**Pure Mathematics 理论数学, 2024, 14(1), 41-52**
Published Online January 2024 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2024.141006

Hans汉斯

# 基于弹性网惩罚的高维部分线性模型的稳健变量选择

**童　画，冯彬娟，袁德美***

重庆工商大学数学与统计学院，重庆

收稿日期：2023年12月5日；录用日期：2024年1月8日；发布日期：2024年1月17日

## 摘　要

高维数据下部分线性模型的变量选择方法大多基于最小二乘回归估计展开，但随着数据复杂性的提升，极端、异常等因素使得以往模型的估计效率直线下滑。为此，本文基于弹性网络法与分位数回归相结合的正则化理论，提出了一种应对高维数据下部分线性模型的稳健变量选择模型。该模型不仅可以有效处理强相关变量组的数据，还可以在面对离群点或存在异方差时仍达到较好的稳健性。此外，理论上证明了在一定条件下模型估计量的相合性和稀疏性，最后通过数值模拟，与其他变量选择方法作比较，进一步表明了该方法的优越性。

## 关键词

高维数据，弹性网，部分线性模型，变量选择，分位数回归

# Robust Variable Selection Based on High-Dimensional Partial Linear Model of Elastic Net Penalty

**Hua Tong, Bingjuan Feng, Demei Yuan***

College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing

Received: Dec. 5th, 2023; accepted: Jan. 8th, 2024; published: Jan. 17th, 2024

## Abstract

**Most of the variable selection methods of some linear models under high-dimensional data are**

*通讯作者。

文章引用: 童画, 冯彬娟, 袁德美. 基于弹性网惩罚的高维部分线性模型的稳健变量选择[J]. 理论数学, 2024, 14(1): 41-52. DOI: 10.12677/pm.2024.141006

**based on least squares regression estimation, but with the increase of data complexity, extreme and abnormal factors make the estimation efficiency of previous model plummet. Therefore, based on the regularization theory combining elastic network method and quantile regression, this paper proposes a robust variable selection model to cope with some linear models under high-dimensional data. The model can not only effectively handle data for groups of strongly correlated variables, but also achieve better robustness in the face of outliers or heteroscedasticity. In addition, the coherence and sparsity of model estimators under certain conditions are theoretically proved, and finally the superiority of this method is further demonstrated by numerical simulation and comparison with other variable selection methods.**

## Keywords

**High-Dimensional Data, Elastic Net, Partial Linear Model, Variable Selection, Quantile Regression**



## 1. 引言

在数字化发展的今天，金融、气象、生物医学等诸多领域的数据中各指标之间的关系变得愈加错综复杂，简单地运用线性模型已经不足以分析各指标的联系。而半参数模型不仅具有强大的拟合能力，且不需要对数据做出过多的假设，十分适用于处理分布未知的复杂数据。部分线性模型[1]则是常见的一类半参数模型，因其同时具有线性模型和非参数模型的优点，吸引了广大统计学者的关注，成为了各领域处理数据的重要工具。

而随着大数据时代的到来，高维数据频繁出现在各个领域。高维数据的变量个数过多，往往在建模分析过程中容易导致“维数灾祸”，传统的变量分析方法已无力应对这种问题。对此，可同时进行变量选择和系数估计的正则化方法被学者们提出，其一般形式为“损失函数 + 惩罚函数”。当采用岭回归[2]作惩罚函数时，其通过有偏估计很好地解决了变量间多重共线性的问题，但这种方法只能限制模型回归系数尽可能地小，不具备稀疏性。LASSO [3]惩罚虽然能实现稀疏变量的目的，但对于具有群组效应的特殊数据结构，它只能勉强选出该变量组的一个变量，容易忽略其他重要变量，对于高维数据中存在多重共线性时，LASSO 的估计效果会变得很差。SCAD [4]惩罚是对称且非凸性的，它虽然可以产生稀疏解，但其迭代算法却极其缓慢，难以适应高维数据情形。而 Zou 和 Hastie [5]提出了弹性网络法，因其结合了 LASSO 和岭回归惩罚项的优势，不仅可以解决多重共线性的问题，而且还能对变量进行有效筛选。

高维数据下的部分线性模型因其需要考虑线性协变量的选择和系数估计、光滑参数的选择、非参数函数的估计等方面的问题，因此如何在高维数据下对部分线性模型进行稳健的变量选择成为了一大挑战。目前，关于高维部分线性模型的变量筛选问题研究还不多。Xie 和 Huang [6]运用 SCAD 惩罚来实现线性部分的稀疏性，并用多项式样条来估计非参数分量。杨宜平等[7]结合样条方法和 Dantzig 或 Lasso 变量选择方法，同时进行变量选择和未知参数估计，并证明了估计误差的非渐近界。Chen 等[8]采用自适应弹性网对部分线性模型进行变量选择，并证明了其 Oracle 性质。Wang 等[9]利用组 Bridge 和最小二乘估计讨论了高维部分线性模型的估计问题。赖秋楠等[10]提出了 profile 贪婪向前回归变量筛选方法，证明了在一定正则条件下所提 PGFR 方法具有筛选相合性。但这些文献大多都是基于最小二乘估计作为损失函数，

而在实践中最小二乘估计对噪音较敏感，稳定性较差，特别是在面对高维数据中含有极端值、离群点或是呈现尖峰、厚尾分布时，其预测效果往往不佳。为了实现更稳健的估计，Koenker 和 Basset [11]提出了分位数回归，因其不需要对误差分布作特定假设仍可获得全局分布的特征，且保持较高的稳健性，进而被不同领域广泛研究运用。

本文就将利用将弹性网络法和分位数回归相结合的正则化方法，对高维部分线性模型进行稳健变量选择。此方法不需要对模型中误差项的分布情况做任何假设，在面对高维数据中出现异常值或极端值时仍保持较好的稳健性，能够丰富具体地反应响应变量的分布特征。除此之外，这种组合的正则化方法还可以灵活处理高维数据中多重共线性与群组效应问题，能对变量间的强相关性信息进行变量筛选，产生稀疏解。理论上将证明该模型估计量的相合性和稀疏性，并通过数值模拟，与其他变量选择方法进行比较，表明该方法更有效更稳健。

## 2. 稳健变量选择过程

考虑如下部分线性模型：

$$\boldsymbol{Y} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + g(\boldsymbol{Z}) + \boldsymbol{\varepsilon}, \tag{1}$$

其中，$\boldsymbol{Y} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ 为 $n$ 维响应变量；$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)^{\mathrm{T}} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \cdots, \tilde{\boldsymbol{x}}_p)$ 为 $n \times p$ 设计矩阵，且 $p \gg n$；$\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}}$ 为 $p$ 维未知参数向量；$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^{\mathrm{T}}$ 为 $n$ 维随机误差向量，与 $(\boldsymbol{X}, Z)$ 独立且期望为 0，方差为 $\sigma^2$；$g(\cdot)$ 为未知光滑函数，为避免维数灾难，假定 $Z$ 为一维随机变量，且在闭区间 $[0,1]$ 上取值。

