

学校代码: 10255

学 号: 2202178

中图分类法: O212.1



专业学位硕士学位论文

高维数据下的单指标期望分位数 回归模型研究

学位申请人: 邓语菲

指导教师: 姜荣副教授

专业学位类别: 应用统计专硕

研究方向: 大数据分析

所在学院: 理学院

提交日期: 2022年5月

University Code: 10255

Student ID: 2202178

CLC Index: O212.1



**PROFESSIONAL MASTER
DISSERTATION**

**SINGLE-INDEX EXPECTILE MODEL FOR
HIGH-DIMENSIONAL DATA**

Author:	Yufei Deng
Supervisor:	Prof. Rong Jiang
Professional Major Type:	Statistics
Research Field:	Data Analysis
College:	College of Donghua
Date of Submit:	May, 2022

东华大学 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。所呈交的学位论文，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结果由本人承担。

学位论文作者签名：邓语菲

日期： 2022 年 5 月 20 日

东华大学 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。本人授权东华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 。

导师签名：

姜宇

学位论文作者签名：

邓语菲

日期： 2022 年 5 月 20 日

硕士学位论文答辩委员会名单

姓名	职称	单位	备注
郑明	教授	复旦大学	主席
李瑜苗	未评职称（统计人员）	妙盈科技	委员
闫理坦	教授	东华大学	委员
张振中	教授	东华大学	委员
刘欣	副教授	东华大学	委员
徐琪	教授	东华大学	委员
田琳琳	讲师	东华大学	秘书

答辩人：邓语菲

答辩地点：线上答辩

答辩日期：2022.5.17

高维数据下的单指标期望分位数回归模型研究

摘要

伴随着计算机科学技术的日新月异,各种数据收集的工具应运而生,数据采集工作也变得越来越容易,这就导致了在生物医学领域、工程领域、经济金融领域等各个领域,高维数据屡见不鲜。高维数据的出现以及其带来的“维数诅咒”问题,让传统的统计和数据分析方法在分析高维数据时显得力不从心,亟需在之前的基础之上寻找或改进为新的数据统计与分析方法来处理高维数据。因此研究高维数据成为当下统计学的热点问题。当我们开始分析高维数据的时候,会遇到了“维数诅咒”问题,必须考虑如何克服“维数诅咒”,提高统计推断的精度,换句话说,就是我们在利用模型对高维数据进行拟合的时候,如何降低所使用的数据的维度。

单指标模型能通过降维的手段来避免高维数据的“维数诅咒”问题,是非常重要的非参数模型。为了解决高维数据下的单指标模型参数估计问题,本文选择结合期望分位数回归估计参数并进行变量选择,即研究一种单指标期望分位数回归模型。期望回归具有很多优势。普通最小二乘回归对正残差和负残差赋予同等权重,并将条件均值函数估计为回归变量或预测变量的线性函数。Newey 和 Powell (1987)为经典线性回归模型提出了一类估计量,其允许对正残差和负残差使用不同的权重,并称为期望分位数回归。指出这个估计量具有同质性、单调性和平移不变性等优良特性。我们选择非对称最小二乘法(ALS)和 Adaptive LASSO 来进行参数估计和变量选择。本文介绍了一种新的惩罚 ALS(PALS)方法,用于对单指标期望分位数回归模型中的指标系数进行有效估计,同时进行变量选择。并纠偏方法与 PALS 结合,可证明所提出的估计具有渐近正态性,可用来构造有效置信区间和作假设检验。

关键词: 纠偏方法; 期望分位数回归; 单指标模型; 变量选择

SINGLE-INDEX EXPECTILE MODEL FOR HIGH-DIMENSIONAL DATA

ABSTRACT

With the rapid development of computer science and technology, various data collection tools came into being, which leads to the frequent occurrence of high-dimensional data. The emergence of high-dimensional data and the "Curse of dimensionality" brought by it make the traditional statistics and data analysis methods unable to analyze high-dimensional data. Therefore, the study of high-dimensional data has become a hot issue in statistics. We must consider how to overcome the "dimension curse" and improve the accuracy of statistical inference. In other words, when we use the model to fit high-dimensional data, how to reduce the dimension of the data used.

The single-index model can avoid the "curse of dimensionality" problem of high-dimensional data by means of dimensionality reduction, and is a very important nonparametric model. In order to solve the problem of single-index model parameter estimation under high-dimensional data, this thesis chooses to combine expectile model to estimate parameters and perform variable selection, that is, to study a single-index expectile model. Expectile model has many advantages. Newey and Powell (1987) proposed a class of estimators for classical linear regression models that allow for the use of different weights for positive and negative residuals and called expected quantile regression. We choose asymmetric least squares (ALS) and Adaptive LASSO for parameter estimation and variable selection. This thesis introduces a new penalized ALS (PALS) method for efficient estimation of indicator coefficients in single-index expectile model with simultaneous variable selection. Combining the bias correction method with PALS, it can be proved that the proposed estimate has asymptotic normality, which can be used to construct effective confidence intervals and test hypotheses.

Deng Yufei (Statistics)
Supervised by Jiang Rong

KEY WORDS: Debiasing Technique, Expectile Regression, Single-index Model, Variable Selection

目录

摘要.....	I
ABSTRACT.....	II
第一章 绪论	1
1.1 研究背景和意义.....	1
1.1.1 高维数据下的单指标期望分位数回归模型背景.....	1
1.1.2 非对称最小二乘法背景.....	3
1.1.3 LASSO 变量选择和纠偏背景	3
1.2 国内外研究现状.....	5
1.2.1 关于单指标模型的研究现状	5
1.2.2 关于期望分位数回归模型的研究现状	6
1.2.3 关于求解期望分位数回归模型方面的研究	7
1.2.4 关于 LASSO 理论方面的研究	8
1.3 本文研究内容与结构安排	9
1.3.1 研究内容	9
1.3.2 结构安排	9
第二章 理论知识介绍	11
2.1 模型介绍.....	11
2.1.1 单指标模型	11
2.1.2 期望分位数回归模型	12
2.2 LASSO 变量选择	13
2.2.1 LASSO 的定义.....	14
2.2.2 ADAPTIVE LASSO 的定义	14
2.2.3 LASSO 和 ADAPTIVE LASSO 比较.....	15
2.3 纠偏理论介绍	15

2.4 本章小结	17
第三章 模型的变量选择和纠偏估计	19
3.1 变量选择	19
3.1.1 非对称最小二乘法的估计步骤	19
3.1.2 ADAPTIVE LASSO 下的惩罚非对称最小二乘估计	20
3.1.3 PALS 估计的渐近性质和 ORACLE 属性	21
3.2 纠偏估计	22
3.3 参数的统计推断	23
3.4 本章小结	24
第四章 模拟实验	26
4.1 数值模拟	26
4.1.1 数据生成	26
4.1.2 评价指标	27
4.1.3 模拟结果	27
4.2 实证分析	31
4.2.1 数据描述	31
4.2.2 实验结果	32
4.3 本章小结	35
第五章 总结与展望	36
5.1 总结	36
5.2 不足与展望	36
参考文献	38
附录 A 渐近性质所需的条件	42
附录 B 定理证明	43
附录 C 代码	48
致谢	69

图录

图 4-1. 模型 2 在不同误差和期望下 95%水平的平均覆盖率.....	31
图 4-2. 波士顿住房数据单指数预期回归估计.....	33
图 4-3. 波士顿住房数据的变量选择和检验（原点表示被选择的变量）.....	34

表录

表 4-1. 模型 1 在不同误差和期望水平下的模拟结果.....	28
表 4-2. 模型 2 在不同误差和期望水平下的模拟结果.....	28
表 4-3. 模型 1 在不同误差和期望下 95%水平的平均覆盖率.....	29
表 4-4. 波士顿住房数据描述.....	32

第一章 绪论

1.1 研究背景和意义

1.1.1 高维数据下的单指标期望分位数回归模型背景

随着数据采集技术的不断更新迭代,以及计算机计算能力的飞速发展提升,在社会实践过程中,我们可以采集越来越多的数据。在看待现实中的具体问题时,我们可以考虑的更全面,同时为了更科学、更准确地建立模型,需要尽可能多地将所有显著的协变量引入其中。带着这样的思想,当我们开始分析高维数据的时候,就遇到了“维数诅咒”问题。我们在研究多元非参数回归问题时,随着变量的维数 P 逐渐增加,高维空间中的数据分析遇到的问题越多,统计推断的精度也越来越低,这就是“维数诅咒”问题。用一个简单的例子来说明:假设有 n 个样本,它们均匀分布在区间 $[0,1]$ 上,那么落在区间 $[0,0.1]$ 上的样本就大概有 $n / 10$ 个。但是,如果是在高维空间上,比如 $[0,1]^{10} = [0,1] \times \dots \times [0,1]$ 的样本上有 n 个均匀分布的样本,那么就只有大概 $n / 10^{10}$ 个样本落在区间 $[0,0.1]$ 上。这个时候,要想 $[0,1]^{10}$ 内有数据可以进行统计推断,就要样本容量 n 充分大。统计学的主要任务就包含通过各种算法模型对数据进行处理、分析。在实际研究中,生物医学领域、工程领域、经济金融领域等各个领域,高维数据已经屡见不鲜。高维数据的出现以及其带来的“维数诅咒”问题,让传统的统计和数据分析方法在分析高维数据时显得力不从心,亟需在之前的基础之上寻找或改进为新的数据统计与分析方法来处理高维数据。因此研究高维数据成为当下统计学的热点问题。所以必须考虑的问题就是如何克服“维数诅咒”,提高统计推断的精度,换句话说,就是我们在利用模型对高维数据进行拟合的时候,如何降低所使用的数据的维度。

统计学家们提出了很多方法来解决“维数诅咒”问题,其中利用回归分析的方法来处理高维数据是一大热点。高维数据带来的最主要的问题就是对于一般的参数估计方

法, 伴随着协变量的增加, 其估计效果会随之变弱。学者们提出了多种方法来克服这一问题, 比如 SIR 法、投影追踪法、图回归等降维方法, 其中还包括单指标模型。我们通常所说的半参数模型是介于参数模型和非参数模型之间的一类模型, 单指标模型就是一种半参数模型。由于可以用非光滑函数作为系数, 单指标模型和经典的回归模型相比较而言, 在降低维数方面就具有天然的优势。同时高维数据的稀疏性特点也能被单指标模型克服, 通过降维的手段, 极大提高统计推断的精度。从 20 世纪 80 年代中后期到如今, 许多学者对单指标模型进行研究, 提出了许多经典的理论。

单指标模型下的参数估计问题以及变量选择问题是当下研究的热点。最近, Jiang, Hu & Yu (2020)^[1]提出一种单指标期望分位数回归模型。这一模型将单指标模型与期望分位数回归模型相结合, 用非对称最小二乘估计的方法来估计单指标模型中的参数并进行变量选择。期望分位数回归具有很多优势。普通最小二乘 (OLS) 回归对正残差和负残差赋予同等权重, 并将条件均值函数估计为回归变量或预测变量的线性函数。Newey 和 Powell (1987)^[2]为经典线性回归模型提出了一类估计量, 其允许对正残差和负残差使用不同的权重, 并称为期望分位数回归。 Y 的总体 τ -Expectile, $\tau \in (0,1)$, 被定义为 $Q_\tau(Y)$ 的值, 它最小化下面的式子:

$$E\{|\tau - I(Y < m)| \cdot (Y - m)^2\}.$$

Newey & Powell (1987) 指出这个估计量具有同质性、单调性和平移不变性等优良特性。Breckling & Chambers (1988)^[3]表明期望分位数是 M-估计的一种特殊形式。Jones (1994)^[4]表明除了原始分布的转换外, 期望分位数确实是分位数。Ziegel (2014)^[5]发现了期望分位数的连贯性和可导性。

基于上述单指标模型和期望分位数回归模型的种种优势, 我们结合二者, 探讨高维数据下的单指标期望分位数回归模型。我们用以下单指标期望分位数回归模型来估计 Y 在给定 X 下的 τ -conditional Expectile:

$$Q_\tau(Y|X) = g_0(x^T \gamma_0), \quad (1-1)$$

其中 Y 是单响应变量, X 是 p 维协变量。函数 $g_0(\cdot)$ 是未指定的非参数平滑函数。 γ_0 是未知的指数向量系数, 以及为了便于识别 (Lin & Kulasekera 2007) ^[6], 我们假设 $\|\gamma_0\| = 1$ 并且 γ_0 的第一个分量是正的。这里的 $\|\cdot\|$ 是欧几里得范数。注意 $g_0(\cdot)$ 和 γ_0 可能依赖于 τ 。为了简便, $g_0(\cdot)$ 和 γ_0 里的 τ 被省略了。

1.1.2 非对称最小二乘法背景

在回归分析中, 传统最小二乘法具有明显的缺陷。它有一个前提假设: 对于模型中的随机扰动因子, 全都服从均值为零、方差一样的正态分布, 以便于对最小二乘估计进行优化, 得到最小方差无偏估计。然而在实际应用中, 这种分布的可能性几乎为没有。例如在数据存在显著异方差的情况下, 或者数据出现厚尾或者尖尾分布, 最小二乘法将发挥不出应有的作用。Newey 和 Powell (1987) ^[2] 提出了期望分位数回归的思想, 同时在此基础之上提出了非对称最小二乘估计 (Asymmetric Least Squares Regression) 的方法, 用来估计线性期望分位数模型的参数。相对于分位数, 期望分位数有两个优势: 一是期望分位数的检测损失函数是处处可微的, 但是分位数的检测损失函数在零点不可微。二是在计算线性协方差时, 计算误差密度函数值对于统计分析有重要意义, 然而线性期望分位数没有计算这一函数值。因此本文在求解期望分位数回归模型时, 主要是基于非对称最小二乘估计 (ALS)。ALS 的估计方法是在传统的最小二乘估计方法的基础上, 将其损失函数拓展为二次形式, 并且是非对称的, 即用 $\tau (0 < \tau < 1)$ 和 $1 - \tau$ 对正负两个部分进行加权, 用回归得到被解释变量的期望分位数估计。

1.1.3 LASSO 变量选择和纠偏背景

变量选择近年来发展迅速, 在模型构建过程中的作用越来越明显。在实践中, 一组数据存在大量的、可用的候选预测变量的情况非常之多, 为了在建模的时候消除潜

在的偏差，在建模的初始阶段将它们包括在内。然而，如果在最终的模型中，我们保留了不相关的预测变量，这是不可取的。因为这会导致我们在对结果模型进行解释的时候异常困难，模型的预测能力也会随之被削弱。为此目的，逐步删除和子集选择实际上是有用的。然而，这些过程忽略了误差积累和计算效率。AIC 和 BIC 原则也被提出来以在一定损失估计精度下控制预测变量的数量。在正则化框架中，变量选择和估计精度同时考虑，并引入多种不同类型的惩罚来实现变量选择。如今高维数据引起了广泛的讨论和研究，基因组学、功能性磁共振成像、经济和金融等越来越多的领域都需要处理这类数据。与此同时，由于高维数据给我们带来了大量的信息，另一个挑战是弄清楚需要什么。这一目标需要我们在回归过程中进行变量选择。Tibshirani(1996)提出使用 l_1 范数作为罚函数，即著名的 LASSO。该算法对数据没有过多的要求。无论变量是一元还是多元、连续还是离散，都可以用 LASSO 的方法对变量进行处理。后来，许多学者通过研究证明，LASSO 方法具有很好的稳定性，同时很好地可以适用于高维数据。但是，LASSO 算法有一个缺陷，那就是 LASSO 算法是有偏的，从而在对每个系数进行惩罚时，使用相同的权重的做法是不科学的，容易造成模型偏差较大。因此，就有学者们提出了 SCAD (smooth clipped absolute deviation) 惩罚函数，该函数被 Fan 和 Li 在 2001 年提出来，SCAD 可以同时进行变量选择和系数估计，同时证明了他们的方法保持了所谓的 Oracle 性质：它产生稀疏解，保证了模型选择的稳定性，并提供了渐近无偏估计。但是，SCAD 惩罚是非凸的，因此不容易解决相应的优化问题。相比之下，Adaptive LASSO 具有不一样的特征：对于具有不同大小的一些估计参数，Adaptive LASSO 在进行变量选择的时候，它给每个惩罚项分配了不一样的权重。也就是说，在 Adaptive LASSO 的方法下，惩罚函数的参数值不是始终不变的，而是依据各个变量的不同值而变化的。在 Adaptive LASSO 的方法下，重要的解释变量它的惩罚系数就会很轻；而对于次要的解释变量，它的惩罚系数就会相对较重，这样的做法就使不重要的变量更容易被模型过滤掉。更令人欣喜的是，Adaptive LASSO 具有 Oracle 的性质。有关于惩罚下期望分位数回归模型的研究还很少。就我

们目前所知,对惩罚单指标期望分位数时间序列模型中变量选择的研究还不够。所以在本文中,我们通过利用 Adaptive LASSO 惩罚来研究单指标期望分位数回归,以填补这一空白。

还有一个更为重要的问题是,如何在高维环境中执行统计推理,如置信区间和假设检验。这个问题的一个主要难点是,像 LASSO 这样的稀疏估计没有一个易于处理的极限分布。人们可以为被选择的变量建立渐近正态性,但对未被选择的变量没有推论结果。因此,正则化方法不容易产生渐近正态分布的估计量,使其难以建立统计推断的结果。与现有文献相比,本文考虑了高维单指标期望分位数回归模型的纠偏方法。具体地说,我们开发了一个渐近无偏估计方法,它应用纠偏方法和非对称最小二乘(ALS)方法,用于单指标期望分位数回归模型的未知参数的估计。

