Advances in Applied Mathematics 应用数学进展, 2025, 14(1), 147-158
Published Online January 2025 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2025.141018

Hans汉斯

高维部分线性可加稳健Expectile回归模型

徐朝丹

西南大学数学与统计学院，重庆

收稿日期：2024年12月22日；录用日期：2025年1月15日；发布日期：2025年1月24日

摘　要

高维数据一般因具有异方差或非齐次协变量而具有异质性，分位数回归和expectile回归是分析异质高维数据的有力工具，但前者由于损失函数非光滑的特性在计算方面存在较大挑战，而后者会因异常值而不稳健。本文利用一类稳健的非对称损失函数来研究部分线性可加模型的稳健expectile回归，用B样条基函数近似非参数部分，利用加入非凸惩罚的正则化方法来实现变量筛选并进行参数估计。该方法的优势在于：(1) 通过取不同分位水平得到响应变量更完整的条件分布，从而探索数据的异质性分布；(2) 部分线性的模型结构兼顾了线性解释变量和非线性解释变量，一方面增加了模型的灵活性，同时也具有一定的模型可解释性；(3) 稳健expectile回归估计比分位数回归方法计算效率高，比expectile回归稳健。数值模拟和实际数据分析均显示了该方法在模型估计和计算效率上的优势。

关键词

稳健，Expectile回归，半参数模型，B-Spline

Partially Linear Additive Robust Expectile Regression in High Dimension

Chaodan Xu

School of Mathematics and Statistics, Southwest University, Chongqing

Received: Dec. 22nd, 2024; accepted: Jan. 15th, 2025; published: Jan. 24th, 2025

Abstract

High-dimensional data are generally heterogeneous due to heteroskedasticity or non-homogeneous covariates. Quantile regression and expectile regression are powerful tools for analyzing heterogeneous high-dimensional data, but the former is a great challenge in calculation due to the non-smooth nature of the loss function, while the latter is unstable due to outliers. In this paper, a class of robust asymmetric loss functions is used to study the robust expectile regression of partial

文章引用: 徐朝丹. 高维部分线性可加稳健 Expectile 回归模型[J]. 应用数学进展, 2025, 14(1): 147-158.
DOI: 10.12677/aam.2025.141018

linear additive models, the B-spline basis function is used to approximate the non-parametric part, and the regularization method with non-convex penalty is used to realize variable screening and parameter estimation. The advantages of this method are: (1) A more complete conditional distribution of response variables can be obtained by taking different quantile levels, so as to explore the heterogeneity distribution of data; (2) The partial linear model structure takes into account both linear explanatory variables and nonlinear explanatory variables, which increases the flexibility of the model on the one hand, and has a certain interpretability of the model; (3) The robust expectile regression estimation score digit regression method has higher computational efficiency and is more robust than the expectile regression. Both numerical simulation and actual data analysis show the advantages of the proposed method in model estimation and computational efficiency.

Keywords

Robust, Expectile Regression, Semiparametric Model, B-Spline

1. 引言

半参数模型拥有线性模型的直观性和非参数模型的灵活性而常被作为处理高维数据的重要工具[1] [2]。本文主要研究半参数模型中的部分线性可加模型(partially linear additive models，简记为 PLAMs)，是广义可加非参数回归模型的特殊类型和多元线性回归模型的推广[3]。PLAMs 的估计和推断理论已有丰富的研究结果[4]-[7]。

在实际研究中，高维数据常有异质性表现，噪声变量的分布呈现重尾型或非对称尾型。分位数回归(quantile regression)是有效分析异质性的方法，在 PLAMs 框架下已有很多研究成果产出。Tadao [8]于 2014 年用核密度法近似非参数部分的部分线性可加分位数回归，使用两步估计的方法并得出估计值的渐近性质；Sherwood 和 Wang [9]于 2016 年提出超高维部分线性可加分位数回归，用带惩罚的分位数回归分析超高维数据的异质性并证明了估计量的渐进正态性。