对模型(1)两边关于 $Z$ 求条件期望，可得

$$E(Y|Z) = \{E(X|Z)\}^{\mathrm{T}}\beta + g(Z), \tag{2}$$

由(2)式可得 $g(Z)$ 的表达式，再将其代入(1)式，经整理得

$$Y - E(Y|Z) = \{X - E(X|Z)\}^{\mathrm{T}}\beta + \varepsilon, \tag{3}$$

若 $E(X|Z)$ 和 $E(Y|Z)$ 已知，则模型(3)即为标准的线性模型。

记 $m_X(Z) = E(X|Z)$，$m_Y(Z) = E(Y|Z)$，令 $\hat{m}_X(Z)$ 和 $\hat{m}_Y(Z)$ 分别为 $m_X(Z)$ 和 $m_Y(Z)$ 的估计量，本文采用核估计，即

$$\hat{m}_X(Z) = \frac{\sum_{i=1}^{n} K\left(\frac{Z_i - Z}{h}\right) X_i}{\sum_{i=1}^{n} K\left(\frac{Z_i - Z}{h}\right)};$$

$$\hat{m}_Y(Z) = \frac{\sum_{i=1}^{n} K\left(\frac{Z_i - Z}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{Z_i - Z}{h}\right)},$$

其中 $K(\cdot)$ 为核函数，$h$ 为窗宽。

记 $\hat{\boldsymbol{Y}} = \boldsymbol{Y} - \hat{m}_Y(Z)$，$\hat{\boldsymbol{X}} = \boldsymbol{X} - \hat{m}_X(Z)$；$\boldsymbol{Y} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n)^{\mathrm{T}}$，$\hat{\boldsymbol{X}} = (\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \cdots, \hat{\boldsymbol{x}}_n)^{\mathrm{T}}$，

则式(3)可改写为

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{4}$$

在面对高维数据时( $p=O\left(n^{a}\right), a>0$ )，为了确保模型的可识别性，提高模型拟合的精确度，假定真实的系数向量 $\boldsymbol{\beta}^{*}$ 为稀疏的，即仅小部分系数非零。在不失一般性的情况下，我们定义 $\boldsymbol{\beta}^{*}=\left(\left(\boldsymbol{\beta}_{1}^{*}\right)^{\mathrm{T}}, 0^{\mathrm{T}}\right)^{\mathrm{T}}$，用 $s$ 表示真实回归系数的非零元素的数量，假定真实模型表示为

$$\mathcal{M}_{*}=\operatorname{supp}\left(\boldsymbol{\beta}^{*}\right)=\{1,2,\cdots,s\},$$

其补集 $\mathcal{M}_{*}^{c}=\{s+1,\cdots,p\}$ 表示噪音协变量，即 $\boldsymbol{\beta}^{*}$ 的前 $s$ 个分量是非零的，设计矩阵可写为

$$\hat{\boldsymbol{X}}=(\boldsymbol{S}, \boldsymbol{Q}),$$

其中，$\boldsymbol{S}=\left(\boldsymbol{S}_{1}, \boldsymbol{S}_{2}, \cdots, \boldsymbol{S}_{n}\right)^{\mathrm{T}}=\left(\hat{\boldsymbol{x}}_{1}, \hat{\boldsymbol{x}}_{2}, \cdots, \hat{\boldsymbol{x}}_{s}\right)$，$\boldsymbol{Q}=\left(\boldsymbol{Q}_{1}, \boldsymbol{Q}_{2}, \cdots, \boldsymbol{Q}_{n}\right)^{\mathrm{T}}=\left(\hat{\boldsymbol{x}}_{s+1}, \hat{\boldsymbol{x}}_{s+2}, \cdots, \hat{\boldsymbol{x}}_{p}\right)$；$\boldsymbol{S}$ 表示非零系数的信号协变量矩阵，$\boldsymbol{Q}$ 表示系数为零的噪音协变量矩阵。本文将设计矩阵 $\hat{\boldsymbol{X}}$ 的每一列标准化使得 $\left\|\hat{\boldsymbol{x}}_{i}\right\|_{2}=\sqrt{n}$， $i=1,2, \cdots, p$。

结合(4)式和线性模型的弹性网惩罚分位数估计[12]，可得求解 $\boldsymbol{\beta}^{*}$ 的模型如下：

$$\hat{\boldsymbol{\beta}}_{Enet}=\underset{\beta \in R^{p}}{\arg \min }\left\{\sum_{i=1}^{n} \rho_{i}\left(\hat{y}_{i}-\hat{\boldsymbol{x}}_{i}^{\mathrm{T}} \boldsymbol{\beta}\right)+n \lambda_{n}\|\boldsymbol{\beta}\|_{1}+n \mu_{n}\|\boldsymbol{\beta}\|_{2}^{2}:=L_{n}(\boldsymbol{\beta})\right\},$$

其中，$\rho_{\tau}(u)=u(\tau-1(u<0))$，$0<\tau<1$ 为分位数回归的损失函数，$\|\boldsymbol{\beta}\|_{1}=\sum_{i=1}^{p}\left|\beta_{i}\right|$ 为 $\boldsymbol{\beta}$ 的 $L_{1}$ 范数，$\|\boldsymbol{\beta}\|_{2}=\sqrt{\sum_{i=1}^{p} \beta_{i}^{2}}$ 为 $\boldsymbol{\beta}$ 的 $L_{2}$ 范数，$n \lambda_{n}\|\boldsymbol{\beta}\|_{1}+n \mu_{n}\|\boldsymbol{\beta}\|_{2}^{2}$ 为弹性网惩罚项，其中 $\lambda_{n}, \mu_{n}>0$ 为正则化系数。

## 3. 统计性质

本节将建立高维部分线性模型参数 $\boldsymbol{\beta}$ 的弹性网估计量 $\hat{\boldsymbol{\beta}}_{Enet}$ 的性质。参照文献 Fan [13]，定义预先知道协变量位置的 Oracle 正则化估计 $\hat{\boldsymbol{\beta}}^{o}=\left(\left(\hat{\boldsymbol{\beta}}_{1}^{o}\right)^{\mathrm{T}}, 0^{\mathrm{T}}\right)^{\mathrm{T}}$，并且满足 $\hat{\boldsymbol{\beta}}^{o}=\underset{\boldsymbol{\beta}}{\arg \min } L_{n}(\boldsymbol{\beta})$，并做出如下假设条件：

(A1) $\varepsilon$ 独立于 $(X, Z)$，且满足 $E(\varepsilon \mid X, Z)=0$，$E\left(\varepsilon^{3} \mid X, Z\right)<\infty$。

(A2) 最佳窗宽 $h=O\left(n^{-1/5}\right)$。

(A3) $\Sigma=\operatorname{cov}(\tilde{X})$ 是一个正定矩阵，$\sup E\left(\|X\|_{2}^{3}\right)<\infty$，$\sup_{Z} E\left(\|X\|_{2}^{3} \mid Z=z\right)<\infty$。

