

学校编码: 10384

学 号: 15420171151969

廈門大學

碩 士 学 位 论 文

基于惩罚高维经验似然的部分函数型线性模型的统计推断

Statistical inference of partial functional linear model

Based on penalized empirical likelihood

文怡方

指导教师姓名: 郭鹏辉 副教授

专业名称: 统 计 学

论文提交日期: 2020 年 3 月

论文答辩日期: 2020 年 4 月

学位授予日期: 2020 年 6 月

2020 年 6 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

本人声明该学位论文不存在剽窃、抄袭等学术不端行为,并愿意承担因学术不端行为所带来的一切后果和法律责任。

声明人(签名): 文怡方

指导教师(签名): 高鹏峰

2020年6月6日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

- () 1. 经厦门大学保密委员会审查核定的涉密学位论文，于 年
月 日解密，解密后适用上述授权。
- () 2. 不涉密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。涉密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）： 文怡方

2020年6月6日

摘要

在现代实验中,当从随机过程和高维标量协变量中采样时,经常同时遇到函数和非函数型数据。本文基于部分函数型线性模型来描述标量响应变量与函数型和非函数型变量之间的回归关系。主要用惩罚经验似然的方法对部分函数型线性回归模型进行变量选择与统计推断,能够识别出重要变量,并提高了估计值的可解释性。结合函数型主成分分析与惩罚经验似然,构造了感兴趣参数的经验对数似然比,并用 SCAD 惩罚函数得到了参数的最大惩罚经验似然估计。根据高维发散情形下惩罚经验似然对一般估计方程的推断,我们建立了该方法在一定条件下 Oracle 性质。仿真模拟的结果也显示,在一定的正则性条件下,本文得到的参数估计具有 Oracle 性质。同时,系数函数的估计效果也较好,处于 10%和 90%分位数函数之间。在参数的假设检验问题中,惩罚经验似然方法通过对部分函数型线性模型使用惩罚函数比单一经验似然法更加有效,因其稀疏性使计算似然函数时模型的复杂度降低,而且构造的置信域也更加稳定。本文以第一个非零参数为例构造了参数的假设检验问题,模拟研究表明了在原假设成立的条件下,其犯第一类错误的概率接近显著性水平,而且在原假设不成立的情况下,其功效也随样本量的增大以及检验值与真值差距的增大而越来越大。同时本文还对比了不同残差分布下的参数估计效果和置信域覆盖率的情况,体现出惩罚经验似然法的优势。最后,本文用上证指数的数据进行实证分析。分析结果显示,本文基于惩罚经验似然的部分函数型线性模型相比于传统的混合回归模型有更好的预测效果。

关键词:函数型数据; 函数型主成分;置信域;惩罚经验似然;变量选择

Abstract

In modern experiments, functional and nonfunctional data are often encountered when sampling from random processes and high-dimensional scalar covariables. This paper describes the regression relationship between scalar response variables and functional and nonfunctional covariables based on partial linear functional models. In this paper, the method of penalized empirical likelihood is mainly used to select variables and make statistical inferences for partial functional linear regression models, which can identify important variables and improve the interpretability of estimated values. Combining functional principal component analysis and penalized empirical likelihood, the empirical logarithm likelihood ratio of the interested parameters is constructed, and the maximum penalized empirical likelihood estimator of corresponding parameters is derived by using SACD penalized function. Based on the inference of the penalized empirical likelihood for the general estimation equation in the case of high-dimensional divergence, we establish the Oracle properties of the method under certain conditions. Simulation results also show that the parameter estimation obtained in this paper has Oracle properties under certain regularity conditions. At the same time, the estimation effect of coefficient function is also good, and the real coefficient function is between 10% and 90% Empirical quantile function. In the parameter hypothesis testing problem, penalized empirical likelihood more simple and effective than the traditional empirical likelihood by using penalty function, and its sparsity reduces the complexity of the model when calculating the likelihood function and makes the constructed confidence field more stable. This paper takes the first non-zero parameter as an example to construct the hypothesis test problem of the parameter and its confidence interval. The simulation results show that the size is close to the confidence level. In addition, when the original hypothesis is not valid, its power increases with the increase of sample size and the gap between the test value and the truth value. Meanwhile the effect of parameter estimation and the coverage rate of confidence fields under different residual and difference distribution

is compared, which shows the advantage of penalized empirical likelihood. Finally, this paper uses the data of Shanghai composite index for empirical analysis. The results show that the partial functional linear model based on penalized empirical likelihood has better prediction effect than the traditional mixed regression model.

Key words: Functional data ; Functional principal component ; Confidence region ; Penalized empirical likelihood ; Variable selection

目录

摘要.....	I
Abstract.....	II
第 1 章 引言.....	1
1.1 研究背景.....	1
1.2 研究意义.....	3
1.3 研究内容框架.....	3
第 2 章 文献综述.....	5
2.1 部分函数型线性回归模型.....	5
2.2 经验似然.....	7
2.3 变量选择与惩罚经验似然.....	10
第 3 章 研究内容与方法论.....	14
3.1 函数型主成分分析.....	14
3.2 模型的惩罚经验似然估计与推断.....	16
第 4 章 算法与调和参数的选择.....	20
4.1 经验似然的嵌套坐标下降算法.....	20
4.2 调和参数的选择.....	23
第 5 章 模拟研究.....	24
5.1 FPCA.....	24
5.2 估计量的 Oracle 性质.....	26
5.3 置信域以及假设检验的功效.....	29
5.4 不同分布假定下的结果比较.....	31
第 6 章 实证.....	35
6.1 上证指数研究的背景与模型.....	35
6.2 数据的来源与处理.....	36
6.3 模型的结果与分析.....	38
第 7 章 总结与扩展.....	42
附录.....	44
参考文献.....	44
致谢.....	53

Contents

Chinese Abstract.....	I
ABSTRACT.....	II
Chapter 1 Introduction.....	1
1.1 Research Background.....	1
1.2 Research Significance.....	3
1.3 Research Framework.....	3
Chapter 2 Literature Review.....	5
2.1 Partial Functional Linear Regression Model.....	5
2.2 Empirical Likelihood.....	7
2.3 Variable Selection And Penalized Empirical Likelihood.....	10
Chapter 3 Research Content And Methodology.....	14
3.1 Functional Principal Component Analysis.....	14
3.2 Penalized Empirical Likelihood Estimation And Inference.....	16
Chapter 4 Selection Of Algorithm And Harmonic Parameters.....	20
4.1 Empirical Likelihood Nested Coordinate Descent Algorithm.....	20
4.2 Selection Of Harmonic Parameters.....	23
Chapter 5 Simulation Study.....	24
5.1 FPCA.....	24
5.2 Oracle Properties Of Estimators.....	26
5.3 Confidence Fields And The Power Of Hypothesis Testing.....	29
5.4 Comparison Of Results Under Different Distribution.....	31
Chapter 6 Empirical Study.....	35
6.1 Background And Model Of Shanghai Composite Index.....	35
6.2 Source And Processing Of Data.....	36
6.3 Results And Analysis Of The Model.....	38
Chapter 7 Summary And Extension.....	42
Appendix.....	44
Reference.....	48
Acknowledgement.....	53

第 1 章 引言

1.1 研究背景

随着计算和测量技术的发展,科学家们在研究中收集到的样本观测数据形态多种多样,通常会面对由曲线、曲面或其他的随连续变量变化的信息数据,而对于这种类型的数据结构我们称之为函数型数据,它引起了各领域极大的兴趣。例如,在化学计量学中由数百个不同波长光谱组成的光谱测量数据;在证券交易市场中,股票一个交易日内每分钟或者一段时间内每小时的交易价格数据;银行每年的日交易数据;可以恢复利用人体不可见器官轮廓的 fMRI 数据;气象站对当地温度、风速、湿度等测量的小时数据;用于研究实体的拓扑、几何或地理属性的空间数据;居民月度用电量、电价的变化数据;物理学中分子运动的轨迹数据;在心理学研究中,收集到的各种脑电波信号数据;儿童连续几年的身高和体重变化的数据;医学诊断中的各种仪器的记录数据等等。这些观测数据不再像以前所研究的那些固定状态的数据,而更应该被看成一个不断变化的量。函数型数据有多种形式,但它们定义的量是由函数组成的,通常但并不总是平滑曲线。它的表现形式为连续的函数,即在数据空间表现出明显的函数型特性,它把每一条函数曲线看作一个样本而非看成序列点。许多不同领域都会产生函数型数据,从考古学家发掘的骨骼形状,到多年来收集的经济数据,再到魔术师手指所描绘的轨迹。多元数据分析允许研究有限组数的观测结果,但现代数据的收集情况可能还会涉及函数数据或产生数据的过程,即函数。函数型数据分析涉及无限维过程或数据。样条理论支持对随机函数样本进行线性建模和主成分分析。样条平滑依赖于将函数空间划分为两个正交子空间,其中一个子空间包含一组观测函数之间变化的明显分量,另一个子空间包含剩余分量。这种划分是通过使用线性微分算子来实现的。函数型数据分析关注的变量的第 i 个观测是一个实函数 $x_i(t), i=1,2,\dots,n$, 其中 $t \in T$, T 一个实时间间隔,因此每个 x_i 是函数空间 H 的一个点。我们希望将多元数据分析中常见的线性模型和主成分分析等概念应用到这个无限维的领域。从函数角度考虑数据分析的实际原因主要有四个方面:

(1) 随着越来越多的研究人员可以利用自动联机数据收集设施，函数型数据观测越来越多地出现在实际环境中，而且平滑和插值过程为产生有限组观测值的函数表示提供了可行性。

(2) 即使只有有限的观测数据，有些模型问题也更自然地从一个函数的角度来考虑问题。

(3) 分析的目标可以是函数型的目标，比如用有限的观测数据来估计整个函数、其导数或其他函数的值，情况也是如此。

(4) 考虑到了函数过程产生多变量数据的平滑性等因素，对我们的分析具有重要意义。

如果我们有一个响应变量，它的值是根据一组独立协变量来预测或近似的，而这些变量中至少有一个是函数型的。对于响应变量 y 与协变量 x 之间存在的联系，一般我们发现它们之间关系的第一步就是建立模型，进行模型推断。可供选择的模型有很多，而回归模型是应用最为广泛的统计模型，其中线性模型，或线性回归分析的函数类似物也是这里的研究重点。本文关心回归系数函数的估计，以便得出关于函数协变量在预测响应值中是沿着 t 轴的哪个位置起重要作用的结论。在许多实际问题中，为了提高预测能力和解释能力，我们通常还会收集大量的非函数型实值变量信息。例如在国家死亡率、发病率和空气污染的应用，这项研究的主要目的是调查空气质量的影响对美国不同城市非意外死亡率的影响，分析了几种大气污染物对死亡率的短期影响，其中包括每日 PM2.5 的浓度测量连同每天对温度和湿度的观察曲线以及美国每个城市的人口普查，个别土地面积，水域、城市人口的比例，比例的人口至少有一个高中文凭，比例的人口至少有一个大学学位，贫困线以下的人口比例等实值变量，这些都能被视为预测变量。还有关于股票价格的预测问题，除了将一段时间的连续波动看作一条曲线外，国家的财政政策和货币政策，公司的经营状况，行业的发展等相关变量对股价的波动也会带来影响。因此，本文引入部分函数型线性回归模型来探索响应变量与这种混合类型的预测因子之间的关系。同时，高维数据现在也越来越普遍，各个领域都涉及到了高维数据的应用。研究表明，当参数的维数越来越大时，数据通常也会变得越来越稀疏。例如上面提到的股价预测模型中还可能涉及到反应投资情绪的文本评论数据或多类别数据编码，那么就很容易遇到高维稀疏的变量。通常在

高维数据的分析过程中，会碰到“维数灾难问题”，其中数据的维数还会依赖于样本量增长，并且能随 $n \rightarrow \infty$ 增长到无穷大。在处理高维数据时，许多以前处理低维数据的传统方法都存在着一些难以解决的困难，如数据不符合正态分布或对数据没有较多先验信息等等。分析高维数据的关键就是希望能从数据中找到感兴趣的变量或者在众多的影响中寻找能够揭示事物本质规律的变量，我们称这种方法为变量选择。当然，实际问题中的也会涉及到感兴趣参数的假设检验问题，由于高维数据带来了协方差函数的估计困难，一般难以构建置信区间或假设检验问题，经验似然的在统计推断中的广泛应用与其效果的优越性恰能很好的解决此问题。因此，本文将综合变量选择与非参数方法去解决上述问题。

1.2 研究意义

基于上述背景可知，本文对高维情形下部分函数型线性模型的研究具有重要意义。尤其，当我们没有关于分布和数据的先验信息时，惩罚经验似然能很好的进行模型的变量选择，参数估计与推断。国内外很多学者把惩罚经验似然的方法应用于各种模型和数据的研究，但对函数型数据的研究甚少。H, Xue, & Feng (2018)^[1]在固定维数下，用经验似然法对部分函数型线性回归模型进行推断，但不涉及高维情形下的变量选择问题。Kong, Xue, Yao, & Zhang (2016)^[2]基于 Fan & Li (2001)^[3]提出的 SCAD 惩罚函数，用成组惩罚最小二乘的方法研究了具有多个函数型变量且非函数型变量高维的部分函数型线性模型，却很难构造估计的置信区间。因此，本文利用 Tang & Leng (2010)^[4]提出的惩罚经验似然法，在参数维度发散的情况下，对高维部分函数型线性模型进行研究大大推广了惩罚经验似然的应用范畴，并使部分函数型线性模型的应用变得更加灵活。

1.3 研究内容框架

本文主要用惩罚经验似然的方法对部分函数型线性回归模型进行变量选择、参数估计与假设检验。主要的研究目标为：

- 1、讨论惩罚经验似然的基本思想和部分线性函数型模型的特点以及其回归模型的估计与推断。

2、用 Tang , Leng (2010)^[4]惩罚高维经验似然的方法将 Hu , Xue , Feng (2018)^[1]对部分线性函数型模型的经验似然法扩展到变量维度发散的高维情形下。

3、利用 Monte Carlo 模拟部分函数型线性模型的数据来验证估计量与统计量的性质以及对比得出模型的优势。

4、用上证指数的数据进行实证研究，得出相应的股价预测效果。

主要的框架即论文的布置如下：

第一章是文章的引言。简要阐述了本文的研究背景和意义以及本文的主要内容与框架。

第二章是文章的文献综述部分。说明了部分函数型线性模型与惩罚经验似然的研究现状，以及本文主要创新之处。

第三章是文章的方法论部分。介绍函数部分的函数型主成分分析以及线性部分的变量选择及参数的惩罚经验似然估计，得到估计与统计量的性质，并以此构造 β 的假设检验问题。

第四章是算法与调和参数的选择。介绍 Cheng & Tong (2014)^[41]的嵌套坐标下降算法以及 Fan & Li (2001)^[3]的局部二次逼近算法；并使用 Wang et al. (2009)^[34]BIC 准则以及 Kong , Xue, Yao, & Zhang (2016)^[2]的 ABIC 准则同时选择截断参数 m ，即用于控制逼近真实函数的函数空间维数以及选择惩罚参数 τ 。

第五章是模拟研究。从数值模拟的角度，验证估计值与统计量的性质以及置信域的覆盖精度，并在不同残差分布的设定下将不同方法的结果与惩罚经验似然法进行对比。

第六章是实证研究。基于部分函数型线性模型研究了我国上证指数的预测问题以及变量重要性的假设检验问题。

第七章是研究结论。总结本文的主要结论以及可拓展方向。

第2章 文献综述

2.1 部分函数型线性回归模型

在数据收集技术不断发展的今天,数据分析方法也应该要不断扩展和改善才能适用于具有不同性质的数据,否则会使模型的估计失真或信息量损失严重从而降低模型的效果。Ramsay & Dalzell (1991)^[6]提出函数型数据分析(FDA)的概念,并对在时间上无限维度的函数型数据提出了一些研究方法和分析的工具。他们首先是将离散的观测数据利用光滑技术进行预处理,提出了函数线性模型和函数型主成分分析的方法来研究变量间的关系。并用加拿大的温度及其降水量的数据进行实证研究。不同的光滑方法对函数型主成分分析产生的影响不同。函数平滑法和函数型主成分回归在研究函数型线性模型上具有巨大贡献。Dauxois & Pousse (1982)^[8]在函数型数据分析中用测度论和纯分析的方法,推导出了关于函数型主成分分析的渐近理论,这为后来函数型主成分回归分析奠定基础。Ramsay & Silverman (1997)^[7]著作出版后,各研究界开始广泛关注起函数型数据分析方法,并在各个领域都掀起了函数型数据分析方法研究热潮。采用函数型数据分析方法的主要优势有:

- (1) 对于无限维空间的曲线数据,函数型数据分析方法显然有很好的效果;
- (2) 可以利用曲线的特性,通过曲线求导或者积分从而能获得更多的数据信息;
- (3) 在一般情况下,它不需要太多的假设条件,从而具有更加广泛的适用性;
- (4) 对于分析的结果可以给出几何图形解释,可以看到量的影响度随时间的变化从而使分析结果更加直观且易于理解;
- (5) 不同观测对象的观测点所对应的观测次数还可以允许不相同;
- (6) 也可以用该方法来提取一些非函数型的数据信息。

部分线性模型中既有参数分量,又同时包含了非参数分量,属于半参数模型的范畴。部分线性模型一方面解决了与完全非参数模型相关的维数问题,便于解释与线性部分相关的协变量所带来的影响。另一方面,当某些协变量与响应变量呈现非线性关系时,相比标准线性回归它们会更灵活。这体现出了混合模型的优

势。在统计研究的另一个方向,即最近对函数型数据的统计模型的兴趣有所增加,因为在许多实验中,函数型数据作为观测的基本单位出现。函数型数据分析作为多元数据分析一个自然的延伸,为解决这些问题提供了有价值的方向。与离散多元分析相比,函数型数据分析考虑了高维协变量的光滑性,并经常对以前没有发现的问题提出新的解决方法。正如上面提到的优势,函数型方法也可以为非函数型数据提供新的视角。本文的目标就是将参数和非参数方法与函数型主成分回归相结合,对函数型部分线性模型进行推断。

在过去的二十年中,函数型数据越来越受欢迎,而函数型数据分析也引起了广泛的关注。由于无限大的维度和预测因子之间存在很强的相互关系,多元统计方法未能分析函数型数据。为了克服这些问题,Ramsay & Silverman (1997)^[7]介绍了一些基本模型以及一些函数型数据的分析工具。