1.2 国内外研究现状

1.2.1 关于单指标模型的研究现状

我们在对单指标模型进行研究时,主要问题组成有两部分,即估计指标函数和指标参数。尤其是未知参数的估计引起了大量科学家的关注。为了在高维数据的情况下实现非参数多元回归,单指标模型就被 Friedman^[7](1981)在研究投影寻踪回归问题时提出来。Ichimura^[8]在 1993 年构造了半参数最小二乘估计(SLS),在指标参数给定的情况下,结合非线性最小二乘法估计指标参数。证明了 SLS 估计具有渐近性质,同时还给出了一种加权半参数最小二乘估计(WSLS)。最小平均方差估计(MAVE)在 2001 年由 Xia 等^[9]提出,以往的估计方法通常要对联系函数做欠光滑的处理,而 MAVE 方法在跳过这一步骤的情况下,仍旧在参数估计的时候表现出了较快的收敛速度。在 2001 年 Hristache 等^[10]提出了一种梯度外积估计(OPG),这是一种改进平均导数估计的方法。2006 年 Xia^[11]对梯度外积的估计方法进行了改进,同时还改进了最小平均方差估计,估计的渐近分布也一并给出。Zhu^[12]等在 2012 年对单指标模型在基于“线性

条件下”，提出了两步算法，大大提高了估计的运算速度，并且这一方法适合处理高维数据。2015年 Wang^[13]通过 B 样条近似研究了高维和超高维数据下的单指标模型。单指标模型参数的全局估计被刘阳^[14]在 2018 年通过 Adaptive-LASSO 得到，随后讨论了该模型下的变量选择问题，证明了通过 Adaptive-LASSO 得到的单指标模型的全局估计量满足 Oracle 性质。2019 年，Liu 等^[15]在存在随机缺失的协变量的情况下，对单指标加权复合分位数模型进行估计，得出了模型的估计指标参数和连接函数具有大样本属性的结论。

1.2.2 关于期望分位数回归模型的研究现状

期望分位数的相关理论最早由 Newey & Powell (1987)^[2]提出，它的定义是非对称最小二乘估计(Asymmetric Least Squares Regression)的解，即上面所述公式(1-1)的最小化解。其中，损失函数非对称程度即由 $\tau \in (0,1)$ 决定。2008 年 Taylor^[16]和 2009 年 Kuan 等^[17]在标准期望分位数回归模型的基础上，提出了 CARE(Conditional Auto-Regression 期望分位数 Models)条件自回归期望分位数模型。这一模型在获取收益率的非对称信息方面具有良好的优势，但没有考虑风险因素的动态效应和交互作用。苏辛等^[18]在 2013 年引入杠杆效应进入 CARE 模型中，利用这一模型对开放式基金的收益率进行研究，用来测度其风险。在 Kuan 等^[17]的 CARE 模型之上，谢尚宇等^[19]在 2014 年对其进行拓展，将误差项引入到 ARCH 模型中，建立了一种线性 ARCH-期望分位数模型。这一模型不仅对异方差数据适用，还可以做风险因素的贡献率研究。Xie & Wan (2014)^[20]通过研究反映当前投资环境信息的变量，并采用变系数的设置，以此丰富了条件动态期望模型。Cai, Fang & Tian (2018)^[21]为一类具有部分变系数的动态期望模型提出了一个三阶段的估计程序。极值理论被胡宗义等^[22]在 2018 年引入，在刻画极端风险方面构建一种期望分位数-EVT 模型。Liao 等^[23]在研究住房数据时，对带有 SCAD 惩罚和 Adaptive LASSO 的期望分位数回归模型进行研究。

1.2.3 关于求解期望分位数回归模型方面的研究

期望分位数是一个平方损失函数，对于度量风险是一种很好的方法，可以进行很好的有效估计。同时，期望分位数对极端值非常敏感，且求解更加容易，所以能更好的计算极端损失的大小。在实践中，通常有大量的候选预测变量可用。回归分析的一个主要挑战就是在许多潜在的预测变量中选择出合适的变量。这个目标需要我们在回归过程中进行变量选择。然而，目前对于期望分位数回归模型的变量选择所做工作很少。期望分位数的求解主要是基于非对称最小二乘回归。1976年，Aigner^[24]首次提出 ALS，并由 Newey 和 Powell (1987)^[2]进一步发展。2013年，YiYang 等^[25]使用了一种兼具计算效率高、可扩展性强、可使用于高维数据的优点的算法，称为广义坐标下降算法，解决了期望分位数回归在带有约束条件的情况下的估计问题。Gu & Zou (2016)^[26]在高维环境中考虑了惩罚期望回归，并提出了一种使用两步程序检测异方差的方法。2017年 Zhao 等^[27]研究了一种特殊的期望分位数回归模型，这一模型带有 SCAD 惩罚，在实现变量选择时，采用差分凸算法(DCA:Difference Convex Algorithm)求解模型。然而目前关于惩罚期望分位数回归的研究目前来说还很少。Liao, Park & Choi(2018)^[28]开发了一种惩罚线性期望分位数回归，使用 SCAD 和 LASSO 惩罚函数。Zhao, Chen 和 Zhang(2018)^[29]研究了在回归误差有有限项时，使用 SCAD 惩罚下的正则化期望分位数回归模型研究高维数据的异方差性。Fan 等人(2018)^[30]研究了一种基于预测的单指标模型，用于极高维数据的回归。他们考虑一般加权损失的复合回归，其中包括分位数回归和期望分位数回归。但是他们的方法在时间序列中并不有效。就我们目前有限所知，对于单指标期望分位数时间序列模型的变量选择研究还不够。在这项工作中，我们通过使用 Adaptive LASSO 来填补这一空白 (Zou 2006)^[31]。我们证明了我们的估计器保持了 Oracle 属性：它对噪声变量具有稀疏性，并且对变量具有渐近正态分布。

1.2.4 关于 LASSO 理论方面的研究

LASSO 方法由 R. Tibshirani^[32]首次提出,它通过以下方法来实现变量选择或者选出最优子集: LASSO 强制地使系数的绝对值之和小于某个固定值,这样做的目的是使残差平方和达到最小值,通过构造一个惩罚函数得到一个较为精简凝练的模型。这样得到的模型的部分回归系数就变为了零。LASSO 方法在进行多重共线性的有偏估计时可以发挥很好的作用。但是 Fan^[33]等人也指出来 LASSO 方法的缺陷,就是该方法在进行模型估计时,模型的精确度往往会因为出现较大的偏差而降低,于是 SCAD 方法便由此提出来。SCAD 方法不仅能像 LASSO 一样,对系数进行连续压缩,在系数较大时还可以取得渐近无偏估计。但是 SCAD 也存在缺陷:由于模型过于复杂,所以它迭代的速度非常缓慢。在这以后, Zou 和 Hastie^[34]将岭回归估计的约束方式和 LASSO 估计的约束方式进行结合,提出了一种弹性约束估计 (Elastic Net)。Zou^[31]提出了 Adaptive LASSO 的方法对 LASSO 进行改进:引入权重向量,对不同的系数采取不同的惩罚程度,模型便能够选择最优子集。胡学钢^[35]等提出了迭代式 LASSO 方法,这一方法在对数据量庞大的高维数据进行变量选择时也非常有效。谢仪^[37]在研究时间序列模型时证明了相比于 LASSO, Adaptive LASSO 对于 AR(P)模型的定阶以及参数估计更具有优越性。张婷婷^[37]构建出一种 Adaptive LASSO-Logistic 回归模型,同时将 Adaptive LASSO-Logistic 方法与传统的 Adaptive LASSO 方法进行比较,并用个人信用评级模型进行验证,证明 Adaptive LASSO-Logistic 回归模型有更好的准确率和优越性。显而易见, LASSO 相关理论已经在社会生活的各个领域有了广泛应用于,在处理高维数据时有着自身的独特优势。

还有一个鲜为人知的问题是如何进行统计推断,例如置信区间和假设检验。我们可以为选定的变量建立渐近正态性,但对非选定的变量没有推论结果。因此,正则化方法不容易产生渐近正态分布的估计量,因此难以建立统计推断的结果。有一系列的工作: Zhang & Zhang (2014)^[38]; van de Geer et al. (2014)^[39]; Javanmard & Montanari

(2014)^[40]提出了一种纠偏 LASSO 估计来构建置信区间并验证高维线性模型和广义线性模型中低维坐标的假设。Gueuning & Claeskens (2016)^[41]研究了基于 LS 方法的部分线性单指数模型的纠偏 LASSO 程序。与现有文献相比，这项工作是考虑单指标期望分位数回归模型的纠偏过程。

1.3 本文研究内容与结构安排

1.3.1 研究内容

为了解决高维数据下的单指标模型参数估计问题，本文选择结合期望分位数回归估计参数并进行变量选择，即研究一种单指标期望分位数回归模型。我们选择非对称最小二乘法(ALS)和 Adaptive LASSO 来进行参数估计和变量选择。本文的主要创新点有两点：

(1)介绍了一种新的惩罚 ALS 方法，用于对模型(1-1)中的指标系数进行有效估计，同时进行变量选择。

(2)研究了一种新的纠偏 ALS 方法，它将纠偏方法与 ALS 结合，可证明所提出的估计具有渐近正态性，可用来构造有效置信区间和作假设检验。

1.3.2 结构安排

本文的第一部分绪论，主要介绍了单指标期望分位数回归模型的背景和非对称最小二乘法以及 LASSO 变量选择和纠偏的背景。然后综述了关于单指标模型和期望分位数回归模型以及 LASSO 理论的国内外研究现状。最后是介绍本文的研究内容和结构安排。

第二部分为本文理论部分的研究和阐述。第一小节主要介绍本文所用的单指标模型和期望分位数模型及其具体结合。第二小节对 LASSO 和 Adaptive LASSO 变量选择进行介绍。第三小节介绍了本文所用的纠偏相关理论。

第三部分为求解模型。第一小节里面，我们介绍了模型(1-1)的变量选择的方法，它结合非对称最小二乘估计和 Adaptive LASSO。第一步是给出模型(1-1)中 γ_0 的非对称最小二乘估计的步骤，第二步是结合 Adaptive LASSO，给出 γ_0 的 Adaptive LASSO 惩罚非对称最小二乘估计，最后还证明了 Adaptive LASSO 下的惩罚 ALS 估计的一致性和 Oracle 性质。第二小节里面，我们介绍了惩罚 ALS 方法的纠偏技术，证明了纠偏后的 γ_0 的估计具有渐近正态分布，可以基于此作出统计推断。第三小节里面，我们做了参数的统计推断，给出了纠偏后的 γ_0 估计的假设检验方法和置信区间。

第四部分，我们通过 R 语言程序进行了蒙特卡罗仿真模拟以及运用真实数据集来进行实证分析。

第五部分我们总结了本文所研究的内容，指出本文的研究还存在哪些不足，最后表达了对未来工作发展的期望。

第二章 理论知识介绍

2.1 模型介绍

2.1.1 单指标模型

单指标模型是一种降低协变量的维数的好模型,其可以在保证数据信息不被丢失的情况下抓取到高维数据的特征,具有非常大的灵活性。正是由于单指标模型所具有的这些优点,它引起了学者们的极大关注。通常在经济研究中,单指标模型被认为是介于严格参数化模型和灵活但难以估计的非参数模型之间的一种平衡。

单指标模型是由一类线性模型推广得到的。假设 $\{(Y_i, X_i), i = 1, \dots, n\}$ 来自线性模型

$$Y = X^T \gamma_0 + \varepsilon, \quad (2-1)$$

其中 ε 是随机误差,是独立的。 $X_i = (X_i^{(1)}, \dots, X_i^{(p)})^T \in \mathbb{R}^p$ 是 p 维设计向量, γ_0 是真实参数向量。单指标模型即是线性模型(2-1)的推广。它通过这样的转换来降低数据的维数:将线性模型中的线性组合 $X^T \gamma$ 投影到一个线性空间上,然后用一个一元函数 $g(X^T \gamma)$ 去在这个一维空间上拟合,从而得到一个单指标参数 $X^T \gamma$,它是多元变量 X 的转变。单指标模型的具体形式如下:

$$Y = g_0(X^T \gamma_0) + \varepsilon, \quad (2-2)$$

其中 Y 是响应变量。 X 是一个 p 维的协方差, $g_0(\cdot)$ 是未知函数, γ_0 是未知的指标参数。考虑到模型的可识别性,假设 $\|\gamma_0\|=1$, $\|\cdot\|$ 表示向量的模长,并且 γ_0 的第一项是正数, ε 是误差向量。

2.1.2 期望分位数回归模型

期望分位数的概念最早是由 Aigner 等提出, 由 Newey 和 Powell (1987)^[2]进一步给出了具体的理论方法。

假设随机变量 Y 表示资产收益, 它的分布函数是 F_Y 。对于 $\theta \in (0,1)$, 定义如下损失函数

$$Q_\theta(y) = \begin{cases} (1-\theta)y^2 & y \leq 0, \\ \theta y^2 & y > 0. \end{cases}$$

y 的期望分位数值 $v(\theta)$ 可通过下式得到

$$v(\theta) = \arg \min_{|a| < \infty} E\{Q_\theta(Y - a)\}, \quad (2-3)$$

显然, 当 $\theta = 0.5$ 时, (2-3) 即为标准最小二乘目标函数。由于 $Q_\theta(y)$ 具有连续一阶导数, 所以 $v(\theta)$ 满足

$$E\{L_\theta(Y - v(\theta))\} = 0,$$

其中

$$L_\theta(y) = \begin{cases} (1-\theta)y & y \leq 0, \\ \theta y & y > 0. \end{cases}$$

继而有

$$\frac{\int_{-\infty}^v |y - v| dF_Y(y)}{\int_{-\infty}^v |y - v| dF_Y(y) + \int_v^{\infty} |y - v| dF_Y(y)} = \frac{\int_{-\infty}^v |y - v| dF_Y(y)}{\int_{-\infty}^{\infty} |y - v| dF_Y(y)} = \theta.$$

上面式子也可以证明下式:

$$v(\theta) = \gamma E[Y|Y > v(\theta)] + (1 - \gamma) E[Y|Y \leq v(\theta)],$$

其中 $\gamma = \theta[1 - F_Y(v(\theta))]/\{\theta[1 - F_Y(v(\theta))] + (1 - \theta)F_Y(v(\theta))\}$ 。

期望分位数的一些性质也被 Newey 和 Powell (1987)^[2]给出:

定理 2.1. 假设随机变量 Y 的期望 $E[y]$ 存在, 则对于任意 $\theta \in (0,1)$, 存在唯一的 $v(\theta)$ 使得 (2-3) 成立, 并且:

(i) 函数 $v(\theta): (0,1) \rightarrow \mathbb{R}$ 是严格单调递增的。

(ii) $v(\theta)$ 的值域是 I_F 。

(iii) 对于 $\tilde{Y} = sY + t$, \tilde{Y} 的期望分位数满足:

$$\tilde{v}(\theta) = \begin{cases} sv(\theta) + t & s > 0, \\ sv(1 - \theta) + t & s < 0. \end{cases}$$

证明: 详细证明过程请参考 Newey 和 Powell (1987)^[2]。

单指标期望分位数回归模型是结合了单指标模型和期望分位数模型的一种模型, 它结合了两种模型的优势, 既能够用来研究高维数据, 同时又具有期望分位数模型的同质性、单调性和平移不变性等优良特性。

2.2 LASSO 变量选择

目前, 诸如相关分析、回归分析、聚类分析等常用的一些统计方法, 它们都属于系统分析的方法。这些统计方法为研究社会经济、生物医疗等做出了巨大贡献, 在统计学的一部分领域也占据领先地位, 但也存在一定的局限性:

1、这些统计方法都是基于一定的假设, 通常需要大量的甚至是超大量、海量的数据来进行支撑, 如果数据量不符合要求, 就不能得出有用的统计规律;

2、在一些特殊的情形下, 还可能出现一类棘手的情况: 定量结果与定性分析不一致。这一情况会导致所选取的样本与原系统中的总体变量关系产生不一致的现象。

而 LASSO 方法在筛选出对因变量影响较大的解释变量的同时还可以提供筛选出的变量的系数, 消除多重共线性的影响。因此相比较于上述的一些统计学方法, LASSO 方法可以弥补一些不足。基于 LASSO 的这两点优势, 本文采用 LASSO 中常用的一种方法——Adaptive LASSO 来同时实现这两个过程。

2.2.1 LASSO 的定义

通常对于一个自变量 y 可能伴随着很多影响因素。但是我们只想要影响因素里面的重要的解释变量，这就要求模型中不重要的自变量的系数为零。这一过程通常被叫做变量选择。变量选择就是通过获取的数据来确定哪些影响因素的系数为零，从而去除变量，得到一个稀疏模型。LASSO 方法就是通过这一变量选择的过程得到一个稀疏模型。LASSO 方法是最小二乘方法的一种衍生方法，它在回归系数 β 上施加了一种 L_1 范数。它考虑了以下问题：

$$\begin{aligned} \min_{\beta} \sum_{i=1}^N (y_i - x_i' \beta)^2, \\ \text{s. t. } \sum |\beta| \leq t, t \geq 0, \end{aligned}$$

其中，调节参数 t 越小对于参数 β 的限制越强。上述的表达式也可以表示为：

$$\min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda |\beta|,$$

其中 $\lambda \geq 0$ 是惩罚参数。LASSO 的方法就是对所有的系数都进行相同的压缩，这一做法存在一定程度上的缺陷，最后得到的模型精度可能较低，并且 LASSO 的方法不具有 Oracle 性质。相比于 LASSO，Adaptive LASSO 就具有 Oracle 性质。