虽然分位数回归受到研究者们的普遍关注，但其损失函数不光滑，在高维情形的计算负担过重。而非对称的最小平方回归(expectile regression) [10]因其具有光滑的损失函数，使得计算效率大幅提升，是分析异质性数据的另一重要工具，同时它也是风险分析领域的常用工具，在金融风险领域中受到关注[11] [12]。Fabian [13]等人于 2013 年提出部分线性可加 expectile 回归并得出估计值的渐近性质和 Bootstrap 置信区间；Zhao [14]等人于 2019 年提出高维部分线性可加 expectile 回归，并研究模型的估计与变量选择。

虽然 expectile regression 比 quantile regression 计算速度快，却易受异常值的影响造成较大的估计误差，远不如 quantile regression 稳健。可见 expectile regression 和 quantile regression 都是分析数据异质性的典型方法，但都存在不同的缺点。Man 等人[15]等人 2022 年提出一种稳健的 expectile 回归，用 Lipschitz 连续局部平方的函数代替 expectile regression 损失函数的平方项，并调节能随数据自适应的参数 γ 让偏差与稳健达到平衡，解决了 quantile regression 计算速度慢和 expectile regression 不稳健的问题又兼具了二者的优点，于是本文将此种方法推广到更加灵活的 PLAMs。

本文主要研究高维部分线性可加稳健 expectile 回归模型。全文内容安排如下：第 2 节详细介绍部分

线性可加稳健 expectile 回归模型(retire-PLAMs)，第 3 节介绍带非凸惩罚的高维部分线性可加 expectile 回归及算法，第 4 节是数值模拟，第 5 节是实证研究，第 6 节是总结。

2. 高维部分线性可加稳健 Expectile 回归

2.1. 一类非对称的稳健平方损失

Newey 和 Powell [10]受 quantile regression 的启发于 1987 年提出 expectile regression。设 $Z\in\mathbb{R}$ 是一个具有有限矩的随机变量，同 quantile 回归的中 τ-th quantile 思想，Z 的 τ-th expectile 可被定义为

$$e_\tau := \arg\min_{u\in\mathbb{R}} E\left\{\eta_\tau(Z-u)-\eta_\tau(Z)\right\},\qquad \tau\in(0,1) \tag{1}$$

其中，$\eta_\tau(u)=\left|\tau-1(u<0)\right|\cdot\dfrac{u^2}{2}$ 是非对称的平方损失。$\tau=1/2$ 时，有 $e_{1/2}(Z)=E(Z)$，可见 expectile regression 是均值回归的不对称类型，expectile 的命名也来自“expectation”和“quantile”的组合。expectile regression 比 quantile regression 计算效率快，却不如 quantile regression 稳健，对重尾分布敏感，尤其是在高维情形。Man 等人[15]于 2022 年提出稳健 expectile 回归，受文献[16]的启发用 Lipschitz 连续和局部强凸的函数替换 expectile regression 损失函数中的平方部分，让 $\gamma>0$ 作为平衡偏差与稳健的调节参数。对于 $u\in\mathbb{R}$，令 $\ell_\gamma(u)=\gamma^2\ell(u/\gamma)$，其中 $\ell:\mathbb{R}\mapsto[0,\infty)$ 要满足以下三个条件：

(i) $\ell'(0)=0$ 且对于所有的 $u\in\mathbb{R}$ 有 $\left|\ell'(u)\right|\le\min(a_1,|u|)$；(ii) $\ell''(0)=1$ 且对任意的 $|u|\le a_3$ 都有 $\ell''(u)\ge a_2$；(iii) 对于所有的 $u\in\mathbb{R}$ 有 $\left|\ell'(u)-u\right|\le u^2$ 成立，其中 a_1,a_2 和 a_3 都是正常数。