本文考虑以下部分函数型线性模型:

$$Y = Z^T \beta + \int_0^1 \gamma(t)x(t)dt + \varepsilon \quad (1)$$

其中, Y 为响应变量, $Z = (Z_1, Z_2, \dots, Z_p)^T$ 为 p 维的协变量向量; β 是 p 维的未知回归系数, $\gamma(t)$ 是未知的系数函数; $x(t) \in L^2(\mathcal{I})$ 是零均值的随机函数, $L^2(\mathcal{I})$ 为Hilbert空间,是所有在 \mathcal{I} 上平方可积的函数集合,为了研究方便,可令 $\mathcal{I} = [0, 1]$; ε 是均值为0,方差为 σ^2 的随机误差,且与 Z 和 $x(\cdot)$ 独立。

模型(1)灵活的涵盖了几种重要的模型。例如,当 $\gamma(\cdot) = 0$ 时,模型(1)可以简化为经典线性回归模型;令 $m(t) = \int_0^1 \gamma(t)x(t)dt$,模型(1)为Aneiros-Perez & Vieu (2006)^[13]中半函数型部分线性回归模型的特例,也就是将非参数函数模型与标准线性回归的分量相结合;当 $Z = 0$,模型(1)便成为函数型线性回归模型,文献中也有很多相关研究,如Cardot, Ferraty, & Sarda (2003)^[9]首先研究了一种基于函数系数惩罚B-样条展开的估计,它在某种程度上推广了岭回归,还引入了光滑函数型主成分回归。Hall & Hosseini-Nasab (2006)^[10]研究了函数型主成分分析的性质并将其应用到了函数型线性回归模型中。Hall & Horowitz (2007)^[12]详细讨论了函数主成分分析法,即关于特征值和特征函数的谱分解,并证明了在一定条件下,该方法有最优的收敛速度,同时还提出了一种基于二次正则化的方法,

证明其在某些角度上也有一定的优势。鉴于函数型主成分分析和 Karhunen-Loeve 展开具有很好的收敛性，本文可以将其用于模型 (1) 函数型数据部分的处理。

关于部分函数型线性模型的研究, Zhang, Lin, & Sowers (2007)^[14]基于两阶段非参数回归校正方法利用周期平滑样条函数估计模型 (1) 的回归系数函数。Shin (2009)^[15]利用 Karhunen-Loeve 展开和主成分回归的思想近似模型 (1) 并建立了参数和系数函数的估计, 然后研究了其大样本性质。Shin & Lee (2012)^[16]进一步研究了模型 (1) 均方预测误差的收敛速度。Lu, Du, & Sun (2014)^[17]扩展了 Shin (2009)^[15]的研究结果到分位数回归的研究。上述文章主要集中在点估计问题上, 我们还需要构建 β 的置信区域来进一步获得估计的准确性。H, Xue, & Feng (2018)^[11]首先利用 Karhunen-Loeve 展开和主成分回归的思想近似模型 (1), 在给定 β 下, 构建 $\gamma(t)$ 的剖面(profile)最小二乘估计。然后基于 $\gamma(t)$ 的初步估计, 用经验似然的方法分别得到了完备数据下模型 (1) 的参数 β 和系数函数 $\gamma(t)$ 的经验似然估计以及在响应值随机缺失情况下(给定缺失率)的加权经验似然估计, 证明了经验版的 Wilks' 定理, 并以此构建了 β 的置信区域, 最后模拟研究显示经验似然构造的置信区间长度和覆盖率都优于正态近似法。但是他们没有考虑 β 高维的情况, 而实际研究中高维数据普遍存在。Kong, Xue, Yao, & Zhang (2016)^[2]研究了具有多个函数预测子且 β 高维的部分函数型线性模型, 基于 Fan & Li (2001)^[3]提出的平滑剪切绝对偏差(SCAD)惩罚函数, 利用惩罚最小二乘以及成组压缩的方法对该模型同时进行变量选择和估计, 通过仿真研究验证了该方法的性能, 并利用空气污染数据说明了该方法的应用, 同时还根据对函数系数曲线的估计得出了在对温度和家庭所有权进行调整后, 夏季 PM2.5 浓度对美国城市老年居民近期死亡率有显著影响。但是该方法只关注参数估计与模型解释力的问题, 并没有涉及 β 的假设检验问题。

2.2 经验似然

在参数理论中, 一般我们可以最大化似然函数来估计感兴趣的参数, 而对于非参数似然函数, 同样也可以得到非参数极大似然估计。经验似然首先是由 Owen (1988)^[18]提出, 它是一种构造未知参数置信区间/域的非参数统计推断的方法。经验似然方法没有严格的参数分布假设, 但有类似于参数似然的一些优点, 包括经

验似然比服从卡方分布。限制较少的假设条件赋予了经验似然对于模型错误设定的稳健型优势。此外，它还具有将多个信息相结合以及构造面向数据、无形状限制的置信区域的优点。

假设样本 $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F_0$ ，其中 F_0 是未知的分布函数。那么经验分布 F_n 则为 F_0 的极大似然估计，因为对于所有的分布函数 F 有

$$F_n = \arg \max_F L(F) = \prod_{i=1}^n \{F(X_i) - F(X_i-)\} \quad (2)$$

其中， $F_n = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ 。那么我们可以定义经验似然比

$$R(F) = \frac{L(F)}{L(F_n)} \quad (3)$$

如果关心一个统计函数 $\theta = T(F)$ 的统计推断问题，那么我们可以这样定义剖面经验似然比函数

$$\mathfrak{R}(\theta) = \sup_F \{R(F) \mid T(F) = \theta\} \quad (4)$$

很显然，式 (4) 即为经验似然比检验统计量。如果充分小，检验 $H_0: T(F) = \theta_0$ 将会被拒绝。同样的， $T(F_0)$ 置信区域可表示为 $\{\theta \mid \mathfrak{R}(\theta) \geq c\}$ ，其中 c 是某一临界值。由式 (2) - (4) 可知，只有当样本观测的点概率质量

$$\omega_i = F(\{X_i\}) = F(X_i) - F(X_i-) > 0, \quad 1 \leq i \leq n$$

式 (4) 才能极大化，所以只需考虑离散分布。不失一般性，假设数据没有结点，即使有结点，Owen (1988)^[18] 也证明了式 (5) 是合理的，即让

$$\sum_{k: X_k = X_i} \omega_k = F(\{X_i\}),$$

假设 $\sum_{i=1}^n \omega_i = 1$ ，则式 (4) 等价于

$$\mathfrak{R}(\theta) = \sup_F \left\{ \prod_{i=1}^n n\omega_i \mid \sum_{i=1}^n \omega_i = 1, \omega_i \geq 0, T(F) = \theta \right\} \quad (5)$$

其中， $L(F_n) = \prod_{i=1}^n \frac{1}{n}$ 。用拉格朗日乘数法解式 (5) 可得到 $-2 \log \prod_{i=1}^n n\omega_i \rightarrow \chi^2_p$ ，这便是

是 Owen (1990)^[19] 证明的经验似然定理。

本文之所以采用经验似然法是基于以下几个动机。首先，由于函数型数据模型的方差估计比较复杂，而经验似然统计推断法不涉及方差的估计。二是基于经验似然法构造的置信区域的形状和方向是完全由数据决定。此外，它有类似于 bootstrap 的抽样特性，但它是将一个多项式分布族用在随机选择样本点上，而不是通过重抽样。经验似然还能进行 Bartlett 纠偏 Diccio & Romano (1991)^[20]或 Bootstrap 校正 Owen (2001)^[21]来提高覆盖精度，能将中心覆盖误差减少到 $O(n^{-2})$ 。而且，它能很方便的加入一些辅助信息进而能提高估计和推断的有效性。Owen (1991)^[22]表明，当样本分步的某个统计函数已知时，可以通过在式 (5) 中加入这一约束获得改进的推断。

近年来，经验似然法也成功地应用于各种模型，如线性回归模型 Owen (1991)^[22]、变系数模型 Yang, Li, & Pend (2014)^[23]、部分线性单指标模型 Zhu & Xue (2006)^[24]等。对于函数型数据，Lian (2012)^[25]用经验似然法和偏差纠正的经验似然法分别对非参数函数模型和半函数型部分线性模型的非参数部分进行统计推断，并将其与正态近似的方法进行对比。数值模拟的结果显示，不管是纠偏前还是纠偏后，基于经验似然得到的非参数部分的置信区间覆盖率与区间长度都优于正态近似，且不用估计复杂的方差。但是它旨在构造整个非参函数的置信区间，没有对 β 和函数系数进行估计。H, Xue, & Feng (2018)^[1]用经验似然法研究了部分函数型线性模型的参数和函数系数的估计，构建了单个 β 的置信区间，并且分别研究了完备数据与响应变量随机缺失数据下的 β 的经验似然估计及其对应的经验似然比统计量的渐近性质。模拟了参数在不同缺失率下的置信区间长度与覆盖率，并对比了正态近似方法。他们将经验似然方法运用到带有缺失响应的部分函数型线性模型中，避免了估计未知协方差，而协方差的估计对于函数型数据模型来说是相当复杂的。Jiang, Huang & Fan (2019)^[45]将经验似然法运用到高维部分函数型线性模型中，建立了非函数协变量系数的经验对数似然比统计量，并证明了该统计量在一定条件下是服从渐近正态分布的。但它是通过建立参数投影向量的剖面经验似然比统计量来研究感兴趣参数的置信区间长度或置信域的覆盖率，没有涉及变量的选择问题。

2.3 变量选择与惩罚经验似然

在实际问题中,为了减少模型的偏差,通常会引入很多变量,变量选择成了知识发现和模型预测的重要话题。以传统 AIC, BIC 为选择标准的逐步回归和最佳子集挑选在高维数据上往往是计算不可行的。此外,最佳子集回归方法是分别选择模型和估计参数,这样会导致模型选择的不稳定性和样本性质的复杂性。后来,惩罚变量选择作为一种能同时进行参数估计和变量选择的方法就开始被广泛应用。统计学家们就随之不断的提出并研究了各种惩罚最小二乘的方法,其中包括 Tibshirani (1996)^[26]提出的 lasso(最小绝对收缩和选择算子)与 Fu (1998)^[27]的桥回归。Fan & Li (2001)^[3]的研究指明了这些以 L_q 或硬门限为惩罚函数的惩罚方法并没有同时满足以下性质:

1.稀疏性:得到的惩罚估计的阈值规则能自动地将小的估计系数设为零,这样能减少模型的复杂性;

2.连续性:得到的惩罚估计与最小二乘估计连续,从而避免了模型预测时的不稳定性;

3.无偏性:当真实未知参数值足够大时,得到的惩罚估计几乎是无偏的,从而避免了非必要的建模偏差。

因此他们提出平滑剪切绝对偏差 (SCAD) 惩罚函数:

$$p_\lambda(\theta) = \lambda \left\{ I(\theta \leq \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\} \quad (6)$$

其中, $a > 2, \theta > 0$ 。SCAD 惩罚函数具有对称性、在区间 $(0, \infty)$ 非凹并且在原点处有奇解。因此本文主要是基于 SCAD 非凹惩罚函数来进行变量选择与参数估计。

但是,当存在内生性时,或者在进行参数推断时,惩罚最小二乘法不适用。Fan & Li (2001)^[3]又提出了一种基于非凹惩罚似然的变量选择方法,它是一种参数模型,可以用于同时估计参数和选择重要变量,他们还证明了此类方法在参数维度有限固定的情况下是具有 oracle 性质的,即能够以概率 1 识别出真实的模型,并能有效地估计出非 0 系数,就像事先知道真实模型的稀疏性一样。在算法上,提出了局部二次逼近算法。但是大多数模型选择的问题中参数的数量可能会很大,并且还能随着样本量的增加而增加。Fan & Peng (2004)^[28]在正则条件下,建

立了惩罚似然估计的 oracle 性质以及非零真实参数的渐近正态性, 进一步讨论了非凹惩罚似然比统计量, 而且在零假设成立下得出了非凹惩罚似然比统计量的渐近分布, 上述的性质同样适用于变量维度随着样本量的增加而趋于无穷的情况。

在统计分析中, 很多情况下研究人员没有足够的信息来构造参数似然, 分布的错误设定可能会给我们的模型带来偏差估计。同时, 考虑到人们对高维数据和经验似然的兴趣, 有必要评估经验似然在数据维数和样本容量同时增加时的行为。在一般的多元模型下, Chen, Peng, & Qin (2009)^[37]评估了数据维数对高维数据经验似然比的渐近正态性的影响, 指出多元随机向量各分量之间的数据维数和相关性直接通过协方差矩阵的迹和特征值来影响经验似然。综合惩罚似然和经验似然的思想, Tang & Leng (2010)^[4]提出了参数维度发散问题下参数估计及其变量选择的惩罚经验似然方法。通过在经验似然的目标函数中加入 SCAD 惩罚函数, 分别研究了多元随机向量均值以及高维线性模型的惩罚经验似然估计, 证明了他们得出的估计值也是具有 oracle 性质的, 这无疑论证了该方法在高维分析中的稳健型与有效性。如果不指定参数的分布函数, 惩罚经验似然也同样具有生成稀疏模型的优势。而且他们还同样证明了所提出的惩罚经验似然比渐近 χ^2 分布, 这对感兴趣参数的假设检验和产生数据驱动的置信域提供了方便。他们对线性模型的惩罚经验似然推断为后续该方法在不同模型和数据下的扩展提供了研究思路以及证明框架。虽然 Fan & Peng (2004)^[28]提出的惩罚参数似然比检验具有这种性质, 但指定一个高维分布可能是非常具有挑战性的。惩罚经验似然比惩罚似然方法的适用范围更广, 因为它还仍然保留着经验似然法的优点, 如不需要指定分布函数。而且最关键的是它并不涉及到复杂协方差的估计, 因此对于各种复杂模型都能很好的适用, 使统计推断更容易。

作为惩罚似然方法的补充, 惩罚经验似然方法为高维推断中的参数似然比提供了一种可靠的选择, 且具有非参数似然方法效率高、自适应能力强的优点。虽然惩罚经验似然方法没有严格假设分布函数, 但它却能与参数方法一样有效。这也是近几年来惩罚经验似然方法被广泛用于研究的原因。Leng & Tang (2012)^[5]研究了维数增长情况下一般高维估计方程的参数估计和变量选择的惩罚经验似然问题。他们量化了惩罚经验似然的渐近性质, 如证明了一般估计方程下的惩罚经验似然也具有 oracle 性质等。这意味着如果模型找到了适合条件的估计方程,

便可以运用他们的研究框架。最后他们也通过不同模型的仿真研究和数据分析，验证了该方法的有效性。当模型未指定参数似然函数时，估计方程可作为统计推断的工具，经验似然作为一种数据驱动的非参数方法，在合并辅助信息方面是稳健和灵活的。

若假定 $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ 是独立同分布的随机向量， $\boldsymbol{\theta} \in \mathbb{R}^p$ 是未知参数，假设辅助信息是以无偏估计的形式出现，即 $E(g(\mathbf{Z}_i; \boldsymbol{\theta})) = 0$ 。当估计方程的维数大于参数的维数，直接求解方程 $n^{-1} \sum g(\mathbf{Z}_i; \boldsymbol{\theta}) = 0$ 得到参数 $\boldsymbol{\theta}$ 的估计一般是不可行。而 Leng & Tang (2012)^[5] 证明了经验似然法结合高维估计方程却可以得到有效的估计量，其证明的相应经验版本的 Wilks 定理也为假设检验和构造参数的置信域提供了一个稳健的方法。F., L. & Yu (2016)^[29] 对自变量带有测量误差的高维部分线性变系数模型进行了研究，分别在不同大小的测量误差下与 PLS 模型的置信域的覆盖率作对比，并证明了他们运用惩罚经验似然方法得到的估计具有一定的预测精度；Fang, Liu & Lu (2017)^[30] 将惩罚经验似然的方法运用到参数维度发散的高维半参数模型中，得出惩罚似然估计的渐近性质以及以及相应的惩罚似然比统计量在零假设下是渐近 χ^2 分布的结论，模拟研究了部分线性变系数及其带测量误差等半参模型，最后还将该方法下的部分线性逻辑回归模型应用于癌症分类，都取得了很好的效果。Ando & Sueishi (2017, 2018)^[31-32] 分别研究了 SCAD 惩罚经验似然估计的一致收敛率问题和正则化参数的选择问题。证明了在合理的正则化参数下，基于 SCAD 的矩约束模型的惩罚经验似然估计以 $\sqrt{n/p_n}$ 一致收敛。他们还将矩约束模型表示为一组概率分布的形式，然后利用从估计模型到真实数据生成过程的 Kullback-Leibler 信息准则的估计值来评价模型的优劣，提出了一种新的信息准则，即经验似然信息准则，用于选择惩罚经验似然估计量的正则化参数。还有将惩罚经验似然应用于面板数据，纵向数据和删失数据的研究，如 He, Hong & Fan (2018)^[33]，谭晓燕 (2018)^[49] 和侯文 (2012)^[50]。在前人惩罚高维经验似然的研究基础上，韩慧 (2011)^[51] 具体研究了部分线性模型；刘琦 (2012)^[60] 研究了广义线性模型；李吉妮 (2014)^[61] 研究了单指标模型。将但据我所知，将惩罚经验似然用于高维混合函数型数据研究的几乎没有。

