所提交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ，在 _____ 年解密后适用本版权书。
本学位论文属于
不保密 。

学位论文作者签名：吕升
日期：2020年6月22日

指导教师签名：何帮强
日期：2020年6月23日

基于高维半参数回归模型的惩罚经验似然估计 及其应用研究

摘 要

随着目前互联网大数据的不断发展，数据中的信息量越来越多，相对应的数据的维度也越来越高，如何有效的处理高维数据已经成为近些年来研究的热点。在一些实际数据分析当中，人们不仅可以得到基础的样本信息，还可以的得到相关回归系数的信息。通过这些回归系数的讯息，可以有效的提高模型的估计效率。然而现实中，模型的真实结构往往都是未知的，简单的只用参数回归模型或者是非参数回归模型来进行统计推断，得到的结果有很大的偏差，会有错误的结论。在这样的问题背景下，半参数回归模型应运而生，它将参数回归模型的简单性与非参数回归模型的灵活性巧妙地相结合。因此，研究半参数回归模型，对于现实生活的应用有着很大帮助。然而在半参数回归模型中，当协变量的维度随着样本量的增大而增大，即当协变量维度较高时，将会遇到“维数祸根”等问题。在经过大量学者的研究之后，将经验似然方法(EL)与 SCAD 惩罚函数相结合并应用于模型当中，可以有效的解决高维数据情况下的变量选择问题，从而降低模型的复杂度，解决模型在做预测时的不稳定性的问题。因此研究高维半参数回归模型的惩罚经验似然(PEL)估计，不仅是拓展了半参数回归模型的适用范围，也是为研究这类模型的学者提供了科学的参考。

第一章主要介绍了高维半参数回归模型的惩罚经验似然估计的研究历史背景与研究的意义,以及现在半参数回归模型与惩罚经验似然估计的国内外研究状况,最后简单说明了一下文章的内容与结构。

第二章介绍了在误差为鞅差序列的前提下,是如何探究非线性半参数测量误差模型的经验似然推断的。首先是通过逆卷积的方法来消除模型中测量误差所带来的影响,从而得到参数的无偏估计。然后通过引入的辅助函数以此来构造经验似然比统计量及其经验似然置信域,最后证明其该统计量服从渐进卡方分布。

第三章考虑了在误差是鞅差序列下高维部分线性半参数测量误差模型的惩罚经验似然推断,首先消除测量误差的影响,通过逆卷积的方法得到无偏估计,其次应用经验似然的方法,构造经验似然比统计量,然后针对参数数量离散的情况,提出了参数的惩罚经验似然方法,构造了惩罚经验似然比统计量及其惩罚经验似然置信域,并证明了所提出的惩罚经验似然估计量具有 Oracle 性质。最后再根据模型以及模型的假设条件,对参数分量的惩罚经验似然估计进行数值模拟,检验了估计结果的准确性,并进行了实例分析。

第四章研究的是对于带有固定效应的高维部分线性变系数面板数据模型,首先使用了 B 样条的方法得到了非参数的估计,之后采用局部线性虚拟变量的方法消除固定效应,并引入了一个合适的辅助函数,以此构造经验对数似然比函数。最后将 SCAD 函数与惩罚函数的思想相结合,推出惩罚经验似然法并应用于模型当中,构造了 PEL 统计量并得到其置信域,证明了其性质是服从渐进卡方分布的。

关键词: 鞅差序列, 高维数据, 半参数回归模型, 惩罚经验似然, 渐进卡方分布

PENALIZED EMPIRICAL LIKELIHOOD AND ITS
APPLICATION FOR HIGH-DIMENSIONAL
SEMI-PARAMETRIC REGRESSION MODELS

ABSTRACT

With the continuous development of Internet big data, the amount of information in the data is more and more, and the corresponding dimension of the data is also higher and higher, how to effectively handle high-dimensional data has become a hot research in recent years. In some actual data analyses, one can obtain not only the basic sample information, but also information on the relevant regression coefficients. With these messages of regression coefficients, the estimation efficiency of the model can be improved. In reality, the true structure of the model is often unknown, and simply using only parametric regression models or non-parametric regression models to make statistical inferences, the predicted results will be highly biased and wrong results will be obtained. The semi-parametric regression model organically combines the simplicity of a parametric regression model with the flexibility of a

non-parametric regression model, with the advantages of both. The study of semi-parametric regression models is very helpful for real life applications. However, in the semi-parametric regression model, the dimensionality of the covariate increases with the sample size, that is, when the covariate dimension is high, there is a "dimensional curse" problem. A large number of scholars have shown that combining the empirical likelihood method with the SCAD penalty function and applying it to the model can effectively solve the problem of variable selection in the case of high-dimensional data, thereby reducing the complexity of the model and solving the problem of model instability in making predictions. Therefore, the study of high-dimensional semi-parametric regression model for penalty empirical inference not only expands the scope of application of semi-parametric regression models, but also provides scientific reference for scholars who study such models.

The first chapter mainly introduces the historical background and significance of the research on high-dimensional semi-parametric regression models for penalized empirical likelihood estimation, and the current state of domestic and international research on semi-parametric regression models and penalized empirical likelihood estimation, and concludes with a brief description of the content and structure of the article.

In the second chapter, we introduces how to explore the empirical likelihood inference of the nonlinear semi-parametric measurement error model under the premise that the error is a martingale difference sequence. First, the effect of measurement error is eliminated by the method of inverse convolution, so as to obtain an unbiased estimate of the parameters. Then, the introduced auxiliary function is used to construct the empirical likelihood ratio statistic and its empirical likelihood confidence region. Finally, it is proved that it follows the progressive chi-square distribution.

In the third chapter, we consider the penalized empirical likelihood inference for the high-dimensional partial linear semi-parametric measurement error model when the error is a martingale difference sequence. First of all, the effect of measurement error is eliminated by the method of inverse convolution, and an unbiased estimate is obtained. Secondly, the empirical likelihood method is used to construct the empirical likelihood ratio statistic. Then, for the discrete number of parameters, the parameter penalized empirical likelihood (PEL) method is proposed to construct the penalized empirical likelihood ratio statistic and its penalized empirical likelihood confidence region, and proves that the proposed penalized empirical likelihood estimator has Oracle property. Finally, according to the assumptions of the model, the numerical simulation of the punishment empirical likelihood estimation of the

parameter components is carried out to verify the accuracy of the estimation results, and an example analysis is carried out.

In the fourth chapter, for the high-dimensional partial linear varying coefficient panel data model with fixed effects. First of all, the B-spline method is used to obtain non-parametric estimates, and then the local linear dummy variable method is used to eliminate the fixed effects, and a suitable auxiliary function is introduced to construct the empirical log-likelihood ratio function. Finally, the idea of SCAD function and penalty function is combined, the penalized empirical likelihood method is introduced and applied to the model, the penalized empirical likelihood ratio statistic is constructed and the confidence region is obtained from this, which proves that its nature is subject to the progressive chi-square distribution.

Lv Sheng-ri (Applied Mathematics)

Supervised by He Bang-qiang

KEYWORDS: martingale difference sequence, high-dimensional data, semi-parametric regression model, penalized empirical likelihood, asymptotic chi-square distribution

目 录

摘要.....	I
ABSTRACT.....	III
目录.....	VII
第1章 绪论.....	1
1.1 研究背景与研究意义.....	1
1.2 国内外现状研究.....	2
1.3 主要研究内容与论文结构.....	3
第2章 鞅差序列下非线性半参数测量误差模型的经验似然推断.....	4
2.1 模型框架.....	4
2.2 方法与主要结果.....	5
2.3 主要结果证明.....	7
2.4 本章小结.....	13
第3章 鞅差序列下高维半参数测量误差模型的惩罚经验似然推断.....	14
3.1 模型框架.....	14
3.2 方法与主要结果.....	15
3.2.1 经验似然法.....	15
3.2.2 惩罚经验似然法.....	16
3.2.3 计算方法.....	17
3.3 模拟研究.....	18
3.4 实例分析.....	20
3.5 主要结果证明.....	21
3.6 本章小结.....	28
第4章 带有固定效应的高维部分线性变系数面板数据模型的惩罚经验似然推断.....	29
4.1 模型框架.....	29
4.2 方法与主要结果.....	30
4.2.1 经验似然法.....	30
4.2.2 惩罚经验似然法.....	32
4.3 主要结果证明.....	33
4.4 本章小结.....	40
第5章 总结与展望.....	41
参考文献.....	42
硕士期间发表论文.....	47
致谢.....	48

第 1 章 绪论

1.1 研究背景与研究意义

在金融以及医学等领域，人们对现有的数据进行分析处理，从而对将来的发展趋势进行一个预测分析，所以参数回归模型成为了研究自然科学以及社会经济等发展规律的重要工具，并且其理论体系也相对比较丰富完整。在现实问题当中，参数模型过于依赖由现有的历史大量额外信息，所以当模型的假设条件成立时，该模型的推断有较高的精度，然而当模型的假设条件和实际相背离时，这样的模型的拟合效果就会不是很理想了，而如果用非参数模型去处理，则可能会丢失掉现在已有的历史经验所提供的信息，并且当协变量的维数比较大时，该模型将会发生“维数祸乱”等问题。为了解决参数模型和非参数模型所含有问题，在二十世纪八十年代，半参数回归模型便应运而生，半参数回归模型是将参数与非参数模型巧妙的相互结合起来，所以半参数回归模型既含有参数分量，又含有非参数分量，它同时具有参数模型和非参数模型的优点，比单纯的参数模型或非参数模型有着更广的适用范围，同时也有了对未来预测更强的解释能力，以及更准确的预测结果。

如今随着互联网以及大数据技术的不断创新与发展，人们在网上获得的数据将会包含越来越多的信息，在近些年来，在统计学中，对于如何快速有效的处理高维数据一直是一个炙手可热的研究热点。简单的说，在数据分析中模型的偏差会随着引入的自变量变多而变小，但于此同时也会有一些多余的或者是不必要的信息被纳入到统计模型中，进而使模型的复杂度变得很高。为了降低模型的复杂度，如何处理高维数据、如何变量选择自然而然的也就成为了一个非常重要的问题。而惩罚经验似然的思想，可以有效的降低模型的复杂度与模型在做预测时的不稳定性，所以研究惩罚经验似然方法可以有效的解决当数据维度过高时造成的各种问题从而提高模型的精度，并且结合半参数回归模型极为广泛的适用性，研究高维半参数回归模型的惩罚经验似然估计是很有必要的，它将有助于让半参数回归模型的适用范围扩展的更为宽泛，同时也能更好地应用与实际当中，为社会发展作出一些贡献。

1.2 国内外研究现状

自半参数回归模型在 20 世纪 80 年代被首次被 Engle 等^[1]推出以来,在广大学者的研究之后,已经有了一个较为完整的体系,半参数回归模型也是目前来看适用性极广的统计模型,它既有着参数模型简单性又有着非参数模型的灵活性,这也是它相对于参数或非参数模型来说,更为具有广泛应用价值的一面。Yang 和 Li^[2]对具有纵向数据的半参数变系数部分线性模型进行经验似然研究,提出了参数化成分的经验似然比统计量,并证明了 Wilk 定理的非参数化的版本,最后可以构造出参数分量的置信区间,证明了其覆盖概率为正态分布。此外,You 和 Chen^[3], Ahmad 等^[4], Bravo 和 Francesco^[5], Wang 等^[6]以及 Zhou 和 Liang^[7]都对半参数回归模型进行了深入的探讨。

经验似然统计方法是由 Owen^[8]于 20 世纪 80 年代末 90 年代初提出的,目的是为了更加高效的处理密集型的数据,作为一种非参数似然方法,经验似然的在构造置信域方面的优点也很明显,正如王启华^[9]总结的那样,用经验似然方法构造出的置信域具有变换不变性、Bartlett 纠偏性和域保持性,并且置信域的形状完全是由数据本身自行决定并不需要预先给定及无需构造枢轴量等等优点。但是用经验似然的方法处理高维数据,就存在了一个比较突出的问题,Min Taso^[10]发现标准的经验似然函数是不能在一定条件下的高维数据问题中适用的,在高维数据中,经验似然函数会遭遇“凸包”的问题,从而使其失效。为了解决这种问题,Owen 和 Emerson^[11]及 Chen^[12]等都提出了调整经验似然,通过添加伪数据的方法。Bartolucci^[13]和 Taisuke Otsu^[14]通过运用惩罚的思想,重新定义参数的经验似然函数,也解决了高维数据的问题。Leng^{[15][16]}他们提出了惩罚经验似然思想,在处理高维数据问题时,很好的解决了变量选择问题,并且通过数据模拟,得到了很多不错的结果。He 等^[17]研究了基于带有固定效应的面板数据模型中估算模型参数等使用了惩罚经验似然方法。以及 Shen 等^[18]、Fang 等^[19]和 Fan 等^[20]都在自己的模型中使用了惩罚经验似然方法,并且都得到了不错的结果。

从已有文献来说,用惩罚经验似然方法来处理高维数据,国外的研究比国内的稍微早了那么些年。从目前来看,半参数回归模型如何解决各种高维数据,依旧是一个非常火热的研究热点。本文用惩罚经验似然的方法来处理高维半参数回归模型,是一个非常好的方法。

1.3 主要研究内容与论文结构

本课题研究的是高维半参数的惩罚经验似然,首先通过逆卷积的方法得到无偏估计量,给出半参数回归模型的参数分量和非参数分量的经验似然估计,并证明该统计量的渐近正态性。之后给出半参数回归模型的惩罚经验对数似然比函数统计量,构造模型中统计量的惩罚经验似然置信域。再根据模型以及假设条件,对参数分量的惩罚经验似然估计进行数值模拟,检验估计结果的准确性。

具体内容如下:

1、在误差为鞅差序列的前提下,我们是如何研究非线性半参数测量误差模型的经验似然的,首先运用逆卷积的方法得到了非参数的无偏估计,然后引入一个合适的辅助变量,构造了经验似然比统计量,在证明了这个统计量是服从渐进卡方分布的同时,也构造了其经验似然置信域。此章主要是说明半参数回归模型在非线性的情况下经验似然的性质。

2、在前文的基础上,略有不同的是,模型从非线性半参数回归模型变为部分线性半参数回归模型,在用逆卷积方法得到无偏估计以后,根据 Owen 的经验似然方法构造了 EL 统计量,并针对参数数量离散的情况,提出了参数的惩罚经验似然(PEL),构造了惩罚经验似然比统计量,并证明了 PEL 估计量具有 Oracle 属性。此外,进行了仿真研究和实际数据的应用,用来评估我们提出的方法的性能。

3、对于带有固定效应的面板数据模型,首先通过使用 B 样条的方法得到了非参数估计,对固定效应施加一个限制后,采用局部线性虚拟变量的方法来消除固定效应。然后构造辅助函数,通过经验似然法来构造经验对数似然比函数。最后将 SCAD 函数与惩罚函数的思想相结合,提出了惩罚经验似然法并应用于模型当中,构造了惩罚经验似然统计量并借此得到其置信域,证明了其性质是服从渐进卡方分布的。

第 2 章 鞅差序列下非线性半参数测量误差模型的经验似然推断

本章研究了鞅差误差序列下非线性半参数测量误差回归模型,通过逆卷积的方法得到了无偏的非参数估计,并通过一个可估计的辅助随机变量构造了未知参数的经验对数似然比统计量,证明了该统计量服从渐近卡方分布,并借此得出该未知参数的置信域。通过得到的统计量的渐近性质,说明了经验似然方法在鞅差误差序列下非线性半参数测量误差回归模型中是有效的。

2.1 模型框架

考虑如下非线性半参数回归模型:

$$Y = g(X, \beta) + m(U) + \varepsilon, \quad (2.1)$$

其中 Y 为实值响应变量, $\beta \in R^p$ 为 p 维未知参数向量, X 是 d 维协变量, $g(\cdot, \cdot)$ 为已知的可测函数。 U 是取值于 $[0,1]$ 上的随机变量, $m(\cdot)$ 是定义在 $[0,1]$ 上的光滑未知函数, ε 为随机误差。

关于模型(1),当 $g(X, \beta) = X^T \beta$ 时,模型(1)是部分线性模型。对于部分线性模型的研究,在部分线性模型首次被 Engle 等^[1]提出以来,学者们就对部分线性模型做了许多的研究与扩充。一些细节的研究能被发现在 Wooldridge 和 Li^[21], Cui 和 Chen^[22], Wang 和 Jing^[23]以及公徐路和李幸福^[24]等的研究中。Li 和 Xue^[25], 崔恒建^[26]以及李晓妍^[27]等研究了在含有误差变量下部分线性模型的统计推断。当 $g(X, \beta)$ 是 X, β 的非线性函数时,冯三营等^[28]在用经验似然方法得出未知参数置信域的基础上通过模拟研究与最小二乘法得出结果做了优劣比对。以上文献的研究都是在模型误差是独立的条件下进行的,然而在实际中大部分数据并不是独立的,特别是对于连续收集的经济数据,其误差具有明显的相依结构。而鞅差结构是一种相对较弱的相依数据结构,并且在实际中也是大量存在的。Chen 等^[29]将经验似然方法应用于带有鞅差序列的误差模型中,并将通过经验似然法与最小二乘法得到的结果通过模拟研究进行比较。李国亮等^[30]与 Fan 等^[31]分别研究了带有鞅差序列模型中相关估计的强相和性,与误差是鞅差序列的异方差半参数回