2.2.2 Adaptive LASSO 的定义

LASSO 估计就是用一个定值 λ ，即所有变量都被同一个参数所限制。也正是因为使用相同的参数，LASSO 估计得到的是一个有偏估计。而对于 Adaptive LASSO，有

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2 + \lambda \sum_{i=1}^n \omega_i |\beta|,$$

其中 ω 是权重向量。从上式可以看出，Adaptive LASSO 会根据估计参数的大小，给不同的惩罚项赋予不同的权重，也就是说惩罚函数的参数是可以变化的。于是更多的估计变量就可以被模型所选择，同时还能保障良好的拟合效果。

2.2.3 LASSO 和 Adaptive LASSO 比较

LASSO 算法可以处理多种多样的数据类型, 换句话说, 其对数据的要求非常低, 无论是连续型变量还是离散型变量, 也无论是一元变量还是多元变量, 都可以用 LASSO 算法进行处理。经过一代又一代的学者们对 LASSO 进行研究, 证明了 LASSO 方法具有许多的优点。它在变量选择中展现出了出色的稳定性。但是, 研究也表明了 LASSO 算法并不满足 Oracle 性质, 这就使得 LASSO 算法具有很大的缺陷, 所得到模型的精确度大打折扣。而对比之下, Adaptive LASSO 就具有 Oracle 性质: 它根据所得的参数估计的大小, 对不同的惩罚项赋予不同的权重。对于重要的解释变量, Adaptive LASSO 的算法会赋予它较小的惩罚系数, 而对于次要的解释变量, Adaptive LASSO 则是赋予其较大的惩罚系数, 这样一来, 模型在筛选解释变量的时候就更加容易而准确。

2.3 纠偏理论介绍

在高维稀疏模型中, 关于构造置信区间、假设检验和分配不确定性的工作做得很少。这些问题的一个主要困难是, 类似 LASSO 这样的稀疏处理器没有一个易于处理的极限分布, 即便已经在低维设置中, 它依旧依赖于未知参数并且收敛到的极限是不均匀的。van de Geer et al.^[40](2014)提出了一种在对模型的稀疏性和设计矩阵进行假设时具有最优性的方法。这一方法以套索为基础, 与相应的 KKT 条件“反演”。这就产生了一个非稀疏估计量, 它具有一个高斯(有限)分布。van de Geer et al.^[40](2014)证明, 在稀疏线性模型设置中, 估计量在达到半参数效率界限的意义上是最优的。该过程可用于高维稀疏线性和广义线性模型, 并用于分析具有一般凸(稳健)损失函数的回归问题。具体方法如下:

考虑一个高维线性模型:

$$Y = X\beta^0 + \varepsilon,$$

其中 $X = [X_1, \dots, X_p]$ (X_j 是 $n \times 1$ 维), X 是 $n \times p$ 的矩阵, $\varepsilon \sim N_n(0, \sigma_\varepsilon^2 I)$ 并独立于 X , 未知回归向量 β^0 是 $p \times 1$ 维。假设 $p > n$ 并且在渐近结果中 $\frac{\log(p)}{n} = o(1)$, 用 $S_0 := \{j; \beta_j^0 \neq 0\}$ 表示变量的活动集及其基数 $s_0 := |S_0|$ 。van de Geer et al. [40] (2014) 的主要目标是对参数向量 β_j^0 ($j = 1, \dots, p$) 的分量进行逐点统计推断, 还讨论了参数 $\beta_G^0 := \{\beta_j^0; j \in G\}$ 的同时推断, 其中 $G \subseteq \{1, \dots, p\}$ 是任意组。

定义 LASSO 为:

$$\hat{\beta} = \hat{\beta}(\lambda) := \arg \min_{\beta \in \mathbb{R}^p} (\|Y - X\beta\|_2^2/n + 2\lambda\|\beta\|_1),$$

显然 $\hat{\beta}$ 满足 KKT 条件:

$$-X^T(Y - X\hat{\beta})/n + \lambda\hat{\kappa} = 0, \quad (2-4)$$

$$\|\hat{\kappa}\|_\infty \leq 1 \quad \hat{\kappa}_j = \text{sign}(\hat{\beta}_j), \quad \hat{\beta}_j \neq 0.$$

向量 $\hat{\kappa}$ 来自于 $\|\beta\|_1$ 的次微分, 根据(2-4)我们总是可以把它表示为:

$$\lambda\hat{\kappa} = X^T(Y - X\hat{\beta})/n. \quad (2-5)$$

KKT 条件可以被写为: $\hat{\Sigma} = X^T X/n$:

$$\hat{\Sigma}(\hat{\beta} - \beta^0) + \lambda\hat{\kappa} = X^T\varepsilon/n.$$

假设 $\hat{\Theta}$ 是 $\hat{\Sigma}$ 的逆的一个合理的近似, 那么

$$\hat{\beta} - \beta^0 + \hat{\Theta}\lambda\hat{\kappa} = \hat{\Theta}X^T\varepsilon/n - \Delta/\sqrt{n}, \quad (2-6)$$

其中

$$\Delta := \sqrt{n}(\hat{\Theta}\hat{\Sigma} - I)(\hat{\beta} - \beta^0).$$

Δ 在一定的稀疏性假设下渐近可忽略, 运用(2-5)式, 于是便有以下估计:

$$\hat{b} = \hat{\beta} + \hat{\Theta}\lambda\hat{\kappa} = \hat{\beta} + \hat{\Theta}X^T(Y - X\hat{\beta})/n.$$

利用(2-6), 当 Δ 可以忽略不计时, 我们可以立马得到一个渐近枢轴:

$$\sqrt{n}(\hat{b} - \beta^0) = W + o_p(1),$$

$$W|X \sim N_p(0, \sigma_\varepsilon^2 \hat{\Theta}\hat{\Sigma}\hat{\Theta}^T).$$

β_j^0 的渐近逐点置信区间即是：

$$[\hat{b}_j - c(\alpha, n, \sigma_\varepsilon), \hat{b}_j + c(\alpha, n, \sigma_\varepsilon)],$$

$$c(\alpha, n, \sigma_\varepsilon) := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sigma_\varepsilon \sqrt{\hat{\Theta} \hat{\Sigma} \hat{\Theta}^T}_{jj} / n.$$

其中 $\Phi(\cdot)$ 是标准正态分布。如果 σ_ε 未知，就用一个一致估计量代替。

2.4 本章小结

本章主要为本文所做单指标期望分位数回归模型研究的理论部分的介绍。

在第一小节里面，详细地介绍了本文所使用的单指标期望分位数回归模型的由来，分别给出了原始的单指标模型和期望分位数回归模型的定义和形式，给出了二者具有的优良性质。鉴于单指标模型和期望分位数回归模型的优势将两者结合，介绍了单指标期望分位数回归模型，该模型同时具有单指标和期望分位数的优势，在处理高维数据方面具有很大的优势。

在第二小节里面，详细地介绍了在对单指标期望分位数回归模型进行变量选择的时候所用到的方法：Adaptive LASSO。本节首先阐明了 LASSO 选择相对于其它传统的统计分析方法所具有的优势，接下来对 LASSO 理论作了介绍。之后引出 Adaptive LASSO，它是 LASSO 估计的一种衍生，给出了 LASSO 和 Adaptive LASSO 算法，将二者进行比较，介绍了 Adaptive LASSO 相对于 LASSO 的优势，即 Adaptive LASSO 能很好的满足 Oracle 性质。

在第三小节里面介绍了本文所使用的纠偏算法。我们参考了 van de Geer et al.^[40] (2014) 里面所提出的纠偏算法，预备将其应用在变量的纠偏中，使其具有易于处理的极限分布，以便进行下一步的统计推断。

在下一章节中, 将给出模型(1-1)中 γ_0 的具体估计步骤, 以及其具体的纠偏估计和统计推断结果。

第三章 模型的变量选择和纠偏估计

3.1 变量选择

3.1.1 非对称最小二乘法的估计步骤

在本文中，我们令 Y 为资产回报， X 为可能包含滞后收益或者其他经济金融因素的风险因素。我们假设所有 (X, Y) 中的序列 $\{X_t, Y_t\}_{t=1}^T$ 都是严格平稳过程，并且满足强混合(α -混合)条件，具有有限二阶矩。理论上来说，模型(1-1)中的真参数向量 γ_0 就等于下面式子的最小化解：

$$\gamma_0 = \arg \min_{\|\gamma\|=1, \gamma_1 > 0} E[\rho_\tau(Y - g_0(X^T \gamma))], \quad (3-1)$$

其中 $\rho_\tau(\lambda) = |\tau - I(\lambda < 0)| \cdot \lambda^2$, $\tau \in (0, 1)$ 是损失函数，上标 T 代表向量或者矩阵的转置。(3-1)右侧是期望损失，它等同于：

$$E[\rho_\tau(Y - g_0(X^T \gamma))] = E\{E[\rho_\tau(Y - g_0(X^T \gamma)) | X^T \gamma]\}. \quad (3-2)$$

我们通过局部线性平滑估计来估计(1-1)中的指数函数 $g_0(\cdot)$ 。因为 $X^T \gamma$ “靠近” u ，因此 $g(X_t^T \gamma)$ 可以被近似线性估计为：

$$g(X_t^T \gamma) \approx g(u) + g'(u)(X_t^T \gamma - u) \equiv a + b(X_t^T \gamma - u), \quad (3-3)$$

其中 $a \equiv g(u)$ 并且 $b \equiv g'(u)$ 。紧接着(3-3)， $E[\rho_\tau(Y - g_0(X^T \gamma)) | X^T \gamma = u]$ 可以被估计为：

$$\sum_{t=1}^T \rho_\tau\{Y_t - a - b(X_t^T \gamma - u)\} K\left(\frac{X_t^T \gamma - u}{h}\right),$$

其中 $K(\cdot)$ 是核函数， h 是带宽。通过平均化 u ，我们可以得到(3-2)的经验近似函数是：

$$\sum_{t'=1}^T \sum_{t=1}^T \rho_\tau\{Y_t - a_{t'} - b_{t'}(X_t - X_{t'})^T \gamma\} K_{t,t'}, \quad (3-4)$$

其中 $K_{t,t'} = K_h((X_t - X_{t'})^T \gamma) / \sum_{t=1}^T K_h((X_t - X_{t'})^T \gamma)$ ， $K_h(\cdot) = K(\cdot/h)/h$ 。通过(3-1)，

(3-2)，(3-4)， γ_0 的期望分位数回归估计是：

$$\hat{\gamma} = \arg \min_{\|\gamma\|=1, \gamma_1 > 0} \sum_{t'=1}^T \sum_{t=1}^T \rho_\tau\{Y_t - a_{t'} - b_{t'}(X_t - X_{t'})^T \gamma\} K_{t,t'}. \quad (3-5)$$

由于在(3-5)中, $a_{t'}$ 和 $b_{t'}$, $t' = 1, \dots, T$ 是未知的, (3-4)的最小化应该由两个简单问题的迭代求解得到, 一个是关于 $a_{t'}$ 和 $b_{t'}$, $t' = 1, \dots, T$, 另一个关于 γ 。 γ_0 的期望分位数回归估计步骤如下:

初始化步骤: 根据 Hardle & Xia(2006)^[42]所述, 我们通过最小均方估计 MAVE 获得初始估计 $\hat{\gamma}^0$, 即 $\tau=0.5$ 时的 ALS 估计。

第一步: 给定 $\hat{\gamma}^0$, 通过求解下面的式子获得 $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^T$,

$$\min_{(a_{t'}, b_{t'})} \sum_{t=1}^T \rho_{\tau}\{Y_t - a_{t'} - b_{t'}(X_t - Y_{t'})^T \hat{\gamma}^0\} K_h((X_t - Y_{t'})^T \hat{\gamma}^0), \quad (3-6)$$

其中 $K(\cdot)$ 为核心函数, h 为带宽, $K_h(\cdot) = K(\cdot/h)/h$ 。

第二步: 给定步骤一中的 $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^T$, 通过求解下面的式子获得 $\hat{\gamma}$,

$$\min_{\|\gamma\|=1, \gamma_1 > 0} \sum_{t'=1}^T \sum_{t=1}^T \rho_{\tau}\{Y_t - \hat{a}_{t'} - \hat{b}_{t'}(X_t - Y_{t'})^T \gamma\} K_{t,t'}, \quad (3-7)$$

其中 $K_{t,t'} = K_h((X_t - X_{t'})^T \gamma) / \sum_{t=1}^T K_h((X_t - X_{t'})^T \gamma)$ 由 $\hat{\gamma}^0$ 和 h 所估计。

第三步: 对第一步和第二步不断进行重复, 直到收敛。

3.1.2 Adaptive LASSO 下的惩罚非对称最小二乘估计

Adaptive LASSO 可以看做是 LASSO 惩罚的一种推广。基本上, 这个想法是通过使用自适应权值来惩罚不同的水平上的不同协变量的系数。然而, 很容易看到, 指标参数的估计过程(3-7)是一个计算非常复杂繁琐的过程, 计算代价非常高, 这一过程涉及到最小化 T^2 项之和。而 Liu & Yang(2017)^[43]中所提出的算法过程可以将 T^2 项简化为 T 项, 极大地简化了计算过程, 提高了计算效率。因此, 我们使用 Liu & Yang(2017)中介绍的程序, 它由 $\hat{g}(X_t^T \gamma | \gamma)$ 在步骤二中的 γ_0 附近的 \sqrt{T} 一致估计 $\hat{\gamma}$ 使用局部线性逼近构成。有下式:

$$\hat{g}(X_t^T \gamma | \gamma) \approx \hat{g}(X_t^T \gamma | \gamma) + \hat{g}'(X_t^T \gamma | \gamma) X_t^T (\gamma - \hat{\gamma}),$$

其中 $\hat{g}(X_t^T \gamma | \gamma)$ 和 $\hat{g}'(X_t^T \gamma | \gamma)$ 可以通过最小化(3-6)式得到。之后，我们通过最小化下面的式子来估计 γ_0 ：

$$\min_{\gamma} \sum_{t=1}^T \rho_{\tau}\{Y_t - \hat{g}(X_t^T \hat{\gamma} | \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma} | \hat{\gamma}) X_t^T (Y - \hat{\gamma})\}.$$

因此，模型(1-1)的 Adaptive LASSO 惩罚最小二乘法(penalized asymmetric least squares, PALS)估计，记为 $\hat{\gamma}^P$ ，即下面式子的最小值

$$R_T(\gamma) = \sum_{t=1}^T \rho_{\tau}\{Y_t - \hat{g}(X_t^T \gamma | \gamma) - \hat{g}'(X_t^T \gamma | \gamma) X_t^T (Y - \gamma)\} + \lambda_T \sum_{j=1}^p v_j |\gamma_j|. \quad (3-8)$$

由于 Adaptive LASSO 的有限样本性能很大程度上依赖于 v 和 λ_T ，因此权重向量 $v = (v_1, \dots, v_p)$ 和调优参数 λ_T 的选择就至关重要。在 v 的选择上，根据 Zou^[33](2006)中所建议，我们选择 $v_j = |\hat{\gamma}_j^{-1}|$ ， $j = 1, \dots, p$ ，它能够满足所有的理论性质，并在研究中能够产生良好的有限样本性能。对于 λ_T 的选择，我们使用贝叶斯信息准则(BIC)：

$$\text{BIC}(\lambda_T) = \log(T^{-1} \sum_{t=1}^T \rho_{\tau}\{Y_t - \hat{g}(X_t^T \hat{\gamma}^P | \hat{\gamma}^P)\}) + \frac{\log T}{T} \text{df}(\lambda_T),$$

其中 $\text{df}(\lambda_T)$ 是 $\hat{\gamma}^P$ 中所有非零系数的总和。我们可以选择 $\hat{\lambda}_T = \arg \min_{\lambda_T} \text{BIC}(\lambda_T)$ 。

3.1.3 PALS 估计的渐近性质和 Oracle 属性

为了证明 $\hat{\gamma}^P$ 的渐近性质，我们定义 $s_1 = \max\{v_j, j \in \Lambda\}$ ， $s_2 = \min\{v_j, j \in \Lambda^c\}$ ，其中 $\Lambda = \{j: \gamma_{0,j} \neq 0\}$ ， Λ^c 是补集。 s_1 是显著系数的最大惩罚， s_2 是不显著系数的最小惩罚。下面的两个定理证明了 PALS 估计的一致性和 Oracle 属性。

Theorem 3.1(一致性): 假设附录中的(C1)-(C5)均成立，并且 $\lambda_T s_1 / \sqrt{T} = o(1)$ ，那么

$$\hat{\gamma}^P - \gamma_0 = O_p\left(T^{-\frac{1}{2}}\right).$$

Theorem 3.2(Oracle 性质): 假设附录中的(C1)-(C5)均成立，如果 $\lambda_T / \sqrt{T} \rightarrow 0$ ， $s_1 = O(1)$ ， $s_2 \lambda_T / \sqrt{T} \rightarrow \infty$ ， $T \rightarrow \infty$ ， $h \rightarrow 0$ 以及 $Th^2 \rightarrow \infty$ ，那么

(i)选择的一致性:

$$P(\{j: \hat{\gamma}_j^P \neq 0\} = \Lambda) \rightarrow 1,$$

(ii)渐近正态性:

$$\sqrt{T}(\hat{\gamma}_\Lambda^P - \gamma_{0,\Lambda}) \xrightarrow{L} N\left(0, \{\tilde{S}^{-1}\tilde{\Sigma}\tilde{S}^{-1}\}_{\Lambda\Lambda}\right),$$