在实际中实现经验似然需要密集的计算，尤其是在数据维度增加的情形下，这可以看作是放弃完全参数分布假设和随后的优点所付出的代价。关于经验似然

的算法研究, Owen (2001)^[21]是将经验似然法定义为一个约束优化问题, 然后利用对偶法求解了该约束优化问题, 提出了牛顿拉格朗日算法以及 Nested 算法解对偶问题, 对于估计方程的经验似然问题, Nested 算法应用到各种模型都产生了较好的效果。当变量的维数较高时, 牛顿型方法收敛性较好但稳定性可能会有问题, 因为目标函数通常是非凸的且 Hessian 矩阵会在迭代过程中出现不可逆等问题。对于惩罚函数的正则化方法, 普遍使用的便是牛顿梯度下降法, 但高维下大矩阵的逆很难计算。近年来, 各种应用包括计算统计学和机器学习中的应用都产生了一些问题, 坐标下降(CD)方法在性能上与享有信誉的替代方法相比具有竞争力。坐标下降方法是一种近乎通用的算法优化方法的原型, 即通过解决一系列较简单的优化问题来解决目标优化问题, 它的显著性及其在许多情况下的可接受性可能是其在从业者中具有长期吸引力的原因。坐标下降法的思想就是将多变量优化问题分解为一系列单变量优化的子问题。目标函数沿着每个坐标上山或下山, 其他所有坐标一起固定。它因其简单和稳定而受到称赞。坐标下降作为一种方便、稳健的算法, 在已有的文献中被证明是一种方便有效的优化算法, 且在已有的文献中得到了很好的应用。例如, Wu & Lange, Kenneth (2008)^[39]将坐标下降法用于 Lasso 惩罚的回归模型。经验似然作为一种非参数方法已知有许多可取的优点。虽然它在方法学研究方面得到了深入的发展, 但由于需要进行深入的优化, 其实际应用却很少得到探索。因此, 高效稳定的算法是经验似然实际实现的迫切需要。Fu.WJ (1998)^[43]在求解 L1 正则化问题时, 重点研究了坐标下降算法。坐标下降算法是解决 L1-正则化问题的重要方法, 尤其是应用在 Friedman, Hastie, Hoefling & Tibshirani (2007)^[44]和 Wu, Lange & Kenneth (2008)^[39]的高维数据中。算法的收敛性根据 Ruszczyński A (2006)^[41]和 Tseng P (2001)^[42]中的结果可以很容易地证明。我们的动机是使用方便的坐标下降算法来解决经验似然问题。然而, 这样的研究并不直接, 因为多层优化求解经验似然问题引入了一个很大的参数空间。为了保证高效稳定的计算性能, Cheng & Tong (2014)^[40]提出了一种嵌套坐标下降算法来解决一般的惩罚经验似然问题。所提出的嵌套算法在构造置信区域等方面非常有用, 他们将此计算方法应用于经验似然的一些典型场景中, 证明了嵌套坐标下降算法是稳定有效的。最后还将修改后的嵌套坐标下降算法与 MM 算法相结合, 进一步简化了计算。嵌套坐标下降算法与高维数据中基于剖面估计和

变量选择的推断是自然的、完美的结合。两层嵌套算法是由一层对给定参数的经验似然求值的内层构造的,该内层嵌入经验似然对参数的优化中。这两层优化都是通过坐标下降的循环来解决的,所提出的嵌套算法在构造置信区域时非常有用。

第3章 研究内容与方法论

由于在一些情况下,能同时遇到高维混合数据,此时部分函数型线性回归模型的统计推断问题会显得不那么容易。Shin (2009)^[15]对部分函数型线性模型的研究虽然没有涉及到高维情形下的变量选择,但它对于函数型数据部分的处理确是十分有效。因此本文也采用函数型主成分分析的方法,3.1节中便介绍了该方法的主要原理。又因为惩罚经验似然法能同时进行变量选择与统计推断,将原本复杂的问题转化成求解带约束的非线性优化问题,而根据目标函数的渐进分布还能构造参数的假设检验问题,方法简单却能达到好的效果。所以在高维情形下对函数型数据部分的处理之后的模型采用惩罚经验似然法能够很好地解决问题。3.2节便是本文模型的惩罚经验似然的运用过程。

3.1 函数型主成分分析

正如 Ramsay & Silverman (2005)^[46]所述,我们首先回顾了广泛使用的函数型主成分分析(FPCA)技术。函数型主成分分析(FPCA)在函数型数据分析(FDA)中起着重要作用,它的目的是用来描述一个函数目标的几何形状。FPCA的基本思想是将曲线空间分解为主要的变化方向。在函数型主成分分析中,每个特征值对应的是一个特征函数,而不再是一个向量。这些特征函数描述了主要的变分分量。我们想知道数据中主要的变化模式是什么,其中有多少是实质性的。在多元统计中,双变量协方差函数 $K_X(s, t)$ 的特征根是描述这些主成分重要性的指标,绘制特征值是一种确定需要多少主成分才能对数据进行合理汇总的方法。函数型数据分析的第一步便是数据的函数表示,即平滑与插值。对于模型

$$Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + \int_0^1 \gamma(t) x_i(t) dt + \varepsilon_i \quad (7)$$

中函数型数据（积分部分）的处理，本文首先利用傅里叶基函数展开将离散的数据转化为函数曲线。因为我们通常观测到的是一定区间内的稠密点，需要拟合成光滑的曲线，才能提取函数特征。我们可以在保证数据周期性的前提下进行一定程度的平滑，方法是将傅里叶级数拟合到每条观测上：一个常数加上几对正弦和余弦项就可以很好地处理一些数据。然后本文采用 $K-L$ 展开和函数型主成分回归的思想来近似上述模型（7），根据 *Mercer's* 定理 (*R & Sz. 1955*):

$$K_X(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t)$$

和 *Karhunen-Loève* 展开:

$$x(t) = \sum_{j=1}^{\infty} \xi_j \phi_j(t)$$

其中 $K_X(s, t)$ 是 $x(t)$ 的协方差函数，*Mercer's* 定理就是对这个函数做特征分解，而 $K-L$ 展开则是说随机过程 $x(t)$ 可以由 $K_X(s, t)$ 的特征函数线性表示，其中 ξ_j 为随机变量，且 $E(\xi_j) = 0$ ， $Var(\xi_j) = \lambda_j$ ， $\lambda_1 > \lambda_2 > \dots > 0$ 为特征值。

对于函数型主成分分析，我们首先是想找到 $\phi_1(t)$

$$\begin{aligned} \max \quad & var\left[\int \phi_1(t) x_i(t) dt\right] \\ \text{s.t} \quad & \int \phi_1(t)^2 dt = 1 \end{aligned}$$

然后找到 $\phi_2(t)$

$$\begin{aligned} \max \quad & var\left[\int \phi_2(t) x_i(t) dt\right] \\ \text{s.t} \quad & \int \phi_2(t)^2 dt = 1 \\ & \int \phi_1(t) \phi_2(t) dt = 0 \end{aligned}$$

以此类推，可以找到正交基 $\{\phi_1(t), \phi_2(t), \dots\}$ 作为主成分，而 $\lambda_j / \sum \lambda_j$ 是表示第 j 个主成分的方差解释度。那么 $x_i(t)$ 的第 j 个主成分得分为

$$u_{ij} = \int \phi_j(t) x_i(t) dt = \langle x_i(t), \phi_j(t) \rangle$$

若选取前 m 个主成分， m 充分大，再带入 $K-L$ 展开的公式，模型（7）可近似为

$$Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + \mathbf{u}_i^T \boldsymbol{\gamma} + \varepsilon_i \quad (8)$$

其中 $\mathbf{u}_i = (\langle x_i(t), \phi_1(t) \rangle, \dots, \langle x_i(t), \phi_m(t) \rangle)^T$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^T$, $\gamma_j = \langle \gamma(t), \phi_j(t) \rangle$ 。

但实际中，特征函数并不知道，可用

$$\hat{K}_X(s, t) = \frac{1}{n} \sum_{i=1}^n x_i(s) x_i(t) = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{\phi}_j(s) \hat{\phi}_j(t)$$

估计。

3.2 模型的惩罚经验似然估计与推断

对于模型 (8)，给定 $\boldsymbol{\beta}$ ，得到 $\boldsymbol{\gamma}$ 的剖面最小二乘估计作为一个初始的估计

$$\tilde{\boldsymbol{\gamma}}(\boldsymbol{\beta}) = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \mathbf{Z}_i^T \boldsymbol{\beta} - \mathbf{u}_i^T \boldsymbol{\gamma})^2 = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T (\mathbf{Y} - \mathbf{Z} \boldsymbol{\beta})$$

那么我们可以构造估计方程：

$$E(\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})) = \mathbf{0} \quad (9)$$

$$\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) = \mathbf{Z}_i [Y_i - \mathbf{Z}_i^T \boldsymbol{\beta} - \mathbf{u}_i^T \tilde{\boldsymbol{\gamma}}(\boldsymbol{\beta})]$$

其中 $\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})$, $i = 1, 2, \dots, n$ 相互独立，式 (9) 由于最小二乘估计的无偏性以及截断参数 m 的选择得到。

那么可以定义剖面(profile)经验对数似然函数：

$$l(\boldsymbol{\beta}) = \sup \left\{ \sum_{i=1}^n \log(\omega_i) \mid \omega_i \geq 0, \sum_{i=1}^n \omega_i = 1, \sum_{i=1}^n \omega_i \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) = \mathbf{0} \right\}$$

用拉格朗日乘数法解上述问题得到：

$$\left\{ \begin{array}{l} \omega_i = \frac{1}{n[1 + \boldsymbol{\lambda}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})]} \\ \sum_{i=1}^n \frac{\boldsymbol{\lambda}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} = 0 \\ l(\boldsymbol{\beta}) = - \sum_{i=1}^n \{ \log(n) + \log(1 + \boldsymbol{\lambda}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})) \} \end{array} \right.$$

经验似然估计即为解如下优化问题：

$$\begin{aligned}
\min Q(\boldsymbol{\beta}) &= \sum_{i=1}^n \log[1 + \lambda^T \tilde{\eta}_i(\boldsymbol{\beta})] \\
\text{s.t.} \quad & \sum_{i=1}^n \frac{\lambda^T \tilde{\eta}_i(\boldsymbol{\beta})}{1 + \lambda^T \tilde{\eta}_i(\boldsymbol{\beta})} = 0 \\
& 1 + \lambda^T \tilde{\eta}_i(\boldsymbol{\beta}) \geq \frac{1}{n}, \quad 1 \leq i \leq n
\end{aligned} \tag{10}$$

可得

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{el} &= \underset{\boldsymbol{\beta}}{\operatorname{argmin}} Q(\boldsymbol{\beta}) \\
\Rightarrow \hat{\boldsymbol{\gamma}}_{el} &= (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T (\mathbf{Y} - \mathbf{Z} \hat{\boldsymbol{\beta}}_{el}) \\
\Rightarrow \hat{\boldsymbol{\gamma}}_{el}(t) &= \hat{\boldsymbol{\phi}}(t)^T \hat{\boldsymbol{\gamma}}_{el}, \quad \hat{\boldsymbol{\phi}}(t) = (\hat{\phi}_1(t), \dots, \hat{\phi}_m(t))^T
\end{aligned}$$

再根据 H, Xue & Feng (2018)^[1]中完备数据下的经验似然估计, 我们有如下模型的假设:

- A1: 平方可积的随机函数 $x(t)$ 满足 $E(\|x(t)\|^4) < \infty$;
- A2: 对每个 j 都有 $E(u_j^4) < C\lambda_j^2$;
- A3: $E(\|Z\|_{R^p}^4) < \infty, E(\varepsilon^2 | Z = z_i, X = x_i) = \sigma^2 < \infty$
- A4: 对每个特征值 λ_j , 存在一个常数 C 和 $a > 1, C^{-1}j^{-a} \leq \lambda_j \leq Cj^{-a}, \lambda_j - \lambda_{j+1} \geq Cj^{-a-1}, j \geq 1$;
- A5: 对于系数 γ_j , 存在一个常数 C 和 $b, \gamma_j \leq Cj^{-b}, j \geq 1, b > a/2 + 1$;
- A6: 存在存在一个常数 C , 对于每个 $k, \left| \langle K_{z_i, x}, \phi_j \rangle \right| \leq Cj^{-(b+a)}, j \geq 1$;
- A7: 对于调和参数 m , 假定 $m \sim n^{1/(a+2b)}$;

在高维数据分析中, 当参数的维数较大时, 即 $p \rightarrow \infty$, 一个合理的假设是, 只有一个参数子集是非零的。为了选择出有影响力变量子集并在稀疏的假设下对 $\boldsymbol{\beta}$ 进行统计推断, 我们在 (10) 式后面添加一个惩罚函数项, 那么惩罚经验似然的目标函数定义为:

$$\tilde{Q}(\boldsymbol{\beta}) = \sum_{i=1}^n \log[1 + \lambda^T \tilde{\eta}_i(\boldsymbol{\beta})] + n \sum_{j=1}^p p_\tau(|\beta_j|) \tag{11}$$

这里 $p_\tau(|\beta_j|)$ 是惩罚函数, 其中 τ 是调和参数。在文献中有许多常用的惩罚函数, 例如, L_1 和 L_2 惩罚函数; Tibshirani (1996)^[26]的 LASSO 惩罚; 的 Fan & Li (2001)^[3]提出的平滑剪切绝对偏差惩罚(SCAD); Zhang & Lu (2007)^[48]自适应 LASSO。在

本节中，我们利用 Fan & Li (2001)^[3]中的 SCAD 惩罚函数来研究部分函数型线性模型的惩罚经验似然估计(PEL)，这里 SCAD 惩罚函数一阶导数满足

$$p'_\tau(t) = \tau \left\{ I(t \leq \tau) + \frac{(a\tau - t)_+}{(a-1)\tau} I(t > \tau) \right\}$$

其中 $I(\cdot)$ 是示性函数且 $a > 2$ ，类似于 Fan & Li (2001)^[3]，本文也设置 $a = 3.7$ 。

令 $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \dots, \beta_{0p})^T \in \mathbb{R}^p$ ，且定义 $\wp = \{j: \beta_{0j} \neq 0\}$ ，其基数 $d = |\wp|$ 。不失一般性，让 $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ ，其中 $\boldsymbol{\beta}_1 \in \mathbb{R}^d$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p-d}$ 分别对应非零系数部分和零系数部分。因此真实系数可以写为 $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}^T, \mathbf{0}^T)^T$ ，则惩罚经验似然估计即最小化式(11)得 $\hat{\boldsymbol{\beta}}_{pel} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T)^T$ 。

根据 Jiang, Huang & Fan (2019)^[45]对高维部分函数型线性模型的经验似然研究以及 Leng & Tang (2012)^[5]对于一般估计方程的惩罚经验似然的估计与统计量性质的证明框架，我们可以在添加以下模型的假设下得出本文相应的渐近性质：

$$\text{A8: 令 } \boldsymbol{\eta}(\boldsymbol{\beta}) = \varepsilon \left(\mathbf{Z} - \sum_{i=1}^{\infty} \frac{\langle K_{\mathbf{Z}\mathbf{X}}, \phi_j \rangle \langle \mathbf{X}, \phi_j \rangle}{\lambda_j} \right), \text{ 那么对于整数 } k \geq 4, \text{ 存在常数 } C,$$

当 $n \rightarrow \infty$ 时，有 $E\left(\|\boldsymbol{\eta}(\boldsymbol{\beta})/\sqrt{p}\|^k\right) < C$;

A9: 让 $\boldsymbol{\Sigma}(\boldsymbol{\beta}) = E[\boldsymbol{\eta}_i(\boldsymbol{\beta})\boldsymbol{\eta}_i(\boldsymbol{\beta})^T]$ ，存在 $0 < d \leq D < \infty$ ，使得 $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ 的特征值在它们之间；

A10: 随着 $n \rightarrow \infty$, $p^5/n \rightarrow 0$ ，且对于某个 $y, 0 < C_0 < y < 1$, 有 $p/r \rightarrow y$ 。

A11: 随着 $n \rightarrow \infty$, $\tau(n/p) \rightarrow \infty$ 且 $\min \beta_{0j}/\tau \rightarrow \infty$ 。

$$\text{A12: } \max_{j \in \wp} p'_\tau(|\beta_{0j}|) = o\left\{(np)^{-1/2}\right\} \text{ 且 } \max_{j \in \wp} p'_\tau(|\beta_{0j}|) = (p)^{-1/2}.$$

这里 A9 是高维情形下的正则化条件，A10 和 Leng & Tang (2012)^[5]中所提及的那样要求 $p = o(n^{1/5})$ 。A12 用来控制惩罚项对非零分量的影响，而在给定 A11 时，Fan & Li (2001)^[3]的 SCAD 惩罚函数是满足这个条件的，因为当 n 足够大时，

$$\max_{j \in \wp} p'_\tau(|\beta_{0j}|) = 0.$$

定理 1: 假设(A1) - (A12)成立，那么随着 $n \rightarrow \infty$ ，惩罚经验似然估计 $\hat{\boldsymbol{\beta}}_{pel}$ 满足：

(a) (选择一致性)：

$$P(\hat{\boldsymbol{\beta}}_2 = \mathbf{0}) \rightarrow 1, \quad n \rightarrow \infty;$$

(b) (渐近有效性)：

$$\sqrt{n}\mathbf{B}_n\boldsymbol{\Omega}_p^{-1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10}) \xrightarrow{L} N(0, \mathbf{G}), \quad n \rightarrow \infty$$

其中 \mathbf{B}_n 是一个 $l \times d$ 的矩阵且 $\mathbf{B}_n\mathbf{B}_n^T \rightarrow \mathbf{G} \in \mathbb{R}^{l \times l}$ 。

定理 1 (a) 说明了本文得到的惩罚经验似然估计 $\hat{\boldsymbol{\beta}}_{pel}$ 具有如 Fan & Peng (2004)^[28] 所提到的 Oracle 性质, 也就是说, 惩罚的经验似然方法在模型选择上是一致的, 并且执行起来就像真正的稀疏模型是已知的一样有效。且定理 1 (b) 说明了非零真实参数的惩罚经验似然估计 $\hat{\boldsymbol{\beta}}_1$ 的 \sqrt{n} 一致性及其渐近正态性。经验似然的一个显著优势就在于构造假设检验问题或者 $\boldsymbol{\beta}$ 的置信区域。在实际应用中, 未知参数的有限组合的置信域是非常重要的。这里我们考虑如下线性检验假设的问题:

$$H_0: \mathbf{L}_n\boldsymbol{\beta}_0 = \mathbf{0} \quad \text{vs} \quad \mathbf{L}_n\boldsymbol{\beta}_0 \neq \mathbf{0}$$

这里 $\mathbf{L}_n \in \mathbb{R}^{q \times p}$ 且对于固定 q 有 $\mathbf{L}_n\mathbf{L}_n^T = \mathbf{I}_q$, 其中 \mathbf{I}_q 是 q 维的单位矩阵。基于式 (11), 可以构造一个惩罚经验似然比统计量

$$\tilde{Q}(\mathbf{L}_n) = -2 \left\{ \tilde{Q}(\hat{\boldsymbol{\beta}}_{pel}) - \min_{\boldsymbol{\beta}, \mathbf{L}_n\boldsymbol{\beta}=\mathbf{0}} \tilde{Q}(\boldsymbol{\beta}) \right\} \quad (12)$$

那么惩罚经验似然比统计量遵循下面的定理:

定理 2: 假设(A1) - (A12)成立, 那么随着 $n \rightarrow \infty$,

$$\tilde{Q}(\mathbf{L}_n) \xrightarrow{L} \chi_q^2。$$

定理 2 将 Leng & Tang (2012)^[5] 的结论应用到高维部分函数型线性模型。通过定理 2, 我们能够构造 $\mathbf{L}_n\boldsymbol{\beta}_0$ 的渐近置信域, 即

$$\ell_\alpha = \left\{ \mathbf{v}: -2 \left\{ \tilde{Q}(\hat{\boldsymbol{\beta}}_{pel}) - \min_{\boldsymbol{\beta}, \mathbf{L}_n\boldsymbol{\beta}=\mathbf{v}} \tilde{Q}(\boldsymbol{\beta}) \right\} \leq \chi_{q, 1-\alpha}^2 \right\},$$

这里 $\chi_{q, 1-\alpha}^2$ 是 χ_q^2 分布的 $1-\alpha$ 下侧分位数, 且当 $n \rightarrow \infty$ 时, $\mathbf{L}_n\boldsymbol{\beta}_0$ 的渐近置信域满足 $P\{\boldsymbol{\beta}_0 \in \ell_\alpha(\boldsymbol{\beta})\} \rightarrow 1-\alpha$ 。

本文研究了高维部分函数型线性模型的惩罚经验似然的统计推断问题, 但该框架仅适用于样本容量大于参数维度的情况。当这种情况被违反时, 可以使用 SIS 等初步筛选的方法将维度由高降至中等规模即低于样本量大小。

第 4 章 算法与调和参数的选择

4.1 经验似然的嵌套坐标下降算法

首先, 我们描述了如何处理经验似然提出的优化问题。由于经验似然比的非凸性, 计算经验似然估计量也需要合适的处理与算法。此外, 3.2 节还涉及到了不可微的惩罚函数, 从而也为惩罚经验似然估计的计算增加了难度。这里, 我们采用 Owen (2001)^[21]的 Newton-Lagrange 算法解 3.2 节中式 (10) 的经验似然优化问题。由于 Newton-Lagrange 算法需要计算大矩阵的逆, 因此当变量的维度较小时, 可以很快速的计算出估计值, 若给定一个比较合适的初始值, 一般只需要 1-2 次迭代即可收敛。式 (9) 的经验似然优化问题可等价求解如下问题:

$$\min_{\beta, \lambda} Q^*(\beta, \lambda) = \sum_{i=1}^n \log^*[1 + \lambda^T \tilde{\eta}_i(\beta)]$$

其中

$$\log^*(v) = \begin{cases} \log(v), & v \geq \varepsilon \\ \log(\varepsilon) - 1.5 + \frac{2v}{\varepsilon} - \frac{v^2}{2\varepsilon^2}, & v \leq \varepsilon \end{cases}, \quad \varepsilon > 0.$$

这里本文将 $\log(v)$ 替换为 $\log^*(v)$ 是根据 Owen (2001)^[21]提到的方便我们加入 $\omega_i \geq 0$ 的约束。 $Q^*(\beta, \lambda)$ 是关于 λ 二次连续可微且对于每个 β , 在 λ 上是凸的。如果我们知道优化问题的解一定要满足 $\max_i n p_i \leq A$, 那么我们可以选择 $\varepsilon = 1/A$ 并且通过 $\log^*(v)$ 得到解而非通过在 $\log(v)$ 下添加约束项, 即由于 $\omega_i = 1/\{n[1 + \lambda^T \tilde{\eta}_i]\} \leq 1$ 可取 $\varepsilon = 1/n$ 。然后通过牛顿法, 即解以下等式

$$\begin{pmatrix} Q_{\lambda}^* \\ Q_{\beta}^* \end{pmatrix} = \mathbf{0} \in \mathbb{R}^{2p}$$

相应的 Hessian 矩阵为

$$\begin{pmatrix} Q_{\lambda\lambda}^* & Q_{\lambda\beta}^* \\ Q_{\beta\lambda}^* & Q_{\beta\beta}^* \end{pmatrix}.$$

这里

$$\mathbf{Q}_\lambda^* = \frac{\partial}{\partial \lambda} Q^*(\boldsymbol{\beta}, \lambda) = \sum_{i=1}^n \log^*(v) \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})$$

$$\mathbf{Q}_{\lambda\lambda}^* = \frac{\partial^2}{\partial \lambda \partial \lambda^T} Q^*(\boldsymbol{\beta}, \lambda) = \sum_{i=1}^n \log^*(v) \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})^T$$

然而，对于式 (11) 的惩罚经验似然的优化问题，我们采用 Cheng & Tong (2014)^[40] 的嵌套坐标下降算法 (Nested coordinate descent algorithms) 解高维稀疏的惩罚经验似然问题。我们可以看到，除了感兴趣的参数 $\boldsymbol{\beta}$ ，拉格朗日乘数 λ 也同样参与了优化。与统计推断中的那些典型优化问题不同，参数和拉格朗日乘数使得搜索空间增大，尤其是当变量的维数增加甚至随样本量变化时，对高效稳定的优化算法提出了新的挑战。另一方面，循环坐标下降算法在求解这类问题时具有较好的优势。在循环坐标下降过程中，所有参数依次循环和更新。因此，多元优化问题归结为一系列单变量优化，只需要标量计算，完全避免了矩阵运算。如上面提到的 Newton-Lagrange 算法，直接一起搜索 $(\boldsymbol{\beta}^T, \lambda^T)^T$ ，这样的策略在直觉上似乎很显然，但它可能非常不稳定，且参数维数很高时会导致矩阵的计算困难。因为 λ 本质上取决于给定值 $\boldsymbol{\beta}$ ，因此在固定 λ 下更新 $\boldsymbol{\beta}$ 的坐标可能导致激烈的目标函数值的变化。为了在惩罚经验似然方法中加入这种优化特性，并克服不稳定性带来的困难，本文使用嵌套坐标下降算法。嵌套算法主要分为两层，其中内层是在固定 $\boldsymbol{\beta}$ 下求解拉格朗日乘子 λ ，外层则是关于 $\boldsymbol{\beta}$ 的求解。也就是说， $\lambda = \lambda(\boldsymbol{\beta})$ 可看作是关于 $\boldsymbol{\beta}$ 的函数。式 (11) 的惩罚经验似然优化问题可等价求解如下问题：

$$\min_{\boldsymbol{\beta} \in \Theta_0} \max_{\lambda \in \Lambda_n(\Theta_0)} \tilde{Q}(\boldsymbol{\beta}, \lambda) = \sum_{i=1}^n \log^*[1 + \lambda^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})] + n \sum_{j=1}^p p_\tau(|\beta_j|)。$$

由于其中 SCAD 惩罚函数 $p_\tau(|\beta_j|)$ 在原点处为奇异函数，且它不具有连续的二阶导数。然而，它可以被一个二次函数局部近似，即使用 Fan & Li (2001)^[3] 的局部二次逼近 (LQA) 算法。先给定一个与式 (11) 最小值解很接近的初始值 $\boldsymbol{\beta}^0$ ，那么对于第 k 步 β_j 的估计值 $\beta_j^{(k)}$ ，如果 $\beta_j^{(k)}$ 非常接近 0 ($\beta_j^{(k)} \leq 10^{-6}$)，可以直接设置 $\beta_j^{(k)} = 0$ ；若 $\beta_j^{(k)} \neq 0$ ，那 $p_\tau(|\beta_j|)$ 可以局部近似为

$$p_\tau(|\beta_j^{(k)}|) + \frac{1}{2} \frac{p_\tau'(|\beta_j^{(k)}|)}{|\beta_j^{(k)}|} (\beta_j^2 - (\beta_j^{(k)})^2),$$

重复这个过程直到收敛。

以上过程涉及到的主要导数为

$$\begin{aligned}
\tilde{\mathbf{Q}}_{\lambda}^* &= \mathbf{Q}_{\lambda}^* \\
\tilde{\mathbf{Q}}_{\lambda\lambda}^* &= \mathbf{Q}_{\lambda\lambda}^* \\
\tilde{\mathbf{Q}}_{\beta}^* &= \sum_{i=1}^n \log^*(v) \frac{\partial \tilde{\eta}_i^T}{\partial \beta} \lambda + n \left\{ \frac{p'(|\beta_j^{(k)}|)}{|\beta_j^{(k)}|} \beta_j \right\}_{j=1,2,\dots,p} \\
\tilde{\mathbf{Q}}_{\beta\beta}^* &= \sum_{i=1}^n \log^*(v) \frac{\partial \tilde{\eta}_i^T}{\partial \beta} \lambda \lambda^T \frac{\partial \tilde{\eta}_i}{\partial \beta} + n \left\{ \frac{p'(|\beta_j^{(k)}|)}{|\beta_j^{(k)}|} \right\}_{j=1,2,\dots,p} \\
\tilde{\mathbf{Q}}_{\beta\lambda}^* &= \tilde{\mathbf{Q}}_{\lambda\beta}^* = \sum_{i=1}^n \log^*(v) \frac{\partial \tilde{\eta}_i^T}{\partial \beta} \lambda \tilde{\eta}_i(\beta)^T + \sum_{i=1}^n \log^*(v) \frac{\partial \tilde{\eta}_i}{\partial \beta}
\end{aligned}$$

其中 $\partial \tilde{\eta}_i / \partial \beta = -\mathbf{Z}_i \mathbf{Z}_i^T + \mathbf{Z}_i \mathbf{u}_i^T (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{Z}$ 。根据 Owen (2001)^[21]在嵌套坐标下降算法中外层的二阶导数通过计算

$$\mathbf{R}_{\beta\beta} = \tilde{\mathbf{Q}}_{\beta\beta}^* - \tilde{\mathbf{Q}}_{\beta\lambda}^* (\tilde{\mathbf{Q}}_{\lambda\lambda}^*)^{-1} \tilde{\mathbf{Q}}_{\lambda\beta}^*$$

的对角线元素得到。那么我们可以总结嵌套坐标下降算法更新 β 和 λ 过程的伪代码如下：

步骤 1：首先设置迭代计数器 $k = 0$ ，以及参数初始值 $\beta^{(0)}$ 和 $\lambda^{(0)}$ ；

步骤 2：定义函数 $\tilde{\eta}_i(\beta)$ ；

步骤 3：对于 $j = 1, 2, \dots, p$ ：

(1) 用 $\hat{\beta}_j^{(k+1)} = \hat{\beta}_j^{(k)} - \tilde{\mathbf{Q}}_{\beta}^{*(k)}(j) / \mathbf{R}_{\beta\beta}^{*(k)}(j, j)$ 更新 β_j

(2) 更新 λ 通过计算 $\lambda^{(k+1)} = \lambda^{(k)} - \tilde{\mathbf{Q}}_{\lambda}^{*(k)} (\tilde{\mathbf{Q}}_{\lambda\lambda}^{*(k)})^{-1}$

步骤 4：如果 $\max_{1 \leq j \leq p} (|\beta_j^{(k+1)} - \beta_j^{(k)}|) < \gamma$ ，迭代结束。

步骤 5：否则重复步骤 3 和步骤 4。

这里 γ 是一个阈值，例如设置 $\gamma = 10^{-5}$ 。

4.2 调和参数的选择

本文涉及到的调和参数主要有函数型主成分分析的主成分个数 m 和 3.2 节中惩罚函数中的 τ 。对于主成分个数，太少会导致信息提取不充分，太多会加入一些无用变量，增加估计偏差，从而影响预测效果。而我们研究模型的最终目的就是它预测与解释的效果要好。因此根据 H, Xue, & Feng (2018)^[1]提到的用 CV 选择截断参数 m ，即通过最小化下面的 CV 得分

$$CV(m) = \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_{pel}^{(-i)} - \mathbf{u}_i^T \hat{\boldsymbol{\gamma}}_{pel}^{(-i)})^2$$

这里 $\hat{\boldsymbol{\beta}}_{pel}^{(-i)}$ 和 $\hat{\boldsymbol{\gamma}}_{pel}^{(-i)}$ 为去除第 i 个样本后用剩下的样本所得到的惩罚经验似然估计。为了减少计算的复杂对，我们先画出函数型主成分分析的碎石图，根据碎石图确定最可能的 m 的取值，再用 CV（留一交叉验证）选择预测误差最小的 m 。

为了选出惩罚参数 τ ，类似于 Variyath et al. (2010)，我们使用下面的贝叶斯信息准则(BIC):

$$BIC(\tau) = 2\mathbf{Q}(\boldsymbol{\beta}_\tau) + df_\tau \log(n)$$

其中 $\boldsymbol{\beta}_\tau$ 是给定调和参数 τ 的情况下 $\boldsymbol{\beta}$ 的惩罚经验似然估计， df_τ 则为 $\boldsymbol{\beta}_\tau$ 中非零参数的个数。为了在每次的循环中能减少计算量，本文先不断尝试选出适合数据的 τ 的最小变动区间，以减少计算的复杂度。但这样在留一交叉验证下再嵌套惩罚参数的选择，计算复杂度还是很高，尤其是当 n 很大时。因此本文采用 Kong, Xue, Yao, & Zhang (2016)^[2]的成对最小化 ABIC 的方法选择最优的调和参数对 (m^*, τ^*) ,

$$ABIC(m, \tau) = 2\tilde{\mathbf{Q}}(\boldsymbol{\beta}_{m,\tau}) + df_{m,\tau} \log(n) + 2m$$

这里 $df_{m,\tau} = I(\boldsymbol{\beta}_{m,\tau}) + \sum_{j=1}^m \lambda_j$ ，其中 $I(\boldsymbol{\beta}_{m,\tau})$ 表示 $\boldsymbol{\beta}_{m,\tau}$ 中非零参数的个数， λ_j 表示第 j 个主成分的特征值。

接下来，我们利用估计方程对部分函数型线性模型惩罚经验似然法的推理进行仿真研究，验证它们推理的性质，并对比说明惩罚经验似然方法有效性。

第 5 章 模拟研究

为了验证上述定理和方法的有效性, 本文从模型

$$\mathbf{Y} = \mathbf{Z}^T \boldsymbol{\beta} + \int_0^1 \gamma(t) \mathbf{x}(t) dt + \boldsymbol{\varepsilon}$$

生成数据, 其中 $\boldsymbol{\beta}_{10} = (2, -1, 1, 1.5)^T$, $\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{B})$, $\mathbf{B} = \{0.25^{k-l}\}_{k,l=1,2,\dots,p}$, $\varepsilon_i \stackrel{iid}{\sim} N(0, 1)$,

$$\gamma(t) = \sqrt{2} \sin\left(\frac{\pi t}{2}\right) + 3\sqrt{2} \sin\left(\frac{3\pi t}{2}\right),$$

$$\mathbf{x}(t) = \sum_{j=1}^{\infty} u_j \phi_j(t), t \in [0, 1],$$

$$\phi_j(t) = \sqrt{2} \sin((j - 0.5)\pi t),$$

$$u_j \sim N(0, \lambda_j), \lambda_j = ((j - 0.5)\pi)^{-2}.$$

这里, $x_i(t)$ 的观测值在 $t \in [0, 1]$ 区间内的 100 个等间距点上取值, 样本量的大小取 $(n, p) = (200, 10), (300, 20), (400, 30), (500, 40)$, 且所有的试验都重复 500 次。

5.1 FPCA

我们能观测到的函数型数据是离散的点, 虽然具有函数特性, 但能记录的也是密集的等间距 (或不等间距) 的点。所以我们先对函数型部分的数据即 $x_i(t)$ 用傅里叶基函数展开的方法光滑化, 从而可以从曲线的角度去描述模型 (1)。图 1 便是傅里叶基函数光滑化后的 ($n=200$) 曲线, 它们的均值曲线在 0 附近。那么我们可以对这 n 条曲线基于 K-L 展开进行函数型主成分分析, 画出它的碎石图如图 2。

图 2 的碎石图中的左图是每个主成分曲线的贡献度, 即 $\lambda_j / \sum \lambda_j$, 右图则是主成分曲线的累积方差贡献度。可以看出前两个主成分曲线的方差贡献度达到了 91%, 前三个则接近 95% 的累积贡献度, 且从第四个主成分曲线开始, 它们的方差贡献度都接近零。因此我们可以取截断参数 $m=2$ 或 $m=3$ 作为 4.2 节中提到的初始参数值, 具体的选择再结合模型的拟合效果, 即第 4 章提到的 ABIC 选择的 $m=2$ 。

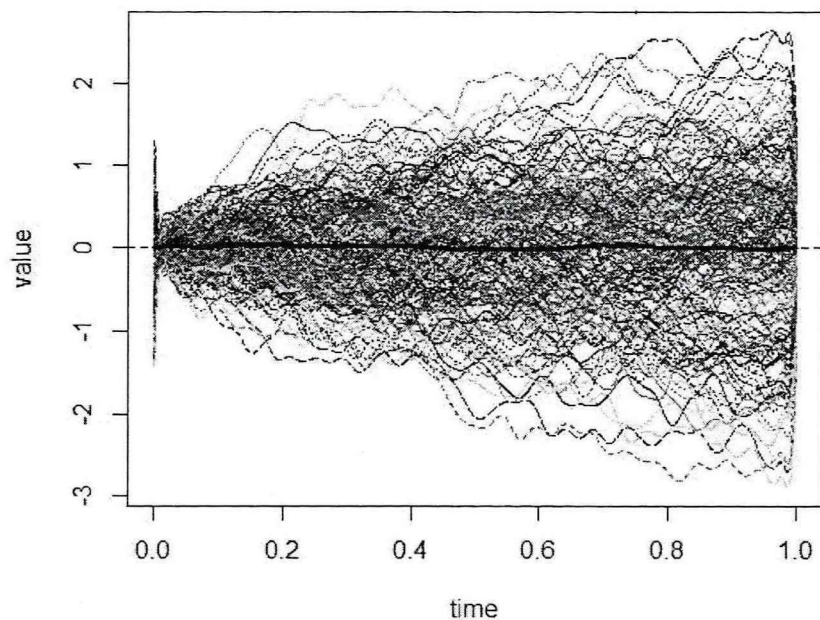


图 1: 函数型数据的光滑曲线图

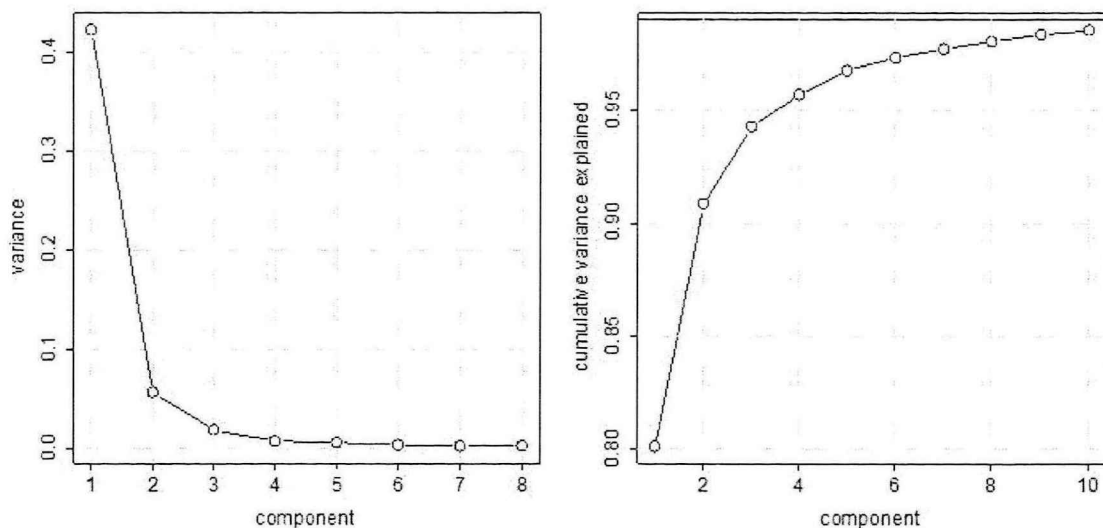


图 2: 碎石图 (左: 主成分的方差贡献度); (右: 主成分的累积方差贡献度)