归模型的经验似然估计。本文在此基础上研究了当测量误差是协变量 U 时,模型误差为鞅差序列的非线性半参数测量误差模型的经验似然推断。通过一个可估计的辅助随机变量构造了未知参数的经验对数似然比统计量,证明了统计量服从渐近卡方分布,最后给出了未知参数的置信域。

2.2 方法与主要结果

令数据 $\{X_i, Z_i, Y_i\}_{i=1}^n$ 是来自 $\{X, Z, Y\}$ 的一组可观测的独立同分布的随机样本,即

$$\begin{cases} Y_i = g(X_i, \beta) + m(U_i) + \varepsilon_i, \\ Z_i = U_i + e_i, \end{cases} \quad (2.2)$$

其中 $\{e_i\}_{i=1}^n$ 是服从均值为0,方差为 σ_e^2 的独立同分布的测量误差。且 e_i 与 $(X_i, U_i, \varepsilon_i)$ 相互独立, $\{\varepsilon_i, F_i, i \geq 1\}$ 是鞅差序列,即 ε_i 其在 F_i -是可测的且 $E(\varepsilon_i | F_{i-1}) = 0$ 。

首先假设 β 是给定的,模型(1)可以变为下面的非参数回归模型

$$Y - g(X, \beta) = m(U) + \varepsilon,$$

因此若 β 已知,即当 U 可观测时,有 $m(U) = E(Y - g(X, \beta) | U)$,定义 $m(\cdot)$ 的核估计为

$$\begin{aligned} \hat{m}(u) &= \frac{\sum_{i=1}^n K\left(\frac{u - U_i}{h}\right) (Y_i - g(X_i, \beta))}{\sum_{i=1}^n K\left(\frac{u - U_i}{h}\right)} \\ &\triangleq \sum_{i=1}^n W_i(u) (Y_i - g(X_i, \beta)) \end{aligned} \quad (2.3)$$

其中

$$W_i(u) = \frac{1}{nh} K\left(\frac{u - U_i}{h}\right) / \hat{f}(u)$$

$K(\cdot)$ 为已知的核函数, h 是光滑参数。 $\hat{f}(u) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{u - U_i}{h}\right)$ 为协变量 U 的密度函数 $f(u)$ 的核估计。将式(2.3)代入式(2.2)可以得到一个近似残差

$$\tilde{\varepsilon}_i(\beta) = Y_i - g(X_i, \beta) - \hat{m}(U_i) = \tilde{Y}_i - \tilde{g}(X_i, \beta),$$

其中 $\tilde{Y}_i = Y_i - \sum_{j=1}^n W_j(U_i) Y_j$, $\tilde{g}(X_i, \beta) = g(X_i, \beta) - \sum_{j=1}^n W_j(U_i) g(X_j, \beta)$, (2.4)

定义辅助随机变量

$$\zeta_i(\beta) = \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} [\tilde{Y}_i - \tilde{g}(X_i, \beta)],$$

当 β 是真实参数时, $E(\zeta_i(\beta)) = 0$ 。由此我们来构造经验对数似然比函数

$$l_n(\beta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \zeta_i(\beta) = 0 \right\}, \quad (2.5)$$

然而, 由于式(2.4)中 U_i 是不可直接观测的, 实际上观测的是 Z_i , 所以 β 的置信域不能直接用 $l_n(\beta)$ 来构造。故而, 我们需对 $m(\cdot)$ 进行重新估计, 结合 Fan 和

Truong^[32] 逆卷积的方法我们得到 $m(\cdot)$ 的估计:

$$m_n(u) \hat{=} m_n(u, \beta) = \sum_{i=1}^n W_{ni}(u) (Y_i - g(X_i, \beta)),$$

其中 $W_{ni}(\cdot) \hat{=} \frac{1}{nh} K_n \left(\frac{\cdot - Z_i}{h_n} \right) / \hat{f}_n(\cdot)$, $K_n(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-isu) \frac{\varphi_K(s)}{\phi_e(s/h)} ds$. $\varphi_K(\cdot)$ 为核函数 $K(\cdot)$ 的 Fourier 的变换, $\phi_e(\cdot)$ 为测量误差变量 e 的特征函数, i 是虚数单位, 分别

用 $m_n(z)$, $W_{nj}(z)$ 替换式(2.3)中的 $\hat{m}(u)$, $W_j(u)$ 可得估计的辅助随机变量

$$\hat{\zeta}_i(\beta) = \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} [\tilde{Y}_i - \tilde{g}(X_i, \beta)],$$

以此可以构造出一个估计的经验对数似然比函数

$$\hat{l}_n(\beta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\zeta}_i(\beta) = 0 \right\},$$

运用 Lagrange 乘子法, 可以得到估计的经验对数似然比统计量

$$\hat{l}_n(\beta) = 2 \sum_{i=1}^n \log(1 + \lambda^T \hat{\zeta}_i(\beta)), \quad (2.6)$$

其中 λ 为 Lagrange 乘子, 满足:

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda^T \hat{\zeta}_i(\beta)} = 0. \quad (2.7)$$

记

$$\frac{\partial g(X_i, \beta)}{\partial \beta_l} = h_l(U_i, \beta) + v_{il}(\beta), \quad 1 \leq i \leq n, \quad 1 \leq l \leq p,$$

其中 $h_l(U_i, \beta) = E \left(\frac{\partial g(X_i, \beta)}{\partial \beta_l} \mid U_i \right)$ 。则当 $0 < \sup_i E[v_{il}^2(\beta)] < \infty$ 时, 结合大数定律有

$$\frac{1}{n} \sum_{i=1}^n v_i(\beta) v_i^T(\beta) \rightarrow \Omega(\beta), a.s., n^{-3/4} \left| \sum_{i=1}^n v_{il}(\beta) \right| = O(1), a.s. \quad (2.8)$$

对于所有的 $l=1,2,\dots,p$ 均成立, $v_i(\beta) = (v_{i1}(\beta), \dots, v_{ip}(\beta))^T$,

$$\Omega(\beta) \doteq E \left[\frac{\partial g(X, \beta)}{\partial \beta} - E \left\{ \frac{\partial g(X, \beta)}{\partial \beta} \mid U \right\} \right]^{\otimes 2} > 0$$

是 $p \times p$ 的正定矩阵, 其中 $A^{\otimes 2} = AA^T$ 。

为了得到 $\hat{l}_n(\beta)$ 的渐近分布, 给出以下几个基本假设条件:

A1 对 $u \in [0,1], m(u), h_l(u, \beta), U$ 的密度函数 $f(u)$ 满足一阶的 Lipschitz 条件

$$\text{且 } 0 < \inf_{0 \leq u \leq 1} f(u) \leq \sup_{0 \leq u \leq 1} f(u) < \infty.$$

A2 对每一个 X , $g(X, \beta)$ 关于 β 都有连续的二阶导数。

A3 (i) 存在 $\gamma > 0$, 使得 $\sup_{n \geq 1} E(|\varepsilon_n|^{2+\gamma} \mid F_{n-1}) < \infty, a.s.$;

(ii) 存在 $0 < \sigma^2 < \infty$, 使得 $E(\varepsilon_i^2 \mid F_{i-1}) = \sigma^2 + o(1), a.s.$

A4 $\sup_{-\infty < e < \infty} E(e_i) < \infty, a.s.$

A5 测量误差 e 的特征函数 $\phi_e(\cdot)$ 非退化, 且分布是普通光滑的。

A6 $K(\cdot)$ 为非负有界 k 阶核函数, 且核函数 $K(\cdot)$ 的 Fourier 的变换满足:

$$\int_{-\infty}^{\infty} |\varphi_K(u) u^\alpha| du < \infty, \int_{-\infty}^{\infty} |\varphi_K(u) u^\alpha|^2 du < \infty, \text{其中 } \alpha \text{ 是 } e \text{ 的分布的光滑阶数。}$$

定理 2.1 假定条件 A1~A6 成立。在 β 是参数真值条件下, 那么 $\hat{l}_n(\beta) \xrightarrow{L} \chi_p^2$, 其中“ \xrightarrow{L} ”表示依分布收敛。

基于定理 2.1, 构造参数 β 的 $1-\alpha_0$ 水平的置信域: $I_{\alpha_0}(\beta) = \{\beta: \hat{l}_n(\beta) \leq c_{\alpha_0}\}$

其中 $0 < \alpha_0 < 1$, 存在 c_{α_0} 使得 $P(\chi_p^2 > c_{\alpha_0}) = \alpha_0$ 。

2.3 主要结果的证明

引理 2.1 在定理 2.1 的条件下

$$\max_{1 \leq i \leq n} \left| m(u_i) - \sum_{k=1}^n W_{nk} m(u_k) \right| = o(n^{-1/2}), a.s.$$

证明：测量误差分布的特征函数 $\phi_e(\cdot)$ 满足 $u \rightarrow \infty$ 时， $d_0|u|^{-\alpha} \leq |\phi_e(u)| \leq d_1|u|^{-\alpha}$ ，其中 d_0, d_1, α 为正常数，存在正常数 M 和 C 使得，当 $|s| > M$ 时， $|\phi_e(s)| \geq C|s|^{-\alpha}$ ，于是由文献[28]得到

$$|K_n(u)| = O(h^{-\alpha}) \text{ 和 } \max_{1 \leq i \leq n} |W_{nm}(u)| = O\left((nh^{1+\alpha})^{-1}\right), \text{ a.s.} \quad (2.9)$$

因为误差为鞅差序列有 $E(\varepsilon_i) = E(E(\varepsilon_i | F_{i-1})) = 0$ ，不难得出

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n W_{nk}(Z_i) \varepsilon_k \right| = O(n^{-1/2}), \text{ a.s.} \quad (2.10)$$

根据文献[32]的定理 4 可得

$$\max_{1 \leq i \leq n} \left| m(U_i) - \sum_{k=1}^n W_{nk}(Z_i) \{m(U_k + \varepsilon_k)\} \right| = o(n^{-1/2}), \text{ a.s.},$$

因此 $\max_{1 \leq i \leq n} \left| m(U_i) - \sum_{k=1}^n W_{nk}(Z_i) m(U_k) \right| = o(n^{-1/2}), \text{ a.s.}$ ，此引理得证。

引理 2.2 在定理 2.1 的条件下，当 β 是真实参数，有

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} \left[\frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} \right]^T \rightarrow \Omega(\beta), \text{ a.s.}, \quad (2.11)$$

其中 $\frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} \triangleq \left(\frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta_1}, \dots, \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta_p} \right)^T$ 。

证明：证明见文献[28]引理 4.3。

引理 2.3 在定理 2.1 的条件下，若 β 是真实参数，则

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\zeta}_i(\beta) \xrightarrow{L} N(0, \sigma^2 \Omega(\beta)), \quad (2.12)$$

$$\begin{aligned} \text{证明：} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\zeta}_i(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_i - g(X_i, \beta) - m_n(Z_i)] \frac{\partial g(X_i, \beta)}{\partial \beta} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{m}(U_i) \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\varepsilon}_i \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta} \\ &\stackrel{\Delta}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n B \varepsilon_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n B \tilde{m}(U_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n B \bar{\varepsilon}_i. \end{aligned}$$

其中 $B = \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta}$, $B_l = \frac{\partial \tilde{g}(X_i, \beta)}{\partial \beta_l}$ 为 B 的第 l 项， $\tilde{m}(U_i) = m(U_i) - \sum_{j=1}^n W_{nj}(Z_i) m(U_j)$ ，

$$\bar{\varepsilon}_i = \sum_{j=1}^n W_{nj}(Z_i) \varepsilon_j.$$

我们先证明 $\sum_{j=1}^n W_{nj}(Z_i) B = o(n^{1/2})$, *a.s.* 成立。考虑其第 l 项,

$$\begin{aligned} \sum_{i=1}^n \sum_{k=1}^n W_{ni}(Z_k) B_l &= \sum_{i=1}^n \sum_{k=1}^n W_{ni}(Z_k) \nu_{kl}(\beta) + \sum_{i=1}^n \sum_{k=1}^n W_{ni}(Z_k) \tilde{h}_l(U_k, \beta) \\ &\quad - \sum_{i=1}^n \left[\sum_{k=1}^n \left\{ \sum_{q=1}^n W_{nq}(Z_k) \nu_{qi}(\beta) \right\} W_{ni}(Z_k) \right] \end{aligned}$$

其中 $\tilde{h}_l(U_k, \beta) = h_l(U_k, \beta) - \sum_{k=1}^n W_{nk}(Z_i) h_l(U_i, \beta)$ 。

由式(2.9)及文献[28]的引理 2.2 容易得到

$$\left| \sum_{i=1}^n \sum_{k=1}^n \tilde{h}_l(U_k, \beta) W_{ni}(Z_k) \right| \leq n \max_{1 \leq k \leq n} \left| \sum_{i=1}^n W_{ni}(Z_k) \right| \max_{1 \leq k \leq n} |\tilde{h}_l(U_k, \beta)| = o(n^{1/2}), \quad a.s.$$

由式(2.8)、式(2.9)及 Abel 不等式, 有

$$\begin{aligned} &\left| \sum_{i=1}^n \left[\sum_{k=1}^n \left\{ \sum_{q=1}^n W_{nq}(Z_k) \nu_{qi}(\beta) \right\} W_{ni}(Z_k) \right] \right| \\ &\leq n \max_{1 \leq k \leq n} \left| \sum_{i=1}^n W_{ni}(Z_k) \right| n \max_{1 \leq k \leq n} \left| \sum_{q=1}^n W_{nq}(Z_k) \nu_{qi}(\beta) \right| = o(n^{1/2}), \quad a.s. \end{aligned}$$

所以

$$\sum_{j=1}^n W_{nj}(Z_i) B = o(n^{1/2}), \quad a.s. \quad (2.13)$$

由引理 2.1, 引理 2.2 有

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n B \tilde{m}(U_i) \right\| \leq \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |\tilde{m}(U_i)| \sum_{i=1}^n \|B\| = o(1),$$

因为误差为鞅差序列 $E(\varepsilon_j | F_{j-1}) = 0$, $E(\varepsilon_j) = E(E(\varepsilon_j | F_{j-1})) = 0$ 。

由式(2.13)和条件 A3(i)得

$$E \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \bar{\varepsilon}_j \sum_{i=1}^n B \right\|^2 \leq \frac{1}{n} \sum_{j=1}^n \left(\left\| \sum_{i=1}^n B W_{nj}(Z_i) \right\| \right)^2 E \varepsilon_j^2 = o(1).$$

即

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\varepsilon}_i B = o(1), \quad a.s. \quad (2.14)$$

所以要证引理 3.3 成立, 只需证 $\frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i B \xrightarrow{L} N(0, \sigma^2 \Omega(\beta))$,

设 $S_n = \frac{1}{n} \sum_{i=1}^n BB^T$, 由 $\Omega(\beta)$ 的定义和引理 2.2 及 Cramer Wold 策略, 只需证对

与任意的 $p \times 1$ 维向量 d , $d^T \frac{1}{\sqrt{n}} S_n^{-1/2} \sum_{i=1}^n \varepsilon_i B \xrightarrow{L} N(0, \sigma^2 \|d\|^2)$, *a.s.*

令 $a_{ni} = d^T \frac{1}{\sqrt{n}} S_n^{-1/2} B$, $\xi_{ni} = a_{ni} \varepsilon_i$, 根据鞅差中心极限定理, 只需证明

$$\sum_{i=1}^n E(\xi_{ni}^2 | F_{i-1}) \xrightarrow{p} \sigma^2 \|d\|^2, \quad (2.15)$$

对于任何 $\sigma > 0$,

$$\sum_{i=1}^n E(\xi_{ni}^2 I(|\xi_{ni}| > \sigma) | F_{i-1}) \xrightarrow{p} 0, \quad (2.16)$$

使用条件 A3(ii), 有

$$\begin{aligned} \left| \sum_{i=1}^n E(\xi_{ni}^2 | F_{i-1}) - \sigma^2 \|d\|^2 \right| &= \left| d^T \left(S_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i=1}^n BB^T E(\varepsilon_i^2 | F_{i-1}) S_n^{-\frac{1}{2}} - \sigma^2 I_p \right) d \right| \\ &= \left| d^T S_n^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{i=1}^n BB^T (E(\varepsilon_i^2 | F_{i-1}) - \sigma^2) \right) S_n^{-\frac{1}{2}} d \right| \\ &\leq d^T S_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i=1}^n BB^T |o_p(1)| S_n^{-\frac{1}{2}} d \\ &\xrightarrow{p} 0. \end{aligned}$$

由此可得式(2.15)。引理 2.2 表明 $\max_{1 \leq i \leq n} \|B\| / \sqrt{n} \rightarrow 0$, 因此 $\max_{1 \leq i \leq n} \|a_{ni}\| \rightarrow 0$ 。

结合式(2.15), 条件 A3 (i) 以及 Markov 不等式, 对任意的 $\delta > 0$, 有

$$\begin{aligned} \sum_{i=1}^n E(\xi_{ni}^2 I(|\xi_{ni}| > \delta) | F_{i-1}) &= \sum_{i=1}^n E(a_{ni}^2 \varepsilon_i^2 I(|a_{ni} \varepsilon_i| > \delta) | F_{i-1}) \\ &= \sum_{i=1}^n a_{ni}^2 E(\varepsilon_i^2 I(|\varepsilon_i| > \delta / |a_{ni}|) | F_{i-1}) \\ &\leq \sum_{i=1}^n a_{ni}^2 E(|\varepsilon_i|^{2+\gamma} | F_{i-1}) \delta^{-\gamma} |a_{ni}|^\gamma \\ &\leq C \delta^{-\gamma} \max |a_{ni}|^\gamma \sum_{i=1}^n a_{ni}^2 \\ &\xrightarrow{p} 0. \end{aligned}$$