(A4) 核函数 $K(\cdot)$ 的支撑集为 $[-1,1]$，存在常数 $0 \leq M_{1}<M_{2}$，核函数满足 $M_{1} \leq K(t)_{t \in[-1.1]} \leq M_{2}$，且 $\int K(t) \mathrm{d} t=1$，$\int t K(t) \mathrm{d} t=0$，$\int t^{2} K(t) \mathrm{d} t \neq 0$。

(A5) $m_{X}(\cdot)$，$m_{Y}(\cdot)$ 和 $g(\cdot)$ 的二阶导数都是一阶 Lipschitz 连续的。

(A6) 存在常数 $c_{1}>0$ 和 $c_{2}>0$，对任意满足 $|u| \leq c_{1}$ 的 $u$，$f_{i}(u)$ 一致有界且不为 0 和 $\infty$，有

$$\left|F_{i}(u)-F_{i}(0)-u f_{i}(0)\right| \leq c_{2} u^{2},$$

其中，$f_{i}(u)$ 和 $F_{i}(u)$ 分别为误差 $\varepsilon_{i}$ 的密度函数和分布函数。

(A7) 定义 $\boldsymbol{H}=\operatorname{diag}\left\{f_{1}(0), \cdots, f_{n}(0)\right\}$，则 $n^{-1} \boldsymbol{S}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{S}$ 的特征值介于 $c_{0}$ 与 $c_{0}^{-1}$ 之间，且

$$\kappa_{n}=\max_{ij}\left|x_{ij}\right|=o\left(\sqrt{n} s^{-1}\right).$$

(A8) 令 $\gamma_{n}=C_{1} \sqrt{s}\left(\lambda_{n}+\mu_{n}+\sqrt{\log n / n}\right)$，$C_{1}$ 为大于零的常数，有

$$\left\|n^{-1} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{S}\right\|_{2, \infty}<\frac{\lambda_{n}}{2 \gamma_{n}},$$

其中，对于矩阵 $\boldsymbol{A}$ 与向量 $\hat{\boldsymbol{x}}$，

$$\|\boldsymbol{A}\|_{2,\infty}=\sup_{\hat{x}\neq 0}\|\boldsymbol{A}\hat{\boldsymbol{x}}\|_{\infty}/\|\hat{\boldsymbol{x}}\|_{2},$$

且对某些常数 $b\in(0,1)$，有 $\log(p)=o(n^{b})$。

条件(A1)~(A5)是研究半参数模型常用的假设条件，其中条件(A2)为核估计最优窗宽，条件(A4)是对核函数的一般假定。条件(A6)对噪音分布进行假设，任意 $f_i(u)$ 在 0 点周围是 Lipschitz 连续的，如常用的柯西分布就满足这一条件。条件(A7)对信号协变量矩阵 $\boldsymbol{S}$ 和设计矩阵 $\hat{\boldsymbol{X}}$ 的规模进行限制，当设计矩阵 $\hat{\boldsymbol{X}}$ 产生某种分布时，$\kappa_n$ 的界以渐近 1 的概率满足。条件(A8)对设计矩阵 $\hat{\boldsymbol{X}}$ 与子矩阵 $\boldsymbol{Q},\boldsymbol{S}$ 的列向量的相关性进行控制。

下面给出定理说明可用 Oracle 信息来确定信号协变量的位置，正则化估计 $\hat{\boldsymbol{\beta}}^{o}$ 以趋于 1 的概率可估计出真实系数向量 $\boldsymbol{\beta}^{*}$。

**定理 1** 若 $\lambda_n\sqrt{s}\kappa_n\to 0$，$\mu_n\sqrt{s}\kappa_n\to 0$，且满足条件(A1)~(A7)，则存在常数 $c_1,c_2>0$，使得

$$P\left(\left\|\hat{\boldsymbol{\beta}}_1^{o}-\boldsymbol{\beta}_1^{*}\right\|_2\le\gamma_n\right)\ge 1-c_1n^{-c_2s},$$

其中，$\gamma_n=C_1\sqrt{s}\left(\lambda_n+\mu_n+\sqrt{\log n/n}\right)$，$C_1$ 为大于零的常数。

**定理 2** 假设条件(A1)~(A8)成立，若

$$\max_j\left|\beta_j\right|=o(\lambda_n/\mu_n),$$

$$\gamma_n s^{2/3}\kappa_n^2(\log_2 n)^2=o(n\lambda_n^2),$$

$$\kappa_n^3\gamma_n^2=o(\lambda_n),$$

且 $\lambda_n>2\sqrt{(1+c)(\log p)/n}$，$c$ 为大于零的常数。那么，存在目标函数 $L_n(\boldsymbol{\beta})$ 的全局最小值 $\hat{\boldsymbol{\beta}}_{Enet}=\left(\left(\hat{\boldsymbol{\beta}}_1^{o}\right)^{\mathrm{T}},\hat{\boldsymbol{\beta}}_2^{\mathrm{T}}\right)^{\mathrm{T}}$，至少以 $1-O\left(c_1n^{-c_2s}\right)$ 的概率满足：

1) $\hat{\boldsymbol{\beta}}_2=0$，
2) $\left\|\hat{\boldsymbol{\beta}}_1^{o}-\boldsymbol{\beta}_1^{*}\right\|_2\le\gamma_n$。

## 4. 定理证明

在证明上述定理前，本节先给出几个引理。

**引理 1** (Chen [14])假设条件(A1)~(A5)成立，令 $h=n^{-1/5}(\log n)^{1/5}$，则有

$$\max_{1\le j\le p}\sup_{z}\left|\hat{m}_j(z,1,h)-m_j(z)\right|=O\left(n^{-2/5}(\log n)^{2/5}\right).$$

**引理 2** (Fan [13])假设条件(A7)成立，则对任意 $t>0$，都有

$$P\left(Z_n(M_n)\ge 4M_n\sqrt{s/n}+t\right)\le\exp\left(-nc_0t^2/\left(8M_n^2\right)\right).$$

**引理 3** (Bühlmann [15])霍夫丁不等式：设独立随机变量 $Z_1,Z_2,\cdots,Z_n$ 来自某个空间 $\Gamma$，$\gamma$ 为 $\Gamma$ 上的实值函数，若存在正实数 $C_1,C_2,\cdots,C_n$，使得对任意的 $i$ 满足 $E(\gamma(Z_i))=0$，$|\gamma(Z_i)|<C_i$，则对所有的 $t>0$，有

$$P\left(\left|\sum_{i=1}^{n}\gamma(Z_i)\right|\ge t\right)\le 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n}C_i^2}\right).$$

**引理 4** 考虑以 $\boldsymbol{\beta}^{*}$ 为中心的球形领域 $R^{s}$ 内：

$$\mathcal{N}=\left\{\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}}\right)^{\mathrm{T}}\in R^P:\boldsymbol{\beta}_2=0,\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right\|_2\le\gamma_n\right\},$$