图 3 还画出了前两个主成分曲线图, 图中红色实线代表第一个主成分曲线, 红色虚线则为其真实的特征曲线, 即 $\phi_1(t)$; 黑色实线代表第二个主成分曲线, 黑色虚线则为其真实的成分曲线, 即 $\phi_2(t)$ 。很明显可以看出实线与虚线很接近, 这进一步验证了 K-L 展开的有效性。这两条特征曲线可以作为函数型数据样本的特征因子来描述数据带来的信息。

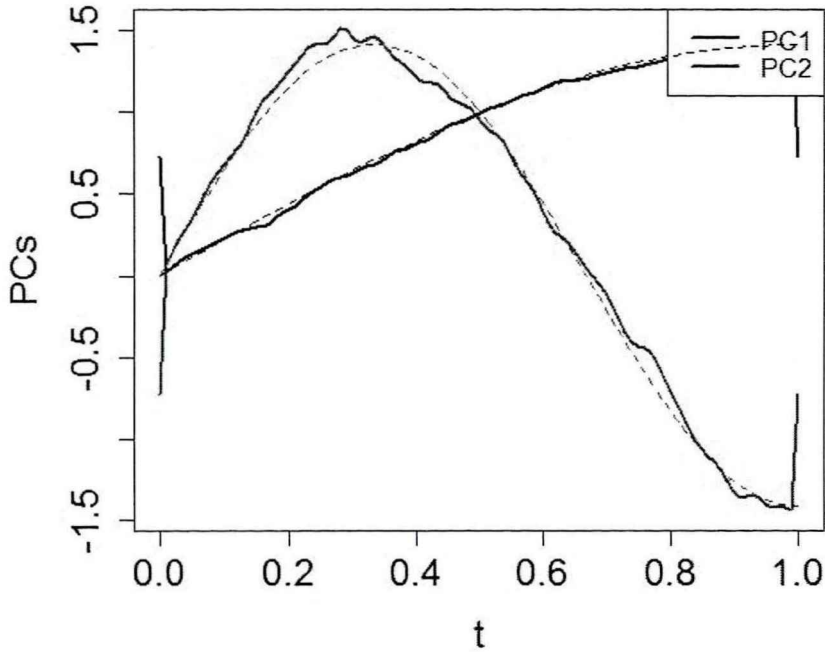


图 3: 前两个函数型主成分曲线与真实曲线

5.2 估计量的 Oracle 性质

变量的 Oracle 性质是以概率 1 识别真实系数，也就是说已知真实为零的系数，来估计非零的系数。因此我们首先根据第 4 章提到的算法估计式 (11) 得到了每个 β 的惩罚经验似然估计值，然后令 $p = 4$ ，即已知 $\beta_2 = 0$ 用经验似然估计式 (10) 得到 β_1 的估计 β_{or} 。同时，在惩罚经验似然估计中，还统计了估计值中被真实压缩为零的系数个数以及被错误压缩为零的系数个数。整个过程重复生成样本 500 次。这里，我们使用参数估计值与真实值之间的根号均方误差(RMSE)以系数函数估计值与真实系数函数之间的根号均方可积误差(RMISE)作为参数和系数函数拟合效果的评价标准，得到了不同样本量和不同变量维度组合下 β 的惩罚经验似然估计值的 RMSE 和 $\gamma(\cdot)$ 估计值的 RMISE 如表 1 所示。同时，表 1 还列出了 500 次模拟的平均零系数压缩的结果。

表 1: 不同设置下的惩罚经验似然估计 $\hat{\beta}_j$ 的 RMSE 和 $\hat{\gamma}(\cdot)$ 的 RMISE 以及模型选择

n	p		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}(\cdot)$	Zero coefficients	
								True	False
200	10	$\hat{\beta}$	0.484313	0.505957	0.555462	0.535438	3.415327	5.44	0.14
		$\hat{\beta}_{Or}$	0.585239	0.54923	0.492682	0.459247	3.329351	--	--
300	20	$\hat{\beta}$	0.354831	0.383017	0.390815	0.356023	2.953471	15.53	0.02
		$\hat{\beta}_{Or}$	0.370958	0.347291	0.338299	0.367368	3.076197	--	--
400	30	$\hat{\beta}$	0.261804	0.288333	0.339286	0.257762	2.817737	25.59	0.08
		$\hat{\beta}_{Or}$	0.240775	0.199416	0.271205	0.275458	2.781849	--	--
500	40	$\hat{\beta}$	0.233538	0.246932	0.268845	0.215567	2.659572	35.78	0.01
		$\hat{\beta}_{Or}$	0.250767	0.225274	0.207382	0.212136	2.722228	--	--

从表 1 可知, 随着样本量的增大且参数维度 p 也随之增大, 惩罚经验似然的估计效果都明显改善, 所以当参数维度发散时, 惩罚经验似然也很有效。 β 的估计 RMSE 也都很小, 更重要的是, 它们与各自的 Oracle 估计量的效果都非常接近。这就证实了惩罚经验似然估计在部分函数型线性模型中的 Oracle 性质。再看零系数的压缩结果, 500 次模拟中被错误压缩的系数几乎为 0 个, 即在 β_1 中几乎没有系数被压缩到零。而且, 真实为零的系数 β_2 中压缩到零的平均系数的个数也大都接近 $p-4$ 。其中压缩结果自然取决于调和参数 τ 的选择结果, τ 越大, 越容易将所有系数压缩到零, 但 τ 越小则会产生估计值的大偏差, 这里我们通过 ABIC 选择的最优的 τ , 设置合适的网格范围, 每次模拟的样本都选择 ABIC 最小的 τ , 平均的来看 τ 大概在 0.12~0.17 之间, 一般大于 0.2 便会使很多不为零的系数也压缩到零。可以看出 SCAD 惩罚函数在这里还是具有一定的变量选择效果的。

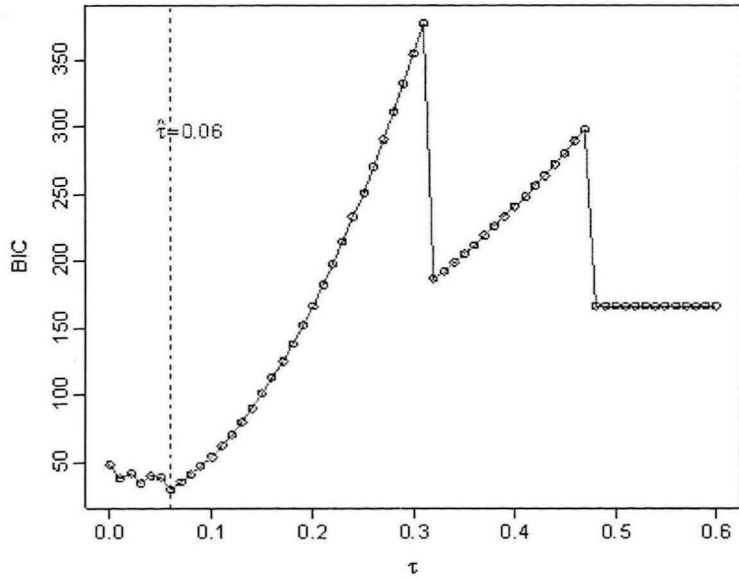


图 4 (上) : $(n, p, \rho) = (200, 21, 0.1)$ 时模拟数据的 BIC 随 τ 变化的散点图

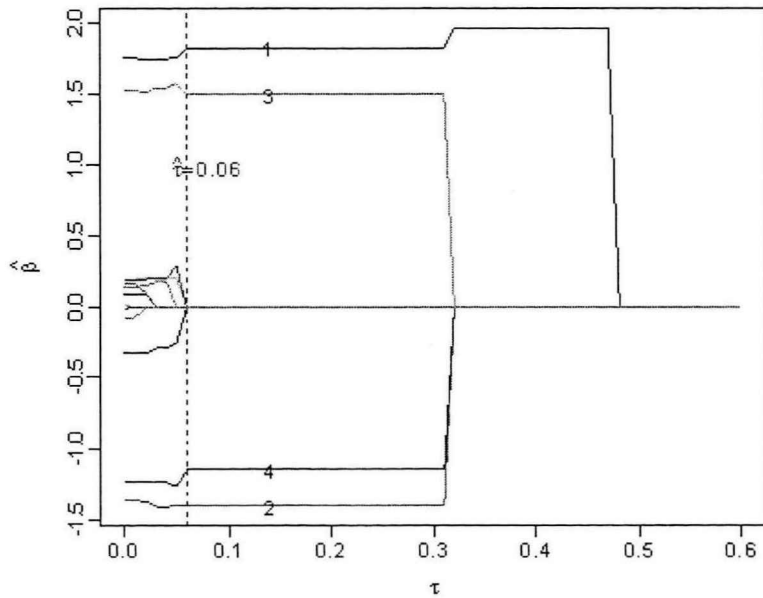


图 4 (下) : $(n, p, \rho) = (200, 21, 0.1)$ 时模拟数据的 PEL 值的解的路径

图 4 分别画出了 $(n, p, \rho) = (200, 21, 0.1)$ 时，一次模拟数据的 BIC 随 τ 变化的散点图和惩罚经验似然估计的关于 τ 的解的路径。其中垂直虚线表示 $\tau = 0.06$ 时 BIC 达到最小。图 4 (下) 中标号 1, 2, 3, 4 分别代表前四个真实非零参数的解的路径。我们观察到相应于非零系数的四个变量与最小 BIC 准则的选择结果一致，说明在 BIC 准则下该方法一定程度上能够识别出真实模型。

对于系数函数的估计, 由于惩罚似然法的结果的优越性以及函数型主成分分析的有效性, 从惩罚经验似然估计和剖面最小二乘得到的 $\hat{\gamma}(\cdot)$ 的 RMISE 也较为合理。下面的图 5 则画出了真实系数函数 $\gamma(\cdot)$ 以及 $(n, p) = (300, 20)$ 下 500 次模拟数据的惩罚经验似然估计 $\hat{\gamma}(\cdot)$ 的 10% 和 90% 经验百分位函数。

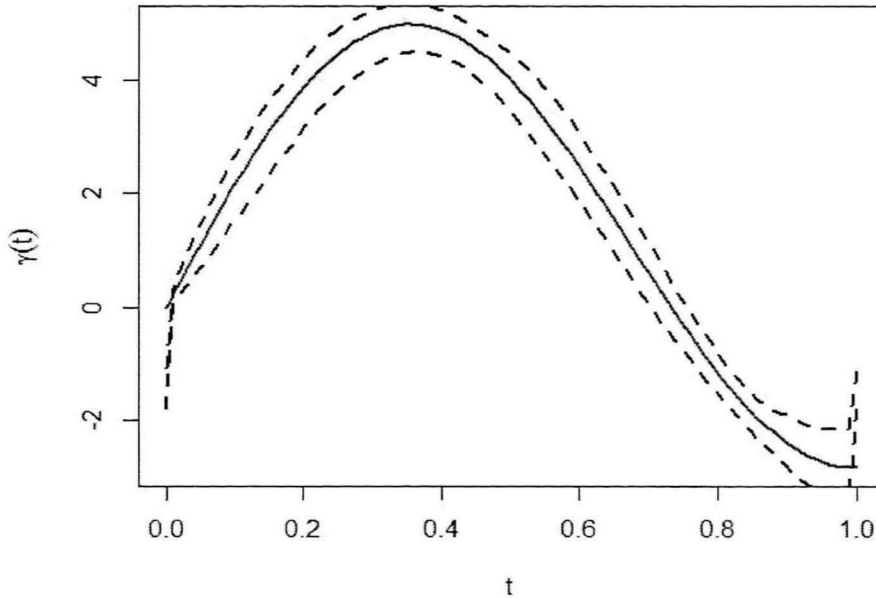


图 5: 真实系数函数 $\gamma(t)$ 及其 $(n, p) = (300, 20)$ 下的 10% 和 90% 的经验分位数函数

图 5 中红色实线则是代表真实的系数函数 $\gamma(t)$, 黑色虚线则分别表示 500 次模拟中 10% 和 90% 的经验分位数函数, 即在每一个点 t 的位置上对 500 次系数函数值的估计值排序, 取第 10% 和第 90% 的数值作为每个 t 点位置经验分位数, 再连接为经验分位数曲线。从图中可以很明显的看出, 真实系数函数完全包括在 10% 和 90% 的经验分位数函数曲线之间, 进一步说明了整个估计过程的有效性。

5.3 置信域以及假设检验的功效

在式 (12) 中, 若取 $\mathbf{L}_n = (1, 1, \dots, 1)$, 为了验证定理 2, 本文记录了 500 次模拟中的惩罚似然函数值, 绘制 $(n, p) = (200, 10)$ 和 $(n, p) = (500, 40)$ 下惩罚经验似然比统计量的卡方 Q-Q 图如图 6 所示。对于 $\mathbf{L}_n = (1, 0, \dots, 0)$, 本文还构造了以 β_1 为例的假设检验/置信区间的问题, 正如 3.3 节提到的那样, 具体的结果如表 2 所示。

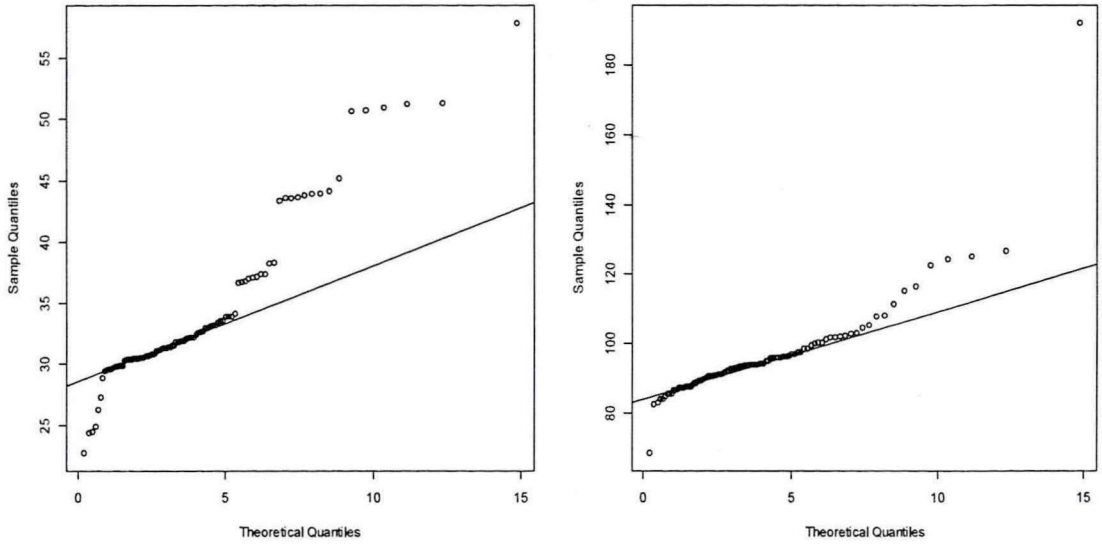


图 6: $(n,p)=(200,10)$ (左)和 $(n,p)=(500,40)$ (右)下惩罚经验似然比统计量的卡方 Q-Q 图

图 6 左图显示样本量大小与参数维度较小时，前面的惩罚似然值在直线上，但尾部偏差较大，且有跳跃，说明其中有系数估计不稳定导致的惩罚似然值有偏差。但是其中右图显示随着样本量大小与参数维度同时增大的情况下，除了尾部的几个值以外，其他所有的惩罚似然值基本在一条直线上，这同时也验证了定理 2 的结论，即当 $n \rightarrow \infty$ ，惩罚经验似然比统计量是渐近卡方分布的，其自由度为 q ，因此，由此作统计推断会有很好的效果。

表 2: 不同设置下给定的 β_1 不落入 95%置信区间的比率

n	p	power		size	power	
		$\beta_1 = 1.8$	$\beta_1 = 1.9$	$\beta_1 = 2$	$\beta_1 = 2.1$	$\beta_1 = 2.2$
200	10	0.3676	0.1171	0.092	0.1288	0.3346
300	20	0.3684	0.2344	0.088	0.2276	0.3885
400	30	0.6017	0.3609	0.076	0.3317	0.6408
500	40	0.9418	0.5438	0.061	0.5732	0.9532

同时，表 2 显示当 β_1 的真实值为 2 时，我们得到 β_1 从 1.8 到 2.2 每间隔 0.1 取值时，500 次模拟中不同 n 和 p 设置下其值不落入 95%置信区间的比率。从表

中可以看出当 $\beta_1 = 2$ 且 $n \rightarrow \infty$ 时, 其值不落入 95% 置信区间的比率接近名义值 $\alpha = 0.05$ 。也就是说, 当原假设成立时, 我们拒绝原假设的概率即在假设检验中犯第一类错误的比率 (size) 接近 $\alpha = 0.05$ 。而当 $\beta_1 \neq 2$, 随着其值增大, 不落入 95% 置信区间的比率越来越大, 且随着 n 增大, 该值也越来越大。这说明了当 $(n, p) = (500, 40)$ 和 $|\beta_1 - 2| = 0.2$ 时, 假设检验中不犯第二类错误的比率 (power) 接近 95%。这无疑验证了定理 2 和假设检验的有效性。