至此，得证式(2.16)，此引理得证。

引理 2.4 在定理 2.1 的条件下，若 β 是真实参数，则

$$\frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) \hat{\zeta}_i^T(\beta) \xrightarrow{P} \sigma^2 \Omega(\beta).$$

证明： 经过简单计算，有

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) \hat{\zeta}_i^T(\beta) &= \frac{1}{n} \sum_{i=1}^n [\varepsilon_i + \tilde{m}(U_i) - \bar{\varepsilon}_i]^T B B^T \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 B B^T + \frac{1}{n} \sum_{i=1}^n \tilde{m}^2(U_i) B B^T + \frac{1}{n} \sum_{i=1}^n \bar{\varepsilon}_i^2 B B^T + \frac{2}{n} \sum_{i=1}^n \varepsilon_i \tilde{m}(U_i) B B^T \\ &\quad - \frac{2}{n} \sum_{i=1}^n \varepsilon_i \bar{\varepsilon}_i B B^T - \frac{2}{n} \sum_{i=1}^n \tilde{m}(U_i) \bar{\varepsilon}_i B B^T \\ &\triangleq C_1 + C_2 + C_3 + 2C_4 - 2C_5 - 2C_6 \end{aligned}$$

由引理 2.2 和条件 A3(ii)可知，对与 C_1 ，

$$\frac{1}{n} \sum_{i=1}^n B B^T E(\varepsilon_i^2 | F_{i-1}) \xrightarrow{P} \sigma^2 \Omega(\beta).$$

进一步，对于任意的 p -维常数向量 d ，将证明：

$$\frac{1}{n} \sum_{i=1}^n d^T B B^T (\varepsilon_i^2 - E(\varepsilon_i^2 | F_{i-1})) d \xrightarrow{P} 0.$$

引理 2.2 表明 $\max_{1 \leq i \leq n} \|B\|/\sqrt{n} \rightarrow 0$ 。又因为 $\{d^T B B^T (\varepsilon_i^2 - E(\varepsilon_i^2 | F_{i-1})) d, F_i, i \geq 1\}$ 是一个鞅差序列，通过运用 Burkholder 不等式^[28]、Stout^[29]的不等式 3.3.14 以及条件 A3(i)和引理 2.2，对于任意的 $0 < \alpha < 1$ ，有

$$\begin{aligned} &E \left| \frac{1}{n} \sum_{i=1}^n d^T B B^T (\varepsilon_i^2 - E(\varepsilon_i^2 | F_{i-1})) d \right|^{1+\alpha} \\ &\leq \frac{n^{\frac{\alpha-1}{2}} \sum_{i=1}^n (d^T B B^T d)^{1+\alpha} E |\varepsilon_i^2 - E(\varepsilon_i^2 | F_{i-1})|^{1+\alpha}}{n^{1+\alpha}} \\ &\leq d \frac{\max_{1 \leq i \leq n} \|B\|^{2\alpha} \left(n^{-1} \sum_{i=1}^n \|B\|^2 \right)}{n^{(1+\alpha)/2}} \\ &= o(n^\alpha) / n^{(1+\alpha)/2} \\ &= o(1) \end{aligned}$$

所以 $C_1 \xrightarrow{p} \sigma^2 \Omega(\beta)$ 。因此为证引理 2.4，只需证 $C_i = o(1)$ ， $i = 2, 3, 4, 5, 6$ 。结合引理 2.1 以及引理 2.2，对于任意的 p -维常数向量 d ，易得

$$|d^T C_2 d| \leq \frac{1}{n} \sum_{i=1}^n |d^T B B^T d (\tilde{m}(U_i))|^2 \leq \max_{1 \leq i \leq n} |\tilde{m}(U_i)|^2 \frac{1}{n} \sum_{i=1}^n |d^T B B^T d| = o(1).$$

上式证明了 $C_2 = o(1)$ 。根据式(2.14)、引理 2.1、引理 2.2 和条件 A3(i)类似于 C_2 的证明，可证 $C_i = o(1)$ ， $i = 3, 4, 5, 6$ 。至此引理 2.4 得证。

引理 2.5 在定理 2.1 的条件下，若 β 是真实参数，则

$$\max_{1 \leq i \leq n} \|\hat{\zeta}_i(\beta)\| = o_p(n^{1/2}). \quad (2.17)$$

$$\|\lambda\| = O_p(n^{-1/2}). \quad (2.18)$$

证明：证明见文献[28]引理 4.6。

定理 1 的证明：将式(2.16)泰勒展开，有

$$\hat{i}_n(\beta) = 2 \sum_{i=1}^n \left(\lambda^T \hat{\zeta}_i(\beta) - (\lambda^T \hat{\zeta}_i(\beta))^2 / 2 \right) + o_p(1) \quad (2.19)$$

根据式(2.17)有

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda^T \hat{\zeta}_i(\beta)} \\ &= \frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) - \frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) \lambda^T \hat{\zeta}_i(\beta) + \frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta) [\lambda^T \hat{\zeta}_i(\beta)]^2}{1 + \lambda^T \hat{\zeta}_i(\beta)} \end{aligned}$$

由引理 2.3 至引理 2.5 得

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|\hat{\zeta}_i(\beta)\|^3 \|\lambda\|^2 |1 + \lambda^T \hat{\zeta}_i(\beta)|^{-1} &\leq O_p(n^{-1}) \max_{1 \leq i \leq n} \|\hat{\zeta}_i(\beta)\| \frac{1}{n} \sum_{i=1}^n \|\hat{\zeta}_i(\beta)\|^2 \\ &= O_p(n^{-1}) o_p(n^{1/2}) O_p(1) \\ &= o_p(n^{-1/2}) \end{aligned}$$

从而， $\lambda = \left(\sum_{i=1}^n \hat{\zeta}_i(\beta) \hat{\zeta}_i^T(\beta) \right)^{-1} \sum_{i=1}^n \hat{\zeta}_i(\beta) + o_p(n^{-1/2})$ 。

代入式(2.19)有

$$\hat{i}_n(\beta) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\zeta}_i(\beta) \right)^T \left(\frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) \hat{\zeta}_i^T(\beta) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\zeta}_i(\beta) \right) + o_p(1)$$

再结合引理 2.3 和引理 2.4，此定理得证。

2.4 本章小结

鞅差序列的运用在医疗以及经济领域有着广泛的应用,本文是非线性半参数测量误差模型在鞅差误差序列方向的延伸。我们运用逆卷积的方法得到了无偏的非参数估计,并应用辅助随机变量构造了未知参数的经验对数似然比统计量,并证明了该经验对数似然比统计量服从渐近卡方分布以及构造出未知参数估计量的置信域。通过得到的统计量的渐近性质,说明了经验似然方法在鞅差误差序列下非线性半参数测量误差回归模型中是有效的,为研究带有鞅差误差序列的非线性半参数测量误差回归模型提供了一种方法与思路。并且为下一步处理误差是鞅差序列下的高维半参数回归模型打下了一定的理论基础。

第 3 章 鞅差序列下高维半参数测量误差模型的惩罚经验似然推断

本文主要研究了具有误差是鞅差序列下的高维半参数回归模型的惩罚性高维经验似然推断。我们为模型中的回归参数提出了经验对数似然比函数，结合 SCAD 惩罚函数，我们针对参数数量离散的情况，提出了参数的惩罚经验似然。构造了惩罚经验似然比统计量，以及证明了惩罚经验似然估计量具有 Oracle 属性。此外，我们在原假设下，得到了惩罚经验似然比统计量其极限分布服从渐近卡方的结论。最后进行了仿真研究和实际数据的应用，从而评估我们提出的方法的性能。

3.1 模型框架

部分线性半参数回归模型经常被用于回归分析中，因为它能很好地实现灵活性与真实性的平衡。在过去的二十年里，各种形式的半参数回归模型被开发出来，包括了有部分线性模型 Engle 等^[1]，变系数部分线性模型 Fan 和 Huang^[35]，部分线性单指标模型 Xia 和 Härdle^[36]，非线性半参数回归模型 Sepanski 等^[37]，等等各类半参数回归模型。

对于一个相对完整的理论研究中，我们考虑了部分线性测量误差模型形式如下：

$$\begin{cases} Y_i = X_i^\tau \beta + g(U_i) + \varepsilon_i, \\ Z_i = U_i + v_i \end{cases} \quad (3.1)$$

其中 Y_i 为实值响应变量， $\beta = (\beta_1, \dots, \beta_p)^\tau$ 为 p 维未知参数向量， τ 表示一个向量或是矩阵的转置， X_i 是协变量， $g(\cdot)$ 是未知光滑函数。 U_i 是取值于 $[0,1]$ 上的随机变量， $\{\varepsilon_i, F_i, i \geq 1\}$ 是鞅差序列。由于 U_i 不能直接观测，含有测量误差，被 $Z_i = U_i + v_i$ 代替， v_i 是服从均值为 0，协方差为 Σ_v 的独立同分布的测量误差。且 v_i 与 $(X_i, U_i, \varepsilon_i)$ 相互独立。假设 Σ_v 已知的情况，详细的说明可以在 You 和 Chen^[3]，Li 和 Xue^[38] 中见到。当 Σ_v 未知时，可以通过代替重复观测 Z 得到它，详情请见 Liang 和 Härdle^[39]。

由于近来数据采集和存储技术的飞速发展,高维数据分析出现在许多当代统计研究中。一个关键特征是未知参数的维数甚至超过了样本量的维度。当维度偏离时,通过正则化进行变量选择已被证明是有效的。Fan 和 Li^[40]提出了通过非凹罚最小二乘的统一方法来自动同时选择变量。Zhao 和 Xue^[41]研究分析了纵向数据下半参数回归模型中的变量选择问题。惩罚性经验似然法具有从非参数似然方法继承的效率和适应性的优点,同时也具有经验似然(EL)相同的优点,仅使用数据确定置信区域的形状和方向,而无需估计复杂度协方差。有几篇与惩罚性经验似然法有关的论文,例如 Bartolucci^[42]提出了一种由损失函数惩罚的概率权重的优化乘积的 PEL 方法。Ren 和 Zhang^[43]使用惩罚经验似然方法来进行变量选择。Wang 和 Xiang^[44]研究了惩罚性经验似然方法的推论。Fang 等^[45]提出了用于参数估计和变量选择的惩罚性经验似然方法并应用于半参数回归模型。

3.2 方法与主要结果

3.2.1 经验似然方法

假设模型(3.1)中 U_i 可直接观测时,当 β 已知,用核函数的方法来估计 $g(\cdot)$,有

$$\hat{g}(u) = \frac{\sum_{i=1}^n K\left(\frac{u-U_i}{h}\right)(Y_i - X_i^T \beta)}{\sum_{i=1}^n K\left(\frac{u-U_i}{h}\right)} \triangleq \sum_{i=1}^n W_i(u)(Y_i - X_i^T \beta) \quad (3.2)$$

其中

$$W_i(u) = \frac{1}{nh} K\left(\frac{u-U_i}{h}\right) / \hat{f}(u)$$

$K(\cdot)$ 为已知的核函数, h 是光滑参数。 $\hat{f}(u) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{u-U_i}{h}\right)$ 为协变量 U 的密度函数 $f(u)$ 的核估计。然而,由于式(3.2)中 U_i 是不可直接观测的,实际上观测的是 Z_i , 故而,我们需对 $g(\cdot)$ 进行重新估计,结合 Fan 和 Truong^[32] 逆卷积的方法,我们得到 $g(\cdot)$ 的无偏估计:

$$\tilde{g}_n(z) \triangleq \tilde{g}_n(z, \beta) = \sum_{i=1}^n W_n(z)(Y_i - X_i^T \beta),$$

其中 $W_n(\cdot) \triangleq \frac{1}{nh} K_n\left(\frac{\cdot - Z_i}{h_n}\right) / \hat{f}_n(\cdot)$, $K_n(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-isu) \frac{\varphi_K(s)}{\phi_\epsilon(s/h)} ds$. $\varphi_K(\cdot)$ 为核函数 $K(\cdot)$ 的 Fourier 的变换, $\phi_\epsilon(\cdot)$ 为测量误差变量 v 的特征函数, i 是虚数单位,

经过计算得到一个近似残差

$$\tilde{\varepsilon}_i(\beta) = \tilde{Y}_i - \tilde{X}_i^T \beta \quad (3.3)$$

其中 $\tilde{Y}_i = Y_i - \sum_{j=1}^n W_j(Z_i) Y_j$, $\tilde{X}_i = X_i - \sum_{j=1}^n W_j(Z_i) X_j$ 。

引入一个辅助随机变量 $\hat{\zeta}_i(\beta) = \tilde{X}_i^T \tilde{\varepsilon}_i$ ，以此可以构造出一经验对数似然比函数：

$$\hat{l}_n(\beta) = -\max \left\{ \sum_{i=1}^n \log(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\zeta}_i(\beta) = 0 \right\}, \quad (3.4)$$

运用 Lagrange 乘子法，可以得到经验对数似然比统计量：

$$\hat{l}_n(\beta) = \sum_{i=1}^n \log(1 + \lambda^T \hat{\zeta}_i(\beta)), \quad (3.5)$$

其中 λ 为 Lagrange 乘子，满足：

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda^T \hat{\zeta}_i(\beta)} = 0. \quad (3.6)$$

3.2.2 惩罚经验似然方法

然而当我们的模型是高维的数据模型时，使用标准的经验似然来估计，将会有模型的偏差，为了平衡高维模型中的建模偏差和变量选择，结合惩罚的思想对我们的模型使用惩罚经验似然方法。我们对惩罚性经验似然(PEL)的定义如下：

$$L_n(\beta) = \hat{l}_n(\beta) + n \sum_{j=1}^p p_\gamma(|\beta_j|) \quad (3.7)$$

其中 p_γ 是一个惩罚函数， γ 为调整参数。该函数的例子见 Fan 和 Li^[40]。在本文中，我们使用 SCAD 惩罚函数，其一阶导数对于一些 $a > 2$ 和 $p'_\gamma(0) = 0$ ，满足

$$p'_\gamma(\theta) = \theta \left\{ I(\theta \leq \gamma) + \frac{a\gamma - \theta}{(a-1)\theta} + I \right\}, \theta > 0, \gamma > 0,$$

我们按照 Fan 和 Li^[40]的建议，取 $a = 3.7$ 。

最大化 PEL 函数式(3.7)等于最小化下式

$$L_n(\beta) = \sum_{i=1}^n \log(1 + \lambda^T \hat{\zeta}_i(\beta)) + n \sum_{j=1}^p p_\gamma(|\beta_j|). \quad (3.8)$$

使 $\mathcal{S} = \{j: \beta_{0j}\}$ 为真参数向量 β_0 的非零分量的集合，并且 $|\mathcal{S}| = d$ ，当 $n \rightarrow \infty$ 时，

其中 d 满足条件 B3。在不失一般性的前提下，记 $\beta = (\beta_1^T, \beta_2^T)^T$ 其中 $\beta_1 \in R^d$ ，

$\beta_2 \in R^{p-d}$ 。因此, 真参数 $\beta_0 = (\beta_{10}^r, 0)^r$ 。并将 $\hat{\beta} = (\hat{\beta}_1^r, \hat{\beta}_2^r)^r$ 称为 PEL 估计, 它是式 (3.8) 的最小值。

定理 3.1 假设 B1~B3 成立。当 $n \rightarrow \infty$ 时, 那么在概率趋于 1 的情况下, PEL 估计 $\hat{\beta}$ 满足以下条件:

(a) (稀释性): $\hat{\beta}_2 = 0$;

(b) (渐进正态性): $n^{1/2}W_n\Phi_g^{-1/2}(\hat{\beta}_1 - \beta_{10}) \xrightarrow{d} N(0, G)$, 其中对于固定的 q , 有 $W_n \in R^{q \times d}$, $G \in R^{q \times d}$, 满足 $W_n W_n^r \rightarrow G$, 且 $\Phi_g = \Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{21}$ 。

PEL 方法的显著优势在于检验假设和构建 β 的置信区间。为了更清楚地理解这一点, 我们考虑线性假设检验:

$$H_0: \Gamma_n \beta_{10} = 0 \quad \text{vs} \quad H_1: \Gamma_n \beta_{10} \neq 0$$

其中 Γ_n 是一个 $q \times s$ 的矩阵, 这样对于一个固定且有限的 q , 满足 $\Gamma_n \Gamma_n^r = I_q$ 。继续构造惩罚经验似然比统计量为:

$$\tilde{L}_n(\hat{\beta}) = -2\{L_n(\hat{\beta}) - \min_{\beta: L_n \beta = 0} L_n(\beta)\} \quad (3.9)$$

我们将检验统计量的性质总结为下面定理。

定理 3.2 在定理 3.1 的条件下。那么在原假设 H_0 下, 当 $n \rightarrow \infty$ 时有

$$\tilde{L}_n(\hat{\beta}) \xrightarrow{d} \chi_s^2$$

基于该定理的结果, $L_n \beta$ 的 $(1-\alpha)$ 水平置信域构造为

$$V_\alpha = \{v: -2\{L_n(\hat{\beta}) - \min_{\beta: L_n \beta = v} L_n(\beta)\} \leq \chi_{s, 1-\alpha}^2\}. \quad (3.10)$$