其中序列$\gamma_n\to0$，假设

$$\sqrt{1+\gamma_n s^{-3/2}}\kappa_n\left(\log_2 n\right)=o\left(\sqrt{n}\lambda_n\right),$$

$$n^{1/2}\lambda_n\left(\log p\right)^{-1/2}\to\infty,$$

$$\kappa_n\gamma_n^2=o\left(\lambda_n\right),$$

则当条件(A6)~(A8)成立时，有

$$P\left(\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left\|\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'\left(\hat{y}-\boldsymbol{S}\boldsymbol{\beta}_1\right)\right\|_\infty\ge n\lambda_n\right)\le o\left(c_1p^{-c_2}\right),$$

其中，$\rho_\tau'(u)=\tau-1\{u\le0\}$，$c_1,c_2$为大于零的常数。

**定理 1 的证明：**

令$\hat{v}_n(\boldsymbol{\beta})=\sum_{i=1}^n\rho_\tau\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)$，则目标函数可写作

$$L_n(\boldsymbol{\beta})=\hat{v}_n(\boldsymbol{\beta})+n\lambda_n\|\boldsymbol{\beta}\|_1+n\mu_n\|\boldsymbol{\beta}\|_2^2.$$

对于给定的一个常数$M_n>0$，定义集合

$$\mathcal{A}_0(M_n)=\left\{\boldsymbol{\beta}\in R^P:\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\right\|_2\le M_n,\operatorname{supp}(\boldsymbol{\beta})\subseteq\operatorname{supp}\left(\boldsymbol{\beta}^*\right)\right\},$$

然后，定义函数

$$Z_n(M_n)=\sup_{\boldsymbol{\beta}\in\mathcal{A}_0(M_n)}\frac{1}{n}\left|\left(\hat{v}_n(\boldsymbol{\beta})-\hat{v}_n\left(\boldsymbol{\beta}^*\right)\right)-E\left(\hat{v}_n(\boldsymbol{\beta})-\hat{v}_n\left(\boldsymbol{\beta}^*\right)\right)\right|.$$

首先，结合引理 1 可知，对任意$\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1^{\mathrm{T}},0^{\mathrm{T}}\right)^{\mathrm{T}}\in\mathcal{A}_0(M_n)$，当$n$充分大时，有

$$E\left[\hat{v}_n(\boldsymbol{\beta})-\hat{v}_n\left(\boldsymbol{\beta}^*\right)\right]\ge\frac{1}{2}c_0n\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right\|_2^2,\tag{5}$$

其中，$M_n=o\left(\kappa_n^{-1}s^{-1/2}\right)$，$c_0$是$f_i(\cdot)$在 0 的某个领域内的一个下界。

接着证明不等式(5)，令$a_i=\left|\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)\right|$，则对于任意$\boldsymbol{\beta}\in\mathcal{A}_0(M_n)$，根据 Hölder 不等式有

$$|a_i|\le\|\boldsymbol{S}_i\|_2\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right\|_2\le\sqrt{s}\kappa_nM_n\to0.$$

此时，再针对$\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)$的符号进行讨论，当$\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)>0$时，依次根据$P(\varepsilon_i<0)=\tau$，Fubini 定理，泰勒展开及条件(A6)可得

$$\begin{aligned}E\left[\rho_\tau(\varepsilon_i-a_i)-\rho_\tau(\varepsilon_i)\right]&=E\left[\int_0^{a_i}I\left(0\le\varepsilon_i\le t\right)\mathrm{d}t\right]\\&=\int_0^{a_i}\left(F_i(t)-F_i(0)\right)\mathrm{d}t\\&=\frac{1}{2}f_i(0)a_i^2+o(1)a_i^2,\end{aligned}\tag{6}$$

其中，$o(1)$对于$i=1,2,\cdots,n$一致成立。

当$\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)<0$时，可得式(6)右侧同样的结果。进一步，由条件(A7)得

$$\begin{aligned}E\left[\hat{v}_n(\boldsymbol{\beta})-\hat{v}_n(\boldsymbol{\beta}^*)\right]&=\sum_{i=1}^{n}E\left[\rho_\tau(\varepsilon_i-a_i)-\rho_\tau(\varepsilon_i)\right]\geq\frac{1}{2}\sum_{i=1}^{n}f_i(0)\,a_i^2\\&=\frac{1}{2}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)^{\mathrm{T}}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{HS}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*)\geq\frac{1}{2}c_0n\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right\|_2^2.\end{aligned}\tag{7}$$

因此，结合式(6)和式(7)以及 $\hat{v}_n(\boldsymbol{\beta})$ 的定义，对于任意 $\boldsymbol{\beta}\in\left(\boldsymbol{\beta}_1^{\mathrm{T}},0^{\mathrm{T}}\right)^{\mathrm{T}}\in\mathcal{A}_0(M_n)$，可得式(5)成立。但所设定的估计器 $\hat{\boldsymbol{\beta}}^o=\left(\left(\hat{\boldsymbol{\beta}}_1^o\right)^{\mathrm{T}},0^{\mathrm{T}}\right)^{\mathrm{T}}$ 可能不在 $\mathcal{A}_0(M_n)$ 集合中，因此，令 $\tilde{\boldsymbol{\beta}}\in\left(\tilde{\boldsymbol{\beta}}_1^{\mathrm{T}},0^{\mathrm{T}}\right)^{\mathrm{T}}$，其中

$$\tilde{\boldsymbol{\beta}}_1=\mu\hat{\boldsymbol{\beta}}_1^o+(1-\mu)\boldsymbol{\beta}_1^*,\quad\mu=M_n\Big/\left(M_n+\left\|\hat{\boldsymbol{\beta}}_1^o-\boldsymbol{\beta}_1^*\right\|_2\right),$$

此时 $\tilde{\boldsymbol{\beta}}\in\mathcal{A}_0(M_n)$，再根据目标函数为凸函数的性质以及 $\hat{\boldsymbol{\beta}}_1^o$ 的定义，可得

$$L_n(\tilde{\boldsymbol{\beta}})\leq\mu L_n\left(\hat{\boldsymbol{\beta}}_1^o,0\right)+(1-\mu)L_n\left(\boldsymbol{\beta}_1^*,0\right)\leq L_n\left(\boldsymbol{\beta}_1^*,0\right)=L_n(\boldsymbol{\beta}^*).\tag{8}$$

结合式(8)与三角不等式，有

$$\begin{aligned}E\left[\hat{v}_n(\tilde{\boldsymbol{\beta}})-\hat{v}_n(\boldsymbol{\beta}^*)\right]&=\left\{\hat{v}_n(\boldsymbol{\beta}^*)-E\hat{v}_n(\boldsymbol{\beta}^*)\right\}-\left\{\hat{v}_n(\tilde{\boldsymbol{\beta}})-E\hat{v}_n(\tilde{\boldsymbol{\beta}})\right\}+L_n(\tilde{\boldsymbol{\beta}})-L_n(\boldsymbol{\beta}^*)\\&\quad+n\lambda_n\left(\left\|\boldsymbol{\beta}_1^*\right\|_1-\left\|\tilde{\boldsymbol{\beta}}_1\right\|_1\right)+n\mu_n\left(\left\|\boldsymbol{\beta}_1^*\right\|_2^2-\left\|\tilde{\boldsymbol{\beta}}_1\right\|_2^2\right)\\&\leq nZ_n(M_n)+n\lambda_n\left\|\boldsymbol{\beta}_1^*-\tilde{\boldsymbol{\beta}}_1\right\|_1+n\mu_n\left(\left\|\boldsymbol{\beta}_1^*\right\|_2^2-\left\|\tilde{\boldsymbol{\beta}}_1\right\|_2^2\right).\end{aligned}\tag{9}$$