5.4 不同分布假定下的结果比较

本文在维度发散情形下对比不同残差分布下的最小二乘法(PLS)和惩罚经验似然法(PEL)的方法的变量选择和统计推断的结果。参数维度 $p_n = \lceil 10(3n)^{0.45} - 18 \rceil$ 随 n 变化。本节数据生成模型同上面一样, 但残差 ε_i 的分布分别取自标准正态分布 $N(0,1)$ 、自由度为 3 的 t 分布 $t(3)$ 、强度为 1 的指数分布和异方差正态分布 $\varepsilon_i \sim (1 + \mathbf{z}_{i1})N(0,1)$ 以及混合正态分布 $\varepsilon_i \sim 0.9N(0,1) + 0.1N(0,49)$, 其中样本大小 $n = 100, 200, 400, 600$, 则 β 的维度分别为 $p = 10, 21, 28, 33$ 。下面给出的结果是基于 500 次重复的结果。

表 3 是异方差和有极端值的设定下不同方法的非零系数估计值的均值和标准差。其中 ENet 表示的 $\alpha = 0.5$ 时的弹性网惩罚最小二乘, AL 则为适应权重为 $1/|\beta_j|^{0.5}$ 适应性 LASSO 惩罚。SCAD 和 PEL 是指带有 SCAD 惩罚项的惩罚最小二乘法和惩罚经验似然法。从表中我们可以看出, 当样本量较小时, PEL 在两种情形下的估计均值都非常接近真实值 $\beta_{01} = (2, -1, 1, 1.5)'$ 。而其他方法都与真值相差较大, 且它们标准差都相差不大。随着 n 的增大, PEL 的均值一直都最接近真值, 且估计值的标准差也逐渐减小。而 SCAD 和 AL 方法下的估计值均值也随着 n 的增大越来越接近真实值, 且 SCAD 的估计效果都一直略优于 AL。这无疑体现了 SCAD 惩罚的优势和 PEL 的估计效果。不论是异方差情形还是含有极端值的混合正态分布下, 惩罚似然估计都具有一定的稳健性和准确性。

表 3: 不同分布下不同惩罚的 PLS 与 PEL 参数估计值的均值与方差

n		$\varepsilon_i \sim (1+z_{i1})N(0,1)$				$\varepsilon_i \sim 0.9N(0,1)+0.1N(0,49)$			
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
100	ENet	1.4871 (0.6068)	-0.3467 (0.5908)	0.5789 (0.5332)	1.2475 (0.5467)	1.7392 (0.7775)	-0.7568 (0.6565)	0.8775 (0.6941)	1.3597 (0.7249)
	SCAD	1.7632 (0.7586)	-0.725 (0.7387)	0.7679 (0.6293)	1.3849 (0.6849)	1.5017 (1.0841)	-0.5479 (0.7724)	0.6768 (0.8003)	1.1834 (0.908)
	AL	1.5398 (0.6657)	-0.4085 (0.5215)	0.6122 (0.5359)	1.2306 (0.6049)	1.2046 (0.9348)	-0.249 (0.4755)	0.4836 (0.6716)	0.9497 (0.7951)
	PEL	2.0299 (0.6008)	-0.9961 (0.6423)	1.0034 (0.6179)	1.4983 (0.5972)	1.9352 (0.3392)	-1.0058 (0.8582)	1.0368 (0.7909)	1.5427 (0.727)
200	ENet	1.4976 (0.3505)	-0.3923 (0.3462)	0.5948 (0.3552)	1.1769 (0.3)	1.5316 (0.3526)	-0.4434 (0.3768)	0.5921 (0.3173)	1.2047 (0.3192)
	SCAD	1.9656 (0.3067)	-0.8999 (0.3451)	0.931 (0.3318)	1.4668 (0.3094)	1.9915 (0.2165)	-0.9249 (0.3374)	0.8896 (0.3585)	1.5047 (0.2767)
	AL	1.8466 (0.3085)	-0.7222 (0.3379)	0.8011 (0.3263)	1.3957 (0.3134)	1.8684 (0.254)	-0.7606 (0.3192)	0.7754 (0.3434)	1.4132 (0.3114)
	PEL	1.9226 (0.3435)	-0.9179 (0.3948)	0.9451 (0.3218)	1.439 (0.3671)	1.9626 (0.2445)	-1.0442 (0.2301)	1.0166 (0.2177)	1.4595 (0.2527)
400	ENet	1.4109 (0.2953)	-0.3039 (0.3184)	0.5203 (0.2324)	1.1444 (0.2431)	1.4109 (0.2953)	-0.3039 (0.3184)	0.5203 (0.2324)	1.1444 (0.2431)
	SCAD	1.9978 (0.1198)	-0.9916 (0.1305)	0.9794 (0.1336)	1.5102 (0.1389)	1.9978 (0.1198)	-0.9916 (0.1305)	0.9794 (0.1336)	1.5102 (0.1389)
	AL	1.9536 (0.1305)	-0.9036 (0.1567)	0.9091 (0.1617)	1.4842 (0.1485)	1.9536 (0.1305)	-0.9036 (0.1567)	0.9091 (0.1617)	1.4842 (0.1485)
	PEL	1.9954 (0.1643)	-1.0033 (0.1767)	1.0001 (0.1748)	1.4953 (0.1766)	1.9954 (0.1643)	-1.0033 (0.1767)	1.0001 (0.1748)	1.4953 (0.1766)

表 4 是相对应于表 3 的模型选择的结果, 表中 False 指的是被错误压缩的平均系数个数, True 指的是真实为零的系数 β_2 中压缩到零的平均系数个数。这里相应于样本量的 $p-4 = 6, 17, 24$ 。可以看出当样本量较少且 $\varepsilon_i \sim 0.9N(0,1)+0.1N(0,49)$ 时, SCAD 和 AL 模型很不稳定, 易将真实不为 0 的参数压缩为 0。ENet 虽然被错误压缩的平均系数的个数较少, 但是真实压缩到零的平均系数的个数远小于 $p-4$, 模型选择的效果不佳。 $\varepsilon_i \sim (1+z_{i1})N(0,1)$ 时模型选择的结果相差不大, 但随着 $n \rightarrow \infty$, 两种情形下模型选择的结果均有改善。

表 4: 不同分布下不同惩罚的 PLS 与 PEL 模型选择的结果

n		$\varepsilon_i \sim (1+z_{i1})N(0,1)$		$\varepsilon_i \sim 0.9N(0,1) + 0.1N(0,49)$	
		True	False	True	False
100	ENet	3.15	0.27	1.51	0.23
	SCAD	4.01	0.43	4.07	1.28
	AL	4.56	0.65	5.16	1.62
	PEL	4.33	0.43	4.94	0.46
200	ENet	12.38	0.14	12.78	0.20
	SCAD	14.09	0.09	14.43	0.08
	AL	14.86	0.21	14.82	0.15
	PEL	14.76	0.08	14.05	0.13
400	ENet	21.28	0.16	10.48	0.02
	SCAD	22.01	0.01	19.03	0.20
	AL	22.00	0.01	21.98	0.24
	PEL	22.23	0.09	22.44	0.08

表 5 给出了不同残差分布假定下, PEL 和 PLS 的参数估计值与真实值的根号均方误差 RMSE。在名义置信水平 $1-\alpha = 0.95$ 下, 本文还比较了不同情形下特定参数组合 $\beta_1, (\beta_1, \beta_2)$ 和 $(\beta_1, \beta_2, \beta_3)$ 的置信域的覆盖率, 结果如表 5 所示。从这些仿真结果中, 我们得出以下结论: 在表 5 中, 所有的估计值都接近于参数的真值, 随着 n 增大, RMSE 依次减少, 且 PEL 的效果略高于 PLS。当变量个数随样本量增大时, 正态分布下的残差比 t 分布和指数分布下的情形效果较好。

表 5: 不同分布下 PLS 和 PEL 的参数估计值的 RMSE

$\varepsilon_i \sim$	n	PLS				PEL			
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$N(0,1)$	100	0.4850	0.6208	0.5883	0.5023	0.4801	0.6101	0.5735	0.5014
	200	0.2563	0.2653	0.2706	0.2588	0.2507	0.2618	0.2677	0.2568
	400	0.1713	0.1624	0.1735	0.1598	0.1691	0.1513	0.1749	0.1600
	600	0.1311	0.1236	0.1377	0.1220	0.1285	0.1222	0.1389	0.1187
$t(3)$	100	0.4464	0.5426	0.4814	0.4666	0.4440	0.5125	0.4801	0.4605
	200	0.2734	0.2769	0.2593	0.2643	0.2751	0.2547	0.2614	0.2626
	400	0.2734	0.2769	0.2593	0.2643	0.2751	0.2547	0.2614	0.2626
	600	0.1902	0.2054	0.1844	0.1927	0.1899	0.1916	0.1847	0.1935
$exp(1)$	100	0.4766	0.5522	0.5107	0.4841	0.4726	0.5212	0.5093	0.4832
	200	0.1524	0.1505	0.1494	0.1610	0.1509	0.1510	0.1491	0.1650
	400	0.2316	0.2470	0.2678	0.2803	0.2309	0.2416	0.2611	0.2836
	600	0.2316	0.2470	0.2678	0.2803	0.2309	0.2416	0.2611	0.2836

这里 PLS 估计为 $\hat{\beta}_{GLS} = [Z'(I_m - S_m)Z]^{-1} Z'(I_m - S_m)Y$, 其中 $S_m = U(U'U)^{-1}U'$, 且其对应的置信区域为

$$I_{\alpha}(GLS) = \left\{ \beta : n(\hat{\beta}_{GLS} - \beta)' L_n (L_n' V^{-1} L_n)^{-1} L_n (\hat{\beta}_{GLS} - \beta) \leq \chi_{p,(1-\alpha)}^2 \right\}$$

这里 $V = 1/(n\hat{\sigma}^2)Z'(I_m - S_m)Z$, $\hat{\sigma}^2 = 1/(n-p)e'e$, 其中 e 则为 PLS 估计的样本残差。

从表 6 可以发现, 在名义水平 $\alpha = 0.05$ 和每一个样本量下, PEL 方法的覆盖概率随着样本量的增加而增加, 特别是在样本量合适的情况下, 覆盖概率更接近名义水平。此外, 我们发现基于 PEL 的置信区域始终比基于 PLS 的置信区域具有更好的覆盖概率。总之, 本文强烈推荐我们提出的 PEL 方法。PEL 具有更高的覆盖概率这意味着渐近方差的估计对于置信区间/域的构造有一定的影响, 而 PEL 不需要估计渐近方差, 就能得出较好覆盖率。

表 6: 不同分布下 PEL 方法和 PLS 的置信区间/置信域的覆盖率

$\varepsilon_i \sim$	n	PLS			PEL		
		β_1	(β_1, β_2)	$(\beta_1, \beta_2, \beta_3)$	β_1	(β_1, β_2)	$(\beta_1, \beta_2, \beta_3)$
$N(0,1)$	100	0.936	0.934	0.912	0.944	0.938	0.940
	200	0.944	0.946	0.968	0.942	0.934	0.930
	400	0.952	0.946	0.946	0.952	0.957	0.951
	600	0.938	0.944	0.948	0.965	0.961	0.962
$t(3)$	100	0.932	0.934	0.932	0.947	0.941	0.943
	200	0.942	0.956	0.934	0.951	0.955	0.948
	400	0.949	0.932	0.933	0.952	0.959	0.947
	600	0.948	0.948	0.930	0.958	0.954	0.963
$exp(1)$	100	0.916	0.920	0.912	0.946	0.920	0.912
	200	0.924	0.944	0.926	0.939	0.932	0.930
	400	0.924	0.936	0.928	0.944	0.938	0.941
	600	0.941	0.937	0.929	0.954	0.958	0.950

第 6 章 实证

6.1 上证指数研究的背景与模型

上海证券交易所于 1991 年 7 月 15 日起正式发布了上海证券综合指数, 其样本股包括上海证券交易所的所有上市股票, 也可以说它是反映股票行市变动情况的价格平均数, 是国内最具影响力的股票指数。上证指数的股票覆盖面最广, 基本反映了中国整体经济的发展, 投资上证指数也就相当于投资了中国经济的未来; 其次, 它在风格上更加均衡, 股票市场不断变化, 有时大盘股表现好, 有时小盘股表现好, 它能全面把握大中小盘股票的表现; 上证指数的行业分布也更均衡, 从而使得指数整体波动性更小。国际主流媒体也将它作为最能代表中国股票市场的主要指数, 同时它也是指数型基金追踪之一。投资者可以用它来实时的预

测股票市场的大方向。企业家和政界人物也通常会以此为参考指标，来预测社会经济与政治发展形势。因此用有效的方法预测上证指数，对股票市场和金融行业以及国家经济都有重要指导价值和实际意义。综合指数的走势是非线性的，有很多因素都会影响上证指数的变化，例如居民消费的能力，国家政策，人民币汇率，市场经济以及公司或国际交易信息等都会对上证指数产生一些影响。还有技术分析方法中的指标法，从股票价格波动规律中计算描述波动趋势的指标对未来收盘价的预测也会有一定的影响。

投资者可以参考一些股票价格预测模型进行交易，其中股票价格作为预测的响应变量又受到多方面因素的影响，而找到一个能充分运用信息减少预测误差的合适模型并不简单。由于股票价格具有很强的自相关性，所以很自然的会用到时间序列模型。但是大多时序模型需要事先检验样本数据的平稳性或者差分等变换后的平稳性。但不是所有的时序数据都是满足这个平稳性质的。对于上证指数价格的预测，也有用多元回归和支持向量回归等方法进行预测，也有用神经网络的方法进行非线性拟合预测的应用，但它们的样本数据都只是离散的变量。陈丽琼(2014)^[52]将自回归模型和非参数方法结合，用一个混合数据模型来预测未来股票的开盘价。模型中的自回归部分反映了过去的开盘价信息，而非参数部分是一个前交易日的价格波动拟合曲线与其相应系数函数的积分值，即样本除了日交易价格与其滞后项还有函数形式的观测。最后她用沪深指数进行了实证分析，对比得出了该混合模型比传统的自回归模型预测效果更好。程丽娟(2016)^[57]也用函数型数据主成分分析的方法和部分函数型线性回归模型对上证指数的日收益率进行了预测，并模拟了其他变量，取得了不错的效果。但实际上，股票价格的波动涉及的范围非常广泛，受许多因素的影响。经济大环境的变化，政策因素，投资者的预期变化和心理等等，在这样多元的因素的影响下势必要引入更多变量来避免偏差，同时，有的变量又是交叉影响股价波动。因此本文利用前面几部分提到的部分函数型线性模型和惩罚经验似然的方法，对上证综合指数进行了预测。

6.2 数据的来源与处理

本文主要针对 2018-01-03 至 2019-09-30 (共计 425 个工作日) 期间的上证综合指数每日收盘价数据，基于惩罚经验似然的方法建立了上海证券综合指数每日

收盘价的预测模型。本文以昨日收盘价即因变量滞后一期的数据作为其中一个自变量。股票内在价值和一些宏观经济情况是评估股票投资价值并做出交易决策的重要因素。所以本文选用一些宏观经济政策和经济指标，如居民消费价格指数(CPI)，固定资产投资累计增速，财政收入增速，财政支出增速，货币供给量 M_1 ，货币流通量 M_0 ，存款准备金利率等。同时还考虑到了一些常用的技术指标，如AR(5)（最近5日内最高价与收盘价差值的和/开盘价与最低价差值的和）、AR(10)、BR(5)、BR(10)、BIAS(5)、BIAS(10)、J指标、MACD、PSY(5)、PSY(10)、RSI(6)等。其中技术指标的计算方式如下：

(1) 用开盘价的相对位置表示人气的指标：

$$AR(N) = \frac{\sum_{i=N}^i (\text{最高价} - \text{收盘价})}{\sum_{i=N}^i (\text{开盘价} - \text{最低价})}$$

(2) 用今日相对于昨日收盘价的波动范围来表示意愿的指标：

$$BR(N) = \frac{\sum_{i=N}^i \text{强势和}}{\sum_{i=N}^i \text{弱势和}}$$

其中强势表示近 N 日内最高价与前一天的收盘价的正差值累加，弱势和则表示为近 N 日内前一天的收盘价与最低价的正差值累加和。

(3) 反映一定时期内股票价格和对应的平滑移动平均值偏离程度的指标：

$$BIAS(N) = \text{当日收盘价} - N\text{日平滑移动平均值}$$

(4) 根据股票价格分布的相对位置来预测趋势的指标：

$$J = 3 * D - 2 * K$$

$$RSV(N) = (\text{当日收盘价} - \text{近}N\text{日最低价}) / (\text{近}N\text{日最高价} - \text{最近}N\text{日最低价})$$

其中 K 是 RSV 近 N_1 日的移动平均， D 为 K 的近 N_2 日移动平均。

(5) 利用长、短两条平滑曲线计算的离差值：(若该指标由负变为正，一般可作为“买入”参考，若由正到负，则可作为“卖出”参考)

$$MACD = 2 * (DIFF - DEA)$$

其中 $DIFF$ 为收盘价的近 12 天内平滑移动平均值与 26 天内平滑移动平均值之差，而 DEA 则为 $DIFF$ 值 N 天内的平滑移动平均值

(6) 反映市场心理的超买超卖心理线指标：

$$PSY(N) = \text{近}N\text{日内上涨的天数} / N$$

(7) 反映股票市场在一定时期内的景气程度的相对强弱指数:

$$RSI(N)=N\text{日内收盘价上涨幅度总和}/\text{上涨下跌幅度总和}$$

本文的模型中一共选用以上所提及的 19 个自变量, 再考虑到日内波动因素对收盘价的影响, 还引入函数型预测子即昨日交易的每分钟的成交价格数据 $x_i(t)$, 再减去根据每日交易量得到加权均值作为模型的输入数据。本文的数据主要来源于锐思数据库和财富数据中心。

6.3 模型的结果与分析

本文首先将 425 个交易日数据的前 250 个数据作为训练集, 后面的 175 个交易日数据作为预测集。再根据陈丽琼(2014)^[52]的研究, 以平均绝对相对误差 MARE 和均方相对误差 MRSE 作为模型预测结果的评估指标, 即