其中 $\chi_{s, 1-\alpha}^2$ 是 χ^2 分布的 $1-\alpha$ 水平分位数, 当 $n \rightarrow \infty$ 时, $P(\Gamma_n \beta_{10} \in V_\alpha) \rightarrow 1-\alpha$ 。据定理 3.2, 可以通过 V_α 给出 $\Gamma_n \beta_1$ 的 $1-\alpha$ 水平分位数下近似置信区间。

3.2.3 计算方法

本文采用局部二次逼近算法, 得到由式(3.8)定义的 PEL 比值的最小值。具体来说, 在 β_j 的初始值 β_{j0} 不接近于 0 时, 对于每个 $j=1, \dots, p$, 可以通过定义

为 $[p_\gamma(|\beta_j|)]' = p_\gamma(|\beta_j|) \text{sgn}(\beta_j) \approx \{p_\gamma(|\beta_{j_0}|/|\beta_{j_0}|)\}' \beta_j$ 的二次函数来局部逼近；当初始值接近 0 时，我们设置 $\beta_j = 0$ 。也就是说，在给定的非零 β_{j_0} 的领域中，设 $p_\gamma(|\beta_j|) \approx p_\gamma(|\beta_{j_0}|) + 2^{-1} \{p_\gamma(|\beta_{j_0}|/|\beta_{j_0}|)\} (\beta_j^2 - \beta_{j_0}^2)$ 。然后，我们应用经验似然算法(见 Owen)，通过非线性优化得到最小值。这个过程重复进行，直到收敛。

从定理 3.1 和 3.2 可以看出，参数 β 的估计值的渐近特性依赖于调节参数 γ 。遵循 Wang, Li 和 Tsai^[46]，我们采用 BIC 型准则来选择调节参数 γ ，其定义为

$$BIC(\gamma) = -2L_n(\beta_\gamma) + \log(n)df_\gamma,$$

其中， β_γ 为 β 的估计值， df_γ 为非零估计参数的数量。那么，最优调优参数为 BIC 的最小值。

3.3 模拟研究

在本节中，我们进行了一些蒙特卡洛研究，以研究所提出的有限样本估计器的性能。考虑下面的误差为鞅差序列的半参数测量误差模型：

$$\begin{cases} Y_i = X_i^T \beta + g(U_i) + \varepsilon_i, \\ Z_i = U_i + v_i \end{cases}$$

其中 $\beta = (3, 1.5, 0, 0, 2, 0, \dots, 0)^T$ ， $g(U_i) = \cos(2\pi U_i)$ ， $U_i = i/n$ ， X_i 和 v_i 服从 $N(0, 1)$ 。当 $\{\varepsilon_i, F_i, i \geq 1\}$ 是鞅差序列，先取鞅差序列 $\{\xi_i, \sigma(\xi_i), i \geq 1\}$ 并且令 $\varepsilon_i = \xi_{i+1} - \xi_i$ 。在此模拟中，我们生成第一个随机数 $\xi_1 \sim N(0, 1)$ ，然后根据条件分布 $\xi_{i+1} | \xi_i \sim N(\xi_i, 0.2^2), i = 1, \dots, n$ ，得到 ξ_2, \dots, ξ_{n+1} 。

核函数为 $K(u) = (\sqrt{2\pi})^{-1} \exp(-u^2/2) \{1 - \frac{1}{9h^2}(u^2 - 1)\}$ ，并采用“leave-one-subject-out”的交叉验证带宽法选择最优带宽 h_{opt} 。在模拟过程中，样本量 n 的取值分别为 100、200、300、500，其中 p 的值分别为 10, 20, 30, 40。并且我们将惩罚经验似然法选择出的变量与事先知道的模型稀疏性的最小二乘法得出的变量做了的均方根误差的对比，通过这样对比，从而可以更好的展示我们惩罚经验似然方法对于变量选择方面可以有优秀的性质。当 $n=200$ ， $p=30$ 和 $n=500$ ， $p=40$ 时， β_1 的 QQPlot 图和 β_2 的箱型图，都证实了样本数据的正态性，如下图 3-1、图 3-2 所示。

图 3-1 β_1 的 QQPlot 图，左图为当 $n=200, p=30$ 时；右图为当 $n=500, p=40$ 时

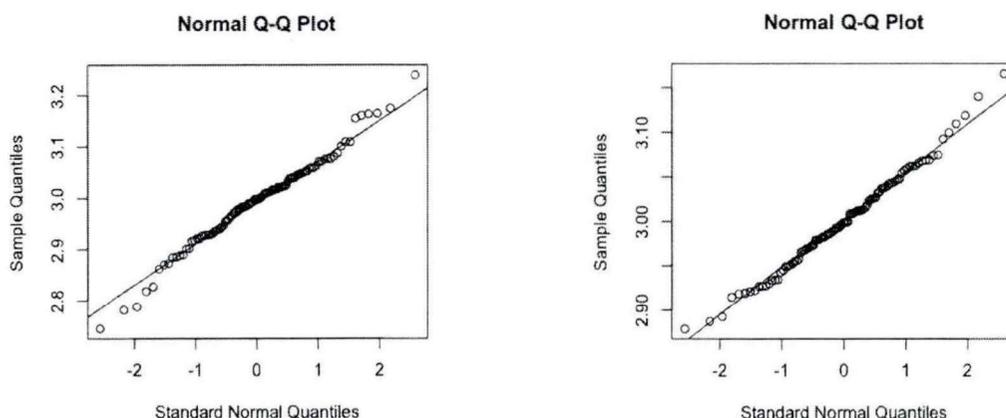
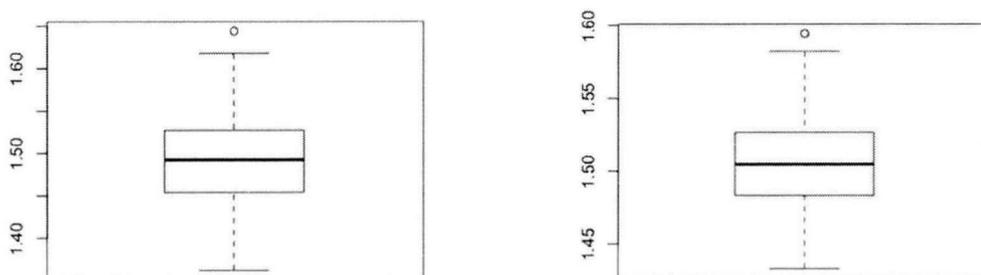


图 3-2 β_2 的箱型图左图为当 $n=200, p=30$ 时；右图为当 $n=500, p=40$ 时



从表 3-1 中可以看出，PEL 方法不会错误地将非零系数设为零。在所有的设置中，估计的零系数的平均数也都接近 $p-3$ 。这表明该模型在非零系数方面与真实模型非常接近。将 PEL 估计值 $\hat{\beta}$ 与事先已经知道模型的真实稀疏度的最小二乘估计法给出的 Oracle 估计值 $\hat{\beta}_{or}$ 进行比较。可以看出，惩罚经验似然的估计值的均方根误差略大，但是与 Oracle 估计的均方根误差相当接近。这印证了定理 3.1 中的理论结果，即 PEL 估计值近似地达到了 Oracle 估计值的有效性。另一方面，图 3-1 中表示，估计值的分布合理地拟合了正态分布，并且当样本量越大时，这种拟合度越好。

表 3-1: PEL 法在鞅差误差序列的半参数测量误差模型中估计回归系数的仿真结果

		均方根误差			0 系数	
		β_1	β_2	β_3	正确	错误
p=10,n=50	$\hat{\beta}$	0.114	0.117	0.109	6.832	0
	$\hat{\beta}_{or}$	0.106	0.114	0.101	—	—
p=20,n=100	$\hat{\beta}$	0.078	0.075	0.071	16.864	0
	$\hat{\beta}_{or}$	0.076	0.074	0.070	—	—
p=30,n=200	$\hat{\beta}$	0.049	0.052	0.049	26.918	0
	$\hat{\beta}_{or}$	0.047	0.050	0.048	—	—
p=40,n=500	$\hat{\beta}$	0.033	0.034	0.032	36.972	0
	$\hat{\beta}_{or}$	0.032	0.032	0.031	—	—

表 3-2 β_1 的值不属于式(3.2)所构造的 95%置信度内的经验频率, β_1 的真值为 3

β_1	2.8	2.9	3.0	3.1	3.2
p=10,n=50	0.202	0.174	0.074	0.112	0.252
p=20,n=100	0.385	0.172	0.072	0.173	0.399
p=30,n=200	0.674	0.225	0.069	0.236	0.640
p=40,n=500	0.944	0.494	0.063	0.477	0.949

在式(3.1)中的 $\Gamma_n = (1, 0, \dots, 0)^T$, 考察了由式(3.2)构造的 $\beta_1 \notin V_\alpha$ 的经验拒绝频率, 其中 β_1 是 β 的第一个元素。由表 3-2 可知, 当 β_1 为模拟的真实值时, 模拟中的经验排斥率接近于 0.05 的置信水平。这也恰巧证实了定理 3.2 的结果。

3.4 实例分析

在这一节中, 我们分析了一组环境数据, 包括 1994 年 1 月 1 日至 1997 年 12 月 31 日期间香港的污染物和其他环境因素的日平均测量值。在 1461 天的数

据中,有 8 个记录的变量。它们主要是 NO_2 : 氮氧化物(g/m^3); Tem : 空气温度(摄氏度); Rsp : 可吸入颗粒物(g/m^3); SO_2 : 二氧化硫(g/m^3); O_3 : 臭氧(g/m^3); Hum : 相对湿度(%); Resp : 呼吸系统疾病每日住院总人数之和等。为了研究空气污染物与呼吸系统疾病每日住院总人数之间的关系,我们考虑以下模型:

$$\text{Resp} = \beta_1 \text{NO}_2 + \beta_2 \text{SO}_2 + \beta_3 \text{Rsp} + \beta_4 \text{O}_3 + \beta_5 \text{Hum} + g(u) + \varepsilon$$

在表 3-3 中报告了参数的 PEL 和最小二乘法(LS)估计值。表 3-3 显示, NO_2 在影响每日住院总人数中占据了非常重要的作用。其中, SO_2 和 Rsp 的系数显著为负, O_3 的系数由 PEL 估计为 0。此外,我们用 PEL 比值来检验 O_3 是否对 Resp 有显著的推导作用,最后对于下面的原假设 $H_0: \beta_4 = 0$, 得到 p 值为 0.4882。因此,我们得出结论, O_3 对 Resp 没有显著性的影响或者说是没有影响的,这与表 3-3 中的 PEL 估计值一致。

表 3-3 PEL 方法得到环境数据的系数估计值与 LS 方法对比

	PEL	LS
NO_2	0.6105	0.7669
SO_2	-0.5664	-0.5465
Rep	-0.2617	-0.2407
O_3	0	0.0494
Hum	0.3445	0.3681

3.5 主要结果证明

在进行主要结果的证明之前需要给出一些假设条件

B1 对于引理 3.1 中所述的 δ , 满足条件 $p^2/n^{1-(4k)} \rightarrow 0, p^{1-2\delta}/n^{1/2-2\delta} \rightarrow 0$, 当

$n \rightarrow \infty, p \rightarrow \infty$ 时。

B2 当 $n \rightarrow \infty$ 时, 调节参数 γ 满足当 $\gamma \rightarrow \infty$ 时, $\gamma(n/p)^{1/2-\delta} \rightarrow \infty$ 。

B3 $\max_{j \in \mathcal{S}} p_\gamma'(|\beta_{0j}|) = o\{(\sqrt{np})^{-1}\}, \max_{j \in \mathcal{S}} p_\gamma''(|\beta_{0j}|) = o(p^{-1/2+\delta} n^{-\delta})$ 。

引理 3.1 假设 B1~B3 成立, 定义 $D_n = \{\beta : \|\beta - \beta_0\| \leq ch_n\}$, $h_n = (p/n)^{1/2-\delta}$,

当 $\beta \in D_n$ 时, 有 $\|\lambda_\beta\| = O_p(h_n)$ 。

证明: 首先, 有:

$$\rho \left\{ \theta^T T_n(\beta) \theta - \max_{1 \leq i \leq n} \|\hat{\zeta}_i(\beta)\| n^{-1} \left| \sum_{i=1}^n \theta^T \hat{\zeta}_i(\beta) \right\} \leq n^{-1} \left| \sum_{i=1}^n \theta^T \hat{\zeta}_i(\beta) \right| \right\} \quad (3.11)$$

其中 $T_n(\beta) = \frac{1}{n} \sum_{i=1}^n \hat{\zeta}_i(\beta) \hat{\zeta}_i^T(\beta)$ 。对于 $\beta \in D_n$, 令 $\lambda_\beta = \rho\theta$ 以及 $\|\theta\|=1$ 为单位向量,

结合式(3.6)以及 $\lambda_\beta \in R^p$ 有

$$\begin{aligned} 0 &= \left| \frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda_\beta^T \hat{\zeta}_i(\beta)} \right| = \frac{1}{n} \left| \theta^T \left\{ \sum_{i=1}^n \hat{\zeta}_i(\beta) - \rho \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta) \theta^T \hat{\zeta}_i(\beta)}{1 + \rho \theta^T \hat{\zeta}_i(\beta)} \right\} \right| \\ &\geq \frac{\rho}{n} \left| \theta^T \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta) \theta^T \hat{\zeta}_i(\beta)}{1 + \rho \theta^T \hat{\zeta}_i(\beta)} \theta - \frac{1}{n} \sum_{i=1}^n \theta^T \hat{\zeta}_i(\beta) \right| \\ &\geq \rho \theta^T T_n(\beta) \theta \left\{ 1 + \rho \max_{1 \leq i \leq n} \|\hat{\zeta}_i(\beta)\| \right\}^{-1} - \frac{1}{n} \left| \sum_{i=1}^n \theta^T \hat{\zeta}_i(\beta) \right| \end{aligned}$$

X_i 满足式(3.8), 且鞅差序列误差 ε_i 和 X_i 是独立分布的, 根据鞅差的性质以及构造的辅助变量简单计算可得

$$\|\lambda_{\beta_0}\| = O_p(\sqrt{p/n})。$$

除了 β_0 以外, 对于 $\beta \in D_n$, 容易推导出

$$\frac{1}{n} \left| \sum_{i=1}^n \theta^T \hat{\zeta}_i(\beta) \right| = \frac{1}{n} \left| \sum_{i=1}^n \theta^T \left\{ \hat{\zeta}_i(\beta_0) - \tilde{X}_i \tilde{X}_i^T (\beta - \beta_0) \right\} \right| \quad (3.12)$$

由于

$$T_n(\beta) - T_n(\beta_0) = -2 \frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^T \tilde{X}_i^T (\beta - \beta_0) \varepsilon_i + \frac{1}{n} \sum_{i=1}^n \tilde{X}_i \tilde{X}_i^T \left\{ \tilde{X}_i^T (\beta - \beta_0) \right\}^2 \quad (3.13)$$

可得

$$\theta^T T_n(\beta) \theta = \theta^T T_n(\beta_0) \theta + o_p(1) = O_p(1)$$

$$\max_{1 \leq i \leq n} \|\tilde{X}_i\| = o(\sqrt{n})$$

$$\max_{1 \leq i \leq n} \|\tilde{X}_i \tilde{X}_i^\tau (\beta - \beta_0)\| \leq ch_n \max_i \|\tilde{X}_i\|^2 = o_p(1)$$

至此结合条件 B1 得到

$$\max_i \|\hat{\zeta}_i(\beta)\| \frac{1}{n} \left| \sum_{i=1}^n \theta^\tau \hat{\zeta}_i(\beta) \right| = o_p(1) \quad (3.14)$$

结合式(3.12)~(3.14), 此引理得证。

引理 3.2 假设 B1~B3 成立, 当 $n \rightarrow \infty$ 时, 依概率 1 有, $L_n(\beta)$ 在 D_n 内有最小值。

证明: 对于 $\beta \in D_n$, 有

$$Q_{1n}(\beta, \lambda_\beta) = \frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda_\beta^\tau \hat{\zeta}_i(\beta)} = 0$$

根据引理 3.1, 有 $\lambda_\beta^\tau \hat{\zeta}_i(\beta) = o_p(1)$ 。对 $Q_{1n}(\beta, \lambda_\beta)$ 泰勒展开, 得到

$$0 = n^{-1} \sum_{i=1}^n \hat{\zeta}_i(\beta) - T_n(\beta) \lambda_\beta + r_n,$$

其中 $r_n = n^{-1} \sum_{i=1}^n \hat{\zeta}_i(\beta) \{\lambda_\beta^\tau \hat{\zeta}_i(\beta)\}^2 (1 + \xi_i)^{-3}$, $|\xi_i| \leq |\lambda_\beta^\tau \hat{\zeta}_i(\beta)|$ 。

则 $\lambda_\beta = T_n^{-1}(\beta) \bar{\zeta}(\beta) + T_n^{-1}(\beta) r_n$, 其中 $\bar{\zeta}(\beta) = n^{-1} \sum_{i=1}^n \hat{\zeta}_i(\beta)$, 将 λ 代入式(3.5),