定义事件

$$A_n=\left\{Z_n(M_n)\leq 2M_nn^{-1/2}\sqrt{s\log n}\right\},$$

根据引理 2，可得

$$P(A_n)\geq 1-\exp\left(-c_0s(\log n)/8\right).$$

又由 Cauchy-Schwarz 不等式，有

$$n\lambda_n\left\|\boldsymbol{\beta}_1^*-\tilde{\boldsymbol{\beta}}_1\right\|_1\leq n\lambda_n\sqrt{s}M_n,$$

$$n\mu_n\left(\left\|\boldsymbol{\beta}_1^*\right\|_2^2-\left\|\tilde{\boldsymbol{\beta}}_1\right\|_2^2\right)\leq 2C_Mn\mu_n\left\|\boldsymbol{\beta}_1^*-\tilde{\boldsymbol{\beta}}_1\right\|_1\leq 2C_Mn\mu_n\sqrt{s}M_n,$$

其中，$C_M=\max\left\{\beta_1^*,\cdots,\beta_s^*,\tilde{\beta}_1,\cdots,\tilde{\beta}_s\right\}$。因此，在事件 $A_n$ 上，根据不等式(9)有

$$E\left[\hat{v}_n(\tilde{\boldsymbol{\beta}})-\hat{v}_n(\boldsymbol{\beta}^*)\right]\leq\left(2\sqrt{sn(\log n)}+n\lambda_n\sqrt{s}+2C_Mn\mu_n\sqrt{s}\right)M_n.$$

取 $M_n=2\sqrt{s/n}+\lambda_n\sqrt{s}+\mu_n\sqrt{s}$，由假设 $\lambda_n\sqrt{s}\kappa_n\to 0$，$\mu_n\sqrt{s}\kappa_n\to 0$ 和条件(A7)可知 $M_n=o\left(\kappa_n^{-1}s^{-1/2}\right)$，再结合式(5)，则在事件 $A_n$ 上有

$$\frac{1}{2}c_0n\left\|\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1^*\right\|_2^2\leq\left(2\sqrt{sn(\log n)}+n\lambda_n\sqrt{s}+2C_Mn\mu_n\sqrt{s}\right)M_n,$$

进一步，得

$$\left\|\tilde{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1^*\right\|_2\leq O\left(\sqrt{s(\log n)/n}+\lambda_n\sqrt{s}+\mu_n\sqrt{s}\right).$$

根据 $\left\|\boldsymbol{\beta}_1^*-\tilde{\boldsymbol{\beta}}\right\|_2\leq M_n$，可得 $\left\|\hat{\boldsymbol{\beta}}_1^o-\boldsymbol{\beta}_1^*\right\|_2\leq 2M_n$，故在事件 $A_n$ 上，有

$$\left\|\hat{\boldsymbol{\beta}}_1^o-\boldsymbol{\beta}_1^*\right\|_2 \le O\left(\sqrt{s(\log n)/n}+\lambda_n\sqrt{s}+\mu_n\sqrt{s}\right),$$

可知事件 $A_n$ 发生的概率大于等于 $1-c_1n^{-c_2s}$，定理 1 得证。

**定理 2 的证明：**

此处先给出引理 4 的证明，其是定理 2 证明的基石。对于固定的 $j\in\{s+1,s+2,\cdots,p\}$，$\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}}\right)^{\mathrm{T}}\in\mathcal{N}$。定义

$$\gamma_{\boldsymbol{\beta},j}\left(\hat{\boldsymbol{x}}_i,\hat{y}_j\right)=\hat{x}_{ij}\left[\rho_\tau'\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)-\rho_\tau'(\varepsilon_i)-E\left[\rho_\tau'\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)-\rho_\tau'(\varepsilon_i)\right]\right],$$

其中，$\hat{\boldsymbol{x}}_i^{\mathrm{T}}=\left(\hat{x}_i,\cdots,\hat{x}_{ip}\right)$ 是设计矩阵 $\hat{\boldsymbol{x}}$ 的第 $i$ 行。

首先，需对引理 4 做如下分解：

$$\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left\|\frac{1}{n}\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'\left(\hat{\boldsymbol{y}}-\boldsymbol{S}\boldsymbol{\beta}_1\right)\right\|_\infty \le \sup_{\boldsymbol{\beta}\in\mathcal{N}}\left\|\frac{1}{n}\boldsymbol{Q}^{\mathrm{T}}E\left[\rho_\tau'\left(\hat{\boldsymbol{y}}-\boldsymbol{S}\boldsymbol{\beta}_1\right)-\rho_\tau'(\varepsilon)\right]\right\|_\infty$$
$$+\left\|\frac{1}{n}\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'(\varepsilon)\right\|_\infty+\max_{j>s}\sup_{\boldsymbol{\beta}\in\mathcal{N}}\frac{1}{n}\sum_{i=1}^n\left|\gamma_{\boldsymbol{\beta},j}\left(\hat{\boldsymbol{x}}_i,\hat{y}_i\right)\right|,$$

根据上式可知，若能证明以下各式至少以 $1-O\left(c_1p^{-c_2}\right)$ 的概率成立，那么引理 4 可得证。

$$I_1=\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left\|\frac{1}{n}\boldsymbol{Q}^{\mathrm{T}}E\left[\rho_\tau'\left(\hat{\boldsymbol{y}}-\boldsymbol{S}\boldsymbol{\beta}_1\right)-\rho_\tau'(\varepsilon)\right]\right\|_\infty=o(\lambda_n), \tag{10}$$

$$I_2=\frac{1}{n}\left\|\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'(\varepsilon)\right\|_\infty=o(\lambda_n), \tag{11}$$

$$I_3=\max_{j>s}\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{1}{n}\sum_{i=1}^n\gamma_{\boldsymbol{\beta},j}\left(\hat{\boldsymbol{x}}_i,\hat{y}_i\right)\right|=o_p(\lambda_n), \tag{12}$$

首先证明 $I_1$ 成立，将式(10)改写为

$$I_1=\max_{j>s}\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{1}{n}\sum_{i=1}^n\hat{x}_{ij}E\left[\rho_\tau'(\varepsilon_i)-\rho_\tau'\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right]\right|. \tag{13}$$

根据条件(A6)，有

$$E\left[\rho_\tau'(\varepsilon_i)-\rho_\tau'\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right]=F_i\left(\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)\right)-F_i(0)=f_i(0)\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)+\tilde{I}_i,$$

其中，$F_i(t)$ 是 $\varepsilon_i$ 的累计分布函数，

$$\tilde{I}_i=F_i\left(\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)\right)-F_i(0)-f_i(0)\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right),$$