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i}$$

$$MSRE = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2$$

本文的每个样本观测代表该交易日的数据, 再以昨日连续的分钟数据曲线即第 i 天的价格变动函数作为函数型预测子。选取的变化时间点是从早上 9 点半到 11 点半以及下午 1 点到 3 点一共 240 分钟的连续成交价格数据。我们首先可以将每天的上证指数收盘价数据用傅里叶基展开拟合成光滑的曲线, 再利用函数型主成分分析从前一日的价格波动曲线中提取出有用的信息, 用于改进预测结果。当然, 这里提取出来的主要信息是一条或者多条特征函数曲线。根据特征值计算的方差解释度, 我们取了前两个主成分, 其方差解释度达到了 94%。再将其代入部分函数型线性模型中, 求解惩罚似然估计值。这里本文设置网格搜索使预测均方相对误差达到最小的调和参数值, 即 $\tau = 0.12$ 。结果显示财政收支, 货币流通量 M_0 , J 指标和 $RSI(6)$ 等指标被自动压缩到 0, 说明这些变量的解释力相对较弱或者它们的影响力在其他变量有体现, 所以相对而言不重要。其中一些主要变量的参数估计值与检验结果如表 7 所示。图 7 还画出了所估计的系数函数曲线拟合图。

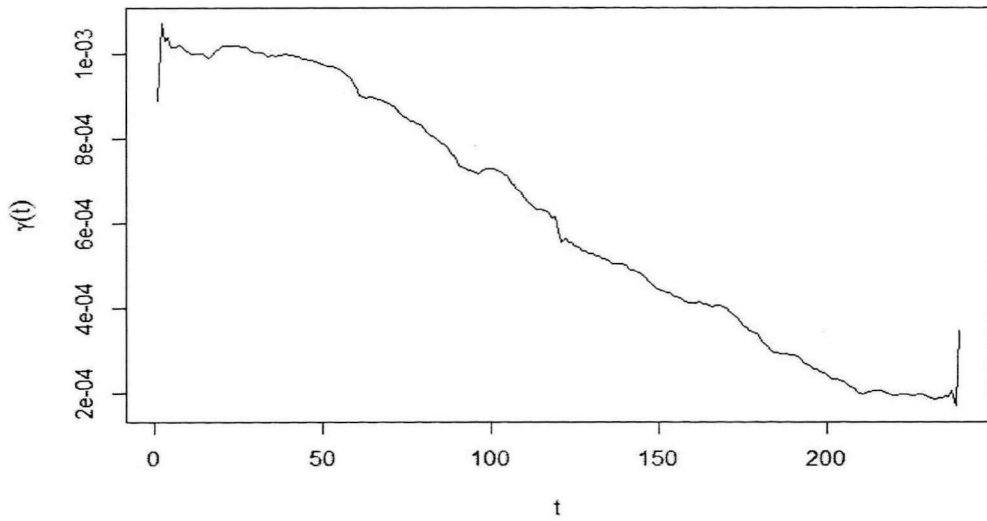


图 7：日内波动曲线的系数函数曲线估计

从图 7 我们可以看到，该系数函数曲线整体是呈现缓慢下降的趋势，前面的数值比较大，越往后数值就越越来越小。所以，我们可以认为随着一天时间的推移，该时刻的交易价格对后一日收盘价的影响力逐渐减小，即在前一日刚开盘时影响力是最大的，且到快要开盘和快要收盘时的几分钟内交易价格的影响力会突然变大。由此可知，如果要分析前一天内价格的变化对后一天收盘价的预测影响，我们可以主要把握上午的交易价格，特别是每天刚开盘和快要收盘的几分钟内的交易价格。下面本文将对此例中部分函数型线性模型的预测效果进行评估，用惩罚经验似然的方法求解混合模型。再对比没有加入函数型预测子的多元回归模型与加入函数型预测子但用最小二乘方法求解的部分函数型线性模型，模型的对比结果如表 7 所示。这里的预测方法是采用更新数据去预测接下来一天的股票价格，即每次只向后预测一天，再得到新的一天更新数据拟合模型接着预测下一天。以此类推，得到第 251 天到第 425 天的收盘价的预测值。

表 7: 不同模型与方法的预测结果的均方相对误差与平均绝对相对误差

模型	MSRE	MARE
LS: $Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + \varepsilon_i$	0.968‰	1.22‰
FDA+LS: $Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + \int_0^T \gamma(t)x_i(t)dt + \varepsilon_i$	0.178‰	0.96‰
FDA+PEL: $Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + \int_0^T \gamma(t)x_i(t)dt + \varepsilon_i$	0.115‰	0.44‰

从表 7 可以看出, FDA+LS 即基于最小二乘的部分函数型线性模型比 LS 线性回归模型的均方相对误差与平均绝对相对误差较小, 即引入前一日价格波动曲线作为函数预测子对预测偏差有一定的纠正作用。而且 FDA+PEL 即基于惩罚经验似然得到的部分函数型线性模型的预测效果最好。图 8 还画出了运用本文的 FDA+PEL 模型得到的预测结果和真实结果的比较图, 更直观的展示出了该模型的预测效果。其中黑色实线表示的是第 251 天到第 425 天的真实收盘价走势, 而红色虚线是 FDA+PEL 模型对第 251 天到第 425 天的收盘价的预测值。从图中可以看出红色虚线与黑色实线的相似度较高, 这说明了本文模型的预测结果比较接近真实值。而且图 8 的预测曲线与真实曲线的升降趋势大都一致, 对于投资者“买入”和“卖出”决策也具有很好的参考价值。

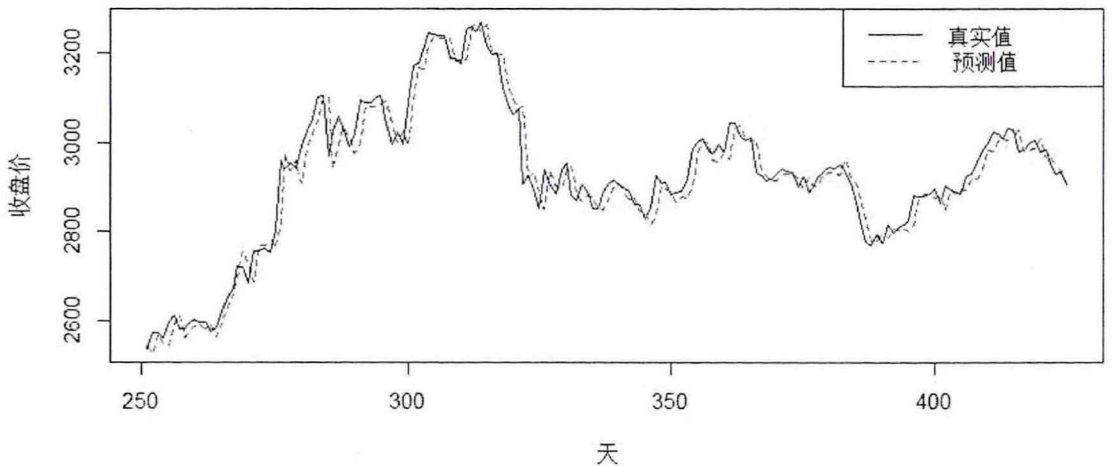


图 8: FDA+PEL 的预测结果与真实值比较

同时，根据模型参数估计的结果，我们还能作变量的显著性检验，即构造假设检验 $H_0: \mathbf{L}_n \boldsymbol{\beta}_0 = \mathbf{0}$ vs $\mathbf{L}_n \boldsymbol{\beta}_0 \neq \mathbf{0}$ 。这里 $\mathbf{L}_n = (0, 0, \dots, 1, 0, \dots)'$ ，其中 1 所处的位置即对应需要检验的变量。再根据 3.3 节中的内容构造一个惩罚经验似然比统计量 $\tilde{Q}(\mathbf{L}_n)$ 。据前面的仿真模拟结果，惩罚经验似然比统计量遵循 $\tilde{Q}(\mathbf{L}_n) \xrightarrow{L} \chi_q^2$ 。因此，我们可以据此得到判定准则，即拒绝域为

$$\omega_\alpha = \left\{ v: -2 \left\{ \tilde{Q}(\hat{\boldsymbol{\beta}}_{pel}) - \min_{\boldsymbol{\beta}, \mathbf{L}_n \boldsymbol{\beta} = v} \tilde{Q}(\boldsymbol{\beta}) \right\} > \chi_{q, 1-\alpha}^2 \right\},$$

这里 $\chi_{1, 1-\alpha}^2 = 3.84146$ 是 χ_q^2 分布的 95% 的下侧分位数。根据判定准则，我们将模型的估计和假设检验的结果展示到表 8 中。

从表 8 可以看出，昨日收盘价显而易见是高度显著变量，收盘价的预测和其滞后一期的值有强相关性。还有固定资产投资的累计增速的增加也会引起上证指数的同向变化，同样居民消费价格指数(CPI)和货币供给量都会显著影响上证指数的变化。技术指标 AR(10)和 BR(5)对上证指数的解释力较强，即市场买卖人气和意愿程度的高低对追踪上证股价指数未来动向作用显著。

表 8: 模型主要参数的解释力

变量	参数估计	$\tilde{Q}(\mathbf{L}_n)$	$\mathbf{L}_n \boldsymbol{\beta} \in \omega_\alpha$
CPI	0.841568	670.2357	TRUE
固定资产增速	3.046004	9.795258	TRUE
货币供给	-0.043496	638.0386	TRUE
昨收盘价	0.888245	NaN	TRUE
AR(5)	23.25333	-19.48233	FALSE
AR(10)	-61.97191	165.8325	TRUE
BR(5)	2.120828	22.5398	TRUE
BR(10)	9.195921	-35.2080	FALSE
RSV	2.024114	-28.7501	FALSE
PSY(5)	-8.819956	-16.38944	FALSE
PSY(10)	-1.226211	-14.44534	FALSE

第7章 总结与扩展

本文在前人对惩罚经验似然和对部分函数型线性模型已有研究成果的基础上进行综合的研究分析。在惩罚最小二乘的基础上把惩罚经验似然运用到部分函数型线性模型中。本文主要结合了惩罚似然的思想和经验似然思想考察部分函数型线性模型的参数估计及其假设检验问题,针对模型结合函数型主成分分析和惩罚高维经验似然方法建立了部分函数型线性模型统计推断的框架,其中惩罚函数选用 SCAD 惩罚函数。惩罚经验似然优化问题的求解采用嵌套坐标下降法收敛到较好的估计值,对于 SCAD 惩罚函数二阶不可导的问题采用局部二次逼近法得到近似解。其中调和参数的选择问题由于涉及到两个参数,且分开选择会导致算法复杂度增加。因此,本文用前人提到且验证过的 ABIC 原则选择出了最优的参数。仿真模拟的结果显示,在参数维度发散的情形下,部分函数型线性模型的参数估计具有 Oracle 性质。同时,系数函数的估计效果也较好,处于 10%和 90%分位数函数之间。在高维参数的假设检验问题中,惩罚经验似然方法比不加惩罚函数的经验似然方法更加有效,因其稀疏性使计算似然函数时的模型的复杂度降低,而且对于构造的置信域也更加稳定。本文以第一个非零参数为例构造了参数的假设检验问题和它的置信区间,模拟研究表明了在原假设成立的条件下,其犯第一类错误的概率接近显著性水平 α ,而且在原假设不成立的情况下,其功效也随样本量的增大以及检验值与真值差距的增大而越来越大。同时,本文在不同的残差分布下,对比了不同惩罚函数下的惩罚最小二乘估计与置信域覆盖率的结果,体现出了惩罚经验似然法的优势。

惩罚经验似然方法能够有效的节约成本,实用性强,从而具有较高的推广价值。本文针对部分函数型线性模型运用惩罚经验似然的方法是在已有方法的基础上取长补短,并且本文也通过模拟研究的结果表明了这种方法在处理高维数据时是一种不错的选择。但是本文考虑到的问题也并非全面,因此,还有可以进一步研究的问题。例如,与经验似然具有相同渐近性质的经验欧氏似然,它是从欧氏距离出发,使其在计算上会相较简单,就是说如果采用经验欧氏似然来代替经验似然来进行统计推断或许有利于提高计算密度,从而能够降低计算的复杂度。还有,本文只是涉及到了一个函数型变量的情形,但若是有多多个函数型的预测子,

对其进行函数型主成分分析必然也会得到很多主成分。因此变量选择问题可能还要涉及到成组压缩。最后便是现今的缺失数据、删失数据、面板数据或变量带有测量误差等特殊情形下的部分函数型线性模型又该如何进行高维惩罚经验似然的统计推断也有待研究。此处限于篇幅和精力便不再进一步讨论。但是，本文所讨论的思想与方法还是可以用来处理经济领域的收益问题和医学仪器测量数据的分析以及一些天气，气温等问题。股票市场是我国资本市场的重要组成部分，它能推动我国经济发展进程。为了更好地了解股票的内在价值从而获得更高的收益，股市的预测就成了很多投资者和学术研究者研究的热点问题。而上证指数在一定程度上反映了我国的经济实力，它也被称为宏观经济的晴雨表，同时又是分析微观经济的一个重要指标，所以本文最后将模型与方法应用到上海证券综合指数上，并得出了很好的预测效果，还能方便运用参数显著性检验得到重要变量的解释度。所以，本文研究高维数据的部分函数型线性模型具有非常广泛现实意义和使用价值。

附录

引理 1: 根据假设 A1-A9, 我们有

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \xrightarrow{P} 0$$

$$\frac{1}{n} \sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})^T \xrightarrow{P} \Sigma(\boldsymbol{\beta}).$$

证明: 根据假设 A1-A9 以及 H, Xue, & Feng (2018)^[1]中对引理 A.2.的证明可得

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{Z}_i [Y_i - \mathbf{Z}_i^T \boldsymbol{\beta} - \mathbf{u}_i^T \boldsymbol{\gamma}(\boldsymbol{\beta})] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \left(\mathbf{Z}_i - \sum_{j=1}^m \frac{\langle \hat{K}_{\mathbf{Z}\mathbf{X}}, \hat{\phi}_j \rangle \langle x_i(t), \hat{\phi}_j \rangle}{\hat{\lambda}_j} \right) \langle \boldsymbol{\gamma}, x_i(t) \rangle \right. \\ &\quad + \left(\sum_{j=1}^{\infty} \frac{\langle K_{\mathbf{Z}\mathbf{X}}, \phi_j \rangle \langle x_i(t), \phi_j \rangle}{\lambda_j} - \sum_{j=1}^m \frac{\langle \hat{K}_{\mathbf{Z}\mathbf{X}}, \hat{\phi}_j \rangle \langle x_i(t), \hat{\phi}_j \rangle}{\hat{\lambda}_j} \right) \boldsymbol{\varepsilon}_i \\ &\quad \left. + \left(\mathbf{Z}_i - \sum_{j=1}^{\infty} \frac{\langle K_{\mathbf{Z}\mathbf{X}}, \phi_j \rangle \langle x_i(t), \phi_j \rangle}{\lambda_j} \right) \boldsymbol{\varepsilon}_i \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{R_{i1} + R_{i2} + R_{i3}\} \end{aligned}$$

其中 $\hat{K}_{\mathbf{Z}\mathbf{X}} = n^{-1} \sum_{i=1}^n \mathbf{Z}_i x_i(t)$, $K_{\mathbf{Z}\mathbf{X}} = \text{cov}(\mathbf{Z}\mathbf{X}(t))$ 。

根据 Shin (2009)^[15]中定理 1.1 的证明, 我们有

$$n^{-1/2} \sum_{i=1}^n R_{i1} \xrightarrow{P} 0$$

根据 Shin (2009)^[15]中引理 A.2.的证明, 我们有

$$n^{-1/2} \sum_{i=1}^n R_{i2} \xrightarrow{P} 0$$

又

$$\begin{aligned} E \left(n^{-1/2} \sum_{i=1}^n R_{i3} \right) &= n^{-1/2} \sum_{i=1}^n E \left(\left(\mathbf{Z}_i - \sum_{j=1}^{\infty} \frac{\langle K_{\mathbf{Z}\mathbf{X}}, \phi_j \rangle \langle x_i(t), \phi_j \rangle}{\lambda_j} \right) \boldsymbol{\varepsilon}_i \right) \\ &= n^{-1/2} \sum_{i=1}^n E(\boldsymbol{\eta}_i(\boldsymbol{\beta})) = 0 \end{aligned}$$

$$\text{var}\left(n^{-1/2}\sum_{i=1}^n R_{i3}\right) = n^{-1}\sum_{i=1}^n E\left(\boldsymbol{\eta}_i(\boldsymbol{\beta})\boldsymbol{\eta}_i(\boldsymbol{\beta})^T\right) = \Sigma(\boldsymbol{\beta})$$

所以 $n^{-1/2}\sum_{i=1}^n R_{i3} \xrightarrow{P} 0$ ，因而有 $n^{-1/2}\sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) = o_p(1)$ 。

$$\begin{aligned} n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})^T &= n^{-1}\sum_{i=1}^n (R_{i1}+R_{i2}+R_{i3})(R_{i1}+R_{i2}+R_{i3})^T \\ &= n^{-1}\sum_{i=1}^n \left(\sum_{k=1}^3 R_{ik}R_{ik}^T + \sum_{k<l}^3 R_{ik}R_{il}^T \right) \end{aligned}$$

因为 $R_{i3} = \boldsymbol{\eta}_i(\boldsymbol{\beta})$ 是独立的随机变量，根据大数定律我们有 $n^{-1}\sum_{i=1}^n R_{i3}R_{i3}^T \xrightarrow{P} \Sigma(\boldsymbol{\beta})$ ，再根据 H, Xue, & Feng (2018)^[1] 中对引理 A.2. 的证明可知其余项均为 $o_p(1)$ ，引理 1 得证。

引理 2: 让 $a_n = (p/n)^{1/2-\delta}$ ， $D_n = \{\boldsymbol{\beta}: \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq ca_n\}$ ，其中 $p^{1-\delta}/n^{1/2-\delta} \rightarrow 0$ ， $\delta, c > 0$ 为常数，那么对于 $\boldsymbol{\beta} \in D_n$ ， $\|\boldsymbol{\lambda}_\beta\| = O_p(a_n)$ ，且随着 $n \rightarrow \infty$ ， $\tilde{Q}(\boldsymbol{\beta})$ 在 D_n 中有最小值依概率 1 成立。