容易得到:

$$2\hat{l}_n(\beta) = n \bar{\zeta}(\beta)^\tau T_n^{-1}(\beta) \bar{\zeta}(\beta) - nr_n^\tau T_n^{-1}(\beta) r_n + \frac{2}{3} \sum_{i=1}^n \{\lambda_\beta^\tau \hat{\zeta}_i(\beta)\}^3 (1 + \xi_i)^{-4}$$

注意到 $\hat{\zeta}_i(\beta) = \tilde{X}_i \tilde{\varepsilon}_i$, 对于 $\beta \in \partial D_n$, 其中 ∂D_n 是 D_n 的边界。有了这些, 不难证明下式:

$$2\hat{l}_n(\beta) - 2\hat{l}_n(\beta_0) = n \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{X}_i \tilde{X}_i^\tau) (\beta - \beta_0) \right\}^\tau T_n^{-1}(\beta) \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{X}_i \tilde{X}_i^\tau) (\beta - \beta_0) \right\} = O_p(nh_n^2)$$

这意味着对于任何给定的 C 来说, 当 $n \rightarrow \infty$ 时, 有 $P\left[2\hat{l}_n(\beta) - 2\hat{l}_n(\beta_0) > C\right] \rightarrow 1$ 。

并且, 当 n 足够大时

$$\begin{aligned} L_n(\beta) - L_n(\beta_0) &= \hat{l}_n(\beta) - \hat{l}_n(\beta_0) + n \sum_j \left\{ p_\gamma(|\beta_j|) - p_\gamma(|\beta_{0j}|) \right\} \\ &\geq \hat{l}_n(\beta) - \hat{l}_n(\beta_0) + n \sum_{j \in \mathcal{G}} \left\{ p_\gamma(|\beta_j|) - p_\gamma(|\beta_{0j}|) \right\} \geq \hat{l}_n(\beta) - \hat{l}_n(\beta_0) \end{aligned}$$

其中, 由于条件 B3 和 SCAD 惩罚函数的无偏属性, 最后一个不等式成立。这样, 也就是当 n 足够大的时候, $j \in \mathcal{G}$, $p_\gamma(|\beta_{0j}|) = p_\gamma(|\beta_j|)$ 。因此, 当 $\beta \in \partial D_n$ 时, 有 $P\{L_n(\beta) - L_n(\beta_0)\} \rightarrow 1$ 。至此本引理成立。

定理 3.1 的证明: 由引理 3.2, 我们注意到 $L_n(\beta)$ 在 D_n 内有最小值, 那么对于 $\beta \in D_n$ 的每个组成部分

$$\frac{1}{n} \frac{\partial L_n(\beta)}{\partial \beta_j} = \frac{1}{n} \sum_{i=1}^n \frac{-\lambda_\beta^r \tilde{X}_i \tilde{X}_{ij}}{1 + \lambda_\beta^r \hat{\zeta}(\beta)} + p'_\gamma(|\beta_j|) \text{sign}(\beta_j) = I_j + \Pi_j,$$

根据引理 3.1, 我们有 $\max_{1 \leq i \leq n} |\lambda_\beta^r \tilde{X}_i| = o_p(1)$, 通过选择一个足够大的 n , 我们可以得到:

$$\max_{j \in \mathcal{G}} |I_j| \leq \max_i |\lambda_\beta^r \tilde{X}_i| \max_{j \in \mathcal{G}} \left| n^{-1} \sum_{i=1}^n \tilde{X}_{ij} \right| \leq \max_i |\lambda_\beta^r \tilde{X}_i| \left\| n^{-1} \sum_{i=1}^n \tilde{X}_i \right\| = o_p(\sqrt{p/n}),$$

结合引理 3.1 以及条件 B3, 可以得出 $\gamma(n/p)^{1/2-\delta} \rightarrow \infty, P(\max_{j \in \mathcal{G}} |I_j| > \gamma/2) \rightarrow 0$ 。

另外, 当 $j \notin \mathcal{G}, \gamma/h_n \rightarrow \infty$, 以及 $\Pi_j = p'_\gamma(|\beta_j|) \text{sign}(\beta_j) = \gamma \text{sign}(\beta_j)$ 时, 可以看出,

对于所有的 $j \notin \mathcal{G}, p'_\gamma(|\beta_j|) \text{sign}(\beta_j)$ 渐近地主导 $\frac{\partial L_n(\beta)}{\partial \beta_j}$ 的符号。并且当 $n \rightarrow \infty$ 时,

对于任意的 $j \notin \mathcal{G}$, 依概率 1 有:

$$\frac{\partial L_n(\beta)}{\partial \beta_j} \geq 0, \beta \in (0, ch_n); \frac{\partial L_n(\beta)}{\partial \beta_j} \leq 0, \beta \in (-ch_n, 0),$$

这意味着, 依概率趋于 1 有 $\hat{\beta}_2 = 0$ 。至此, 定理 3.1 的(a)部分已经被证明。

接着证明(b)部分, 令 Ψ_2 是 $\Psi_2 \beta = \beta_2$ 这样的矩阵。当 $n \rightarrow \infty, P(\hat{\beta}_2 = 0) \rightarrow 1$ 时, 用拉格朗日乘子法找到 $L_n(\beta)$ 的极小值, 可以等效的通过解决以下目标函数的最小化问题来得到

$$\frac{1}{n} \sum_{i=1}^n \log \{1 + \lambda^T \hat{\zeta}_i(\beta)\} + \sum_{j=1}^p p_\gamma(|\beta_j|) + \nu^T \Psi_2 \beta.$$

其中 ν 是另一个拉格朗日乘子的 $p-s$ 维列向量。定义

$$\begin{aligned} \tilde{Q}_{1n}(\beta, \lambda, \nu) &= -\frac{1}{n} \sum_{i=1}^n \frac{\hat{\zeta}_i(\beta)}{1 + \lambda^T \hat{\zeta}_i(\beta)}, \\ \tilde{Q}_{2n}(\beta, \lambda, \nu) &= -\frac{1}{n} \sum_{i=1}^n \frac{\tilde{X}_i \tilde{X}_i^T \lambda}{1 + \lambda^T \hat{\zeta}_i(\beta)} + b(\beta) + \Psi_2^T \nu, \\ \tilde{Q}_{1n}(\beta, \lambda, \nu) &= \Psi_2 \beta. \end{aligned}$$

其中 $b(\beta) = \{p_\gamma(|\beta_1|) \text{sign}(\beta_1), p_\gamma(|\beta_2|) \text{sign}(\beta_2), \dots, p_\gamma(|\beta_s|) \text{sign}(\beta_s), 0, \dots, 0\}^T$ 。

并将 $\tilde{Q}_{jn}(\beta, \lambda, \nu)$ 在 $(\beta_0, 0, 0)$ 的扩展成如下式:

$$\begin{pmatrix} \tilde{Q}_{1n}(\beta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\Sigma & -\Sigma & 0 \\ -\Sigma & 0 & \Psi_2^T \\ 0 & \Psi_2 & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \\ \hat{\nu} \end{pmatrix} + R_n, \quad (3.15)$$

其中 $R_n = \sum_{k=1}^5 R_n^{(k)}$, $R_n^{(1)} = (R_{1n}^{r(1)}, R_{1n}^{r(2)}, 0)^T$, $R_{jn}^{(1)} \in R^p$, 并且当 $j=1, 2$ 时, $R_{jn}^{(1)}$ 的第 k 个分量由下式给出:

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

$\eta = (\beta, \lambda)^T$, $\tilde{\eta} = (\tilde{\beta}, \tilde{\lambda})^T$ 满足 $\|\tilde{\eta} - \eta_0\| \leq \|\hat{\eta} - \eta_0\|$, $\|\tilde{\lambda}\| \leq \|\hat{\lambda}\|$. $R_n^{(2)} = \{0, b^T(\beta_0), 0\}^T$ 和接下

来的 $R_n^{(3)} = \{0, \{b^T(\beta_0)(\hat{\beta} - \beta_0)\}^T, 0\}^T$, $R_n^{(4)} = \{ \{(T_n(\beta_0) - \Sigma)\hat{\lambda}\}^T, 0, 0\}^T$, 以及最后的

$R_n^{(5)} = \{ \{(T_n(\beta_0) - \Sigma)\hat{\lambda}\}^T, \{(T_n(\beta_0) - \Sigma)(\hat{\beta} - \beta_0)\}^T, 0\}^T$ 。结合条件 A9 以及 SCAD 惩罚函数, 可以得到 $\|R_n^{(2)}\| = o_p(1/\sqrt{n})$, $\|R_n^{(3)}\| = o_p(1/\sqrt{n})$ 。

根据 $\|T_n(\beta_0) - \Sigma\| = O_p(\sqrt{p/n})$ 及 $\|\hat{\beta} - \beta_0\| = O_p(h_n)$, $\|\hat{\lambda}_\beta\| = O_p(h_n)$, 可以得出 $\|R_n^{(4)}\| = o_p(1/\sqrt{n})$, $\|R_n^{(5)}\| = o_p(1/\sqrt{n})$ 。因此有 $\|R_n^{(k)}\| = o_p(1/\sqrt{n})$, $k=1, \dots, 5$ 。

接着定义 $K_{11} = -\Sigma$, $K_{12} = [-\Sigma, 0]$ 以及 $K_{21} = K_{12}^T$,

$$K_{22} = \begin{pmatrix} 0 & \Psi_2^r \\ \Psi_2 & 0 \end{pmatrix}, K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

并使 $\omega = (\beta^r, v^r)$ 。通过对式(3.15)的简单求逆运算, 有

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\omega} - \omega_0 \end{pmatrix} = K^{-1} \left\{ \begin{pmatrix} -\tilde{Q}_{1n}(\beta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} - R_n \right\},$$

由于矩阵 K 分为四个块, 因此可以按以下方式逐块进行求逆,

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

其中 $A = K_{22} - K_{21}K_{11}^{-1}K_{12} = \begin{bmatrix} \Sigma^{-1} & \Psi_2^r \\ \Psi_2 & 0 \end{bmatrix}$ 。因此, 得到有

$$\hat{\omega} - \omega_0 = A^{-1}K_{21}K_{11}^{-1}\tilde{Q}_{1n}(\beta_0, 0, 0) + o_p(1/\sqrt{n}).$$

矩阵 A 也可以使用求逆公式逐块进行求逆, 即

$$A^{-1} = \begin{bmatrix} \Sigma - \Sigma\Psi_2^r(\Psi_2\Sigma\Psi_2^r)^{-1}\Psi_2\Sigma & \Sigma\Psi_2^r(\Psi_2\Sigma\Psi_2^r)^{-1} \\ (\Psi_2\Sigma\Psi_2^r)^{-1}\Psi_2\Sigma & -(\Psi_2\Sigma\Psi_2^r)^{-1} \end{bmatrix}$$

此外, 还有

$$\hat{\beta} - \beta_0 = \left\{ \Sigma^{-1} - \Sigma^{-1}\Psi_2^r(\Psi_2\Sigma\Psi_2^r)^{-1}\Psi_2\Sigma^{-1} \right\} \left(n^{-1} \sum_{i=1}^n \tilde{X}\tilde{\varepsilon}_i + R_n \right),$$

其中 R_n 是 R_n 以及 $\|R_n\| = o_p(1/\sqrt{n})$ 中的对应分量。接下来是 $\hat{\beta}_1$ 的展开,

$$\hat{\beta} - \beta_0 = \left\{ \Psi_1\Sigma^{-1} - \Psi_1\Sigma^{-1}\Psi_2^r(\Psi_2\Sigma\Psi_2^r)^{-1}\Psi_2\Sigma^{-1} \right\} \left(n^{-1} \sum_{i=1}^n \tilde{X}\tilde{\varepsilon}_i + R_n \right). \quad (3.16)$$

$\sqrt{n}\hat{\beta}_1 = \sqrt{n}\Psi_1\hat{\beta}_1$ 的渐近协方差为:

$$\Phi_g = \Psi_1\Sigma^{-1}\Psi_1^r - \Psi_1\Sigma^{-1}\Psi_1^r(\Psi_2\Sigma^{-1}\Psi_2^r)^{-1}\Psi_2\Sigma^{-1}\Psi_1^r.$$

通过分解 $\Sigma^{-1} = \begin{pmatrix} \dot{\Sigma}_{11} & \dot{\Sigma}_{12} \\ \dot{\Sigma}_{21} & \dot{\Sigma}_{22} \end{pmatrix}$, 其中 $\dot{\Sigma}_j = \Psi_j\Sigma^{-1}\Psi_j^r$, 得到有 $\Phi_g = \dot{\Sigma}_{11} - \dot{\Sigma}_{12}\dot{\Sigma}_{22}^{-1}\dot{\Sigma}_{21}$ 。然

后, 在 Σ 上应用块逆矩阵, 我们得到

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}F^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{11}F^{-1} \\ -F^{-1}\Sigma_{21}\Sigma_{11}^{-1} & F^{-1} \end{pmatrix},$$

其中 $F = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ 。因此 $\Phi_g = \Sigma_{11}^{-1}$ ，结合 $\|\sqrt{n}W_n\Phi_g^{-1/2}R_{1n}\| = o_p(1)$ 和 Lindeberg-Feller 中心极限定理，(b)的部分也证明完毕。

定理 3.2 的证明：令 $\hat{\beta}$ 为式 (3.9) 中最小值，并泰勒展开有

$$L_n(\beta) = \sum_{i=1}^n \hat{\lambda}^r \hat{\zeta}_i(\hat{\beta}) - \sum_{i=1}^n \frac{(\hat{\lambda}^r \zeta_i(\hat{\beta}))^2}{2} + \sum_{i=1}^n \frac{(\hat{\lambda}^r \zeta_i(\hat{\beta}))^3}{3(1+\xi_i)^4} + o_p(1).$$

其中 $|\xi_i| \leq |\hat{\lambda}^r \zeta_i(\hat{\beta})|$ 以及 $o_p(1)$ 是由惩罚函数决定。并将 $\hat{\beta}$ 的扩展和 $\hat{\lambda}$ 代入到 $|\xi_i| \leq |\hat{\lambda}^r \zeta_i(\hat{\beta})|$ 中，我们可以得到

$$2L_n(\beta) = n\bar{\zeta}(\beta)^r \Psi_2^r (\Psi_2 \Sigma^{-1} \Psi_2^r)^{-1} \Psi_2 \bar{\zeta}(\beta) + o_p(1). \quad (3.17)$$

在原假设下，由于 $\Gamma_n \Gamma_n^r = I_q$ ，则存在 $\tilde{\Psi}_2$ 且 $\tilde{\Psi}_2 \beta = 0$ 以及 $\tilde{\Psi}_2 \tilde{\Psi}_2^r = I_{p-d+q}$ 。现在，通过重复定理 3.1 的证明，我们确立了在原假设下， β 的估计值可以通过最小化式 (3.16) 被估计出，可以轻易得到

$$2L_n(\tilde{\beta}) = n\bar{\zeta}(\beta)^r \tilde{\Psi}_2^r (\tilde{\Psi}_2 \Sigma^{-1} \tilde{\Psi}_2^r)^{-1} \tilde{\Psi}_2 \bar{\zeta}(\beta) + o_p(1).$$

并结合式(3.17)，有

$$L_n = n\bar{\zeta}(\beta)^r \Sigma^{-1/2} \{P_1 - P_2\} \Sigma^{-1/2} \bar{\zeta}(\beta) + o_p(1).$$

其中 $P_1 = \Sigma^{-1/2} \tilde{\Psi}_2^r (\tilde{\Psi}_2 \Sigma^{-1} \tilde{\Psi}_2^r) \tilde{\Psi}_2 \Sigma^{-1/2}$ 和 $P_2 = \Sigma^{-1/2} \Psi_2^r (\Psi_2 \Sigma^{-1} \Psi_2^r) \Psi_2 \Sigma^{-1/2}$ 是两个同幂矩阵。由于 $P_1 - P_2$ 的维度为 q ， $P_1 - P_2$ 可以被写为 $\Omega_n^r \Omega_n$ ，其中 Ω 是 $p \times q$ 维矩阵则 $\Omega_n^r \Omega_n = I_q$ ，更多的，得到有

$$\sqrt{n}\Omega_n \Sigma^{-1/2} \bar{\zeta}(\beta) \xrightarrow{d} N(0, I_q).$$

之后得到

$$n\bar{\zeta}(\beta)^r \Sigma^{-1/2} (P_1 - P_2) \Sigma^{-1/2} \bar{\zeta}(\beta) \xrightarrow{d} \chi_q^2,$$

至此，定理 3.2 证明完毕。

3.6 本章小结

本文研究了误差是鞅差序列下的高维半参数测量误差模型的惩罚经验似然推断。在提出的模型中首先得到了参数的无偏估计，随后应用经验似然的方法构造经验似然统计量，并结合惩罚的思想，推出惩罚经验似然法应用于模型当中，构造了惩罚经验似然统计量并借此得到其置信域，以及通过使用适当的方法，证明了惩罚经验似然估计器具有 oracle 属性。同时，通过数值模拟和实例分析的应用来评估我们提出的方法的样本性能，数值模拟跟实例分析的结果都印证了将惩罚经验似然法应用于高维半参数回归模型是有着令人满意的结果的。

第 4 章 带有固定效应的高维部分线性变系数面板数据模型的惩罚经验似然推断

对于带有固定效应的高维部分线性变系数面板数据模型，本文使用了 B 样条的方法估计非参数，并在采用局部线性虚拟变量的方法消除固定效应后，引入了一个辅助函数，以此构造经验对数似然比函数，然后结合惩罚函数的思想提出了具有参数数量发散的惩罚经验似然(PEL)和变量选择过程，通过使用适当的惩罚函数，证明了惩罚经验似然估计量具有 Oracle 属性。