对任意 $j>s$，有

$$\sum_{i=1}^n\hat{x}_{ij}E\left[\rho_\tau'(\varepsilon_i)-\rho_\tau'\left(\hat{y}_i-\hat{\boldsymbol{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right]=\sum_{i=1}^n\left(f_i(0)\hat{x}_{ij}\boldsymbol{S}_i^{\mathrm{T}}\right)\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)+\sum_{i=1}^n\hat{x}_{ij}\tilde{I}_i.$$

结合式(13)及 Cauchy-Schwarz 不等式，可推出

$$I_1\le\left\|\frac{1}{n}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{S}\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right)\right\|_\infty+\max_{j>s}\left|\frac{1}{n}\sum_{i=1}^n\hat{x}_{ij}\tilde{I}_i\right|. \tag{14}$$

依次考虑不等式(14)右边的两项，由条件(A8)可得第一项的上界被限制为

$$\left\| \frac{1}{n} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{HS} \left( \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right) \right\|_{\infty} \le \left\| \frac{1}{n} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{HS} \right\|_{2,\infty} \left\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right\|_2 < \frac{\lambda_n}{2}, \tag{15}$$

再根据条件(A6)~(A7)及$\left|\tilde{I}_i\right| \le C\left(\boldsymbol{S}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*\right)\right)^2$，可得第二项的上界被限制为

$$\max_{j>s} \left| \frac{1}{n} \sum_{i=1}^{n} \hat{x}_{ij} \tilde{I}_i \right| \le \frac{\kappa_n}{n} \sum_{i=1}^{n} \left| \tilde{I}_i \right| \le C \frac{\kappa_n}{n} \sum_{i=1}^{n} \left( \boldsymbol{S}_i^{\mathrm{T}} \left( \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right) \right)^2 \le C\kappa_n \left\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right\|_2^2 .$$

又因为$\boldsymbol{\beta} \in \mathcal{N}$，由假设条件$\lambda_n^{-1}\kappa_n\gamma_n^2 = o(1)$有

$$\max_{j>s} \left| \frac{1}{n} \sum_{i=1}^{n} \hat{x}_{ij} \tilde{I}_i \right| \le C\kappa_n\gamma_n^2 = o\left(\lambda_n\right), \tag{16}$$

再将上述不等式(16)与式(15)代入式(14)，可得$I_1$成立。接下来证明$I_2$，根据引理 3，如果$\lambda_n > 2\sqrt{(1+c)(\log p)/n}$，则有

$$P\left( \left\| \boldsymbol{Q}^{\mathrm{T}} \rho_\tau'(\varepsilon) \right\|_{\infty} \ge n\lambda_n \right) \le \sum_{j=s+1}^{p} 2\exp\left( -\frac{n^2\lambda_n^2}{4\sum_{i=1}^{n} \hat{x}_{ij}^2} \right) = 2\exp\left[ \log(p-s) - \frac{n\lambda_n^2}{4} \right] \le O\left(c_1 p^{-c_2}\right),$$

因此，$I_2$至少以概率$1-O\left(c_1 p^{-c_2}\right)$成立。

现将利用 Bühlmann [15]中的推论 14.4 证明$I_3$成立。首先，对每个固定的$j$，定义函数空间$\Gamma_j = \left\{ \gamma_{\boldsymbol{\beta},j} : \boldsymbol{\beta} \in \mathcal{N} \right\}$，则对任意的$\gamma_{\boldsymbol{\beta},j} \in \Gamma_j$，有$E\left[ \gamma_{\boldsymbol{\beta},j}\left( \hat{\boldsymbol{x}}_i, \hat{y}_j \right) \right] = 0$。又因函数$\rho_\tau'$有界，可得

$$\left\| \gamma_{\boldsymbol{\beta},j} \right\|_n \equiv \left( \frac{1}{n} \sum_{i=1}^{n} \gamma_{\boldsymbol{\beta},j}^2 \left( \hat{\boldsymbol{x}}_i, \hat{y}_i \right)^2 \right)^{1/2} \le 2 .$$

此外，对任意的$\boldsymbol{\beta} \in \left( \boldsymbol{\beta}_1^{\mathrm{T}}, \boldsymbol{\beta}_2^{\mathrm{T}} \right)^{\mathrm{T}} \in \mathcal{N}$和$\tilde{\boldsymbol{\beta}} \in \left( \tilde{\boldsymbol{\beta}}_1^{\mathrm{T}}, \tilde{\boldsymbol{\beta}}_2^{\mathrm{T}} \right)^{\mathrm{T}} \in \mathcal{N}$，由条件(A6)和均值定理可得

$$\begin{aligned} & E\left[ \rho_\tau'\left( \hat{y}_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}} \boldsymbol{\beta} \right) - \rho_\tau'\left( \varepsilon_i \right) \right] - E\left[ \rho_\tau'\left( \hat{y}_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}} \tilde{\boldsymbol{\beta}} \right) - \rho_\tau'\left( \varepsilon_i \right) \right] \\ & = F_i\left( \boldsymbol{S}_i^{\mathrm{T}} \left( \tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^* \right) \right) - F_i\left( \boldsymbol{S}_i^{\mathrm{T}} \left( \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right) \right) = f_i\left( a_{1i} \right) \boldsymbol{S}_i^{\mathrm{T}} \left( \boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1 \right), \end{aligned} \tag{17}$$

其中，$a_{1i}$介于$\boldsymbol{S}_i^{\mathrm{T}}\left( \tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^* \right)$与$\boldsymbol{S}_i^{\mathrm{T}}\left( \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^* \right)$之间。又因$f_i'(u)$一致有界，结合条件(A7)和式(17)，有

$$\begin{aligned} & \left| \hat{x}_{ij} E\left[ \rho_\tau'\left( \hat{y}_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}} \boldsymbol{\beta} \right) - \rho_\tau'\left( \varepsilon_i \right) \right] - \hat{x}_{ij} E\left[ \rho_\tau'\left( \hat{y}_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}} \tilde{\boldsymbol{\beta}} \right) - \rho_\tau'\left( \varepsilon_i \right) \right] \right| \\ & \le C\left| \hat{x}_{ij} \boldsymbol{S}_i^{\mathrm{T}} \left( \boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1 \right) \right| \le C\left\| \hat{x}_{ij} \boldsymbol{S}_i \right\|_2 \left\| \boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1 \right\|_2 \le C\sqrt{s}\kappa_n^2 \left\| \boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1 \right\|_2, \end{aligned} \tag{18}$$

其中，$C$是大于零的常数。

由 Bühlmann [15]中引理 14.27 可知$R^s$空间的球$\mathcal{N}$可被$\left(1+4\gamma_n/\delta\right)^s$个半径为$\eta$的球所覆盖，又因$\rho_\tau'\left( \hat{y}_i - \hat{\boldsymbol{x}}_i^{\mathrm{T}} \boldsymbol{\beta} \right) - \rho_\tau'\left( \varepsilon_i \right)$只取 3 种结果，结合式(18)可得$\Gamma_j$的覆盖数为$N\left( 2^{2-k}, \Gamma_j, \|\cdot\|_2 \right) = 3\left( 1 + C^{-1} 2^k \gamma_n s^{1/2} \kappa_n^2 \right)^s$。因此，