证明: 让 $\boldsymbol{\lambda}_\beta = \rho\boldsymbol{\theta}$ ， $\rho = \|\boldsymbol{\lambda}_\beta\|$ ， $\boldsymbol{\theta}$ 为单位向量，那么

$$\begin{aligned} 0 &= \left\| \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\lambda}_\beta^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} \right\| = \left\| \frac{1}{n} \sum_{i=1}^n \frac{\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})}{1 + \rho\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} \right\| \\ &\geq \left| \frac{1}{n} \sum_{i=1}^n \frac{\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})}{1 + \rho\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} \right| = \frac{1}{n} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) - \sum_{i=1}^n \frac{\rho\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})^T \boldsymbol{\theta}}{1 + \rho\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} \right| \\ &\geq \frac{\rho}{n} \sum_{i=1}^n \frac{\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})^T \boldsymbol{\theta}}{1 + \rho\boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} - \frac{1}{n} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right| \\ &\Rightarrow \\ &\frac{1}{n} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right| \geq \rho\boldsymbol{\theta}^T \mathbf{S}_n(\boldsymbol{\beta})\boldsymbol{\theta} \{1 + \rho \max_{1 \leq i \leq n} \|\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\|\}^{-1} \\ &\quad \geq \rho \left\{ \boldsymbol{\theta}^T \mathbf{S}_n(\boldsymbol{\beta})\boldsymbol{\theta} - \max_{1 \leq i \leq n} \|\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\| \frac{1}{n} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right| \right\} \end{aligned}$$

由引理 1 可得

$$n^{-1} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right| \leq n^{-1} \left\| \sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right\| = O_p(\sqrt{p/n})$$

$$\boldsymbol{\theta}^T \mathbf{S}_n(\boldsymbol{\beta})\boldsymbol{\theta} = O_p(1)$$

再根据 A8,A9 以及 Tang & Leng (2010)^[4]中引理 3 的证明, 我们有

$$\begin{aligned} \max_{1 \leq i \leq n} \|\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\| &= O_p(\sqrt{p}) + o_p\left(p^{-\left(\frac{2k-1}{4k}\right)} n^{\frac{1}{4k}}\right) \\ \Rightarrow \max_{1 \leq i \leq n} \|\tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})\| \frac{1}{n} \left| \sum_{i=1}^n \boldsymbol{\theta}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) \right| &= O_p(p/\sqrt{n}) + o_p\left((p/\sqrt{n})^{\frac{1}{4k}}\right) = o_p(1) \end{aligned}$$

其中 $k \geq 1$ 所以 $\rho = \|\boldsymbol{\lambda}_\beta\| = O_p(\sqrt{p/n})$ 。

其余证明通过引理 1 和 Tang & Leng (2010)^[4]中引理 2 和引理 4 的证明可知成立。

定理 1 的证明: 由引理 1 到 2 可知, 若 $\boldsymbol{\beta} \in D_n$,

$$\begin{aligned} \frac{1}{n} \frac{\partial \tilde{Q}(\boldsymbol{\beta})}{\partial \beta_j} &= \frac{1}{n} \sum_{i=1}^n \frac{\boldsymbol{\lambda}^T \partial \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}) / \beta_j}{1 + \boldsymbol{\lambda}^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})} + P'_\tau(|\beta_j|) \text{sign}(\beta_j) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\boldsymbol{\lambda}^T \partial \boldsymbol{\eta}_i(\boldsymbol{\beta}) / \beta_j}{1 + \boldsymbol{\lambda}^T \boldsymbol{\eta}_i(\boldsymbol{\beta})} + o_p(1) + P'_\tau(|\beta_j|) \text{sign}(\beta_j) \\ &= I_j + II_j + o_p(1) \end{aligned}$$

那么当 n 充分大时,

$$\begin{aligned} \max_{j \notin \wp} |I_j| &\leq \max_{j \notin \wp} \left| \frac{1}{n} \sum_{i=1}^n \boldsymbol{\lambda}^T \partial \boldsymbol{\eta}_i(\boldsymbol{\beta}) / \beta_j \right| \\ &\leq \|\boldsymbol{\lambda}_\beta\| \max_{j \notin \wp} \left| \frac{1}{n} \sum_{i=1}^n \partial \boldsymbol{\eta}_i(\boldsymbol{\beta}) / \beta_j \right| \\ &\leq \|\boldsymbol{\lambda}_\beta\| \left\| \frac{1}{n} \sum_{i=1}^n \partial \boldsymbol{\eta}_i(\boldsymbol{\beta}) / \beta_j \right\| = o_p(1) \end{aligned}$$

因为在 Shin (2009)^[15]定理 1.1 的证明过程中可知

$$n^{-1} \sum_{i=1}^n \partial \boldsymbol{\eta}_i(\boldsymbol{\beta}) / \beta_j = \sum_{j=1}^{\infty} \langle \hat{K}_{\mathbf{Z}\mathbf{X}}, \hat{\phi}_j \rangle \langle x_i(t), \hat{\phi}_j \rangle / \hat{\lambda}_j - n^{-1} \mathbf{Z}^T \mathbf{Z} = O_p(1),$$

且在 A11-A12 下, 当 $j \notin \wp$, $II_j = P'_\tau(|\beta_j|) \text{sign}(\beta_j) = \tau \text{sign}(\beta_j)$ 。也就是说 $n^{-1} \partial \tilde{Q}(\boldsymbol{\beta}) / \beta_j$ 的符号完全取决于 β_j 的符号。因而, 当 $j \notin \wp$, 与 Leng & Tang (2012)^[5]类似, 我们有 $\hat{\boldsymbol{\beta}}_2 = \mathbf{0}$ 依概率 1 成立。

定理 1 (b) 的证明可在 (a) 的结论上即在添加 $\mathbf{H}_2 \boldsymbol{\beta} = \mathbf{0}$ 的条件下最小化原来的目标函数, 等价于

$$\min \quad \tilde{Q}(\boldsymbol{\beta}) = \sum_{i=1}^n \log[1 + \lambda^T \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta})] + n \sum_{j=1}^p p_{\tau}(|\beta_j|) + \boldsymbol{\mu}^T \mathbf{H}_2 \boldsymbol{\beta}$$

$\boldsymbol{\mu} \in \mathbb{R}^{p-d}$ 是拉格朗日乘子。其余的证明类似于 Leng & Tang (2012)^[5]对定理 1 的证明。最后我们得到

$$\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{10} = \mathbf{H}_1 (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{H}_1 \left(\mathbf{V} - \mathbf{V} \mathbf{H}_2^T (\mathbf{H}_2 \hat{\mathbf{V}} \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{V} \right) \{ \hat{\boldsymbol{\Omega}}^T \hat{\boldsymbol{\Sigma}}^{-1} \bar{\boldsymbol{\eta}}(\boldsymbol{\beta}_0) + o_p(1) \}$$

其中 $\hat{\boldsymbol{\Omega}} = n^{-1} \sum_{i=1}^n \partial \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta}$, $\boldsymbol{\Omega} = E(\partial \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta})$, $\hat{\boldsymbol{\Sigma}} = n^{-1} \sum_{i=1}^n \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}_0) \tilde{\boldsymbol{\eta}}_i(\boldsymbol{\beta}_0)^T$ 。

那么

$$\boldsymbol{\Omega}_p = \mathbf{H}_1 \left(\mathbf{V} - \mathbf{V} \mathbf{H}_2^T (\mathbf{H}_2 \mathbf{V} \mathbf{H}_2^T)^{-1} \mathbf{H}_2 \mathbf{V} \right) \mathbf{H}_1^T$$

其中 $\mathbf{V} = \left[E(\partial \tilde{\boldsymbol{\eta}}(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta})^T \boldsymbol{\Sigma}(\boldsymbol{\beta}_0)^{-1} E(\partial \tilde{\boldsymbol{\eta}}(\boldsymbol{\beta}_0) / \partial \boldsymbol{\beta}) \right]^{-1}$ 。

定理 2 的证明: 由于引理 1 和引理 2 以及定理 1 成立, 本文的估计方程满足 Leng & Tang (2012)^[5]中的一般估计方程, 有理由认为定理 2 成立, 证明过程类似。

参考文献

- [1] Hu, Y. , Xue, L., & Feng, S. Empirical likelihood inference for partial functional linear model with missing responses [J]. *Communications in Statistics - Theory and Methods*, 2018, 47:19, 4673-4691.
- [2] Kong, D., Xue, K., Yao, F. , & Zhang, H. H. . Partially functional linear regression in high dimensions [J]. *Biometrika*, 2016, 103(1), asv062.
- [3] Fan, J. and Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *J. Amer. Statist. Assoc.*, 2001, 96, 1348-1360
- [4] Tang, C. Y. , & Leng, C. . Penalized high-dimensional empirical likelihood [J]. *Biometrika*, 2010, 97(4), 905-920.
- [5] Leng, C., & Tang, C.Y.. Penalized empirical likelihood and growing dimensional general estimating equations [J]. *Biometrika*, 2012, 99(3), 703-716.
- [6] Dalzell J O R J . Some Tools for Functional Data Analysis [J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991, 53(3):539-572.]
- [7] Ramsay, J. O., and B. W. Silverman. *Functional data analysis* [M]. New York: Springer, 1997.
- [8] Dauxois, J, Pousse, A, Romain, Y. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference [J]. *Journal of Multivariate Analysis*, 1982, 12(1):136-154.
- [9] Cardot, H., F. Ferraty, and P. Sarda. Spline estimators for the functional linear model [J]. *Statistica Sinica* , 2003, 13:571–92.
- [10] Hall, P., and M. Hosseini-Nasab. On properties of functional principal components analysis [J]. *Journal of the Royal Statistical Society*, 2006, 68:109–26.
- [11] Cai, T., and P. Hall. Prediction in functional linear regression [J]. *The Annals of Statistics*, 2007, 34:2159–79.
- [12] Hall, P., and J. L. Horowitz. Methodology and convergence rates for functional linear regression [J]. *Journal of the American Statistical Association* , 2007, 35:70–91.
- [13] Aneiros-Pérez, G., and P. Vieu. Semi-functional partial linear regression [J]. *Statistics and Probability Letters*, 2006, 76:1102–10.

- [14] Zhang, D., Lin, X., Sowers, M.F. Assessing the effects of reproductive hormone profiles on bone mineral density using functional two-stage mixed models [J]. *Biometrics* , 2007, 63, 351–362.
- [15] Shin, H.. Partial functional linear regression [J]. *Journal of Statistical Planning and Inference*, 2009, 139:3405–18.
- [16] Shin, H., and M. H. Lee. On prediction rate in partial functional linear regression [J]. *Journal of Multivariate Analysis* ,2012, 103:93–106.
- [17] Lu, Y., J. Du, and Z. Sun. Functional partially linear quantile regression model [J]. *Metrika* ,2014, 77:317–32.
- [18] Owen, A. B. Empirical likelihood ratio confidence intervals for a single functional [J]. *Biometrika*, 1988, 74:237–49.
- [19] Owen, A. B. Empirical likelihood confidence regions [J]. *The Annals of Statistics* ,1990, 18:90–120.
- [20] Diccio, T. , Hall, P. , & Romano, J. . Empirical likelihood is Bartlett-correctable [J]. *The Annals of Statistics*, 1991, 19(2), 1053-1061.
- [21] Owen, A. B. *Empirical likelihood*[M]. New York: Chapman and Hall, 2001.
- [22] Owen, A. B. Empirical likelihood for linear models [J]. *The Annals of Statistics* ,1991, 19:1725–47.
- [23] Yang, Y. P., G. R. Li, and H. Peng. Empirical likelihood of varying coefficient errors-in-variables models with longitudinal data [J]. *Journal of Multivariate Analysis* ,2014, 127:1–18.
- [24] Zhu, L. X., and L. G. Xue. Empirical likelihood confidence regions in a partially linear single-index model [J]. *Journal of the Royal Statistical Society*, 2006, 68:549–70.
- [25] Lian, H. Empirical likelihood confidence intervals for nonparametric functional data analysis [J]. *Journal of Statistical Planning and Inference*, 2012, 142:1667–77.
- [26] Tibshirani, R. J. Regression Shrinkage and Selection via the LASSO [J]. *Journal of the Royal Statistical Society*, 1996, 58,267–288.
- [27] Fu, W. J.. Penalized Regression: The Bridge Versus the LASSO [J]. *Journal of Computational and Graphical Statistics*, 1998, 7, 397–416.

- [28] Peng, H., & Fan, J. . Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, 2004, 32(3), 928-961.
- [29] Guoliang, F. , Hanying, L., & Yu, S. Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors [J]. *Journal of Multivariate Analysis*, 2016, 147, 183-201.
- [30] Fang, J. , Liu, W. , & Lu, X. Penalized empirical likelihood for semiparametric models with a diverging number of parameters [J]. *Journal of Statistical Planning & Inference*, 2017, 186, 42-57.
- [31] Ando, T. , & Sueishi, N.. On the convergence rate of the scad-penalized empirical likelihood estimator [J]. *Social Science Electronic Publishing* ,2017.
- [32] Ando, T. , & Sueishi, N. . Regularization parameter selection for penalized empirical likelihood estimator [J]. *Social Science Electronic Publishing* ,2018.
- [33] He, B. Q. , Hong, X. J. , & Fan, G. L. . Penalized empirical likelihood for partially linear errors-in-variables panel data models with fixed effects [J]. *Statistical Papers*, 2018, 1-31.
- [34] Wang, H., Li, B. & Leng, C. Shrinkage tuning parameter selection with a diverging number of parameters [J]. *J. R. Statist. Soc. B*, 2009, 71, 671-83.
- [35] Qin, J., and J. Lawless. Empirical likelihood and general estimating equations [J]. *The Annals of Statistics* ,1994, 22:300–25.
- [36] Johnson, D. . *Statistical Signal Processing* [J]. *Statistical signal processing*. M. Dekker ,2011.
- [37] Chen, S. X. , Peng, L. , & Qin, Y. L. . Effects of data dimension on empirical likelihood [J]. *Biometrika* ,2009, 96(3), 711-722.
- [38] Levitin D J, Nuzzo R L, Vines B W, et al. Introduction to functional data analysis [J]. *Canadian Psychology* ,2007, 48(3):135-155.
- [39] Lange K. Coordinate Descent Algorithms for Lasso Penalized Regression [J]. *Annals of Applied Statistics* ,2008, 2(1):224-244.
- [40] Cheng Y T, Tong T W. Nested coordinate descent algorithms for empirical likelihood [J]. *Journal of Statistical Computation & Simulation*, 2014, 84(9):1917-1930.
- [41] Ruszczyński A. *Nonlinear optimization* [J]. Princeton: Princeton University Press, 2006.
- [42] Tseng P. Convergence of block coordinate descent method for nondifferentiable maximization [J]. *J. Optim. Theory Appl* ,2001, 109:473–492.

- [43] Fu WJ. Penalized regressions: the bridge versus the lasso [J]. *J. Comput. Graph. Stat*, 1998, 7:397–416.
- [44] Friedman J, Hastie T, Hoefling H, Tibshirani R. Pathwise coordinate optimization [J]. *Ann. Appl. Stat*, 2007, 2:302–332.
- [45] Jiang Z, Huang Z, Fan G. Empirical likelihood for high-dimensional partially functional linear model [J]. *Random Matrices: Theory and Applications*, 2019.
- [46] Ramsay J O, Silverman B W. Functional linear models [M]. *Functional Data Analysis*. Springer New York, 2005.
- [47] Berrendero J R, Justel A, Svarc M. Principal components for multivariate functional data [J]. *Computational Statistics and Data Analysis*, 2011, 55(9):2619-2634.
- [48] Zhang H, Lu W. Adaptive Lasso for Cox's proportional hazards model [J]. *Biometrika*, 2007, 94(3):691-703.
- [49] 谭晓燕. 纵向数据中两种模型下的经验似然与惩罚经验似然 [D]. 陕西师范大学, 2018.
- [50] 侯文, 宋立新, & 黄玉洁. 删失数据下非线性半参数回归模型中参数的经验似然推断 [J]. *生物数学学报*, 2012, (2), 371-383.
- [51] 韩慧. 关于部分线性模型的惩罚高维经验似然 [D]. 辽宁师范大, 2011.
- [52] 陈丽琼. 函数型数据分析方法在股票价格预测上的应用 [D]. 厦门大学, 2014.
- [53] 吕红, 费文龙, 秦伟良. 基于奇异谱分析的上证指数预测模型 [C]. 江苏省现场统计研究会第八次学术年会, 2003.
- [54] 邱剑, 艾立翔. 基于多元线性回归和层次分析法的案例推理上证指数预测模型的开发及应用 [J]. *商场现代化*, 2011, (34):54-57.
- [55] 金宇, 阎璐. 引入投资者情绪的上证指数预测模型实证研究 [J]. *当代经济*, 2015, (32):148-151.
- [56] 张翔宇, 王富森, 杨凌杰. 基于支持向量机的上证指数预测研究 [J]. *商业经济*, 2011, (3):104-106.
- [57] 程丽娟. 基于部分函数型线性回归模型的上证指数预测 [J]. *泰山学院学报*, 2016, 38(3).

- [58] 胡志军, 沈根祥. 中国股市资产的价格跳跃行为--基于上证综指、深证成指的高频数据研究 [J]. 上海经济研究, 2013, (04):90-101.
- [59] 张金仙, 闫二乐, 杨拴强. 基于自适应 BP 神经网络的上证指数预测模型的研究 [J]. 长春大学学报, 2016, (3):26-30.
- [60] 刘琦. 广义线性模型的惩罚高维经验似然 [D]. 辽宁师范大学, 2012.
- [61] 李吉妮. 单指标模型的高维惩罚经验似然 [D]. 广西师范大学, 2014.

致谢

随着毕业论文的完成，我三年的研究生学习生活也终于要画上了句号。首先，感谢我的导师。虽然老师有一年时间一直在国外，但我一直记得第一次见郭老师时，老师给我提的论文要求，正是郭老师严谨的科学态度与精益求精的作风，让我一直严格要求自己，我的研究论文才能顺利进行。从刚开始选题的迷茫到开题答辩，再到论文写作的整个过程，郭老师都给了我正确的指引。再次向我的导师郭鹏辉老师表示深深的敬意与感谢！还要感谢我的学姐，以她的经历不厌其烦地回答了我许多问题，不论是论文还是找工作都让我了解到很多信息，从而有了前进的方向。也要感谢所有研究生课程的任课老师，不仅教授了专业知识，更是拓宽了我的知识面，给我的论文选题也有很大帮助。最后还要感谢 2017 级所有统计系的研究生同学，他们在日常的学习生活中给我树立了榜样，同时也给予了我很多的帮助和关心，谢谢！