4.1 模型框架

在大量的统计问题中，灵活和完善的统计模型被广泛寻求。很多研究都是为了推导出未知参数或函数的估计和检验。Xia 等^[47]研究了附加限制条件的变系数模型；He^[48]近来研究了半参数变系数模型的面板计数数据的半参数回归分析。考虑如下带固定效应的部分线性变系数面板数据模型：

$$Y_{it} = X_{it}^T \beta + Z_{it}^T \alpha(U_{it}) + \mu_i + \varepsilon_{it}. \quad (4.1)$$

其中 Y_{it} 是响应变量， $(X_{it}, U_{it}) \in R^p \times R$ 是严格的外生变量， $\beta = (\beta_1, \dots, \beta_p)^T$ 为 p 维未知参数向量， $\alpha(\cdot) = (\alpha_1(\cdot), \dots, \alpha_q(\cdot))^T$ 是一个未知的 q 维系数函数向量， ε_{it} 是随机的模型误差，以及 $Z_{it} \in R^q$ 是协变量。 μ_i 是一个未知的与 (X_{it}, Z_{it}, U_{it}) 相关的结构。本模型是一个固定效应的面板数据模型，面板数据的分析是计量经济学中最活跃和最具创新性的文献之一。与纯粹的时间序列或横断面数据集相比，面板数据集具有各种优点，其中最重要的一点可能是面板数据为研究人员提供了一种灵活的方法，既可以模拟横断面单元之间的异质性，也可以模拟随着时间推移可能发生的结构变化。对于半参数面板数据模型，大量学者都进行了研究，如 Chen 等^[49]研究的是带有固定效应的部分线性单指标面板数据模型，他们提出了一个虚拟变量的方法去消除固定效应，并建立了一种半参数最小二乘平均方差估计方法；Su 和 Ullah^[50]提出了一类具有随机效应的非参数面板数据的两步估计；以及更多学者的研究可以见 Rodriguez-Poo^[51]，Li 等^[52]以及 Hu^{[53][54]}等等。

4.2 方法与主要结果

4.2.1 经验似然法

设 $B_k(\cdot)$, $k=1, \dots, K$, 在区间 $[0,1]$ 上 m 次 B 样条函数的基, 且 $u_i = 1/N_n$ 为节点序列, $i=0, \dots, N_n$, $K = N_n + m$ 。用 B 样条估计方法来估计未知非参数有着许多优点, 这些优点在许多学者的文章中都有介绍, 如 Chen 和 Wang^[55], Tang^[56], Jin 等^[57]。假设每个 $\alpha_s(t)$ 都可以被样条函数 $\alpha_s^*(t)$ 很好地近似, 于是类似 De^[58] 的文中 $\sup_{t \in D} |\alpha_s(t) - \alpha_s^*(t)| = 0$, 即

$$\alpha_s(t) \approx \alpha_s^*(t) = \sum_{k=1}^K \ell_{ks} B_k(u), s=1, \dots, d.$$

其中 ℓ 是样条系数。

为了方便运算, 令 $Y = (Y_1^T, \dots, Y_n^T)^T$, $X = (X_1^T, \dots, X_n^T)^T$, $Z = (Z_1^T, \dots, Z_n^T)^T$, $U = (U_1^T, \dots, U_n^T)^T$, $\mu_0 = (\mu_1^T, \dots, \mu_n^T)^T$, $\varepsilon = (\varepsilon_1^T, \dots, \varepsilon_n^T)^T$ 为 $nT \times 1$ 向量。更加的, 其中 $Y_i = (Y_{i1}, \dots, Y_{iT})^T$, $X_i = (X_{i1}, \dots, X_{iT})^T$, $Z_i = (Z_{i1}, \dots, Z_{iT})^T$, $U_i = (U_{i1}, \dots, U_{iT})^T$, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})^T$, 以及 $D_0 = I_n \otimes i_T$, \otimes 代表 Kronecker 积, I_n 为 $n \times n$ 的单位矩阵, i_n 为 $n \times 1$ 维 1 的向量。至此将模型 4.1 以矩阵形式重写为:

$$Y = X\beta + V\ell + D_0\mu_0 + \varepsilon. \quad (4.2)$$

其中 $V_i = Z_i \otimes B(u)$ 阵形, $V = (V_1, \dots, V_n)^T$, $B(u) = (B_1(u), \dots, B_K(u))^T$ 。本文施加了一个限制 $\sum_{i=1}^n \mu_i = 0$, 再令 $D = [-i_{n-1} \quad I_{n-1}] \otimes i_T$ 以及 $\mu = (\mu_2, \dots, \mu_n)^T$ 。将模型(4.2)改写为:

$$Y = X\beta + V\ell + D\mu + \varepsilon. \quad (4.3)$$

令 $\theta = (\mu^T, \beta^T)^T$, 假设 θ 已给定, 借此可以通过解决接下来的式子来得到 ℓ 的估计:

$$\sum_{i=1}^n \sum_{t=1}^T \{Y_{it} - X_{it}^T \beta - \mu_i - Z_{it}^T \sum_{s=1}^d \sum_{k=1}^K B_k(U_{it}) \ell_{ks}\}^2. \quad (4.4)$$

至此可以得到 $\alpha(\cdot)$ 的估计为:

$$\hat{\alpha}(u) = B\ell = B(V^r V)^{-1}V(Y - X\beta - D\mu). \quad (4.5)$$

其中 $\ell = (V^r V)^{-1}V(Y - X\beta - D\mu)$ 。

在参数面板数据分析中,可以考虑一种由最小二乘虚拟变量模型驱动的去除未知固定效应的方法,类似的解决以下优化问题:

$$\hat{\theta} = \arg \min_{\theta} [Y - X\beta - D\mu - S(Y - X\beta - D\mu)]^r [Y - X\beta - D\mu - S(Y - X\beta - D\mu)].$$

其中 $S = (Q_1^r Z_1, \dots, Q_1^r Z_n)^r$, 且 $Q_i = (I \ 0)B(V^r V)^{-1}V$, 令 $\tilde{Y} = (I - S)Y$, $\tilde{X} = (I - S)X$,

$\tilde{Z} = (I - S)Z$, $\tilde{D} = (I - S)D$ 。可得固定效应 μ 的估计 $\tilde{\mu} = (\tilde{D}^r \tilde{D})^{-1} \tilde{D}^r (\tilde{Y} - \tilde{X}\beta)$,

至此加入一个投影矩阵 $H = I_{nT} - \tilde{D}(\tilde{D}^r \tilde{D})^{-1} \tilde{D}^r$, 显而易见的 $H\tilde{D}\mu = 0$, 至此固定效应解决。引入一个合适的辅助函数:

$$\eta_i(\beta) = \sum_{t=1}^T \tilde{X}_{it}^r H(\tilde{Y}_{it} - \tilde{X}_{it}^r \beta), \quad i = 1, \dots, n \quad (4.6)$$

应用 Owen 的 EL 方法, 可以构造出参数 β 的经验对数似然比函数(ELR)如下:

$$l_n(\beta) = -\max \left\{ \sum_{i=1}^n \ln(np_i) \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \eta_i(\beta) = 0 \right\}. \quad (4.7)$$

通过 Lagrange 乘子法, 显然 $p_i = 1/\{n(1 + \lambda^r \eta_i(\beta))\}$ 以及 $l_n(\beta)$ 可以被写为

$$l_n(\beta) = \sum_{i=1}^n \ln(1 + \lambda^r \eta_i(\beta)). \quad (4.8)$$

其中 λ 为 Lagrange 乘子, 满足:

$$\frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^r \eta_i(\beta)} = 0. \quad (4.9)$$

定理 4.1 在条件 C1 到 C5 成立的假设下, 当 $8 \leq r$ 或者 $E(\varepsilon_1^3 \mid U_1, X_1, Z_1) = 0, \text{a.s.}$ 时, β_0 是的真实值 β , $p^{3+2/(r-2)}/n \rightarrow 0$, 那么 $(2p)^{-1/2}(2l_n(\beta_0) - p) \xrightarrow{d} N(0, 1)$, 当 $n \rightarrow \infty$ 时。

定义 $\hat{\beta}_{ME} = \arg \min_{\beta} l_n(\beta)$, 它是参数 β 的最大经验似然(MEL)估计值, 并且令

$$Y(u) = E(Z_1 Z_1^r | U = u), \quad \mathfrak{U}(u) = E(Z_1 X_1^r | U = u)。$$

定理 4.2 在定理 4.1 的条件下, 当 $E(\varepsilon_1^2 | U_1, X_1, Z_1) = \delta^2$, a.s.时, 可以有

$$n^{1/2} \delta^2 A_n \Omega^{1/2} (\hat{\beta}_{ME} - \beta_0) \xrightarrow{d} N(0, \Delta)。$$

其中 A_n 是 $s \times p$ 维矩阵, 且有 $A_n A_n^r \rightarrow \Delta$, Δ 是一个带有固定的 s 的 $s \times s$ 维非负对称矩阵, $\Omega = E(X_1 X_1^r) - E\{\mathfrak{U}^r(U_1) Y^{-1}(U_1) \mathfrak{U}(U_1)\}$ 。

4.2.2 惩罚经验似然法

通过将 Owen 的经验似然法(EL)与平滑裁剪绝对偏差(SCAD)惩罚函数互相结合的思想, 使用的 SCAD 惩罚函数, 其一阶导数对于一些 $a > 2$ 和 $p'_\gamma(0) = 0$, 满足

$$p'_\gamma(\theta) = \theta \left\{ I(\theta \leq \gamma) + \frac{a\gamma - \theta}{(a-1)\theta} + I \right\}, \theta > 0, \gamma > 0,$$

并按照 Fan 和 Li^[40]的建议, 取 $a = 3.7$ 。

当协变量的维度随着样本量的扩大而扩大时, 定义惩罚经验似然(PEL)函数如下:

$$L_n(\beta) = \hat{l}_n(\beta) + n \sum_{j=1}^p p_\gamma(|\beta_j|) \quad (4.10)$$

其中 p_γ 是一个惩罚函数, γ 为调整参数。该函数的例子见 Fan 和 Li^[40]。

最大化 PEL 函数式(4.10)等于最小化下式

$$L_n(\beta) = \sum_{i=1}^n \log\{1 + \lambda^r \eta_i(\beta)\} + n \sum_{j=1}^p p_\gamma(|\beta_j|)。 \quad (4.11)$$

使 $\mathcal{J} = \{j: \beta_{0j}\}$ 为真参数向量 β_0 的非零分量的集合, 并且 $|\mathcal{J}| = d$, 当 $n \rightarrow \infty$ 时,

其中 d 满足条件 C3。在不失一般性的前提下, 记 $\beta = (\beta_1^r, \beta_2^r)^r$ 其中 $\beta_1 \in R^d$,

$\beta_2 \in R^{p-d}$ 。因此, 真参数 $\beta_0 = (\beta_{01}^r, 0)^r$ 。并将 $\hat{\beta} = (\hat{\beta}_1^r, \hat{\beta}_2^r)^r$ 称为 PEL 估计, 它是式(4.11)的最小值。

定理 4.3 在定理 4.1 的条件下。当 $p^5/n \rightarrow \infty$ 时, 那么在概率趋于 1 的情况下, PEL 估计 $\hat{\beta}$ 满足以下条件:

(c) (稀释性): $\hat{\beta}_2 = 0$;

(d) (渐进正态性): $n^{1/2}W_n\Phi_g^{-1/2}(\hat{\beta}_1 - \beta_{10}) \xrightarrow{d} N(0, G)$, 其中对于固定的 q , 有 $W_n \in R^{q \times d}, G \in R^{q \times d}$, 满足 $W_n W_n^T \rightarrow G$, 且 $\Phi_g = \Phi_{11} - \Phi_{12}\Phi_{22}^{-1}\Phi_{21}$ 。

PEL 方法的显著优势在于检验假设和构建 β 的置信区间。为了更清楚地理解这一点, 我们考虑线性假设检验:

$$H_0: \Gamma_n \beta_{10} = 0 \quad \text{vs} \quad H_1: \Gamma_n \beta_{10} \neq 0$$

其中 Γ_n 是一个 $q \times s$ 的矩阵, 这样对于一个固定且有限的 q , 满足 $\Gamma_n \Gamma_n^T = I_q$ 。继续构造惩罚经验似然比统计量为:

$$\tilde{L}_n(\hat{\beta}) = -2\{L_n(\hat{\beta}) - \min_{\beta: L_n \beta = 0} L_n(\beta)\} \quad (4.12)$$

我们将检验统计量的性质总结为下面定理。

定理 4.4 在定理 4.1 的条件下。那么在原假设 H_0 下, 当 $p^5/n \rightarrow \infty$ 时有

$$\tilde{L}_n(\hat{\beta}) \xrightarrow{d} \chi_s^2$$

基于该定理的结果, $L_n \beta$ 的 $(1-\alpha)$ 水平置信域构造为

$$V_\alpha = \{v: -2\{L_n(\hat{\beta}) - \min_{\beta: L_n \beta = v} L_n(\beta)\} \leq \chi_{s, 1-\alpha}^2\}. \quad (4.13)$$

其中 $\chi_{s, 1-\alpha}^2$ 是 χ^2 分布的 $1-\alpha$ 水平分位数, 当 $n \rightarrow \infty$ 时, $P(\Gamma_n \beta_{10} \in V_\alpha) \rightarrow 1-\alpha$ 。据定理 4.3, 可以通过 V_α 给出 $\Gamma_n \beta_1$ 的 $1-\alpha$ 水平分位数下近似置信区间。

4.3 主要结果证明

本文使用一个矩阵 A 的 Frobenius 法则, 定义为 $\|A\| = \{\text{tr}(A^T A)\}^{1/2}$ 。在给出证明的细节之前, 我们先提出一些正则条件。在进行主要结果的证明之前需要给出一些假设条件:

C1 $\{\alpha_j(\cdot), j=1, \dots, q\}$ 在有界的 \wp 上有连续二阶导。

C2 矩阵 $Y(u)$ 对于每个 $u \in \wp$ 是非奇异的, $E(X_1 X_1^T | U_1 = u)$, $Y^{-1}(u)$ 和 $\mathfrak{U}(u)$ 是全 Lipschitz 连续的并且 $Y^{-1}(u)$ 和 $\mathfrak{U}(u)$ 的每一个元素都有界。

C3 $(\mu_i, X_{it}, Z_{it}, \varepsilon_{it}), i=1, \dots, n, t=1, \dots, T$ 独立同分布。 $E(\varepsilon | X, Z, \mu) = 0, \text{a.s.}$:

对于一些整数 $k \geq 4$, $E(\|X\varepsilon\|^k) \leq \infty$, $E(\|X\|^k) \leq \infty$, $E(|\varepsilon|^k) \leq \infty$.

C4 当 $n \rightarrow \infty$ 时, 调节参数 γ 满足当 $\gamma \rightarrow \infty$ 时, $\gamma(n/p)^{1/2-\delta} \rightarrow \infty$.

C5 $\max_{j \in \mathcal{S}} p'_\gamma(|\beta_{0j}|) = o\{(\sqrt{np})^{-1}\}$, $\max_{j \in \mathcal{S}} p''_\gamma(|\beta_{0j}|) = o(p^{-1/2+\delta} n^{-\delta})$.