满足上述条件的稳健损失函数有 Huber 损失：$\ell(u)=\min\left\{u^2/2,|u|-1/2\right\}$，以及光滑近似 Huber 损失的函数，例如 pseudo-Huber 损失：$\ell(u)=\sqrt{1+u^2}-1$ 和 $\ell(u)=\log\left(\mathrm{e}^u/2+\mathrm{e}^{-u}/2\right)$，smoothed Huber 损失：$\ell(u)=\min\left\{u^2/2-|u|^3/6,|u|/2-1/6\right\}$。本文选择以下函数作为稳健 expectile 回归的损失函数：

$$L_{\tau,\gamma}(u)=\left|\tau-\mathbf{I}(u<0)\right|\cdot\begin{cases}u^2/2, & |u|<\gamma,\\ \gamma u-\gamma^2/2, & |u|>\gamma.\end{cases} \tag{2}$$

其中，$\tau\in(0,1)$，$\gamma>0$ 是能被数据自适应校准的稳健参数。一阶导和二阶导为：

$$L'_{\tau,\gamma}(u)=\begin{cases}-(1-\tau)\gamma, & u<-\gamma,\\ (\tau-1)u, & -\gamma\le u<0,\\ \tau u, & 0\le u\le\gamma,\\ \tau\gamma, & u>\gamma.\end{cases}\qquad L''_{\tau,\gamma}(u)=\begin{cases}\tau-1, & -\gamma\le u<0,\\ \tau, & 0\le u<\gamma,\\ 0, & \text{其他.}\end{cases}$$

2.2. 高维部分线性可加稳健 Expectile 回归

若第 i 个样本观察值为 $\{y_i,\boldsymbol{x}_i,\boldsymbol{z}_i\}$，其中 $y_i\in\mathbb{R}$ 为响应变量，$\boldsymbol{x}_i$ 和 $\boldsymbol{z}_i$ 为解释变量，$\boldsymbol{x}_i=(x_{i1},\cdots,x_{ip_n})'$ 是 p_n 维的向量，p_n 可随 n 的变化发散，$\boldsymbol{z}_i=(z_{i1},\cdots,z_{id})'$ 是 d 维向量，d 是固定常数，$i=1,\cdots,n$。对于一个给定的分位水平 $\tau\in(0,1)$，本文考虑如下部分线性可加 expectile 回归模型：

$$y_i=x_i^{\mathrm{T}}\boldsymbol{\beta}_0+g_0(\boldsymbol{z}_i)+\varepsilon_i \tag{3}$$

其中，$\boldsymbol{\beta}_0$ 是维数为 p_n 的线性回归系数向量；$g_0(\boldsymbol{z}_i)=g_{00}+\sum_{j=1}^{d}g_{0j}(z_{ij})$ 是可加的非参数部分，$g_{00}\in R$，并假设 $\mathbb{E}\left[g_{0j}(z_{ij})\right]=0$；$\varepsilon_i$ 为独立的随机噪声变量，满足 $e_\tau(\varepsilon_i)=0$。模型(1)允许 $\boldsymbol{\beta}_0$ 随不同的 τ 而变化，因此能得出给定 $\boldsymbol{x}_i$ 和 $\boldsymbol{z}_i$ 后 y_i 较全的条件分布情况，进而探究解释变量和反映变量之间更完整的关系[17]。

高维统计推断一般假设系数向量 $\boldsymbol{\beta}_0=\left(\beta_{01},\cdots,\beta_{0p_n}\right)'$ 是稀疏的，记 $A=\left\{j:\beta_{0j}\neq 0,1\le j\le p_n\right\}$ 是系数不为零的变量下标集合，个数 $q_n=|A|$，A 在被估计前不是明确的，不失一般性地记 $\boldsymbol{\beta}_0=\left(\boldsymbol{\beta}_A',\mathbf{0}'_{p_n-q_n}\right)'$，其中前 q_n 个系数不为零而剩下的 p_n-q_n 个系数为零，记 $X=\left(X_1,\cdots,X_{p_n}\right)$ 为 $n\times p_n$ 矩阵，与重要线性变量相关的矩阵 X_A 是由 X 前 q_n 列组成的子矩阵，设 $\boldsymbol{x}_i$ 来自均值为 $\mathbf{0}$ 的分布，$z_{ij}\in[0,1]$，$\forall i,j$。值得指出的是，允许 q_n 随着 n 增加，意味有更多的数据被收集时模型可随着信息量的增多而变化以更好地反映事实规律。针对非参数部分 $g_0(\cdot)$，首先引入如下定义：