其中 \xrightarrow{L} 表示分布中的收敛, \tilde{S}^{-1} 是 \tilde{S} 的广义逆, 且

$$\tilde{S} = 2E\{\tau - I\{\gamma \leq g_0(X^T\gamma_0)\} | \{g'_0(X^T\gamma_0)\}^2 \tilde{X}\tilde{X}^T\},$$

$$\tilde{\Sigma} = 4E\left[\left\{|\tau - I\{\gamma \leq g_0(X^T\gamma_0)\}|(Y - g_0(X^T\gamma_0))g'_0(X^T\gamma_0)\right\}^2 \tilde{X}\tilde{X}^T\right],$$

其中 $\tilde{X} = X - E(X|X^T\gamma_0)$ 。

3.2 纠偏估计

像 LASSO 和其他正则化估计一样, PALS 是有偏的, 并且不具有易于处理的极限分布。所以我们需要对 PALS 进行纠偏, 以便于做统计推断。

我们首先设计一个基于期望分位数损失的纠偏估计器, 使其能够构造稳健的置信区间, 并且在高维的情况下能检验假设。实际上, 公式(3-8)中 $\hat{\gamma}^P$ 的优化问题的一阶条件是

$$-2 \sum_{t=1}^T \omega_\tau(\tilde{e}_t^*) \hat{g}'(X_t^T \gamma | \gamma) X_t \{Y_t - \hat{g}(X_t^T \gamma | \gamma) - \hat{g}'(X_t^T \gamma | \gamma) X_t^T (\hat{\gamma}^P - \hat{\gamma})\} + \lambda_T H = 0,$$

其中 $H = (u_1 \text{sign}(\hat{\gamma}_1^P), \dots, u_p \text{sign}(\hat{\gamma}_p^P))^T$, $\tilde{e}_t^* = Y_t - \hat{g}(X_t^T \gamma | \gamma) - \hat{g}'(X_t^T \gamma | \gamma) X_t^T (\hat{\gamma}^P - \hat{\gamma})$,

之后, 我们可以得到

$$\begin{aligned} & \hat{\gamma}^P - \gamma_0 + \frac{1}{2T} \bar{S}^{-1} \lambda_T H \\ & = \bar{S}^{-1} \frac{1}{T} \sum_{t=1}^T \omega_\tau(\tilde{e}_t^*) \hat{g}'(X_t^T \gamma | \gamma) X_t \{Y_t - \hat{g}(X_t^T \gamma | \gamma) - \hat{g}'(X_t^T \gamma | \gamma) X_t^T (\gamma_0 - \hat{\gamma})\}, \end{aligned} \quad (3-9)$$

其中 $\bar{S}^{-1} = T^{-1} \sum_{t=1}^T \omega_{\tau}(\tilde{\epsilon}_t^*) \{\hat{g}'(X_t^T \hat{\gamma} | \hat{\gamma})\}^2 X_t X_t^T$ 。(3-9)式中, $\hat{\gamma}^P$ 的偏倚主要由 $\bar{S}^{-1} \lambda_T H / (2T)$ 决定。因为根据大样本定律, (3-9)式的右项趋于 0。如果 $\max_{1 \leq j \leq p} \lambda_j \lambda_T / T$ 距离 0 很远, 那么我们在计算时就不能忽略 $\bar{S}^{-1} \lambda_T H / (2T)$ 。在这种情况下, 尽管我们可以用 $\hat{\gamma}^P$ 去做系数的统计推断, 但是不能保证推断的精度, 容易造成偏差。为了克服这一缺点, 一个直接而有效的方法就是对 $\hat{\gamma}^P$ 的估计进行修正。我们采用 van de Geer et al.^[40] (2014) 所提出的想法, 在(3-9)式的基础上, 我们有一个校正估计来对 $\hat{\gamma}^P$ 进行校正, 得到

$$\begin{aligned} \hat{\gamma}^{AP} &= \hat{\gamma}^P + \frac{1}{2T} \bar{S}^{-1} \lambda_T H \\ &= \bar{S}^{-1} \frac{1}{T} \sum_{t=1}^T \omega_{\tau}(\tilde{\epsilon}_t^*) \hat{g}'(X_t^T \gamma | \gamma) X_t \{Y_t - \hat{g}(X_t^T \gamma | \gamma) + \hat{g}'(X_t^T \gamma | \gamma) X_t^T \hat{\gamma}\}. \end{aligned} \quad (3-10)$$

下面的定理证明了 $\hat{\gamma}^{AP}$ 是渐近正态分布, 所以我们可以基于 $\hat{\gamma}^{AP}$ 做统计推断。

Theorem 3.3: 在定理 3.2 成立的情况下, 有

$$\sqrt{T}(\hat{\gamma}^{AP} - \gamma_0) \xrightarrow{L} N(0, \tilde{S}^{-1} \tilde{\Sigma} \tilde{S}^{-1}).$$

注意到 $\hat{\gamma}^{AP}$ 和 $\hat{\gamma}_{\Lambda}^P$ 的渐近方差是一样的, 所以相较于 $\hat{\gamma}_{\Lambda}^P$, 我们的纠偏 PALS 估计 $\hat{\gamma}^{AP}$ 并没有失去原有的效率。 $\hat{\gamma}^{AP}$ 最大的贡献之一是 $\hat{\gamma}_{\Lambda}^P$ 的统计推断可以在 $\hat{\gamma}^{AP}$ 的基础上建立, 但是在 $\hat{\gamma}^P$ 的基础上不行。

3.3 参数的统计推断

根据前一节所述, 我们可知 $\hat{\gamma}^{AP}$ 具有良好的无偏性质, 以及根据定理 3.3 我们可以得到 $\hat{\gamma}^{AP}$ 服从渐近正态分布, 于是接下来我们可以对单指标期望分位数回归模型中的未知参数 γ_0 做出统计推断。根据 Jiang, Hu & Yu(2020)^[1] 的 2.3 部分, $\tilde{S}^{*-1} \tilde{\Sigma}^* \tilde{S}^{*-1}$ 是 $\tilde{S}^{-1} \tilde{\Sigma} \tilde{S}^{-1}$ 的一致估计量。其中

$$\tilde{S}^* = 2T^{-1} \sum_{t'=1}^T \sum_{t=1}^T |\tau - I\{Y_t \leq \hat{g}(X_t^T \hat{\gamma})\}| \hat{g}'^2(X_{t'}^T \hat{\gamma}) (X_t - X_{t'}) (X_t - X_{t'})^T \tilde{R}_{tt'},$$

$$\tilde{\Sigma}^* = 4T^{-1} \sum_{t'=1}^T \sum_{t=1}^T \left[\left[\tau - I\{Y_t \leq \hat{g}(X_t^T \hat{\gamma})\} \right] \{Y_t - \hat{g}(X_t^T \gamma)\} \hat{g}'(X_{t'}^T \hat{\gamma}) \hat{R}_{t,t'} \right]^2 \cdot (X_t - X_{t'})(X_t - X_{t'})^T.$$

随后, $\gamma_{0,j}$, $j \in \{1, \dots, p\}$ 的 $1 - \alpha$ 置信区间被定义为

$$CI_j = \left[\hat{\gamma}_j^{AP} - \frac{\hat{\sigma}_{jj}}{\sqrt{T}} \phi^{-1}(1 - \alpha/2), \hat{\gamma}_j^{AP} + \frac{\hat{\sigma}_{jj}}{\sqrt{T}} \phi^{-1}(1 - \alpha/2) \right],$$

其中 $\hat{\sigma}_{jj}(\tilde{S}^{*-1} \tilde{\Sigma}^* \tilde{S}^{*-1})_{jj}^{1/2}$ 和 $\phi(\cdot)$ 是标准正态累计分布函数。定理 3.3 也可以用来检验下面变量的显著性, 对于 $\gamma_{0,j}$, $j \in \{1, \dots, p\}$, 我们假设检验

$$H_{0,j}: \gamma_{0,j} = 0, \quad H_{1,j}: \gamma_{0,j} \neq 0$$

我们给检验 $H_{0,j}$ 构造 P 值 $P_j = 2\{1 - \phi(\sqrt{T}|\hat{\gamma}_j^{AP}|/\hat{\sigma}_{jj})\}$ 。该 P 值基础上的决策规则是

$$T_j = \begin{cases} 1, & \text{if } P_j \leq \alpha \quad (\text{reject } H_{0,j}) \\ 0, & \text{otherwise (accept } H_{0,j}) \end{cases} \quad (3-11)$$

α 是显著性水平。

3.4 本章小结

第三章具体介绍了本文所研究的单指标期望分位数回归模型的求解过程。

在本章第一节中, 首先介绍了单指标期望分位数回归模型中变量 γ_0 的具体估计步骤。Newey 和 Powell (1987)^[2] 在介绍期望分位数回归模型时给出了模型的求解方法即非对称最小二乘估计法, 在本文中我们用同样的方法的求解单指标期望分位数回归模型。但是在该方法中, 指标参数的估计过程是一个计算代价非常高的过程, 涉及到最小化 T^2 项之和。所以接下来我们运用 Adaptive LASSO 的方法对其进行改进。参考 Liu & Yang(2017)^[43] 所介绍的 Adaptive LASSO 过程, 将 T^2 项简化为 T 项。因此权重向量 $\mathbf{v} = (v_1, \dots, v_p)$ 和调优参数 λ_T 的选择非常重要。在 \mathbf{v} 的选择上, 并分别根据 Zou^[33](2006)

的建议和贝叶斯信息准则(BIC), 对 Adaptive LASSO 下的权重向量 $\mathbf{v} = (v_1, \dots, v_p)$ 和调优参数 λ_T 进行选择, 确保 Adaptive LASSO 下的模型的良好性能。随后给出了在 Adaptive LASSO 下的惩罚非对称最小二乘(PALS)估计具有的良好的渐近性质和 Oracle 属性。

在第二节中, 我们主要介绍了 PALS 的纠偏算法。在第一节里面我们已经证明了 PALS 估计下的 $\hat{\gamma}^P$ 具有的良好统计特性, 但是仅仅如此并不能直接将 $\hat{\gamma}^P$ 用于统计推断, 一方面是因为它不具有易于处理的极限分布, 另一方面是因为 $\hat{\gamma}^P$ 是有偏的。因此, 我们要在 Adaptive LASSO 惩罚下进一步对 $\hat{\gamma}^P$ 进行纠偏估计。首先我们设计一个基于期望分位数损失的纠偏估计器, 使其能够构造稳健的置信区间, 并且在高维的情况下能检验假设。随后采用 van de Geer et al.^[40] (2014) 提出的想法, 在 $\hat{\gamma}^P$ 的优化问题的一阶条件下, 得到了 $\hat{\gamma}^P$ 校正估计 $\hat{\gamma}^{AP}$, 并证明 $\hat{\gamma}^{AP}$ 具有渐近正态分布, 所以我们可以基于 $\hat{\gamma}^{AP}$ 做统计推断。

第三节就是基于上述结论, 给出了单指标期望分位数回归模型中未知参数 γ_0 的统计推断。包括其估计的置信区间和假设检验。

在下一章里面, 将在蒙特卡罗模拟实验和真实数据集中运用所提出的求解的步骤, 来说明所提出的程序具有可操作性。

第四章 模拟实验

在本章中，我们首先使用蒙特卡罗模拟来评估所提出程序的有限样本性能，然后用波士顿住房数据演示所提出方法的应用。本章利用 R 语言进行程序编写。

4.1 数值模拟

在本节中，我们使用蒙特卡罗模拟实验来评估所提出的 PALS 估计方法在高维度样本的情况下的有限样本性能，包括上一章中构建的置信区间和假设检验的性能。

4.1.1 数据生成

该模拟研究的数据是从以下两个模型生成的：

模型一：

$$Y_t = 5 \sin(\pi X_t^T \gamma_0) + \{1 + (X_t^T \gamma_0)^2\} \varepsilon_t, t = 1, \dots, T. \quad (4-1)$$

其中 $\gamma_0 = (\gamma_1, \dots, \gamma_{10})^T$, $(\gamma_1, \gamma_2) = (0.8, 0.6)$, 其它的 $\gamma_j = 0, j = 3, \dots, 10$ 。协变量向量 $X_t = 0.4X_{t-1} - 0.5X_{t-2} + \varepsilon_t^*$, $\varepsilon_t^*(t = 1, \dots, T)$ 是区间 $[-1, 1]$ 上的均匀分布。

模型二：

$$Y_t = (X_t^T \gamma_0)^2 + \varepsilon_t, t = 1, \dots, T. \quad (4-2)$$

其中 $\gamma_0 = (\gamma_1, \dots, \gamma_{100})^T$, $(\gamma_1, \dots, \gamma_{10}) = (1, \dots, 1)/\sqrt{10}$ 并且其他的 $\gamma_j = 0, j = 11, \dots, 100$ 。协变量向量 $X_t = 0.5X_{t-1} + \varepsilon_t^*$, $\varepsilon_t^* \sim N(2, 1)$ 。

本文考虑两种不同的误差分布：标准正态分布 $N(0, 1)$ 和自由度为 5 的 t 分布 $t(5)$ 。我们对每个设置进行 100 次模拟，样本大小为 200。

4.1.2 评价指标

为了评估这些方法的性能，采取以下评估指标：

1、MSE：模型求解得到的估计参数 $\hat{\gamma}^P$ 的预测值和设定的真实参数 γ_0 之间的均方误差：

$$\sqrt{(\hat{\gamma}^P - \gamma_0)^T (\hat{\gamma}^P - \gamma_0)}.$$

2、RMSE：估计参数 $\hat{\gamma}^P$ 相对于未经过 Adaptive LASSO 惩罚所得到的估计 $\hat{\gamma}$ 的相对均方误差：

$$\sqrt{(\hat{\gamma}^P - \gamma_0)^T (\hat{\gamma}^P - \gamma_0)} / \sqrt{(\hat{\gamma} - \gamma_0)^T (\hat{\gamma} - \gamma_0)}.$$

3、False variables：被包括在模型中的错误的非零协变量。

4、True variables：被包括在模型中的正确的非零协变量。

4.1.3 模拟结果

表1和表2总结了不同误差和 $\tau = 0.10、0.25、0.50、0.75、0.90$ 期望水平下的模型的MSE、RMSE、False variables和True variables的模拟结果。考虑到MSE和RMSE两项，系数估计非常好，并且 $\hat{\gamma}^P$ 的性能优于 $\hat{\gamma}$ 的性能。此外，我们可以从表4-1和表4-2中看到True variables分别等于2和10，因此所有信号变量都被选择了，这意味着我们选择的模型包含基础真实模型。模型也可能会以某种方式选择一些噪声变量，这个缺点在变量选择中很常见，目前还没有被克服。然而，我们的模拟结果中False variables的都很小。

表 4-1. 模型 1 在不同误差和期望水平下的模拟结果

Table 4-1 The simulation results under different errors and expectile levels for model 1.

Error	τ	MSE	RMSE	False variables	True variables
N(0,1)	0.10	0.0183 (0.0150)	0.3413 (0.2513)	0.3400 (0.8435)	2.0000
	0.25	0.0146 (0.0112)	0.3013 (0.2167)	0.1800 (0.5752)	2.0000
	0.50	0.0172 (0.0131)	0.3405 (0.2481)	0.2900 (0.7693)	2.0000
	0.75	0.0157 (0.0130)	0.3338 (0.2653)	0.2500 (0.7961)	2.0000
	0.90	0.0178 (0.0139)	0.3195 (0.2114)	0.1900 (0.5449)	2.0000
t(5)	0.10	0.0220 (0.0148)	0.3465 (0.2215)	0.2600 (0.6908)	2.0000
	0.25	0.0300 (0.0863)	0.3642 (0.2413)	0.3700 (0.9173)	2.0000
	0.50	0.0210 (0.0173)	0.3564 (0.2660)	0.3600 (0.9301)	2.0000
	0.75	0.0233 (0.0218)	0.3768 (0.2940)	0.3600 (0.9590)	2.0000
	0.90	0.0231 (0.0182)	0.3617 (0.2589)	0.2600 (0.6296)	2.0000

表 4-2. 模型 2 在不同误差和期望水平下的模拟结果

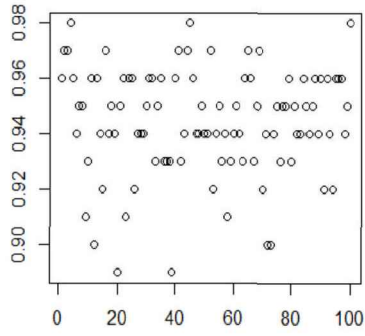
Table 4-2 The simulation results under different errors and expectile levels for model 2.