引理 4.1 假设条件 C1~C5 都成立, 有

$$\frac{1}{n} \tilde{X}^\tau H \tilde{X} \xrightarrow{d} \Omega \{1 + o_p(c_n)\}.$$

其中 Ω 的定义在定理 4.2 中。

证明: 此引理证明可见 He 等^[17]引理 A.3。

引理 4.2 在定理 4.1 的条件下, 如果 β_0 是真值参数, 则

$$\|\lambda(\beta_0)\| = O_p(a_n).$$

$$\frac{\left\{n^{-1/2} \sum_{i=1}^n \eta_i^\tau(\beta_0)\right\} \Omega^{-1} \left\{n^{-1/2} \sum_{i=1}^n \eta_i(\beta_0)\right\} - p}{\sqrt{2p}} \xrightarrow{d} N(0, 1).$$

证明: 借鉴于 Fan 等^[20]的引理 A.4 的证明过程, 不难证明这个引理成立

引理 4.3 假设 C1~C5 成立, 定义 $D_n = \{\beta : \|\beta - \beta_0\| \leq ch_n\}$, $h_n = (p/n)^{1/2-\delta}$,

当 $\beta \in D_n$ 时, 有 $\|\lambda(\beta)\| = O_p(h_n)$ 。

证明: 具体证明见引理 3.1。

引理 4.4 假设 C1~C5 成立, 当 $n \rightarrow \infty$ 时, 依概率 1 有, $L_n(\beta)$ 在 D_n 内有最小值。

证明: 具体证明见本文引理 3.2。

定理 4.1 证明: 令 $T_i = \lambda^r \eta_i(\beta_0)$, 式(4.9)运用泰勒展开有

$$0 = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) \left(1 - T_i + \frac{T_i^2}{1+T_i}\right) = \bar{\eta}(\beta_0) - J(\beta_0)\lambda + \delta_n. \quad (4.14)$$

其中 $J(\beta_0) = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) \eta_i^\tau(\beta_0)$, $\delta_n = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) T_i^2 - \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) \frac{T_i^3}{1+T_i}$ 。

结合引理 4.2 及引理 4.3 有

$$\max_{1 \leq i \leq n} |T_i| \leq \|\lambda(\beta)\| \max_{1 \leq i \leq n} \|\eta_i(\beta_0)\| = O_p(p/n^{1/2-1/r}). \quad (4.15)$$

类似于 Li 等^[59](A.19)的证明,可以得到 $\|\delta_n\| = o_p\left(p_n^{5/2}n^{-1}(n^{-1/2} + c_n^2)\right) + o_p\left(p_n^2n^{-1}c_n\right)$,

并且根据式(4.14),可得 $\lambda = J(\beta_0)^{-1}\bar{\eta}(\beta_0) + J(\beta_0)\delta_n$ 。应用泰勒展开有

$$\ln(1+T_i) = T_i - T_i^2/2 + T_i^3/3(1+\zeta_i)^4, \quad (4.16)$$

对于 ζ_i 有 $|\zeta_i| \leq |T_i|$ 。至此,结合引理 4.2 以及一些简单运算,可以有

$$\begin{aligned} 2l_n(\beta_0) &= 2\sum_{i=1}^n \ln(1+T_i) \\ &= n\bar{\eta}^\tau(\beta_0)\Omega^{-1}\bar{\eta}(\beta_0) + n\bar{\eta}^\tau(\beta_0)\left\{J(\beta_0)^{-1} - \Omega^{-1}\right\}\bar{\eta}(\beta_0) - n\delta_n^\tau J(\beta_0)^{-1}\delta_n \\ &\quad + \frac{2}{3}\mathfrak{R}_n\{1 + o_p(1)\}. \end{aligned} \quad (4.17)$$

其中 $\mathfrak{R}_n = \sum_{i=1}^n T_i^3$,通过运用 Li 等^[59]中引理 B.6 以及式(A.22)的证明方法,可以

得出有 $n\delta_n^\tau J(\beta_0)^{-1}\delta_n = o_p(p^{1/2})$ 和 $n\bar{\eta}^\tau(\beta_0)\{J(\beta_0)^{-1} - \Omega^{-1}\}\bar{\eta}(\beta_0) = o_p(p^{1/2})$ 。至此,

结合上述结论以及引理 2,此定理得证。

定理 4.2 证明:首先令 $\hat{\beta}_{ME}$ 和 $\hat{\lambda}_{ME} = \lambda(\hat{\beta}_{ME})$ 都满足 $H_{1n}(\hat{\beta}_{ME}, \hat{\lambda}_{ME}) = 0$ 和 $H_{2n}(\hat{\beta}_{ME}, \hat{\lambda}_{ME}) = 0$ 。

$$\text{其中 } H_{1n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^\tau \eta_i(\beta)} \text{ 和 } H_{2n}(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^\tau \eta_i(\beta)} \left(\frac{\partial \eta_i(\beta)}{\partial \beta^\tau} \right)^\tau \lambda.$$

令 $\varphi = (\beta^\tau, \lambda^\tau)^\tau$, $\varphi_0 = (\beta_0^\tau, 0)^\tau$ 以及 $\hat{\varphi}_{ME} = (\hat{\beta}_{ME}^\tau, \hat{\lambda}_{ME}^\tau)^\tau$ 。则有:

$$0 = H_{jn}(\hat{\beta}_{ME}, \hat{\lambda}_{ME}) = H_{jn}(\beta_0^\tau, 0) + \frac{\partial \eta_i(\beta, 0)}{\partial \beta} (\hat{\beta}_{ME} - \beta_0) + \frac{\partial \eta_i(\beta_0, 0)}{\partial \lambda} (\hat{\lambda}_{ME} - 0) + \delta_{jn}$$

其中 $\delta_{jn} = \frac{1}{2}(\hat{\varphi}_0 - \varphi_0)^\tau H_{jn}''(\varphi)(\hat{\varphi}_0 - \varphi_0)$, $j=1,2$ 。并且 H_{jn}'' 是 H_{jn} 的 Hessian 矩阵。

则:

$$\begin{pmatrix} \hat{\lambda}_{ME} \\ \hat{\beta}_{ME} - \beta_0 \end{pmatrix} = \left[\begin{array}{cc} \frac{\partial H_{1n}(\beta, \gamma)}{\partial \gamma} & \frac{\partial H_{1n}(\beta, \gamma)}{\partial \beta} \\ \frac{\partial H_{2n}(\beta, \gamma)}{\partial \gamma} & \frac{\partial H_{2n}(\beta, \gamma)}{\partial \beta} \end{array} \right]_{(\beta_0, 0)}^{-1} \begin{pmatrix} H_{1n}(\beta_0, 0) + \delta_{1n} \\ H_{2n}(\beta_0, 0) + \delta_{2n} \end{pmatrix}, \quad (4.18)$$

根据引理 4.1,有

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta_0)}{\partial \beta} = \frac{1}{n} \tilde{X}^r H \tilde{X} = \Omega \{1 + o_p(1)\}, \quad (4.19)$$

通过引理 4.3 $\|\hat{\lambda}\| = O_p(a_n)$ 和引理 4.4 $\|\hat{\beta}_{ME} - \beta_0\| = O_p(a_n)$, 并结合 Cauchy-Schwarz 不等式可以得到

$$\|\delta_{1n}\|^2 = \sum_{k=1}^p \delta_{1n,k} \leq c \sum_{k=1}^p \|\hat{\phi}_0 - \phi_0\|^k \|H_{jn}''(\varphi)\|^2 = O_p(a_n^4 p^3), \quad (4.20)$$

结合条件 C5 可以得到有

$$\|\sqrt{n} A_n \Omega^{-1} \delta_{1n}\|^2 \leq n A_n A_n^r \Omega^{-2} \|\delta_{1n}\|^2 = O_p(n a_n^4 p^3) = o_p(1),$$

以及定义的 $H_{1n}(\beta_0, 0) = n^{-1} \sum_{i=1}^n \eta_i(\beta_0)$, 容易有

$$\hat{\beta}_{ME} - \beta_0 = \Omega^{-1} \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0) + o_p(1),$$

类似于 He 等^[17], 有:

$$\text{Cov}(\eta_i(\beta_0)) = E(X_1 X_1^r) - E\{\mathbf{U}^r(U_1) Y^{-1}(U_1) \mathbf{U}(U_1)\}.$$

因此

$$\lim_{n \rightarrow \infty} \text{Cov}(\eta_i(\beta_0)) = E(X_1 X_1^r) - E\{\mathbf{U}^r(U_1) Y^{-1}(U_1) \mathbf{U}(U_1)\} = \Omega$$

并结合 $A_n A_n^r \rightarrow \Delta$, 最后根据 Slutsky 定理以及 Lindeberg-Feller 中心极限定理, 结合以上的式子, 这个定理成立。

定理 4.3 的证明: 由引理 4.4, 我们注意到 $L_n(\beta)$ 在 D_n 内有最小值, 那么当 $j=1, \dots, p$, 对于 $\beta \in D_n$ 的每个组成部分有:

$$\frac{1}{n} \frac{\partial L_n(\beta)}{\partial \beta_j} = \frac{1}{n} \sum_{i=1}^n \lambda^r \frac{\partial \eta_i(\beta)}{\partial \beta_j} \{1 + o_p(1)\} + p'_\gamma(|\beta_j|) \text{sign}(\beta_j) = I_j + \Pi_j,$$

根据引理 4.3, 以及式(4.19), 可以得到

$$\max_{j \in \mathcal{G}} |\lambda^r \Omega \{1 + o_p(1)\}| = O_p(h_n).$$

结合引理 4.3 以及条件 C5, 可以得出 $\gamma(n/p)^{1/2-\delta} \rightarrow \infty, P(\max_{j \in \mathcal{G}} |I_j| > \gamma/2) \rightarrow 0$.

另外, 当 $j \notin \mathcal{G}, \gamma/h_n \rightarrow \infty$, 以及 $\Pi_j = p'_\gamma(|\beta_j|) \text{sign}(\beta_j) = \gamma \text{sign}(\beta_j)$ 时, 可以看出,

对于所有的 $j \in \mathcal{G}$, $p_\gamma(|\beta_j|)\text{sign}(\beta_j)$ 渐近地主导 $\frac{\partial L_n(\beta)}{\partial \beta_j}$ 的符号。并且当 $n \rightarrow \infty$ 时,

对于任意的 $j \in \mathcal{G}$, 依概率 1 有:

$$\frac{\partial L_n(\beta)}{\partial \beta_j} \geq 0, \beta \in (0, ch_n); \frac{\partial L_n(\beta)}{\partial \beta_j} \leq 0, \beta \in (-ch_n, 0),$$

这意味着,依概率趋于 1 有 $\hat{\beta}_2 = 0$ 。至此, 定理 4.3 的(a)部分已经被证明。

接着证明(b)部分, 令 Ψ_2 是 $\Psi_2 \beta = \beta_2$ 这样的矩阵。当 $n \rightarrow \infty, P(\hat{\beta}_2 = 0) \rightarrow 1$ 时, 用拉格朗日乘子法找到 $L_n(\beta)$ 的极小值, 可以等效的通过解决以下目标函数的最小化问题来得到

$$\frac{1}{n} \sum_{i=1}^n \log\{1 + \lambda^r \eta_i(\beta)\} + \sum_{j=1}^p p_\gamma(|\beta_j|) + v^r \Psi_2 \beta. \quad (4.21)$$

其中 v 是另一个拉格朗日乘子的 $p-s$ 维列向量。定义

$$\begin{aligned} \tilde{Q}_{1n}(\beta, \lambda, v) &= -\frac{1}{n} \sum_{i=1}^n \frac{\eta_i(\beta)}{1 + \lambda^r \eta_i(\beta)}, \\ \tilde{Q}_{2n}(\beta, \lambda, v) &= -\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + \lambda^r \eta_i(\beta)} \left(\frac{\partial \eta_i(\beta)}{\partial \beta} \right)^r + b(\beta) + \Psi_2^r v, \\ \tilde{Q}_{3n}(\beta, \lambda, v) &= \Psi_2 \beta. \end{aligned}$$

其中 $b(\beta) = \{p_\gamma(|\beta_1|)\text{sign}(\beta_1), p_\gamma(|\beta_2|)\text{sign}(\beta_2), \dots, p_\gamma(|\beta_s|)\text{sign}(\beta_s), 0, \dots, 0\}^r$ 。

并将 $\tilde{Q}_{jn}(\beta, \lambda, v)$ 在 $(\beta_0, 0, 0)$ 的扩展成如下式:

$$\begin{pmatrix} \tilde{Q}_{1n}(\beta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\Sigma & -\Sigma & 0 \\ -\Sigma & 0 & \Psi_2^r \\ 0 & \Psi_2 & 0 \end{pmatrix} \begin{pmatrix} \hat{\lambda} \\ \hat{\beta} - \beta_0 \\ \hat{v} \end{pmatrix} + R_n, \quad (4.22)$$

其中 $R_n = \sum_{k=1}^5 R_n^{(k)}$, $R_n^{(1)} = (R_{1n}^{r(1)}, R_{1n}^{r(2)}, 0)^r$, $R_{jn}^{(1)} \in R^p$, 并且当 $j = 1, 2$ 时, $R_{jn}^{(1)}$ 的第 k 个分量由下式给出:

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

使得 $\nu = (\beta, \lambda)^\tau, \bar{\nu} = (\bar{\beta}, \bar{\lambda})^\tau$, 满足 $\|\bar{\nu} - \nu_0\| \leq \|\hat{\nu} - \nu_0\|, \|\bar{\lambda}\| \leq \|\hat{\lambda}\|$. $R_n^{(2)} = \{0, b^\tau(\beta_0), 0\}^\tau$,

和接下来的 $R_n^{(3)} = \{0, \{b'(\beta_0)(\hat{\beta} - \beta_0)\}^\tau, 0\}^\tau, R_n^{(4)} = \{\{(T_n(\beta_0) - \Sigma)\hat{\lambda}\}^\tau, 0, 0\}^\tau$, 以及最

后的 $R_n^{(5)} = \{\{(T_n(\beta_0) - \Sigma)\hat{\lambda}\}^\tau, \{(T_n(\beta_0) - \Sigma)(\hat{\beta} - \beta_0)\}^\tau, 0\}^\tau$. 结合条件 C5 以及 SCAD

惩罚函数, 可以得到 $\|R_n^{(2)}\| = o_p(1/\sqrt{n}), \|R_n^{(3)}\| = o_p(1/\sqrt{n})$.

根据 $\|T_n(\beta_0) - \Sigma\| = O_p(\sqrt{p/n})$ 及 $\|\hat{\beta} - \beta_0\| = O_p(h_n), \|\hat{\lambda}_\beta\| = O_p(h_n)$, 可以得出 $\|R_n^{(4)}\| = o_p(1/\sqrt{n}), \|R_n^{(5)}\| = o_p(1/\sqrt{n})$. 因此有 $\|R_n^{(k)}\| = o_p(1/\sqrt{n}), k = 1, \dots, 5$.

接着定义 $K_{11} = -\Sigma, K_{12} = [-\Sigma, 0]$ 以及 $K_{21} = K_{12}^\tau$,

$$K_{22} = \begin{pmatrix} 0 & \Psi_2^\tau \\ \Psi_2 & 0 \end{pmatrix}, K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

并使 $\omega = (\beta^\tau, \nu^\tau)$. 通过对式(4.22)的简单求逆运算, 有

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\omega} - \omega_0 \end{pmatrix} = K^{-1} \left\{ \begin{pmatrix} -\tilde{Q}_n(\beta_0, 0, 0) \\ 0 \\ 0 \end{pmatrix} - R_n \right\},$$

由于矩阵 K 分为四个块, 因此可以按以下方式逐块进行求逆,

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

其中 $A = K_{22} - K_{21}K_{11}^{-1}K_{12} = \begin{bmatrix} \Sigma^{-1} & \Psi_2^\tau \\ \Psi_2 & 0 \end{bmatrix}$. 因此, 得到有

$$\hat{\omega} - \omega_0 = A^{-1}K_{21}K_{11}^{-1}\tilde{Q}_n(\beta_0, 0, 0) + o_p(1/\sqrt{n}).$$

矩阵 A 也可以使用求逆公式逐块进行求逆, 即

$$A^{-1} = \begin{bmatrix} \Sigma - \Sigma\Psi_2^\tau(\Psi_2\Sigma\Psi_2^\tau)^{-1}\Psi_2\Sigma & \Sigma\Psi_2^\tau(\Psi_2\Sigma\Psi_2^\tau)^{-1} \\ (\Psi_2\Sigma\Psi_2^\tau)^{-1}\Psi_2\Sigma & -(\Psi_2\Sigma\Psi_2^\tau)^{-1} \end{bmatrix}$$

此外, 还有

$$\hat{\beta} - \beta_0 = \left\{ \Sigma^{-1} - \Sigma^{-1}\Psi_2^\tau(\Psi_2\Sigma\Psi_2^\tau)^{-1}\Psi_2\Sigma^{-1} \right\} (\Omega\bar{\eta}(\beta_0) + R_n),$$

其中 R_{1n} 是 R_n 以及 $\|R_{1n}\| = o_p(1/\sqrt{n})$ 中的对应分量。接下来是 $\hat{\beta}_1$ 的展开,

$$\hat{\beta}_1 - \beta_{10} = \left\{ \Psi_1 \Sigma^{-1} - \Psi_1 \Sigma^{-1} \Psi_2^\tau (\Psi_2 \Sigma \Psi_2^\tau)^{-1} \Psi_2 \Sigma^{-1} \right\} (\Omega \bar{\eta}(\beta_0) + R_{1n}). \quad (4.23)$$

$\sqrt{n} \hat{\beta}_1 = \sqrt{n} \Psi_1 \hat{\beta}_1$ 的渐近协方差为:

$$\Phi_g = \Psi_1 \Sigma^{-1} \Psi_1^\tau - \Psi_1 \Sigma^{-1} \Psi_1^\tau (\Psi_2 \Sigma^{-1} \Psi_2^\tau)^{-1} \Psi_2 \Sigma^{-1} \Psi_1^\tau.$$

通过分解 $\Sigma^{-1} = \begin{pmatrix} \dot{\Sigma}_{11} & \dot{\Sigma}_{12} \\ \dot{\Sigma}_{21} & \dot{\Sigma}_{22} \end{pmatrix}$, 其中 $\dot{\Sigma}_j = \Psi_j \Sigma^{-1} \Psi_j^\tau$, 得到有 $\Phi_g = \dot{\Sigma}_{11} - \dot{\Sigma}_{12} \dot{\Sigma}_{22}^{-1} \dot{\Sigma}_{21}$ 。然

后, 在 Σ 上应用块逆矩阵, 我们得到

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} F^{-1} \Sigma_{21} \Sigma_{11}^{-1} & -\Sigma_{11}^{-1} \Sigma_{12} F^{-1} \\ -F^{-1} \Sigma_{21} \Sigma_{11}^{-1} & F^{-1} \end{pmatrix},$$

其中 $F = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ 。因此 $\Phi_g = \Sigma_{11}^{-1}$, 结合 $\|\sqrt{n} W_n \Phi_g^{-1/2} R_{1n}\| = o_p(1)$ 和 Lindeberg-Feller 中心极限定理, (b) 的部分也证明完毕。