定义 1 设 m 为正整数，$v\in(0,1)$，$r\equiv m+v$。$\mathcal{H}_r$ 是 $h(\cdot)$ 在 $[0,1]$ 上 m 次可导且 $h^{(m)}(\cdot)$ 满足阶数为 v 的 Hölder 条件的函数集合，即对于任意的 $h(\cdot)\in\mathcal{H}_r$，都存在正整数 C 使得下式成立：

$$\left|h^{(m)}(z')-h^{(m)}(z)\right|\le C|z'-z|,\quad \forall 0\le z',z\le 1 \tag{4}$$

假设(1)中非参数部分的可加项 $g_{0j}\in\mathcal{H}_r$，$r\ge 1.5$，另外 $\pi(t)=\left(b_1(t),\cdots,b_{k_n+l+1}(t)\right)'$ 是阶数为 $l+1$ 且在 $[0,1]$ 上有 k_n 个均匀结点的标准化 B 样条基函数向量。Stone [4]和 Schumarker [18]指出若满足上述假设条件，则 $g_{0j}(\cdot)$ 可以用 B 样条基函数的线性组合 $\Pi(\mathbf{z}_i)=\left(1,\pi(z_{i1})',\cdots,\pi(z_{id})'\right)'$ 拟合，且存在一个向量 $\boldsymbol{\xi}_0=(\xi_{00},\xi_{01},\cdots,\xi_{0d})\in\mathbb{R}^{D_n}$，其中 $D_n=d(k_n+l+1)+1$，使得 $\sup_{\mathbf{z}_i}\left|\Pi(\mathbf{z}_i)'\boldsymbol{\xi}_0-g_0(\mathbf{z}_i)\right|=O\left(k_n^{-r}\right)$ 成立。

当已知 $\boldsymbol{\beta}_0$ 中重要指标集合 A 时，给定 $\tau\in(0,1)$ 是，模型(1)中参数的稳健 expectile 回归估计为：

$$\left(\hat{\boldsymbol{\beta}}_A,\hat{\boldsymbol{\xi}}\right)=\arg\min_{(\boldsymbol{\beta}_A,\boldsymbol{\xi})}\frac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-\boldsymbol{x}'_{A_i}\boldsymbol{\beta}_A-\boldsymbol{\Pi}(\mathbf{z}_i)'\boldsymbol{\xi}\right) \tag{5}$$

其中 $\boldsymbol{x}'_{A_1},\cdots,\boldsymbol{x}'_{A_n}$ 代表 X_A 的行向量。$\boldsymbol{\beta}_0$ 的 oracle 估计值为 $\left(\hat{\boldsymbol{\beta}}'_A,\mathbf{0}'_{p_n-q_n}\right)'$，非参数部分 $g_0(\mathbf{z}_i)$ 的估计值为 $\hat{g}(\mathbf{z}_i)=\hat{g}_0+\sum_{j=1}^{d}\hat{g}_{0j}(z_{ij})$，记 $\hat{\boldsymbol{\xi}}=\left(\hat{\xi}_0,\hat{\boldsymbol{\xi}}_1',\cdots,\hat{\boldsymbol{\xi}}_d'\right)'$，其中 $\hat{\xi}_0\in\mathbb{R}$，$\hat{\boldsymbol{\xi}}_j'\in\mathbb{R}^{k_n+l+1}$，

$\hat{g}_{0j}(z_{ij})=\pi(z_{ij})'\hat{\boldsymbol{\xi}}_j-n^{-1}\sum_{i=1}^{n}\pi(z_{ij})'\hat{\boldsymbol{\xi}}_j$，$\hat{g}_0=\hat{\xi}_0+n^{-1}\sum_{i=1}^{n}\pi(z_{ij})'\hat{\boldsymbol{\xi}}_j$，$j=1,\cdots,d$。注意到 $\hat{g}_{0j}$ 有中心化的步骤，是为了让其满足可识别条件 $E\left[g_{0j}(\mathbf{z}_i)\right]=0$。