Error	τ	MSE	RMSE	False variables	True variables
N(0,1)	0.10	0.0276 (0.0280)	0.5591 (0.1301)	1.7000 (3.2240)	10.0000
	0.25	0.0225 (0.0185)	0.5226 (0.1276)	1.0800 (1.8677)	10.0000
	0.50	0.0211 (0.0197)	0.5035 (0.1182)	1.0500 (2.4675)	10.0000
	0.75	0.0205 (0.0164)	0.5034 (0.1195)	0.9400 (1.4827)	10.0000
	0.90	0.0322 (0.0352)	0.5722 (0.1349)	1.8600 (2.8427)	10.0000
t(5)	0.10	0.0320 (0.0267)	0.5768 (0.1271)	2.2700 (3.3989)	10.0000
	0.25	0.0315 (0.0307)	0.5476 (0.1330)	1.8100 (2.6579)	10.0000
	0.50	0.0230 (0.0164)	0.5001 (0.1159)	1.0900 (1.4914)	10.0000
	0.75	0.0250 (0.0169)	0.5189 (0.1241)	1.3000 (1.8340)	10.0000
	0.90	0.0270 (0.0206)	0.5734 (0.1164)	1.6900 (2.5013)	10.0000

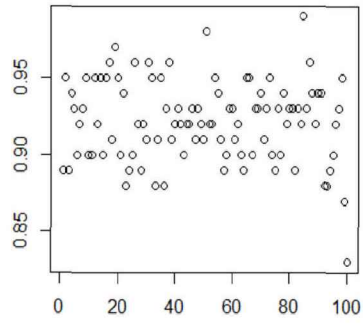
接下来，我们提供模拟来说明置信区间的有限样本属性。表 4-3 和图 4-1 显示了 95% 水平下的不同模型建议区间的覆盖概率。从表 4-3 和图 4-1 可以看出，我们提出的方法对不同的误差、期望水平和模型都有效。

表 4-3. 模型 1 在不同误差和期望下 95%水平的平均覆盖率
 Table 4-3 The mean coverage probability at level 95% under different errors and expectile levels for model 1.

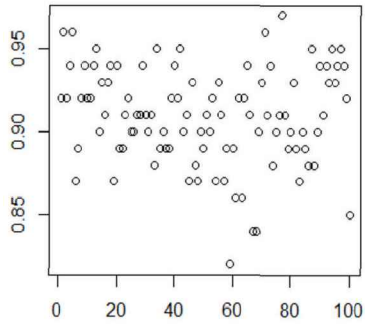
$\tau =$	N(0,1)					t(5)				
	0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90
$\gamma_{0,1}$	0.96	0.99	0.99	0.98	0.98	0.95	0.97	0.97	0.96	0.93
$\gamma_{0,2}$	0.93	0.97	0.97	0.96	0.90	0.92	0.92	0.94	0.92	0.87
$\gamma_{0,3}$	0.84	0.92	0.84	0.93	0.82	0.81	0.88	0.88	0.84	0.80
$\gamma_{0,4}$	0.88	0.89	0.86	0.88	0.82	0.82	0.85	0.90	0.88	0.84
$\gamma_{0,5}$	0.87	0.89	0.93	0.86	0.81	0.85	0.83	0.86	0.91	0.80
$\gamma_{0,6}$	0.82	0.91	0.89	0.93	0.96	0.83	0.84	0.88	0.90	0.78
$\gamma_{0,7}$	0.89	0.89	0.90	0.86	0.81	0.80	0.89	0.86	0.85	0.79
$\gamma_{0,8}$	0.86	0.85	0.93	0.93	0.81	0.80	0.88	0.88	0.87	0.81
$\gamma_{0,9}$	0.89	0.84	0.90	0.87	0.83	0.80	0.89	0.87	0.85	0.84
$\gamma_{0,10}$	0.82	0.89	0.92	0.94	0.84	0.81	0.89	0.90	0.87	0.81



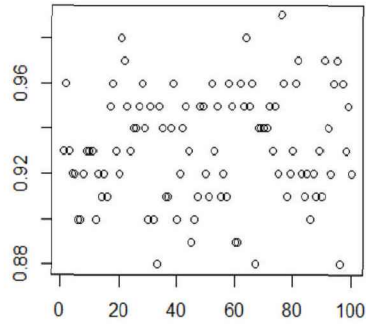
$\tau = 0.1$ and $\varepsilon_t \sim N(0,1)$



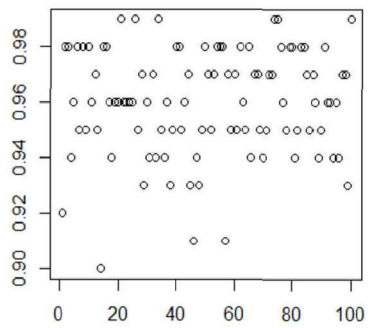
$\tau = 0.25$ and $\varepsilon_t \sim N(0,1)$



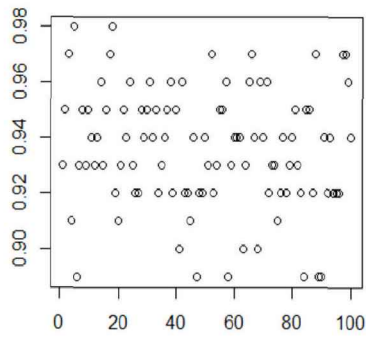
$\tau = 0.5$ and $\varepsilon_t \sim N(0,1)$



$\tau = 0.75$ and $\varepsilon_t \sim N(0,1)$



$\tau = 0.9$ and $\varepsilon_t \sim N(0,1)$



$\tau = 0.1$ and $\varepsilon_t \sim t(5)$

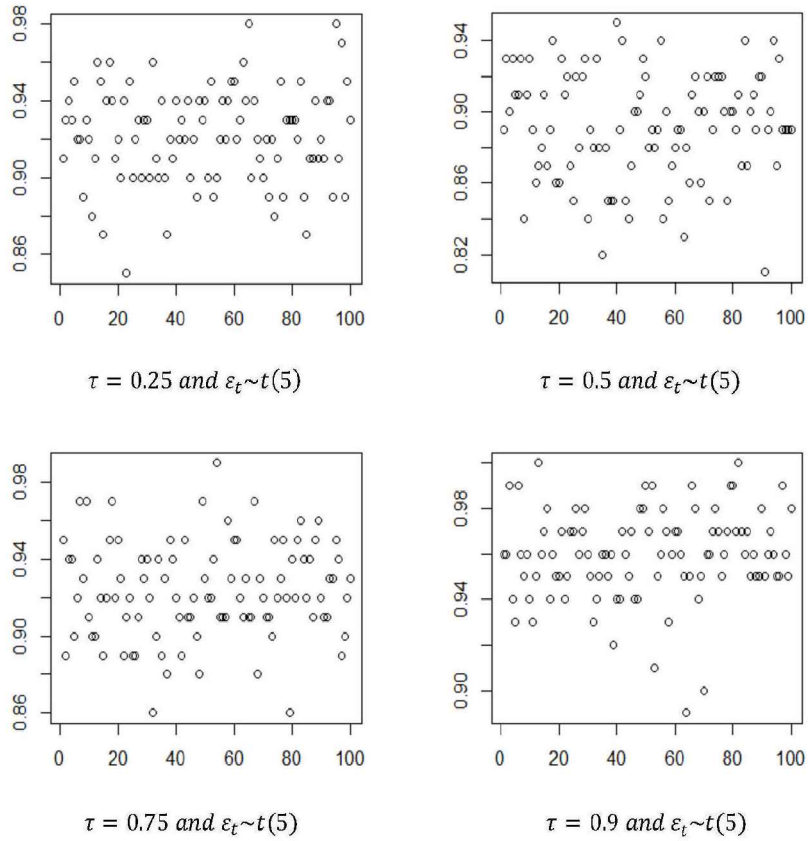


图 4-1.模型 2 在不同误差和期望下 95%水平的平均覆盖率

Fig 4-1 The mean coverage probability at level 95% under different errors and expectile levels for model 2.

4.2 实证分析

4.2.1 数据描述

本文将建议的方法应用于波士顿住房数据。该数据包含对 14 个变量的 506 个观察值，可在 R 中的 MASS 库中找到。输出变量是 MEDV，即自住房以千美元计的平均房价。13 个输入变量是 1970 年人口普查中波士顿郊区 506 个人口普查区的统计测量值，具体如下：

表 4-4. 波士顿住房数据描述
Table 4-4 Boston housing data

序号	变量名	描述
1	RM	住宅平均房间数量
2	CRIM	城镇人均犯罪率
3	ZN	住宅用地超过 25000 英尺的比例
4	INDUS	城镇非零售商用土地比例
5	NOX	一氧化氮浓度
6	AGE	1940 年之前建成的自用房屋比例
7	DIS	到波士顿五个中心区域的加权距离
8	RAD	辐射性公路的接近指数
9	TAX	每 10000 美元的全值财产税率
10	PTRATIO	城镇师生比例
11	B	$1000(\text{BK}-0.63)^2$, 其中 BK 代指城镇中黑人比例
12	LSTAT	人口中地位低下者比例
13	CHAS	查尔斯河虚拟变量

由于变量 CHAS 是分类变量，但我们的方法不适用于分类数据，因此我们使用表 4-4 中的其他 12 个变量进行实验。我们对数据进行标准化并利用单指标期望分位数回归模型(1)在五个不同期望水平 $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$ 下拟合数据。

4.2.2 实验结果

我们首先应用 Jiang, Hu & Yu (2020)^[1] 提出的期望回归模型来分析这些数据。估计的第 10、25、50、75、90 期望值及其 95% 逐点置信区间与 MEDV 和估计指数的散点图一起显示在下图中。这些图表明估计的条件期望函数提供了对数据的良好拟合。

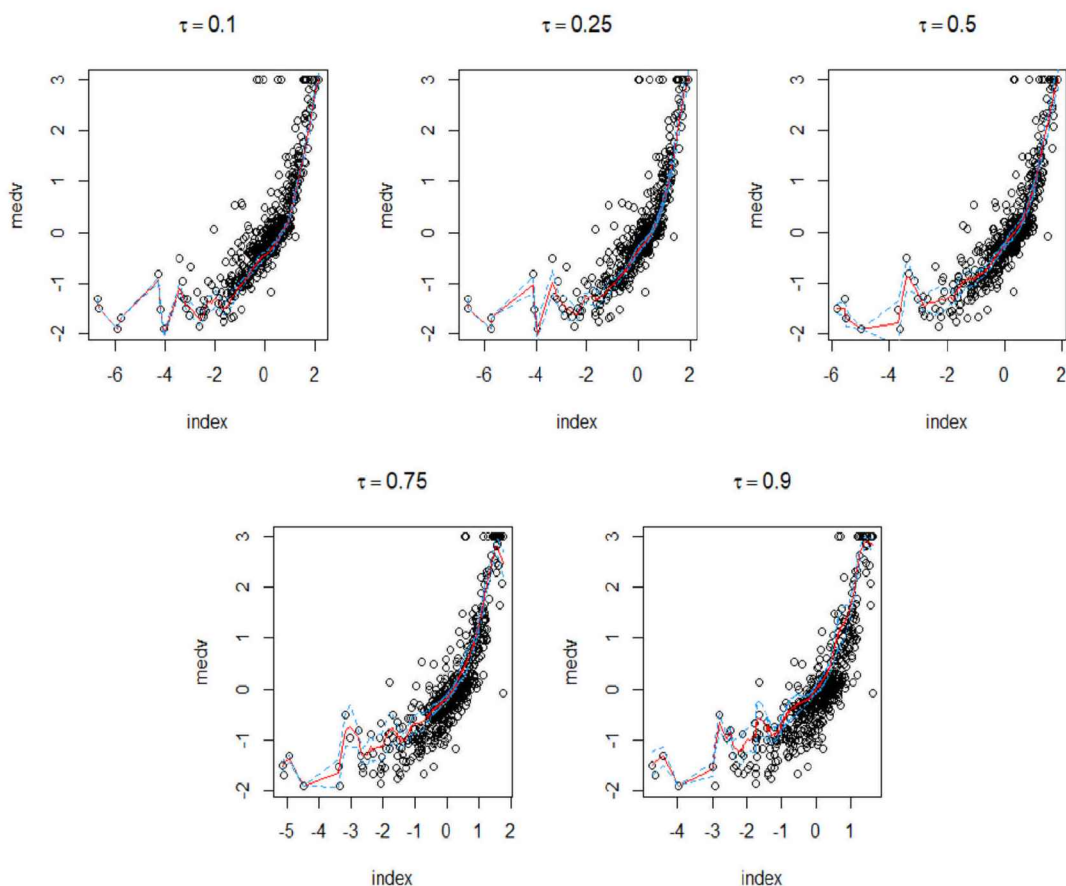


图 4-2. 波士顿住房数据单指数预期回归估计
Fig 4-2 Estimated single index expectile regression for Boston housing data

接下来，使用第三章第一节中提出的变量选择方法和第三节中 5% 显著性水平下的显著性检验(3-11)来选择重要变量。结果如图 4-3 所示。我们可以发现在不同的期望水平下，不同的变量选择方法和检测方法都选择了 RM、CRIM、RAD、PTRATIO 和 LSTAT。进一步分析下：

(1)对于 $\tau = 0.10$ ，显著性检验方法下的所有变量均显著，变量 RM、CRIM、NOX、RAD、TAX、PTRATIO、LSTAT 通过了变量选择。

(2)对于 $\tau = 0.25$ ，两种方法的不同之处在于变量 NOX 和 AGE 被变量选择模型选择了，而变量 B 在显著性检验模型下显著。

(3)对于 $\tau = 0.50$, 显著性检验模型比变量选择模型多选了两个变量, 分别是 NOX 和 B。

(4)对于 $\tau = 0.75$, 变量选择模型比显著性检验模型多选了两个变量, 分别是 ZN 和 NOX。

(5)对于 $\tau = 0.90$, 两种方法的不同之处在于变量 TAX 被变量选择模型选择了, 而变量 ZN 在显著性检验模型下显著。

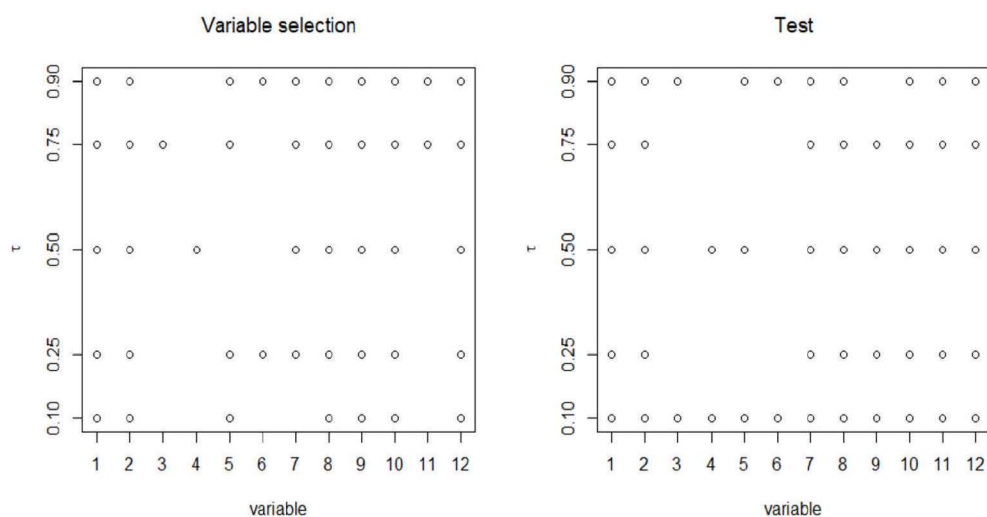


图 4-3. 波士顿住房数据的变量选择和检验 (原点表示被选择的变量)

Fig 4-3 Variable selection and Test for Boston housing data. The dots are the variables being selected

通过以上分析, 我们可以看出两种方法选择的变量并不完全相同。在实践中, 我们可以利用两种方法得到的结果, 并考虑变量的实际背景, 来得到变量是否需要在模型中。

4.3 本章小结

在本章节中,通过蒙特卡罗方法进行数值模拟评估所提出的 PALS 估计器在高维度样本的情况下的有限样本性能以及构建的置信区间和假设检验的性能,并利用波士顿房价真实数据集进行变量选择演示所提出的方法。

在数值模拟实验部分,利用两种不同形式的单指标期望分位数回归模型随机生成两组误差服从不同分布的数据进行模拟,比较了模型求解得到的估计参数 $\hat{\gamma}^P$ 的预测值和设定的真实参数 γ_0 之间的均方误差 MSE、估计参数 $\hat{\gamma}^P$ 相对于未经过 Adaptive LASSO 惩罚所得到的估计 $\hat{\gamma}$ 的相对均方误差 RMSE。通过比对结果得出系数估计得非常好,并且 $\hat{\gamma}^P$ 的性能优于 $\hat{\gamma}$ 的性能。在实证分析部分,使用提出的变量选择方法和显著性检验方法来选择波士顿房价数据集中的重要变量,并对二者进行比较分析。

总结上述实验结果,本文所提出的单指标期望分位数回归模型算法具有良好的拟合精度,在面对不同的误差分布、期望水平、模型时仍旧表现稳定。在对真实数据集进行变量选择时,可结合真实数据背景,结合两种途径筛选出重要变量。

第五章 总结与展望

5.1 总结

本文提出了一种单指标期望分位数回归模型来解决高维数据下的单指标模型参数估计问题。我们提出了一种惩罚 ALS 方法(PALS)来选择重要变量,并开发了一种纠偏方法来构建单指标期望分位数回归模型的有效置信区间和假设检验。

本文所构建的 PALS 估计量具有良好的渐近正态性和 Oracle 性质。对 PALS 估计进行纠偏是本文一个重要的步骤,纠偏后的 PALS 估计具有易于处理的极限分布,可对其进行统计推断。我们首先设计了一个基于期望分位数损失的纠偏估计器,使其能够构造稳健的置信区间,并且在高维的情况下能检验假设。但此时得到的估计量在进行统计推断时不能保证精度,容易造成偏差,因此我们对其进行修正。修正后的估计量在保证统计推断精度的同时仍不失计算效率。最后我们给出了估计量的置信区间和假设检验。