定理 4.4 的证明: 令 $\hat{\beta}$ 为式(4.12)中最小值, 并泰勒展开有

$$L_n(\beta) = \sum_{i=1}^n \hat{\lambda}^\tau \eta_i(\beta) - \sum_{i=1}^n \frac{(\hat{\lambda}^\tau \eta_i(\beta))^2}{2} + \sum_{i=1}^n \frac{(\hat{\lambda}^\tau \eta_i(\beta))^3}{3(1+\xi_i)} + o_p(1).$$

其中 $|\xi_i| \leq |\hat{\lambda}^\tau \eta_i(\hat{\beta})|$ 以及 $o_p(1)$ 是由惩罚函数决定。并将式(4.21)中解出的 $\hat{\beta}$ 和 $\hat{\lambda}$ 代入到 $|\xi_i| \leq |\hat{\lambda}^\tau \eta_i(\hat{\beta})|$ 中, 我们可以得到

$$2L_n(\beta) = n \bar{\eta}(\beta)^\tau \Psi_2^\tau (\Psi_2 \Sigma^{-1} \Psi_2^\tau)^{-1} \Psi_2 \bar{\eta}(\beta) + o_p(1). \quad (4.24)$$

在原假设下, 由于 $\Gamma_n \Gamma_n^\tau = I_q$, 则存在 $\tilde{\Psi}_2$ 且 $\tilde{\Psi}_2 \beta = 0$ 以及 $\tilde{\Psi}_2 \tilde{\Psi}_2^\tau = I_{p-d+q}$ 。现在, 通过重复定理 4.3 的证明, 我们确立了在原假设下, β 的估计值可以通过最小化式(4.23)被估计出, 可以轻易得到

$$2L_n(\check{\beta}) = n \bar{\eta}(\check{\beta})^\tau \tilde{\Psi}_2^\tau (\tilde{\Psi}_2 \Sigma^{-1} \tilde{\Psi}_2^\tau)^{-1} \tilde{\Psi}_2 \bar{\eta}(\check{\beta}) + o_p(1).$$

并结合式(4.24), 有

$$L_n = n \bar{\eta}(\beta)^\tau \Sigma^{-1/2} \{P_1 - P_2\} \Sigma^{-1/2} \bar{\eta}(\beta) + o_p(1).$$

其中 $P_1 = \Sigma^{-1/2} \tilde{\Psi}_2^{\tau} (\tilde{\Psi}_2 \Sigma^{-1} \tilde{\Psi}_2^{\tau}) \tilde{\Psi}_2 \Sigma^{-1/2}$ 和 $P_2 = \Sigma^{-1/2} \Psi_2^{\tau} (\Psi_2 \Sigma^{-1} \Psi_2^{\tau}) \Psi_2 \Sigma^{-1/2}$ 是两个同幂矩阵。由于 $P_1 - P_2$ 的维度为 q , $P_1 - P_2$ 可以被写为 $\Xi_n^{\tau} \Xi_n$, 其中 Ξ 是 $p \times q$ 维矩阵则 $\Xi_n^{\tau} \Xi_n = I_q$, 更多的, 得到有

$$\sqrt{n} \Xi \Sigma^{-1/2} \bar{\eta}(\beta) \xrightarrow{d} N(0, I_q).$$

之后得到

$$n \bar{\eta}(\beta)^{\tau} \Sigma^{-1/2} (P_1 - P_2) \Sigma^{-1/2} \bar{\eta}(\beta) \xrightarrow{d} \chi_q^2,$$

至此, 定理 4.4 证明完毕。

4.4 本章小结

本文研究了带有固定效应的高维半参数变系数面板数据模型的惩罚经验似然推断。在提出的模型中首先通过 B 样条基函数的方法得到了非参数的近似估计, B 样条基函数方法是一种全局光滑法它不仅在计算上有着非常高的效率, 而且有着比较高的估计精度。在消除固定效应过后, 通过一个合适的辅助变量, 给出了参数的经验对数似然比统计量。将 SCAD 函数与惩罚的思想相结合, 推出惩罚经验似然法并应用于模型当中, 构造了惩罚经验似然统计量并借此得到其置信域, 以及通过使用适当的方法, 证明了惩罚经验似然估计器具有 oracle 属性。

第 5 章 总结与展望

本文在在在众多前人对半参数回归模型研究的基础上,主要考察模型误差是半参数回归模型以及其在高维的情况下的参数估计及检验问题。经验似然方法有着不错的优良性质,将经验似然方法结合惩罚函数的思想,针对高维半参数回归模型应用了惩罚经验似然方法。并通过理论证明和模拟结果的显示,在处理高维半参数回归模型下的检验问题以及变量选择的问题当中,惩罚经验似然方法简单而且有效。研究表明用惩罚经验似然法处理的高维半参数回归模型,其性质有着很多优点,如其稀疏性使得模型的复杂度降低,并且稀疏性对构造的置信域也更加的稳定,在实用方面来说,用惩罚经验似然方法做统计推断可以有效的节约成本,值得推广研究及使用。并且我们在最后一章对传统的核估计方法来估计非参数做了一定的改进,我们使用了 B 样条基函数来估计非参数,相对于在计算上相当的耗时的核估计方法来说,它有着更高的效率,节约了时间成本。以下是在今后的研究中,将会继续对这几个方面做进一步的研究:

- 1、本文是对半参数回归模型的参数部分进行了研究,今后还准备对模型的非参数部分进行一定的探究,或者是对模型的误差方差也可以进一步研究一下它所具有的统计性质。

- 2、本文有研究协变量中带有测量误差的数据,进一步的可以研究协变量或者是响应变量缺失的缺失数据,不仅是缺失数据,如生存分析数据,纵向数据等等都有着极其重要的研究意义。

- 3、本文主要是考虑了半参数回归模型,后续可以研究更加复杂的模型,如时变系数半参数回归模型,空间数据半参数回归模型等等,并且单指标模型也有着良好的降维作用,进一步将会考虑研究部分线性单指标模型的惩罚经验似然估计。

参考文献

- [1] Engle R F, Granger C W J, Rice J, etc. semi-parametric estimates of the relation between weather and electricity sales[J]. Journal of the American statistical Association, 1986, 81(394): 310-320.
- [2] Yang H, Li T T. Empirical likelihood for semi-parametric varying coefficient partially linear models with longitudinal data[J]. Statistics & Probability Letters, 80(2):111-121.
- [3] You J, Chen G. Estimation of a semi-parametric varying-coefficient partially linear errors-in-variables model[J]. Journal of Multivariate Analysis, 2006, 97(2):324-341.
- [4] Ahmad, Ibrahim, Leelahanon, etc. Efficient estimation of a semi-parametric partially linear varying coefficient model[J]. Annals of Statistics, 2005, 33(1):258-283.
- [5] Bravo, Francesco. semi-parametric estimation with missing covariates[J]. Journal of Multivariate Analysis, 2015, 139:329-346.
- [6] Wang S S, Cui H J, Li R Z. Empirical likelihood inference for semi-parametric estimating equations[J]. Science China Mathematics, 056(6):1247-1262.
- [7] Zhou Y, Liang H. Statistical Inference for semi-parametric Varying-coefficient Partially Linear Models with Generated Regressors (F06-463)[J]. Annals of Statistics, 2009, 37(1):427-458.
- [8] Owen A B. Empirical likelihood ratio confidence regions[J]. Annals of Statistics, 1990,18(1): 90-120.
- [9] 王启华.经验似然统计推断方法发展综述[J].数学进展,2004(02):141-151.
- [10] Tsao. Bounds on coverage probabilities of the empirical likelihood ratio confidence regions[J]. The Annals of Statistics,2004,32(3): 1215-1221.
- [11] Sarah C. Emerson,Art B. Owen. Calibration of the empirical likelihood method for a vector mean[J]. Electronic Journal of Statistics,2009,3:1161-1192.
- [12] Chen J , Abraham V B . Adjusted empirical likelihood and its properties[J]. Journal of Computational & Graphical Stats, 2008, 17(2):426-443.
- [13] Francesco Bartolucci. A penalized version of the empirical likelihood ratio for

- the population mean[J]. *Statistics and Probability Letters*,2006,77(1):104-110.
- [14] Taisuke Otsu. Penalized empirical likelihood estimation of semi-parametric models[J]. *Journal of Multivariate Analysis*,2007,98(10).
- [15] Leng T C . Penalized high-dimensional empirical likelihood[J]. *Biometrika*, 2010, 97(4):905-919.
- [16] Leng T C. Penalized empirical likelihood and growing dimensional general estimating equations[J]. *Biometrika*,2012, 99(3):703-716.
- [17] He B Q, Hong J X, Fan G L. Penalized empirical likelihood for partially linear errors-in-variables panel data models with fixed effects[J].*Journal of the Korean Statistical Society*,2018,45(3) 395-408.
- [18] Shen Y, Liang H Y, Fan G L. Penalized empirical likelihood for quantile regression with missing covariates and auxiliary information[J]. *Taylor & Francis*,2017,47(8):2001-2021.
- [19] Fang J L, Liu W R, Lu X W. Penalized empirical likelihood for semi-parametric models with a diverging number of parameters[J]. *Journal of Statistical Planning and Inference*,2017:42-57.
- [20] Fan G L, Liang H Y, Shen Y. Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors[J]. *Journal of Multivariate Analysis*,2016,147: 183-201.
- [21] Wooldridge J, Li Q. semi-parametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors[J]. *Econometric Theory*, 2002, 18(3):625-645.
- [22] Cui H J, Chen S X. Empirical likelihood confidence region for parameter in the errors-in-variables models[J].*Journal of Multivariate Analysis*, 2003,84(1): 101-115.
- [23] Wang Q H, Jing B Y. Empirical likelihood for partial linear models with fixed designs[J].*Statistics and Probability Letters*, 1999,41:425-433.
- [24] 公徐路,李幸福.响应变量缺失时纵向数据下部分线性模型的广义经验似然推断[J].*统计与决策*,2019,35(14):13-17.
- [25] Liu A A, Liang H Y. Jackknife empirical likelihood of error variance in partially

- linear varying-coefficient errors-in-variables models[J]. *Statistical Papers*, 2017, 58(1):1-28.
- [26] 崔恒建.有重复观测的部分线性 EV 模型的参数估计[J].*中国科学 A 辑*,2004,34(4):85-100.
- [27] 李晓妍.带线性误差的部分线性 EV 模型的经验似然推断[J].*统计与决策*,2017(19):18-23.
- [28] 冯三营,李高荣,薛留根等.非线性半参数 EV 模型的经验似然置信域[J].*高校应用数学学报*,2010, 25(01):53-63.
- [29] Chen X, Cui H. Empirical likelihood inference for partial linear models under martingale difference sequence[J]. *Statistics & probability letters*, 2008, 78(17): 2895-2901.
- [30] 李国亮,刘禄勤.误差为鞅差序列的部分线性模型中估计的强相和性[J].*数学物理学报*,2007,05:22-35.
- [31] Fan G L, Liang H Y, Xu H X. Empirical likelihood for heteroscedastic partial linear model[J]. *Communications in Statistics - Theory and Methods*,2011,40(8): 1396-1417.
- [32] Fan J, Truong Y K. Nonparametric regression with errors in variables[J]. *Annals of Statistics*, 1993,21(4):1900-1925.
- [33] Hall P, Heyde C C. *Martingale Limit Theory and Its Application*[M]. New York:Academic Press,1980.
- [34] Stout W F. *Almost Sure Convergence*[M]. New York: Academic Press, 1974.
- [35] Fan J Q, Huang T. Profile Likelihood Inferences on semi-parametric Varying-Coefficient Partially Linear Models[J].*Mathematics and Statistics*, 2005, 11(6):1031-1057.
- [36] Xia Y, Härdle W. Semi-parametric estimation of partially linear single-index models[J]. *Journal of Multivariate Analysis*, 2006, 97(5): 1162-1184.
- [37] Sepanski, Jungsywan H, Lee, etc. semi-parametric estimation of nonlinear errors-in-variables models with validation study[J]. *Journal of Nonparametric Statistics*, 1995, 4(4):365-394.
- [38] Li G R, Xue L G. Empirical likelihood confidence region of the parameter in the

- partially linear errors-in-variables model[J]. *Communications in Statistics - Theory and Methods*, 2008, 37(10):1552-1564.
- [39] Liang H, Hardle W, Carroll R J. Estimation in a semi-parametric partially linear errors-in-variables model[J]. *Annals of Statistics*, 1999, 27(5):324-341.
- [40] Fan J, Li R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties[J]. *Journal of the American Statistical Association*, 2001, 96(December):1348-1360.
- [41] Zhao P X, Xue L G. Variable selection in semi-parametric regression analysis for longitudinal data[J]. *Annals of the Institute of Statistical Mathematics*, 2012, 64(1):213-231.
- [42] Francesco B. A penalized version of the empirical likelihood ratio for the population mean[J]. *statistics & probability letters*, 2007, 77(1):104-110.
- [43] Ren Y W, Zhang X S. Variable selection using penalized empirical likelihood[J]. *中国科学:数学*, 2011, 054(009):1829-1845.
- [44] Wang S, Xiang L. Penalized empirical likelihood inference for sparse additive hazards regression with a diverging number of covariates[J]. *Statistics & Computing*, 2017, 27(5):1347-1364.
- [45] Fang J, Liu W, Lu X. Penalized empirical likelihood for semi-parametric models with a diverging number of parameters[J]. *Journal of Statistical Planning and Inference*, 2017, 186:42-57.
- [46] Wang H S, Li R Z. Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method[J]. *Biometrika*, 2007, 94(3):553-568.
- [47] Xia Y F, Zhao Z H, Niu Y J. Estimation of Partially Linear Varying-Coefficient EV Model Under Restricted condition[J]. *journal of statistical & econometric methods*, 2016, 3:1-5.
- [48] He X, Feng X, Tong X, et al. Semiparametric partially linear varying coefficient models with panel count data[J]. *Lifetime Data Analysis*, 2016, 23(3):1-28.
- [49] Chen J, Gao J, Li D, et al. Estimation in Partially Linear Single-Index Panel Data Models with Fixed Effects[J]. *Journal of Business & Economic Statistics*, 2013, 31(3): 315-330.

- [50] Su L, Ullah A. More efficient estimation of nonparametric panel data models with random effects[J]. *economics letters*, 2007, 96(3):375-380.
- [51] Rodriguez-Poo J M, Soberón, Alexandra. Nonparametric estimation of fixed effects panel data varying coefficient models.[J]. *Journal of Multivariate Analysis*, 2014, 133(1):95-122.
- [52] Li D G, Chen J, Gao J T. Non-parametric time-varying coefficient panel data models with fixed effects[J]. *econometrics journal*, 14(3):387-408.
- [53] Hu X M. Semi-parametric inference for semi-varying coefficient panel data model with individual effects[J]. *Journal of Multivariate Analysis*, 2017, 154(C): 262-281.
- [54] Hu X M. Estimation in a semi-varying coefficient model for panel data with fixed effects[J]. *Journal of Systems Science & Complexity*, 2014, 27(03): 594-604.
- [55] Chen G L, Wang Z J. The Multivariate Partially Linear Model with B-Spline[J]. *Chinese Journal of Applied Probability & Statistics*, 2010, 26(2):138-150.
- [56] Tang Q G . B-spline estimation for semi-parametric varying-coefficient partially linear regression with spatial data[J]. *Journal of Non-parametric Statistics*, 2013, 25(1-2):361-378.
- [57] Jin J, Hao C, Ma T. B-spline estimation for partially linear varying coefficient composite quantile regression models[J]. *Communications in Statistics*, 2019, 48(21):5322-5335.
- [58] De Boor C. *A Practical Guide to Splines*[M]. New York: Springer, 2001.
- [59] Li G, Lin L, Zhu L, et al. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters[J]. *Journal of Multivariate Analysis*, 2012, 105(1): 85-111.

硕士期间发表论文

[1]吕升日,何帮强.鞅差序列下非线性半参数测量误差模型的经验似然推断[J].统计与决策,2020,36(07):26-30.

[2]He B Q, Lv S R. Penalized empirical likelihood for high-dimensional partially linear errors-in-function model with martingale difference errors.(修改后录用)

致谢

三年的研究生生涯即将画上句号。在这里，我首先要感谢我的导师何帮强副教授。何老师在我论文的研究方向做出了关键性的指导性意见，并且在当我遇到不论是学术上的或是生活上的困难时，都会给予我很大的帮助，更是在我考博的道路上为我指引方向，在我的身上投入了大量的心血和精力，在此由衷的向何老师表示诚挚的谢意。同时，还要感谢一起呆了三年的学院里给我们上课的和蔼的老师以及可爱的同学们，在这三年里，没有他们的陪伴，缺少了他们的欢声笑语，可想而知这三年将会是多么的枯燥与乏味，感谢有他们的存在，感谢一切发生在生活中无论是好事还是坏事，这都影响了的一生，感谢在我这三年内给予我帮助与支持的人。

其次，还要感谢来我们学校开设讲座、报告的专家们学者们，以及参考文献中的所有的作者们，我是在他们的研究的基础上通过一定的了解跟熟悉，才让我对我的研究课题有了一个很好的思路。

再次，我要感谢数理研 17 班的全体同学，是你们的陪伴让我的研究生生涯如此丰富多彩和回味无穷。特别的要感谢赵颖以及我室友朱庆强，在学习和生活方面对我的理解和支持。

最后，衷心感谢我的家人，感谢他们一直以来给予我的最无私的爱，他们一如既往对我的理解和支持是我不断努力和奋斗的动力源泉。