3. 带非凸惩罚的高维部分线性可加 Expectile 回归及算法

对实际数据进行分析时，通常不清楚维数为 p_n 的解释变量里哪几个 q_n 变量是重要的，在损失函数后加上惩罚项是筛选重要解释变量的常用手段。惩罚函数 Lasso [19]由于其是凸函数便于计算和进行理论分析而倍受欢迎，但不可忽略其会过惩罚较大的系数而导致估计偏差较大的缺点以及设计矩阵需要满足一些较强的条件。相比之下，一些满足条件 1 的非凸惩罚可以解决 Lasso 的问题，例如 SCAD 惩罚[20]。

条件 1 惩罚函数 $P_\lambda(\lambda>0)$ 对任意 $t\ge 0$ 具有表达式 $P_\lambda(t)=\lambda^2p_0(t/\lambda)$，其中函数 $p_0:\mathbb{R}_+\to\mathbb{R}_+$ 满足：(i) $p_0(\cdot)$ 是在 $[0,\infty)$ 上非降的函数且 $p_0(0)=0$；(ii) $p_0(\cdot)$ 在 $(0,\infty)$ 上几乎处处可导且 $\lim_{t\downarrow 0}p_0(t)=1$；(iii) 若 $t_1\ge t_2>0$ 有 $p_0'(t_1)\le p_0'(t_2)$。

SCAD 惩罚和其一阶导数的表达式为：

$$P_\lambda(\theta)=\lambda\theta\cdot\mathrm{I}(\theta\le\lambda)+\frac{a\lambda\theta-\left(\theta^2+\lambda^2\right)}{a-1}\mathrm{I}(\lambda\le\theta\le a\lambda)+\frac{(a+1)\lambda^2}{2}\mathrm{I}(\theta>a\lambda) \tag{6}$$

$$P_\lambda'(\theta)=\lambda\left\{\mathrm{I}(\theta\le\lambda)+\frac{(a\lambda-\theta)_+}{(a-1)\lambda}\mathrm{I}(\theta>\lambda)\right\} \tag{7}$$

其中 $a>2$ 是固定参数，$\theta>0$。于是带非凸惩罚的高维稳健 expectile 回归的估计值可通过以下最优化问题求

$$\left(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\xi}}\right)=\underset{(\boldsymbol{\beta},\boldsymbol{\xi})}{\arg\min}\frac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-\boldsymbol{x}_i'\boldsymbol{\beta}-\Pi\left(\boldsymbol{z}_i\right)'\boldsymbol{\xi}\right)+\sum_{j=2}^{p}P_{\lambda}\left(\left|\beta_j\right|\right) \tag{8}$$

但直接求解(8)会由非凸性而存在计算不稳定的问题，于是本文采用 Zou 和 Li [21]提出的局部线性近似法，每次迭代的惩罚权重由上一次的迭代结果更新。记 $P_{\lambda}'(\cdot)$ 为惩罚函数的一阶导，$\boldsymbol{\beta}^{(0)}$ 为迭代的初始值设为 $\mathbf{0}$，$S(a,b)=(a-b)_{+}-(-a-b)_{+}$ 是软阀值函数，σ 是大于 1 的正数，$\boldsymbol{\xi}^{(t-1)}$ 是第 $t-1$ 次迭代的结果，第 t 次迭代的结果 $\boldsymbol{\beta}^{(t)}$ 由局部线性近似法转为下面的凸优化问题来求

$$\left(\boldsymbol{\beta}^{(t)}\left|\boldsymbol{\xi}^{(t-1)}\right.\right)=\underset{\boldsymbol{\beta}}{\arg\min}\frac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}-\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}^{(t-1)}\right)+\sum_{j=2}^{p}P_{\lambda}'\left(\left|\beta_j^{(t-1)}\right|\right)\left|\beta_j\right| \tag{9}$$