对任意的$0 \le k \le \left( \log_2 n \right)/2$，有

$$\log\left( 1 + N\left( 2^{2-k}, \Gamma_j, \|\cdot\|_2 \right) \right) \le 4\left( 1 + C^{-1}\gamma_n s^{3/2} \kappa_n^2 \right) 2^{2k} .$$

又由 Bühlmann [15]中的推论 14.4 可知，对任意的$t > 0$，有

$$P\left( \sup_{\boldsymbol{\beta} \in \mathcal{N}} \left| \frac{1}{n} \sum_{i=1}^{n} \gamma_{\boldsymbol{\beta},j} \left( \hat{\boldsymbol{x}}_i, \hat{y}_j \right) \right| \ge \frac{8}{\sqrt{n}} \left( 3\sqrt{1 + C^{-1}\gamma_n s^{3/2} \kappa_n^2} \log_2 n + 4 + 4t \right) \right) \le 4\exp\left( \frac{nt^2}{8} \right),$$

令 $t=\sqrt{C(\log p)/n}$，其中 $C>0$ 为充分大的常数，则可得

$$P\left(\max_{j>s}\sup_{\beta\in\mathcal{N}}\left|\frac{1}{n}\sum_{i=1}^{n}\gamma_{\beta,j}\left(\hat{x}_i,\hat{y}_j\right)\right|\geq\frac{24}{\sqrt{n}}\sqrt{1+C^{-1}\gamma_n s^{3/2}\kappa_n^2}\log_2 n\right)\leq 4(p-s)\exp\left(-\frac{C\log p}{8}\right)\to 0.$$

因此，若 $\sqrt{1+\gamma_n s^{3/2}\kappa_n^2}\log_2 n=o\left(\sqrt{n}\lambda_n\right)$，那么 $I_3$ 至少以 $1-O\left(c_1p^{-c_2}\right)$ 的概率成立，从而引理 4 得证。

下面接着证明定理 2，由于 $\hat{\boldsymbol{\beta}}_1^o$ 是 $L_n(\boldsymbol{\beta}_1,0)$ 的最小值点且满足 KTT 条件。若要证明 $\hat{\boldsymbol{\beta}}_{Enet}=\left(\left(\hat{\boldsymbol{\beta}}_1^o\right)^{\mathrm{T}},0^{\mathrm{T}}\right)\in R^P$ 是目标函数 $L_n(\boldsymbol{\beta})$ 的全局最小值，只需验证如下条件

$$\left\|\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'\left(\hat{\boldsymbol{y}}-\boldsymbol{S}\hat{\boldsymbol{\beta}}_1^o\right)\right\|_\infty<n\lambda_n, \tag{19}$$

其中，对任意 $n$ 维向量 $\boldsymbol{u}=(u_1,\cdots,u_n)^{\mathrm{T}}$，有 $\rho_\tau'(\boldsymbol{u})=\left(\rho_\tau'(u_i),\cdots,\rho_\tau'(u_n)\right)^{\mathrm{T}}$，则根据 KKT 条件及 $L_n(\boldsymbol{\beta})$ 为凸函数的性质，可知 $\hat{\boldsymbol{\beta}}_{Enet}$ 是 $L_n(\boldsymbol{\beta})$ 的全局最小值点。定义事件

$$B_1=\left\{\left\|\hat{\boldsymbol{\beta}}_1^o-\boldsymbol{\beta}_1^*\right\|_2\leq\gamma_n\right\},$$

$$B_2=\left\{\sup_{\beta\in\mathcal{N}}\left\|\boldsymbol{Q}^{\mathrm{T}}\rho_\tau'\left(\hat{\boldsymbol{y}}-\boldsymbol{S}\boldsymbol{\beta}_1\right)\right\|_\infty<n\lambda_n\right\},$$

其中，$\gamma_n$ 同定理 1 中所定义，并且

$$\mathcal{N}=\left\{\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}}\right)\in R^P:\left\|\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\right\|_2\leq\gamma_n,\boldsymbol{\beta}_2=0\in\boldsymbol{R}^{P-S}\right\},$$

那么，结合定理 1 和引理 4 可知

$$P\left(B_1\cap B_2\right)\geq 1-o\left(c_1n^{-c_2s}\right).$$

又因为 $\hat{\boldsymbol{\beta}}\in\mathcal{N}$ 属于事件 $B_1$，故在事件 $B_1\cap B_2$ 上不等式(19)成立，定理 2 得证。

## 5. 数值模拟

本节将通过数值模拟来考察所提出方法的有限样本性质。从如下部分线性模型产生数据

$$\boldsymbol{Y}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}+g(\boldsymbol{Z})+\boldsymbol{\varepsilon},$$

其中，真实的回归参数被固定为 $\boldsymbol{\beta}=(0.5,1,2,1.5,0.8,1.75,0.75,0,\cdots,0)^{\mathrm{T}}$，参数数量 $p=400$，样本容量 $n=100$，协变量 $X\sim N_p(0,V)$，其中 $V_{i,j}=0.5^{|i-j|}$；非参数函数 $g(Z)=\sin(2\pi T)$，其中 $T$ 服从 $[-1,1]$ 上的均匀分布。

为了验证所提方法关于厚尾分布的稳健性，考虑如下误差分布：标准正态分布 $\varepsilon\sim N(0,1)$；混合正态分布 MN1: $\varepsilon\sim 0.9N(0,1)+0.1N\left(0,5^2\right)$；自由度为 4 的 $t$ 分布 $\varepsilon\sim t(4)$；Laplace 分布；Cauchy 分布。本节将在这五种误差分布下，将所提的弹性网惩罚分位数回归(Q-EN)与弹性网惩罚最小二乘回归(EN)、Lasso 惩罚分位数回归(Q-lasso)和 Ridge 惩罚分位数回归(Q-ridge)进行比较，其中 $\tau=0.5$。通过计算以下 4 个值来评估所提方法：