通过数值模拟以及实证分析,证明了本文所提出的方法具有良好的拟合精度,在面对不同的误差分布、期望水平、模型时仍旧表现稳定,在实际应用中结合数据真实背景,可准确筛选出重要变量。

5.2 不足与展望

一、目前本文对单指标期望回归模型进行变量选择时,仅仅研究了基于 Adaptive LASSO 惩罚函数下的变量选择与参数估计,但是惩罚函数有很多,例如 Group LASSO、SCAD 等都具有各自的优势。未来可以进一步利用其它类型的惩罚函数来进行变量选择,对比不同惩罚函数下的估计效率和参数,得到估计效率较高的最优参数估计。

二、在对单指标期望回归模型进行参数估计过程中,我们需要重复实验步骤直至模型的参数估计量收敛。本文在模拟实验中进行参数估计时,实验次数设置为 100 次,

两种模型的维度设置为 10 和 100，随着样本量和变量维度的增加，程序运行的时间也随之大大的增加。因此为了节约运行时间，在实验次数和维度的设置上具有一定的主观性。在未来的研究中，可以对实验重复次数、样本量以及维数的设置进行多次尝试，以期望在节省运行时间的同时来获得更好的结果。

三、本文在进行蒙特卡罗模拟时，仅仅模拟了正弦单指标期望分位数回归模型与指数单指标期望分位数回归模型，对于误差分布的选择也只选择了正态分布和 t 分布两种分布。因此在未来的研究中可以进一步探讨其他形式的单指标期望分位数回归模型，并选择不同的误差分布函数进行实验，以不失一般性。

四、在对波士顿住房数据进行实证分析时，我们剔除了一个分类变量，换言之就是我们研究的单指标期望分位数回归模型它的变量选择和参数估计不适用于分类变量。这在实际应用中有很多的局限性。未来可针对分类变量进一步完善模型，让其可以解决更多的实际问题。

参考文献

- [1] Jiang, R., Hu, X. & Yu, K. (2020). Single-index Expectile models for estimating conditional value at risk and expected shortfall[J]. *Journal of Financial Econometrics* doi:10.1093/jfinec/nbaa016.
- [2] Newey, W. & Powell, J. (1987). Asymmetric least squares estimation and testing[J]. *Econometrica* 55, 819–847.
- [3] Breckling, J. & Chambers R. (1988). M-quantiles[J]. *Biometrika* 75, 761–771.
- [4] Jones, M. (1994). Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters* 20, 149–153.
- [5] Ziegel, J. (2014). Coherence and elicibility[J]. *Mathematical Finance* 26, 901–918.
- [6] Lin, W. & Kulasekera, K. (2007). Identifiability of single-index models and additive-index models[J]. *Biometrika* 94, 496–501.
- [7] Friedman J, Stuetzle W. Projection pursuit regression[J]. *Publications of the American Statistical Association*, 1981, 76(376):817-823.
- [8] Ichimura H. Semiparametric least squares (SLS) and weighted SLS estimation of single index models[J]. *Papers*, 1993, 58(1-2):71-120.
- [9] Xia Y, Tong H, Li W, et al. An adaptive estimation of dimension reduction space[J]. *Journal of the Royal Statistical Society*, 2002, 64(3):363-410.
- [10] Hristache M, Spokoiny J. Direct estimation of the index coefficient in a single-index model[J]. *The Annals of Statistics*, 2001, 29(3):595-623.
- [11] Xia Y. Asymptotic Distributions for two estimators of the single-index model[J]. *Econometric Theory*, 2006, 22(6):1112-1137.
- [12] Zhu, L., Huang, M. and Li, R. (2012). Semiparametric quantile regression with high-dimensional covariates[J]. *Statistica Sinica*, 22, 1379-1401.

- [13] Wang, G. and Wang, L. Spline estimation and variable selection for single index prediction models with diverging number of index parameters[J]. *Journal of Statistical Planning and Inference*, 2015, 162: 1-19.
- [14] 刘阳. 通过 Adaptive LASSO 对单指标模型进行变量选择[D]. 郑州大学, 2018.
- [15] Liu H, Yang H, Peng C. Weighted composite quantile regression for single index model with missing covariates at random[J]. *Computational Statistics*, 2019, 34(4): 1711-1740.
- [16] Taylor J W. Estimating value at risk and expected shortfall using expectile[J]. *Journal of Financial Econometrics*, 2008, 6(2): 231-252.
- [17] Kuan C-M, Yeh J-H, Hsu Y-C. Assessing value at risk with CARE, the Conditional Autoregressive expectile models[J]. *Journal of Econometrics*, 2009, 150(2): 261-270.
- [18] 苏辛, 周勇. 条件自回归 期望分位数 模型及其在基金业绩评价中的应用[J]. *中国管理科学*, 2013, 56(06): 22-29.
- [19] 谢尚宇, 姚宏伟, 周勇. 基于 ARCH-期望分位数 方法的 VaR 和 ES 尾部风险测量 [J]. *中国管理科学*, 2014, 22(9): 1-9.
- [20] Xie, S. & Wan, A. (2014). A varying-coefficient expectile model for estimating value at risk[J]. *Journal of Business & Economic Statistics* 32, 576–592
- [21] Cai, Z., Fang, Y. & Tian, D. (2018). Assessing tail risk using expectile regressions with partially varying coefficients[J]. *Journal of Management Science and Engineering* 3, 183–213.
- [22] 胡宗义, 万闯, 李毅. 基于 Expectile 风险建模的原油价格风险测度研究[J]. *统计与信息论坛*, 2018, 23(1): 58-64.
- [23] Liao L, Park C, Choi H. Penalized Expectile regression: an alternative to penalized quantile regression[J]. *Annals of the Institute of Statistical Mathematics*, 2019, 71(2): 409-438.

-
- [24] Aigner D J, Poirier D J. On the estimation of production frontiers : maximum likelihood estimation of the parameters of a discontinuous density function[J]. *International Economic Review*, 1976, 45(12): 377-396.
- [25] YiYang, HuiZou. An Efficient Algorithm for Computing the HHSVM and Its Generalizations[J]. *Journal of Computational & Graphical Statistics*, 2013, 22(2): 396-415.
- [26] Gu, Y. & Zou, H. (2016). High-dimensional generalizations of asymmetric least squares regression and their applications[J]. *Annals of Statistics* 44, 2661–2694.
- [27] Zhao J, Yi Z. Variable Selection In Expetile Regression[J]. *Communication in Statistics- Theory and Methods*, 2017, 47(1): 85-96.
- [28] Liao, L., Park, C. & CHOI, H. (2018). Penalized Expetile regression: an alternative to penalized quantile regression[J]. *Annals of the Institute of Statistical Mathematics* 71, 1–30.
- [29] Zhao, J., Chen, Y. & Zhang, Y. (2018). Expetile regression for analyzing heteroscedasticity in high dimension[J]. *Statistics & Probability Letters* 237, 304–311.
- [30] Fan, Y., H?RDLE, W.K., Wang, W. & Zhu, L. (2018). Single-index-based covar with very high-dimensional covariates[J]. *Journal of Business & Economic Statistics* 36, 212–226.
- [31] Zou, H. (2006). The adaptive lasso ad its oracle properties[J]. *Journal of the American Statistical Association* 101, 1418–1429.
- [32] Tibshirani R. Regression shrinkage and selection via the lasso: A retrospective[J]. *Journal of the Royal Statistical Society*, 1996, 58(1):267-288.
- [33] Jianqing Fan,Runze Li. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties[J]. *Journal of the American Statistical Association*,2001,96(456).
- [34] Hui Zou,Trevor Hastie,and Robert Tibshirani (2006),Sparse Principal Component Analysis[J], *Journal of Computational and Graphical Statistic*, 15:2,265-268.
- [35] 施万锋,胡学钢,俞奎.一种面向高维数据的迭代式 Lasso 特征选择方法[J].*计算机应用研究*, 2011, 28(12):4463-4466.

- [36] 谢仪,高雪,景英川.基于 Lasso 及 Adaptive Lasso 的 AR(p)模型定阶及参数估计[J]. 浙江工业大学学报, 2014, 42(04):463-467.
- [37] 张婷婷,景英川.个人信用评分的 Adaptive Lasso-Logistic 回归分析[J]. 数学的实践与认识, 2016, 46(18):92-99.
- [38] Zhang, C.H. & Zhang, S. (2014). Confidence intervals for low-dimensional parameters with high dimensional data[J]. *Journal of the Royal Statistical Society: Series B* 76, 217–242.
- [39] Van De Geer, S., Bhlmann, P., Ritov, Y. & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models[J]. *Annals of Statistics* 42, 1166–1202.
- [40] Javanmard, A. & Montanari, A. (2014). Confidence intervals and hypothesis testing for high dimensional regression[J]. *Journal of Machine Learning Research* 15, 2869–2909.
- [41] Gueuning, T. & Claeskens, G. (2016). Confidence intervals for high-dimensional partially linear single index models[J]. *Journal of Multivariate Analysis* 149, 13–29.
- [42] Hardle, W.K. & Xia, Y. (2006). Semi-parametric estimation of partially linear single-index models[J]. *Journal of Multivariate Analysis* 97, 1162–1184.
- [43] Liu, H. & Yang, H. (2017). Estimation and variable selection in single-index composite quantile regression[J]. *Communications in Statistics - Simulation and Computation* 46, 7022–7039.

附录 A 渐近性质所需的条件

为了建立所提出的估计量的渐近性质，基于以下技术条件。

C1. 函数 $g_0(\cdot)$ 有连续且有界的二阶可导函数。

C2. 核 $K(\cdot) \geq 0$ 有一个紧致的支撑并且它的一阶导数是有界的。它满足 $\int_{-\infty}^{+\infty} K(u) du = 1$, $\int_{-\infty}^{+\infty} uK(u) du = 0$, $\int_{-\infty}^{+\infty} u^2 K(u) du < \infty$, $|\int_{-\infty}^{+\infty} u^j K^2(u) du| < \infty$, $j = 0, 1, 2$.

C3. 给定 u 下的 Y 的条件密度函数 $f_Y(y|u)$ 是连续的并且对于每一个 y 远离 0 和 ∞ 。

C4. $X^T \gamma$ 的密度函数在 γ_0 的邻域内对 γ 为正且一致连续。此外 $X^T \gamma_0$ 的密度函数 $f_{U_0}(\cdot)$ 是连续的并且在其支撑上远离 0 和 ∞ 。 $f_{U_0, U_1}(\cdot, \cdot)$ 是 (U_0, U_1) 的密度函数且 $f_{U_0, U_1}(\cdot, \cdot)$ 是有界的。

C5. $\{Y_t, X_t\}$ 上的所有过程是 α -混合的，因此对于一些 $\delta > 2$ 和 $c > 1 - 2/\delta$, $\sum_l l^c \{\alpha(l)\}^{1-2/\delta} \leq \infty$ 。更进一步，存在一个正整数序列 S_T , 有 $S_T \rightarrow \infty$, $S_T = o(\sqrt{Th})$, 并且 $\sqrt{T/h} \alpha(S_T) \rightarrow 0$, $T \rightarrow \infty$ 。

备注 A.1: 条件 C1-C5 为标准条件，常用于单指标模型和期望分位数模型，参见 Hardle & Xia (2006) 和 Jiang, Hu & Yu (2020)。

附录 B 定理证明

定理 3.1 证明：令 $\{\gamma_0 + \gamma/\sqrt{T} : \|\gamma\|_2 \leq C\}$ 是 γ_0 周围的一个球，对于 $\|\gamma\|_2 = C$ 和公式 (3-8)，有

$$\begin{aligned}\tilde{R}_T(\gamma) &= R_T\left(\gamma_0 + \frac{\gamma}{\sqrt{T}}\right) - R_T(\gamma_0) = \tilde{Q}_T(\gamma) + \lambda_T \sum_{j=1}^p v_j \left(\left| \gamma_{0,j} + \frac{\gamma_j}{\sqrt{T}} \right| - |\gamma_{0,j}| \right) \\ &\geq \tilde{Q}_T(\gamma) - \frac{\lambda_T S_1}{\sqrt{T}} \sum_{j=1}^p |\gamma_j| \geq \tilde{Q}_T(\gamma) - \frac{\lambda_T S_1}{\sqrt{T}} \|\gamma\|_2 = \tilde{Q}_T(\gamma) - o(\|\gamma\|_2),\end{aligned}$$

其中 $\tilde{Q}_T(\gamma) = \sum_{t=1}^T [\rho_\tau\{\tilde{Y}_t - \hat{g}'(X_t^T \hat{\gamma}) X_t^T \gamma / \sqrt{T}\} - \rho_\tau(\tilde{Y}_t)]$ ， $\tilde{Y}_t = Y_t - \hat{g}(X_t^T \hat{\gamma}) X_t^T (\gamma_0 - \hat{\gamma})$ 。

注意到 $\tilde{Q}_T(\gamma)$ 可以被写为

$$\tilde{Q}_T(\gamma) = E\{\tilde{Q}_T(\gamma)\} - \frac{1}{\sqrt{T}} \sum_{t=1}^T [\rho'_\tau(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T - E\{\rho'_\tau(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T\}] \gamma + o_p(1),$$

并且

$$\begin{aligned}E\{\tilde{Q}_T(\gamma)\} &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T E\{\rho'_\tau(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T\} \gamma \\ &\quad + \gamma^T \frac{1}{2T} \sum_{t=1}^T E[\rho''_\tau(\tilde{Y}_t) \{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_t X_t^T] \gamma + o_p(1),\end{aligned}$$

因此，就有

$$\begin{aligned}\tilde{Q}_T(\gamma) &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T \rho'_\tau(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T \gamma \\ &\quad + \gamma^T \frac{1}{2T} \sum_{t=1}^T E[\rho''_\tau(\tilde{Y}_t) \{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_t X_t^T] \gamma + o_p(1).\end{aligned}$$

第一项是渐近正态分布。因此， $\tilde{Q}_T(\gamma)$ 在 $o_p(\sqrt{4\gamma^T \Sigma \gamma})$ 下是有限的。基于以上所有讨论， $\tilde{R}_T(\gamma)$ 由 $\gamma^T(2T)^{-1} \sum_{t=1}^T E[\rho'_t(\tilde{Y}_t)\{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_t X_t^T] \gamma$ 所支配， C 足够大。之后，对于任意 $\eta > 0$

$$\left\{ \inf_{\|\gamma\|_2=C} R_T(\gamma_0 + \gamma/\sqrt{T}) > R_T(\gamma_0) \right\} \geq 1 - \eta,$$

这一结果显示在概率至少为 $1 - \eta$ 下，球 $\{\gamma_0 + \gamma/\sqrt{T} : \|\gamma\|_2 \leq C\}$ 有一个最小化器，结果是一个最小的 $R_T(\gamma_0 + \gamma/\sqrt{T})$ 使得 $\|\hat{\gamma}^P - \gamma_0\|_2 = O_p(T^{-1/2})$ 。证明完毕。

定理 3.2 证明：我们首先证明渐近正态性。类似于定理 3.1，我们令 $\hat{\gamma}^P = \sqrt{T}(\hat{\gamma}^P - \gamma_0)$ 。之后 $\hat{\gamma}^P$ 即为以下公式的最小值：

$$\begin{aligned} \tilde{R}_T(\tilde{\gamma}^*) &= R_T(\gamma_0 + \tilde{\gamma}^*/\sqrt{T}) - R_T(\gamma_0) \\ &= \bar{Q}_T(\tilde{\gamma}^*) + \lambda_T \sum_{j=1}^p v_j (|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) \end{aligned}$$

且

$$\begin{aligned} \bar{Q}_T(\tilde{\gamma}^*) &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T \rho'_t(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T \tilde{\gamma}^* \\ &\quad + \frac{1}{2T} \sum_{t=1}^T \tilde{\gamma}^{*T} E[\rho''_t(\tilde{Y}_t) \{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_t X_t^T] \tilde{\gamma}^* + o_p(1). \end{aligned}$$

接下来，我们顺着 Zou(2006)中的线索探究 $\lambda_T \sum_{j=1}^p v_j (|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|)$ 的限制行为。如果 $j \in \Lambda$ ，那么 $v_j < \infty$ 且 $\sqrt{T}(|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) \rightarrow \tilde{\gamma}_j^* \text{sign}(\gamma_{0,j})$ 。根据斯卢茨基定理和条件 $\lambda_T/\sqrt{T} \rightarrow 0$ ，有

$$\lambda_T v_j (|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) \xrightarrow{P} 0. \quad (\text{B-1})$$

否则，如果 $j \in \Lambda^c$ ，那么在假设 $s_2 \lambda_T/\sqrt{T} \rightarrow \infty$ 下， $\sqrt{T}(|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) = |\tilde{\gamma}_j^*|$ 和 $v_j \lambda_T/\sqrt{T} \rightarrow \infty$ 有

$$\lambda_T v_j (|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) \begin{cases} \xrightarrow{P} \infty, & \text{if } \tilde{\gamma}_j^* \neq 0, \\ = 0, & \text{if } \tilde{\gamma}_j^* = 0. \end{cases} \quad (\text{B-2})$$

结合(B-1)和(B-2),

$$\lambda_T v_j (|\gamma_{0,j} + \tilde{\gamma}_j^*/\sqrt{T}| - |\gamma_{0,j}|) \xrightarrow{P} \begin{cases} \infty, & \text{if } \tilde{\gamma}_j^* \neq 0 \text{ for some } j \in \Lambda^c \\ 0, & \text{otherwise.} \end{cases}$$