其中 $P_{\lambda}'\left(\left|\beta_j^{(t-1)}\right|\right)\left|\beta_j\right|$ 是非凸惩罚函数 $P_{\lambda}\left(\left|\beta_j\right|\right)$ 在 $\left|\beta_j^{(t-1)}\right|$ 处的局部线性近似。

本文运用两步算法，整体迭代的顺序是先得到不带惩罚的非参数部分的系数 $\boldsymbol{\xi}$，再求带惩罚的线性系数 $\boldsymbol{\beta}$，然后求调节参数 γ。其中非参数部分的迭代算法为经典的坐标下降法(GDBB) [22]，带惩罚项的线性部分为 Fan 等人提出的 Local Adaptive Majorize-Minimization 算法(LAMM) [23]。具体算法描述如表 1~3 所示。

Table 1. Two-step algorithm
表 1. 两步算法

算法 1 两步算法

Step 1. Input $\gamma^{(0)}=\sqrt{n/\log(np)}$，$\varepsilon_1$，$\varepsilon_2$，$\boldsymbol{\beta}^{(0)}=0$，$t=0$

Step 2. 先求非参数部分的系数 $\boldsymbol{\xi}$，再求带惩罚的线性系数 $\boldsymbol{\beta}$

(a) 非参数部分：基于 $t-1$ 次的迭代结果 $\beta^{(t-1)}$，通过以下求解最优化问题得到 $\boldsymbol{\xi}^{(t)}$：

$$\boldsymbol{\xi}^{(t)}=\underset{\boldsymbol{\xi}\in\mathbb{R}^{D_n}}{\arg\min}\frac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}^{(t-1)}-\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}\right)$$

(b) 线性部分：

(b.1) 用第 $t-1$ 次的迭代结果 $\boldsymbol{\beta}^{(t-1)}=\left(\beta_1^{(t-1)},\cdots,\beta_p^{(t-1)}\right)'$ 计算第 t 次迭代权重 $P_{\lambda}'\left(\left|\beta_j^{(t-1)}\right|\right)$

(b.2) 把(a)的计算结果 $\boldsymbol{\xi}^{(t)}$ 带入下面的最优化问题求解 $\beta^{(t)}$：

$$\left(\boldsymbol{\beta}^{(t)}\left|\boldsymbol{\xi}^{(t)}\right.\right)=\underset{\boldsymbol{\beta}}{\arg\min}\frac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}-\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}^{(t)}\right)+\sum_{j=2}^{p}P_{\lambda}'\left(\left|\beta_j^{(t-1)}\right|\right)\left|\beta_j\right|$$

Step 3. 把 $\boldsymbol{\xi}^{(t)}$ 和 $\boldsymbol{\beta}^{(t)}$ 带入下面几式中计算 $r_i^{(t)}$，$\tilde{r}_i^{(t)}$，$mad\left(\tilde{\boldsymbol{r}}^{(t)}\right)$ 和 $\gamma^{(t)}$

$$r_i^{(t)}=y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}^{(t-1)}-\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}^{(t-1)},\tilde{r}^{(t)}=(1-\tau)r_i^{(t)}\mathrm{I}_{r_i^{(t)}\le 0}+\tau r_i^{(t)}\mathrm{I}_{r_i^{(t)}>0}$$

$$mad\left(\tilde{\boldsymbol{r}}^{(t)}\right)=\left\{\Phi^{-1}(0.75)\right\}^{-1}median\left(\left|\tilde{\boldsymbol{r}}^{(t)}-median\left(\tilde{\boldsymbol{r}}^{(t)}\right)\right|\right)$$

$$\gamma^{(t)}=mad\left(\tilde{\boldsymbol{r}}^{(t)}\right)\cdot\sqrt{\frac{n}{\log(np)}}$$

Step 4. 如果 $\left\|\boldsymbol{\beta}^{(t)}-\boldsymbol{\beta}^{(t-1)}\right\|\le\varepsilon_1$，$\left|\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}^{(t)}-\Pi\left(\boldsymbol{z}_i\right)^{\mathrm{T}}\boldsymbol{\xi}^{(t-1)}\right|\le\varepsilon_2$，那么令 $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}^{(t)}$，$\hat{\boldsymbol{\xi}}=\boldsymbol{\xi}^{(t)}$，否则令 $t=t+1$，返回 step 2 继续进行迭代