1) $L_1$ loss：即 $\beta-\beta^*$ 的 $L_1$ 范数。

2) $L_2$ loss：即 $\beta-\beta^*$ 的 $L_2$ 范数。

3) FP：选入模型的噪音协变量的数量。

4) FN：未选入模型的信号协变量的数量。

程序重复运行 100 次，模拟结果如表 1 所示，其中各项数值为其均值。

**Table 1.** Simulation results of relevant covariates
**表 1.** 相关协变量模拟结果

| Error | Method | $L_1$ loss | $L_2$ loss | FP | FN |
|---|---|---|---|---|---|
| $N(0,1)$ | EN | 5.3876 | 8.3052 | 38.14 | 0.672 |
| | Q-EN | **3.1422** | **2.5163** | 124.83 | 0.871 |
| | Q-lasso | 11.4103 | 21.1308 | 12.37 | 2.05 |
| | Q-ridge | 8.6309 | 6.0135 | 392 | 0 |
| MN1 | EN | 5.6724 | 8.9104 | 36.43 | 0.857 |
| | Q-EN | **4.1052** | **3.7861** | 138.56 | 0.672 |
| | Q-lasso | 10.7865 | 23.6209 | 79.82 | 1.54 |
| | Q-ridge | 7.2538 | 6.4618 | 392 | 0 |
| $t_4$ | EN | 9.0324 | 12.0174 | 40.96 | 0.564 |
| | Q-EN | **3.0195** | **2.8739** | 16.38 | 4.35 |
| | Q-lasso | 8.1005 | 16.0534 | 7.5 | 1.782 |
| | Q-ridge | 6.8751 | 3.5213 | 392 | 0 |
| Laplace | EN | 5.1462 | 8.3367 | 48.57 | 0.243 |
| | Q-EN | **3.9788** | **1.8142** | 130.026 | 0.538 |
| | Q-lasso | 12.9032 | 18.9237 | 40.32 | 2.45 |
| | Q-ridge | 8.0914 | 5.7832 | 392 | 0 |
| Cauchy | EN | 9.5261 | 8.6359 | 55.24 | 4.78 |
| | Q-EN | **4.2843** | 6.1308 | 12.78 | 7.56 |
| | Q-Lasso | 21.7933 | 75.4103 | 108.82 | 3.59 |
| | Q-ridge | 10.8637 | **4.6724** | 392 | 0 |

比较 Q-EN 和 EN：在这五种误差分布情形下，Q-EN 的 $L_1$ 损失和 $L_2$ 损失都较 EN 更低。当误差分布为 MN1 时，Q-EN 比 EN 的 FN 更小。当误差分布为 $t_4$ 和 Cauchy 时，Q-EN 较 EN 模型具有更小的 FP。当误差分布为 Laplace 和 $N(0,1)$ 时，EN 的 FP 和 FN 都优于 Q-EN。总体而言，在面对具有厚尾分布和异常值，如 $t_4$ 等情形时，分位数回归较最小二乘估计具有更稳健的特性。

比较 Q-EN 和 Q-lasso、Q-ridge：在前四种误差分布情形下，Q-EN 的 $L_1$ 损失和 $L_2$ 损失都明显更低。当误差分布为 Cauchy 时，虽然 Q-ridge 具有比 Q-EN 更低的 $L_2$ 损失，但它选择了模型中的所有变量，并没有进行有效的变量选择。当误差分布为 $N(0,1)$、MN1、Laplace 和 $t_4$ 时，Q-lasso 具有更小的 FP，表明 Lasso 惩罚项倾向于将真实有效的信号协变量选择到模型中。当误差分布为 $N(0,1)$、MN1 和 Laplace 时，相比 Q-lasso，Q-EN 具有更小的 FN，说明 Q-EN 剔除大量不相关的预测因子，表现更好。

通过在不同的误差分布情形下，与其他变量方法进行比较，表明 Q-EN 的总体性能优于其他四种方法。Q-EN 可以有效的进行变量筛选产生稀疏解，且在面对厚尾分布或异常值时仍表现优异。

## 6. 总结

针对高维数据下的部分线性模型变量选择问题，提出了将弹性网络法和分位数回归相结合的正则化方法。该方法可以有效地处理高维数据中的多重共线性与群组效应问题，还可以在面对数据含有极端值、离群点或是呈现尖峰、厚尾分布时，仍保持较好稳健性。理论结果表明，在一定条件下模型估计量具有相合性和稀疏性，可将不相关变量压缩至零，从而进行有效的变量选择。在数值模拟方面，通过比较不同惩罚函数和不同损失函数在五种误差分布下的估计效果，证实了所提估计方法优于其他变量选择方法，具有更强的稳健性和有效性。在实践应用中，所提模型方法可为生物医学、基因学等领域进行高维数据下的实证分析提供理论支撑。利用此模型方法对高维基因学数据进行基因检测分析，探究与疾病相关程度高的协变量基因，为药物研究提供方向，可推动相关药物研发领域的发展。

## 基金项目

重庆市自然科学基金(CSTB2022NSCQ-MSX1370)。

## 参考文献

[1] Engle, R.F., Granger, C.W.J., *et al.* (1986) Semiparametric Estimates of the Relation between Weather and Electricity Sales. *Journal of the American Statistical Association*, **81**, 310-320. https://doi.org/10.1080/01621459.1986.10478274

[2] Hoerl, A.E., Kennard, R.W., *et al.* (1985) Practical Use of Ridge Regression: A Challenge Met. *Journal of the Royal Statistical Society: Series C*, **34**, 114-120. https://doi.org/10.2307/2347363

[3] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*: *Series B*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[4] Fan, J.Q. and Li, R.Z. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. https://doi.org/10.1198/016214501753382273

[5] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society*: *Series B*, **67**, 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[6] Xie, H.L. and Huang, J. (2009) SCAD-Penalized Regression in High-Dimensional Partially Linear Models. *The Annals of Statistics*, **37**, 673-696. https://doi.org/10.1214/07-AOS580

[7] 杨宜平, 薛留根, 王学娟. 高维部分线性模型中的变量选择[J]. 北京工业大学学报, 2011, 37(2): 291-295.

[8] Chen, B., Yu, Y., *et al.* (2012) Profifiled Adaptive Elastic-Net Procedure for Partially Linear Models with High-Dimensional Covariates. *Journal of Statistical Planning and Inference*, **142**, 1733-1745. https://doi.org/10.1016/j.jspi.2012.02.035

[9] Wang, X.L. and Wang, M.Q. (2017) Adaptive Group Bridge Estimation for High-Dimensional Partially Linear Models. *Journal of Inequalities and Applications*, **2017**, 158. https://doi.org/10.1186/s13660-017-1432-x

[10] 赖秋楠, 李玉杰, 李高容. 超高维部分线性模型的 PGFR 变量筛选[J]. 应用概率统计, 2017, 33(6): 608-624.

[11] Koenker, R. (2005) Quantiles Regression. Cambridge University Press, Cambridge.

[12] Su, M.H. and Wang, W.J. (2021) Elastic Net Penalized Quantile Regression Model. *Journal of Computational and Applied Mathematics*, **392**, Article ID: 113462. https://doi.org/10.1016/j.cam.2021.113462

[13] Fan, J.Q., Fan, Y.Y., *et al.* (2014) Adaptive Robust Variable Selection. *The Annals of Statistics*, **42**, 324-351. https://doi.org/10.1214/13-AOS1191

[14] Chen, B.C., Yu, Y., Zou, H., *et al.* (2012) Profiled Adaptive Elastic-Net Procedure for Partially Linear Models with High-Dimensional Covariates. *Journal of Statistical Planning and Inference*, **142**, 1733-1745. https://doi.org/10.1016/j.jspi.2012.02.035

[15] Bühlmann, P. and Geer, S. (2011) Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer, Heidelberg. https://doi.org/10.1007/978-3-642-20192-9