如果对于一些 $j \in \Lambda^c$, $\tilde{\gamma}_j^* \neq 0$, 定义 $\tilde{R}_T(\tilde{\gamma}^*) \rightarrow \infty$, 并且除此以外

$$\begin{aligned} \tilde{R}_T(\tilde{\gamma}^*) &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T \rho'_t(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_{t,\Lambda}^T \tilde{\gamma}_\Lambda^* \\ &\quad + \frac{1}{2T} \sum_{t=1}^T \tilde{\gamma}_\Lambda^{*T} E[\rho''_t(\tilde{Y}_t) \{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_{t,\Lambda} X_{t,\Lambda}^T] \tilde{\gamma}_\Lambda^* + o_p(1). \end{aligned}$$

最后, $\tilde{R}_T(\tilde{\gamma}^*)$ 有一个独特的最小化器 $(\hat{\gamma}_\Lambda^P, 0)$, 使得

$$\begin{aligned} \sqrt{T}(\hat{\gamma}_\Lambda^P - \gamma_{0,\Lambda}) &= \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^T E[\rho''_t(\tilde{Y}_t) \{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_{t,\Lambda} X_{t,\Lambda}^T] \right\}^{-1} \\ &\quad \times \frac{1}{\sqrt{T}} \sum_{t=1}^T \rho'_t(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}^0) X_{t,\Lambda}^T + o_p(1). \end{aligned}$$

类似于 Jiang, Hu & Yu (2020) 中定理 2.2 的证明, 我们得到

$$\sqrt{T}(\hat{\gamma}_\Lambda^P - \gamma_{0,\Lambda}) \xrightarrow{L} N\left(0, \left\{S^{-1} \sum S^{-1}\right\}_{\Lambda\Lambda}\right).$$

渐近正态性证明完毕。

接下来我们证明一致性。对于 $j \in \Lambda$, $\hat{\gamma}_\Lambda^P$ 的渐近正态性表明 $\hat{\gamma}_j^P \xrightarrow{P} \gamma_{0,j} \neq 0$ 。这是一个显而易见的结果, $P(j \in \tilde{\Lambda}) \rightarrow 1$, $\tilde{\Lambda} = \{j: \hat{\gamma}_j^P \neq 0\}$ 。接下来我们证明 $P(j \in \tilde{\Lambda}^c) \rightarrow 1$, $j \in \Lambda^c$ 。根据定理 3.1, $\hat{\gamma}^P - \gamma_0 = O_p(T^{-1/2})$ 。类似于之前的证法,

$$\frac{1}{\sqrt{T}} \frac{\partial R_T(\gamma)}{\partial \gamma_j} = -\frac{1}{\sqrt{T}} \sum_{t=1}^T \rho'_t(\tilde{Y}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t^T$$

$$\begin{aligned}
& + \frac{1}{\sqrt{T}}(\gamma - \gamma_0) \sum_{t=1}^T E[\rho_t''(\tilde{Y}_t)\{\hat{g}'(X_t^T \hat{\gamma})\}^2 X_t X_t^T] \\
& + \frac{1}{\sqrt{T}} \lambda_T v_j \text{sign}(\gamma_j) + o_p(1).
\end{aligned}$$

显而易见第一项收敛于均值为零的正态分布，第二项是 $O_p(1)$ ，并且

$$\frac{1}{\sqrt{T}} \lambda_T v_j \text{sign}(\gamma_j) \xrightarrow{p} \begin{cases} \infty, & \text{if } \gamma_j > 0, \\ -\infty, & \text{if } \gamma_j < 0. \end{cases}$$

因此，当 $T \rightarrow \infty$ ，如果 $\gamma_j > 0$ ，那么 $\partial R_T(\gamma) / \partial \gamma_j > 0$ ，如果 $\gamma_j < 0$ ，那么 $\partial R_T(\gamma) / \partial \gamma_j < 0$ 。

因此 $\hat{\gamma}_j^p = 0$ ，即 $P(j \in \tilde{\Lambda}^c) \rightarrow 1, j \in \Lambda^c$ 。一致性证明完毕。

定理 3.3 证明：注意到公式(3-10)中

$$\begin{aligned}
& \sqrt{T}(\hat{\gamma}^{AP} - \gamma_0) \\
& = \bar{S}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \omega_\tau(\hat{e}_t) \hat{g}'(X_t^T \hat{\gamma}) X_t \{Y_t - \hat{g}(X_t^T \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma}) X_t^T (\gamma_0 - \hat{\gamma})\} \\
& = A_1 + A_2 + A_3,
\end{aligned}$$

其中

$$\begin{aligned}
A_1 & = \tilde{S}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \omega_\tau(e_{0,t}) \hat{g}'(X_t^T \hat{\gamma}) X_t \{Y_t - \hat{g}(X_t^T \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma}) X_t^T (\gamma_0 - \hat{\gamma})\}, \\
A_2 & = \bar{S}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \{\omega_\tau(\hat{e}_t) - \omega_\tau(e_{0,t})\} \hat{g}'(X_t^T \hat{\gamma}) X_t \{Y_t - \hat{g}(X_t^T \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma}) X_t^T (\gamma_0 - \hat{\gamma})\}, \\
A_3 & = (\bar{S}^{-1} - \tilde{S}^{-1}) \frac{1}{\sqrt{T}} \sum_{t=1}^T \omega_\tau(e_{0,t}) \hat{g}'(X_t^T \hat{\gamma}) X_t \{Y_t - \hat{g}(X_t^T \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma}) X_t^T (\gamma_0 - \hat{\gamma})\}.
\end{aligned}$$

根据 Jiang, Hu & Yu (2020)中定理 2.2 的证明，我们可以得到

$$A_1 \xrightarrow{L} N\left(0, S^{-1} \sum S^{-1}\right),$$

接下来，我们证明 $A_2 = o_p(1)$ 。回忆 $\omega_\tau(\hat{e}_t)$ 和 $\omega_\tau(e_{0,t})$ 的定义，我们有

$$\omega_\tau(\hat{e}_t) - \omega_\tau(e_{0,t}) = \begin{cases} 0, & \text{if } \hat{e}_t > 0, e_{0,t} > 0 \text{ or } \hat{e}_t \leq 0, e_{0,t} \leq 0, \\ 2\tau - 1, & \text{if } \hat{e}_t > 0, e_{0,t} \leq 0, \\ 1 - 2\tau, & \text{if } \hat{e}_t \leq 0, e_{0,t} > 0. \end{cases} \quad (\text{B-3})$$

实际上, 有 $\hat{e}_t - e_{0,t} = g_0(X_t^T \gamma_0) - \hat{g}'(X_t^T \hat{\gamma}) - \hat{g}'(X_t^T \hat{\gamma}) X_t^T (\hat{\gamma}^P - \hat{\gamma})$ 。之后通过定理 3.1 和类似于 Jiang, Hu & Yu (2020) 中定理 2.2 的证明, 有

$$\begin{aligned} \max_{1 \leq t \leq T} |\hat{e}_t - e_{0,t}| &\leq \max_{1 \leq t \leq T} |g_0(X_t^T \gamma_0) - \hat{g}'(X_t^T \hat{\gamma})| \\ &\quad + \max_{1 \leq t \leq T} |\hat{g}'(X_t^T \hat{\gamma}) X_t^T (\hat{\gamma}^P - \hat{\gamma})| = o_p(1), \end{aligned}$$

因此根据(B-3), 有 $\max_{1 \leq t \leq T} |\omega_\tau(\hat{e}_t) - \omega_\tau(e_{0,t})| = o_p(1)$, 随后 $A_2 = o_p(1)$ 。最后我们证明

$A_3 = o_p(1)$ 。注意到

$$0 = \bar{S} \bar{S}^{-1} - I_p = (\bar{S} - \tilde{S}) \bar{S}^{-1} + \tilde{S} (\bar{S}^{-1} - \tilde{S}^{-1}),$$

这是一个合适的大数定理确保 $\bar{S} - \tilde{S} = o_p(1)$, 因此可以得到

$$\|\bar{S}^{-1} - \tilde{S}^{-1}\| \leq \|\tilde{S}^{-1}\| \|\tilde{S} - \bar{S}\| \|\bar{S}^{-1}\| = o_p(1).$$

因此根据 A_3 的定义, 我们得到 $A_3 = o_p(1)$, 证明完毕。

附录 C 代码

数值模拟部分

```
library(SparseM)
library(KernSmooth)
library(MAVE)
library(stats)
library(MASS)
##### simple begin#####
n<-200
m<-100

#p<-10 # Case 1
#beta_true<-c(4,3,rep(0,p-2))/sqrt(25)# Case 1
p<-100# Case 2
beta_true<-c(rep(1,10),rep(0,p-10))/sqrt(10)# Case 2

MSE<-matrix(0,nrow=m,ncol=1)
RMSE<-matrix(0,nrow=m,ncol=1)
FV<-matrix(0,nrow=m,ncol=1)
TV<-matrix(0,nrow=m,ncol=1)
TT<-matrix(0,nrow=m,ncol=1)
Length<-matrix(0,nrow=m,ncol=p)
CV<-matrix(0,nrow=m,ncol=p)
tauall=c(0.1,0.25,0.5,0.75,0.9)
for(tau in 1:5)
```

```

{
  print("tau=")
  tau=tauall[taui]
  print(tau)
for(nn in 1:m)
{
  print(nn)
  #errorg<-rnorm(n)
  errorg<-rt(n,5)
  f<- function(c) sum(abs(tau-(errorg<=c))*(errorg-c)^2)
  cmin <- optimize(f,interval = c(-10,10))
  errortau<-c(cmin$minimum)
  error<-errorg-errortau
  x<-matrix(0,nrow=n,ncol=p)

##### Case 1#####
#x[1,]<-2*runif(p)-1
#x[2,]<-2*runif(p)-1
#for(i in 3:n)
#{ x[i,]<-0.4*x[i-1,]-0.5*x[i-2,]+2*runif(p)-1 }
#y<-5*sin(pi*(x%%beta_true))+(1+(x%%beta_true)^2)*error
##### Case 1#####

##### Case 2#####
x[1,]<-rnorm(p,2,1)
for(i in 2:n)
{ x[i,]<-0.5*x[i-1,]+rnorm(p,2,1) }
y<-(x%%beta_true)^2+error

```

```
##### Case 1#####  
##### simple end#####  
  
##### MAVE begin#####  
dr <- mave(y~x)  
dr.dim <- mave.dim(dr)  
b.mave <- coef(dr.dim)[,1]  
if (b.mave[1]==0)  
{b.mave=ginv(t(x)%*%x)%*%t(x)%*%y}  
else{b.mave <- b.mave*sign(b.mave[1])}  
  
a<-matrix(0,nrow=n,ncol=1)  
b<-matrix(0,nrow=n,ncol=1)  
for(i in 1:n)  
{  
  xn<-x%*%b.mave  
  xi<-xn-xn[i]  
  h<-max(1.06*n^(-0.2)*sd(xn),0.001)  
  k<-15/16*(1-(xi/h)^2)*(abs(xi/h)<=1)  
  z<-cbind(1,xi)  
  zk<-cbind(k,k*xi)  
  ab<-ginv(t(zk)%*%z)%*%t(zk)%*%y)  
  a[i]<-ab[1]  
  b[i]<-ab[2]  
}  
##### MAVE end #####  
  
##### hatgamma begin #####
```

```

gammag<-b.mave
ag<-a
bg<-b
iter<-1
eg<-1
while((iter < 10) & (eg>0.01))
{
##### STEP 1 begin #####
gammagb<-gammag
for(i in 1:n)
{
xn<-x%*%gammag
xi<-xn-xn[i]
h<-max(1.06*n^(-0.2)*sd(xn),0.001)
k<-15/16*(1-(xi/h)^2)*(abs(xi/h)<=1)
z<-cbind(1,xi)
iter1<-1
eg1<-1
while((iter1 < 10) & (eg1>0.01))
{
we<-abs(tau-((y-ag[i]-bg[i]*xi)<=0))
zkw<-cbind(we*k,we*k*xi)
abw<-ginv(t(zkw)%*%z)%*%(t(zkw)%*%y)
eg1<-sum((t(abw)-cbind(ag[i],bg[i]))^2)
ag[i]<-abw[1]
bg[i]<-abw[2]
iter1<-iter1+1
}
}

```

```

}
##### STEP 1 end #####

##### STEP 2 begin #####
iter2<-1
eg2<-1
while((iter2 < 10) & (eg2>0.01))
{
  gamma1g<-matrix(0,nrow=p,ncol=p)
  gamma2g<-matrix(0,nrow=p,ncol=1)
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      xij<-(x[i,]-x[j,])%*%gammag
      weg<-abs(tau-((y[i]-ag[j]-bg[j]*xij)<=0))
      xi<-x%*%gammag-c(x[j,]%*%gammag)
      ki<-sum(15/16*(1-(xi/h)^2)*(abs((xi/h))<=1))
      kij<-15/16*(1-(xij/h)^2)*(abs((xij/h))<=1)/ki
      gamma1g<-c(weg*kij*bg[j]^2)*(x[i,]-x[j,])%*%t(x[i,]-x[j,])+gamma1g
      gamma2g<-c(weg*kij*bg[j])*(x[i,]-x[j,])*(y[i]-ag[j])+gamma2g
    }
  }
  gammag1<-ginv(gamma1g)%*%gamma2g
  gammag1<-gammag1/sqrt(sum((as.vector(gammag1))^2))*sign(gammag1[1])
  eg2<-sum((gammag-gammag1)^2)
  iter2<-iter2+1
  gammag<-gammag1
}

```

```

}
##### STEP 2 end #####
iter<-iter+1
eg<-sum((gammag-gammagb)^2)
}
hatgamma<-gammag
##### hatgamma end #####
##### hatg begin #####
ag<-a
bg<-b
for(i in 1:n)
{
  xn<-x%*%hatgamma
  xi<-xn-xn[i]
  h<-max(1.06*n^(-0.2)*sd(xn),0.001)
  k<-15/16*(1-(xi/h)^2)*(abs((xi/h))<=1)
  z<-cbind(1,xi)
  iterg<-1
  egg<-1
  while((iterg < 10) & (egg>0.01))
  {
    we<-abs(tau-((y-ag[i]-bg[i]*xi)<=0))
    zkw<-cbind(we*k,we*k*xi)
    abw<-ginv(t(zkw)%*%z)%*%(t(zkw)%*%y)
    eg1<-sum(((abw)-cbind(ag[i],bg[i]))^2)
    ag[i]<-abw[1]
    bg[i]<-abw[2]
    iterg<-iterg+1
  }
}

```

```

}
}
##### hatg end #####

##### variable selection begin #####
lambda<-1.1*2*sqrt(n)*qnorm(1-0.05/(2*p))
f1 <-
function(bx){ sum(abs(tau-(y-ag-bg*x%*%(bx[1:p]-hatgamma)<=0))*(y-ag-bg*x%*%(bx
[1:p]-hatgamma))^2)+sum(lambda*abs(bx[1:p])/abs(hatgamma))}
cmin1<-nlm(f1, c(hatgamma))
bpg<-c(cmin1$estimate)
bpgg<-bpg/sqrt(sum((as.vector(bpg))^2))*sign(bpg[1])
bpggg<-((abs(bpgg)>0.01)*bpgg)
hatgammap<-bpggg/sqrt(sum((as.vector(bpggg))^2))*sign(bpggg[1])
##### variable selection end #####

##### bias begin #####
webias<-abs(tau-((y-ag-bg*(x%*%(hatgammap-hatgamma))<=0))
xg<-matrix(0,nrow=n,ncol=p)
for(i in 1:p)
{
  xg[,i]<-webias*bg^2*x[,i]
}
sbias<-t(xg)%*%x/n
hatbg<-matrix(0,nrow=p,ncol=1)
for(i in 1:p)
{
  hatbg[i]<-mean(webias*bg*(y-ag+bg*(x%*%hatgamma))*x[,i])
}

```



```

}
hatbgg<-ginv(sbias)%*%hatbg
hatgammabias<-hatbgg/sqrt(sum((as.vector(hatbgg))^2))*sign(hatbgg[1])
##### bias end #####

##### result #####
MSE[nn]<-sqrt(sum((hatgammabias-beta_true)^2))
RMSE[nn]<-sqrt(sum((hatgammabias-beta_true)^2))/sqrt(sum((hatgamma-beta_true)^2))
FV[nn]<-sum(hatgammabias[11:p]!=0)
TV[nn]<-sum(hatgammabias[1:10]!=0)
TT[nn]<-(sum(hatgammabias[1:10]!=0)+sum(hatgammabias[11:p]==0))/p==1

alefa<-0.05
xb1<-matrix(0,nrow=n,ncol=p)
webias1<-abs(tau-((y-ag-bg*(x%*%(hatgammabias-hatgamma))<=0))
for(i in 1:p)
{
  xb1[,i]<-webias1*bg^2*x[,i]
}
sbias1<-t(xb1)%*%x/n
xb2<-matrix(0,nrow=n,ncol=p)
for(i in 1:p)
{
  xb2[,i]<-webias1*bg*x[,i]*(y-ag-bg*(x%*%(hatgammabias-hatgamma)))
}
sgbias1<-t(xb2)%*%xb2/n
bcov<-ginv(sbias1)%*%sgbias1%*%ginv(sbias1)