Table 2. Step (a) of Algorithm 1
表 2. 算法 1 里的 a 步骤

算法 1.1 Gradient Descent with Barzilai-Borwein Step Size (GDBB)
Step 1. Input $\boldsymbol{\xi}_0$, $0<\varepsilon_a\le 1$, $k=0$
Step 2. $d_k=\dfrac{\partial\left(\dfrac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}^{(t-1)}-\Pi(\mathbf{z}_i)^{\mathrm{T}}\boldsymbol{\xi}_k\right)\right)}{\partial\boldsymbol{\xi}_k}$. if $\|d_k\|<\varepsilon_a$, then stop, return $\boldsymbol{\xi}_k=\boldsymbol{\xi}^{(t)}$; else Step3
Step 3. 计算 $\alpha_k=\min\left\{\dfrac{s_{k-1}^{\mathrm{T}}s_{k-1}}{s_{k-1}^{\mathrm{T}}h_{k-1}},\dfrac{s_{k-1}^{\mathrm{T}}h_{k-1}}{h_{k-1}^{\mathrm{T}}h_{k-1}}\right\}$, where $s_{k-1}=\boldsymbol{\xi}_k-\boldsymbol{\xi}_{k-1}$, $h_{k-1}=d_k-d_{k-1}$,
Step 4. $\boldsymbol{\xi}_{k+1}=\boldsymbol{\xi}_k-\alpha_k d_k$
Step 5. $k=k+1$,, 回到 step 2

Table 3. Step (b.2) of Algorithm 1
表 3. 算法 1 里的 b.2 步骤

算法 1.2 Local Adaptive Majorize-Minimization (LAMM)
Step 1. Input $\boldsymbol{\beta}^{(0)}=\boldsymbol{\beta}^{(t-1)}, 0<\varepsilon_b\le 1, k=1, \phi_0, \phi^{(k-1)}, \sigma$
Step 2. 基于第 $k-1$ 次的迭代结果 $\boldsymbol{\beta}^{(k-1)}=\left(\beta_1^{(k-1)},\cdots,\beta_p^{(k-1)}\right)'$，计算第 k 次迭代相关的权重： $\mathbf{w}^{(k)}=\left(P_\lambda'\left(\left\|\beta_1^{(k-1)}\right\|\right),P_\lambda'\left(\left\|\beta_2^{(k-1)}\right\|\right),\cdots,P_\lambda'\left(\left\|\beta_p^{(k-1)}\right\|\right)\right)^{\mathrm{T}}=\left(w_1^{(k)},\cdots,w_j^{(k)}\right)^{\mathrm{T}}$
Step 3. Initialize: $\phi^{(k)}\leftarrow\max\left\{\phi_0,\sigma^{-1}\phi^{(k-1)}\right\}$ repeat: If $F\left(\boldsymbol{\beta}^{(k)},\mathbf{w}^{(k)}\right)\ge\Psi_{\mathbf{w}^{(k)},\phi^k}\left(\boldsymbol{\beta}^{(k)};\boldsymbol{\beta}^{(k-1)}\right)$, then $\phi^{(k)}\leftarrow\sigma\phi^{(k)}$ until $F\left(\boldsymbol{\beta}^{(k)},\mathbf{w}^{(k)}\right)\le\Psi_{\mathbf{w}^{(k)},\phi^k}\left(\boldsymbol{\beta}^{(k)};\boldsymbol{\beta}^{(k-1)}\right)$, return $\beta_j^{(k)}\leftarrow T_{\mathbf{w}^{(k)},\phi^{(k)}}\left(\beta_j^{(k-1)}\right)$ and $\phi^{(k)}$ Where $F\left(\boldsymbol{\beta}^{(k-1)},\mathbf{w}^{(k)}\right)\equiv\Gamma\left(\boldsymbol{\beta}^{(k-1)}\right)+\sum_{j=2}^{p}w_j^{(k)}\left\|\beta_j^{(k-1)}\right\|$, $\Gamma\left(\boldsymbol{\beta}^{(k-1)}\right)=\dfrac{1}{n}\sum_{i=1}^{n}L_{\tau,\gamma}\left(y_i-x_i^{\mathrm{T}}\boldsymbol{\beta}^{(k-1)}-\Pi(\mathbf{z}_i)^{\mathrm{T}}\boldsymbol{\xi}^{(t)}\right)$, $\beta_j^{(k)}\equiv T_{\mathbf{w}^{(k)},\phi^{(k)}}\left(\beta_j^{(k-1)}\right)=S\left(\beta_j^{(k-1)}-\dfrac{\nabla\Gamma\left(\beta_j^{(k-1)}\right)}{\phi^{(k)}},\dfrac{w_j^{(k)}}{\phi^{(k)}}\right)$ $\Psi_{\mathbf{w}^{(k)},\phi^k}\left(\boldsymbol{\beta};\boldsymbol{\beta}^{(k-1)}\right)\equiv\Gamma\left(\boldsymbol{\beta}^{(k-1)}\right)+\left\langle\nabla\Gamma\left(\boldsymbol{\beta}^{(k-1)}\right),\boldsymbol{\beta}-\boldsymbol{\beta}^{(k-1)}\right\rangle+\dfrac{\phi^{(k)}}{2}\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(k-1)}\right\|_2^2+\sum_{j=2}^{p}w_j^{(k)}\left\|\beta_j^{(k-1)}\right\|$
Step 4. If $\left\|\boldsymbol{\beta}^{(k)}-\boldsymbol{\beta}^{(k-1)}\right\|<\varepsilon_b$ then $\boldsymbol{\beta}^{(t)}=\boldsymbol{\beta}^{(k)}$, otherwise $k=k+1$，返回 step 2 继续进行迭代