```

```
for(i in 1:p)
{
  Length[nn,i]=2*sqrt(bcov[i,i])/sqrt(n)*qnorm(1-alefa/2)

  CV[nn,i]=(beta_true[i]>(hatgammabias[i]-Length[nn,i]/2))*(beta_true[i]<(hatgammabias[i]
  ]+Length[nn,i]/2))
}
print(nn)
}
##### Print results #####
#cbind(beta_true,b.mave,hatgamma,hatgammap,hatgammabias)
MSEr<-c(mean(MSE),sd(MSE))
RMSEr<-c(mean(RMSE),sd(RMSE))
FVr<-c(mean(FV),sd(FV))
TVr<-c(mean(TV),sd(TV))
TTr<-c(mean(TT),sd(TT))
rmean<-c(mean(MSE),mean(RMSE),mean(FV),mean(TV),mean(TT))
rsd<-c(sd(MSE),sd(RMSE),sd(FV),sd(TV),sd(TT))

CV1=colMeans(CV)
pp=seq(from=1,to=100,by=1)

print(round(rmean,4))
print(round(rsd,4))
plot(pp,CV1,xlab="",ylab="")
#print(round(colMeans(CV),4))
```

波士顿住房数据实验

```

library(SparseM)
library(KernSmooth)
library(MAVE)
library(stats)
library(MASS)
library(MultiRNG)
##### simple begin#####
tauall=c(0.1,0.25,0.5,0.75,0.9)
y1<-matrix(Boston[,14],nrow=506,ncol=1)
y<-(y1-mean(y1))/sd(y1)
x1<-Boston[,1]
x2<-Boston[,2]
x3<-Boston[,3]
#x4<-Boston[,4]
x5<-Boston[,5]
x6<-Boston[,6]
x7<-Boston[,7]
x8<-Boston[,8]
x9<-Boston[,9]
x10<-Boston[,10]
x11<-Boston[,11]
x12<-Boston[,12]
x13<-Boston[,13]

x1<-(x1-mean(x1))/sd(x1)
x2<-(x2-mean(x2))/sd(x2)

```

```
x3<-(x3-mean(x3))/sd(x3)
#x4<-(x4-mean(x4))/sd(x4)
x5<-(x5-mean(x5))/sd(x5)
x6<-(x6-mean(x6))/sd(x6)
x7<-(x7-mean(x7))/sd(x7)
x8<-(x8-mean(x8))/sd(x8)
x9<-(x9-mean(x9))/sd(x9)
x10<-(x10-mean(x10))/sd(x10)
x11<-(x11-mean(x11))/sd(x11)
x12<-(x12-mean(x12))/sd(x12)
x13<-(x13-mean(x13))/sd(x13)

x=cbind(x6,x1,x2,x3,x5,x7,x8,x9,x10,x11,x12,x13)
p=12
n<-length(y)
#####

ghatall=matrix(0,nrow=n,ncol=5)
vsall=matrix(0,nrow=p,ncol=5)
testall=matrix(0,nrow=p,ncol=5)
btall=matrix(0,nrow=p,ncol=5)
bptall=matrix(0,nrow=p,ncol=5)
hatgdall=matrix(0,nrow=n,ncol=5)
hatguall=matrix(0,nrow=n,ncol=5)

##### MAVE begin#####
dr <- mave(y~x)
dr.dim <- mave.dim(dr)
```

```

b.mave <- coef(dr.dim)[,1]
a<-matrix(0,nrow=n,ncol=1)
b<-matrix(0,nrow=n,ncol=1)
for(i in 1:n)
{
  xn<-x%*%b.mave
  xi<-xn-xn[i]
  h<-1.06*n^(-0.2)*sd(xn)
  k<-15/16*(1-(xi/h)^2)*(abs(xi/h)<=1)
  z<-cbind(1,xi)
  zk<-cbind(k,k*xi)
  ab<-ginv(t(zk)%*%z)%*%(t(zk)%*%y)
  a[i]<-ab[1]
  b[i]<-ab[2]
}
##### MAVE end #####

```

```

for(tau in 1:5)
{
  tau=tauall[tau]
##### hatgamma begin #####
  gammag<-b.mave
  ag<-a
  bg<-b
  iter<-1
  eg<-1

```

```
while((iter < 10) & (eg>0.001))
{
##### STEP 1 begin #####
gammagb<-gammag
for(i in 1:n)
{
  xn<-x%*%gammag
  xi<-xn-xn[i]
  h<-1.06*n^(-0.2)*sd(xn)
  k<-15/16*(1-(xi/h)^2)*(abs(xi/h)<=1)
  z<-cbind(1,xi)
  iter1<-1
  eg1<-1
  while((iter1 < 10) & (eg1>0.01))
  {
    we<-abs(tau-((y-ag[i]-bg[i]*xi)<=0))
    zkw<-cbind(we*k,we*k*xi)
    abw<-ginv(t(zkw)%*%z)%*%t(zkw)%*%y)
    eg1<-sum((t(abw)-cbind(ag[i],bg[i]))^2)
    ag[i]<-abw[1]
    bg[i]<-abw[2]
    iter1<-iter1+1
  }
}
##### STEP 1 end #####

##### STEP 2 begin #####
iter2<-1
```

```

eg2<-1
while((iter2 < 10) & (eg2>0.01))
{
  gamma1g<-matrix(0,nrow=p,ncol=p)
  gamma2g<-matrix(0,nrow=p,ncol=1)
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      xij<-(x[i,]-x[j,])%*%gammag
      weg<-abs(tau-((y[i]-ag[j]-bg[j]*xij)<=0))
      xi<-x%*%gammag-c(x[j,]%*%gammag)
      ki<-sum(15/16*(1-(xi/h)^2)*(abs((xi/h)<=1))
      kij<-15/16*(1-(xij/h)^2)*(abs((xij/h)<=1))/ki
      gamma1g<-c(weg*kij*bg[j]^2)*(x[i,]-x[j,])%*%t(x[i,]-x[j,])+gamma1g
      gamma2g<-c(weg*kij*bg[j])*(x[i,]-x[j,])*(y[i]-ag[j])+gamma2g
    }
  }
  gammag1<-ginv(gamma1g)%*%gamma2g
  gammag1<-gammag1/sqrt(sum((as.vector(gammag1))^2))
  eg2<-sum((gammag-gammag1)^2)
  iter2<-iter2+1
  gammag<-gammag1
}
##### STEP 2 end #####
iter<-iter+1
eg<-sum((gammag-gammagb)^2)
}

```

```

hatgamma<-gammag
#####

##### hatg begin #####
hatg<-ag
gcov<-matrix(0,nrow=n,ncol=1)
weg<-abs(tau-((y-ag)<=0))
xn<-x%*%hatgamma
h<-1.06*n^(-0.2)*sd(xn)
for(i in 1:n)
{
xi<-xn-xn[i]
k<-15/16*(1-(xi/h)^2)*(abs(xi/h)<=1)
Sx<-sum(weg*k)
Sigx<-sum((weg*k*(y-ag))^2)
gcov[i]<-ginv(Sx)%*%Sigx%*%ginv(Sx)
}

##### ALS end #####

##### variable selection begin #####
lambda<-1.1*2*sqrt(n)*qnorm(1-0.05/(2*p))
f1 <-
function(bx){ sum(abs(tau-(y-ag-bg*x%*%(bx[1:p]-hatgamma)<=0))*(y-ag-bg*x%*%(bx
[1:p]-hatgamma))^2)+sum(lambda*abs(bx[1:p])/abs(hatgamma))}
cmin1<-nlm(f1, c(hatgamma))
bpg<-c(cmin1$estimate)

```



```

bpgg<-bpg/sqrt(sum((as.vector(bpg))^2))
#bpggg<-(abs(bpgg)>0.01/p)*bpgg
hatgammap<-bpgg/sqrt(sum((as.vector(bpgg))^2))
##### variable selection end #####

##### bias begin #####
webias<-abs(tau-((y-ag-bg*(x%*%(hatgammap-hatgamma)))<=0))
xg<-matrix(0,nrow=n,ncol=p)
for(i in 1:p)
{
  xg[,i]<-webias*bg^2*x[,i]
}
sbias<-t(xg)%*%x/n
hatbg<-matrix(0,nrow=p,ncol=1)
for(i in 1:p)
{
  hatbg[i]<-mean(webias*bg*(y-ag+bg*(x%*%hatgamma))*x[,i])
}
hatbgg<-ginv(sbias)%*%hatbg
hatgammabias<-hatbgg/sqrt(sum((as.vector(hatbgg))^2))
##### bias end #####
Test<-matrix(0,nrow=p,ncol=1)
alefa<-0.05
xb1<-matrix(0,nrow=n,ncol=p)
webias1<-abs(tau-((y-ag-bg*(x%*%(hatgammabias-hatgamma)))<=0))
for(i in 1:p)
{
  xb1[,i]<-webias1*bg^2*x[,i]
}

```

```
}
sbias1<-t(xb1)%%/x/n
xb2<-matrix(0,nrow=n,ncol=p)
for(i in 1:p)
{
  xb2[,i]<-webias1*bg*x[,i]*(y-ag-bg*(x%%(hatgammabias-hatgamma)))
}
sgbias1<-t(xb2)%%/xb2/n
bcovg<-ginv(sbias1)%%/sgbias1%%/ginv(sbias1)

for(i in 1:p)
{
  Test[i]<-(2*(1-pnorm(sqrt(n)*abs(hatgammabias[i])/sqrt(bcovg[i,i]))))<=alefa)
}
hatgd<-hatg-sqrt(gcov)*qnorm(1-alefa/2)
hatgu<-hatg+sqrt(gcov)*qnorm(1-alefa/2)

ghatall[,taui]=ag
vsall[,taui]=(abs(hatgammap)>0.00001)
testall[,taui]=Test
btall[,taui]=hatgamma
bptall[,taui]=hatgammap
hatgdall[,taui]=hatgd
hatguall[,taui]=hatgu
}

#par(mfrow=c(3,2),pin = c(2,2))
```

```

xn<-x%*%btall[,1]
fhg<-cbind(xn,y,ghatall[,1],hatgdall[,1],hatgual[,1])
fhg1<-fhg[order(fhg[,1]),]
index<-fhg1[,1]
medv<-fhg1[,2]
hatgp<-fhg1[,3]
hatgdp<-fhg1[,4]
hatgup<-fhg1[,5]
plot(index,medv,type="p")
lines(index,hatgp,col="red",lty=1)
lines(index,hatgup,col=4,lty=2)
lines(index,hatgdp,col=4,lty=2)
title(expression(tau==0.10))

```

```

xn<-x%*%btall[,2]
fhg<-cbind(xn,y,ghatall[,2],hatgdall[,2],hatgual[,2])
fhg1<-fhg[order(fhg[,1]),]
index<-fhg1[,1]
medv<-fhg1[,2]
hatgp<-fhg1[,3]
hatgdp<-fhg1[,4]
hatgup<-fhg1[,5]
plot(index,medv,type="p")
lines(index,hatgp,col="red",lty=1)
lines(index,hatgup,col=4,lty=2)
lines(index,hatgdp,col=4,lty=2)
title(expression(tau==0.25))

```

```
xn<-x%*%btall[,3]
fhg<-cbind(xn,y,ghatall[,3],hatgdall[,3],hatguall[,3])
fhg1<-fhg[order(fhg[,1]),]
index<-fhg1[,1]
medv<-fhg1[,2]
hatgp<-fhg1[,3]
hatgdp<-fhg1[,4]
hatgup<-fhg1[,5]
plot(index,medv,type="p")
lines(index,hatgp,col="red",lty=1)
lines(index,hatgup,col=4,lty=2)
lines(index,hatgdp,col=4,lty=2)
title(expression(tau==0.50))
```

```
xn<-x%*%btall[,4]
fhg<-cbind(xn,y,ghatall[,4],hatgdall[,4],hatguall[,4])
fhg1<-fhg[order(fhg[,1]),]
index<-fhg1[,1]
medv<-fhg1[,2]
hatgp<-fhg1[,3]
hatgdp<-fhg1[,4]
hatgup<-fhg1[,5]
plot(index,medv,type="p")
lines(index,hatgp,col="red",lty=1)
lines(index,hatgup,col=4,lty=2)
lines(index,hatgdp,col=4,lty=2)
title(expression(tau==0.75))
```

```

xn<-x%*%btall[,5]
fhg<-cbind(xn,y,ghatall[,5],hatgdall[,5],hatgual[,5])
fhg1<-fhg[order(fhg[,1]),]
index<-fhg1[,1]
medv<-fhg1[,2]
hatgp<-fhg1[,3]
hatgdp<-fhg1[,4]
hatgup<-fhg1[,5]
plot(index,medv,type="p")
lines(index,hatgp,col="red",lty=1)
lines(index,hatgup,col=4,lty=2)
lines(index,hatgdp,col=4,lty=2)
title(expression(tau==0.90))

#par(mfrow=c(1,2),pin = c(2,1))
vsn<-c(1,2,3,4,5,6,7,8,9,10,11,12)
rvs01<-c(0.1,0.1,0,0,0.1,0,0,0.1,0.1,0.1,0,0.1)#>0.00001
rt01<-rep(0.1,12)

rvs025<-c(rep(0.25,2),rep(0,2),rep(0.25,6),0,0.25)#>0.000005
rt025<-c(0.25,0.25,rep(0,4),rep(0.25,6))

rvs05<-c(0.5,0.5,0,0.5,0,0,0.5,0.5,0.5,0.5,0,0.5)#>0.00001
rt05<-c(0.5,0.5,0,0.5,0.5,0,rep(0.5,6))

rvs075<-c(0.75,0.75,0.75,0,0.75,0,rep(0.75,6))#>0.00001
rt075<-c(0.75,0.75,0,0,0,0,rep(0.75,6))

```

```
rvs09<-c(rep(0.9,2),0,0,rep(0.9,8))#>0.0001
```

```
rt09<-c(rep(0.9,3),0,rep(0.9,4),0,rep(0.9,3))
```

```
rvs<-c(rvs01,rvs025,rvs05,rvs075,rvs09)
```

```
vsn1<-c(vsn,vsn,vsn,vsn,vsn)
```

```
axis(1,at=c(1,2,3,4,5,6,7,8,9,10,11,12))
```

```
axis(2,at=c(0.1,0.25,0.5,0.75,0.9))
```

```
title(expression(Variable~selection))
```

```
rt<-c(rt01,rt025,rt05,rt075,rt09)
```

```
vsn1<-c(vsn,vsn,vsn,vsn,vsn)
```

```
plot(vsn1,rt,type="p",xlim=c(1,12),ylim=c(0.1,0.9),yaxt="n",xaxt="n",xlab="variable",yla
```

```
b=expression(tau))
```

```
axis(1,at=c(1,2,3,4,5,6,7,8,9,10,11,12))
```

```
axis(2,at=c(0.1,0.25,0.5,0.75,0.9))
```

```
title(expression(Test))
```

致谢

写致谢的时候，思绪最先回到了两年前收到东华大学研究生录取通知书的时刻，依旧记得当时那种高兴的心情。在数以万计的考研大军中，我能有幸成为东华学子，有机会继续学习生涯，这都要感谢东华大学给我这个机会。在东华的日子里，我感受到了良好的学习和科研氛围，无论是教学的质量，还是科研的深度、新度都令我终身难忘。研究生的两年生活中，从课上的谆谆教诲，到课下的和睦生活，再到实习和论文的理论与实践体验，都给了我满满的感动。所以，在这里我首先要感谢东华大学，给我机会，给我环境，给我成长。

感谢我的导师姜荣老师。研一下学期姜老师便开始指导我开展毕业论文的工作。经过数月的学习、摸索和努力，最后能够完成了硕士论文的全部内容，全都离不开姜导师的精心指导。在我的脑海里，姜导师是一位既有科研才能、年轻有为，又风趣幽默、亲切学生的好老师，他严谨的科研态度，坚实的科研能力，求精的科研作风对我们学生来说，有一种无形的感染和激励。在最开始的论文方向确定和最后的所有论文内容完成的整个期间，这过程中的每一小步，及每一小阶段的经历，都在刷新着我的认知，这是对新的学习内容和新的学习方式的不断尝试和挑战，姜导师在我论文写作期间给予了超多的包括写作和研究内容上的悉心指导，这对我最终能顺利完成全部论文内容起着关键性的作用，所以，我必须要借此致谢，向姜老师表达真挚的感激之情！感谢导师对包括我和我的论文投入的大量心血和精力。

感谢我的室友们。荏苒的两年研究生生涯中，陪伴我最多的就是室友们，共同携手生活的两年中，有快乐我们一起分享，有困难我们相互扶持。在我脚骨折的日子里，陪我去医院的是你们，帮我带饭、取快递的是你们，扶我上下楼的是你们，这种感情是温暖的，值得我珍藏一生，感谢你们的分享，帮助和陪伴，感谢有你们的日子。

感谢实习单位的领导与同事！在实习期的工作中，感谢你们对我工作上的指导，让我收获知识的同时更让我认识到作为职场新人的不足之处，让我受益匪浅。

感谢我的父母。家人的付出总是最容易被忽视的，老爸老妈虽然对我的学习和生活并不了解，而且有时候会很啰嗦，甚至会因为不理解产生一些小摩擦，但是他们始终在支持和鼓励着我，他们是盾，是树，是山，给我足够的信心，让我在自己选择的人生道路上走的自信，走的安全。

此外，还要感谢对我研究生生涯中提供过帮助的朋友。也要感谢参考文献中作者们的科研工作，正是他们的研究工作，给我们打下了科研基础，同时提供了极其丰富的研究资源。最后还要感谢论文评阅老师们的辛苦工作。

与自己的学习生涯最后说一次告别，未来的日子里我也将始终怀揣一份感恩，带着一份坚定，走好自己的人生道路。