4. 数值模拟

本文分别研究了部分线性可加稳健 expectile 回归(retire-PLAMs)在不带惩罚的低维模型和带惩罚的稀疏高维模型两种情况下数值模拟的表现，在部分线性可加框架下让 retire 与不同的方法比较：(i) huber；(ii) 非对称最小平方回归(sales)；(iii) quantile 回归(qr)。各进行了 100 次随机模拟，采用以下准则来比较各个方法：

1) MSE：估计 $\boldsymbol{\beta}_0$ 的均方误差，即一百次模拟结果中 $\left\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|^2$ 的均值。

2) AE：$\boldsymbol{\beta}_0$的绝对估计误差的均值，即一百次模拟结果中$\sum_{j=1}^{p}\left|\hat{\beta}_j-\beta_j\right|$的均值。

3) AADE：拟合非参数的平均绝对离差的均值，即一百次模拟结果$n^{-1}\sum_{i=1}^{n}\left|\hat{g}\left(\boldsymbol{z}_i\right)-g_0\left(\boldsymbol{z}_i\right)\right|$中的均值。

4) TV：一百次模拟中正确地识别到系数为非零的线性解释变量的次数占比的均值。

5) FV：一百次模拟中错误地识别出系数为非零的线性解释变量的次数占比的均值。

4.1. 不带惩罚的低维模型

首先生成来自多元正态分布$N_{11}\left(\mathbf{0}_{11},\Sigma\right)$的$\tilde{X}=\left(\tilde{X}_1,\cdots,\tilde{X}_{11}\right)'$，其中$\Sigma=\left(\sigma_{jk}\right)_{11\times11}$，$\sigma_{jk}=0.5^{|j-k|}$。让$Z_1=\Phi\left(\tilde{X}_{10}\right)$，$Z_2=\Phi\left(\tilde{X}_{11}\right)$，$\Phi(\cdot)$是标准正态分布的分布函数，$X_l=\tilde{X}_l$，$l=1,\cdots,9$，响应变量$Y_i$有以下回归模型生成，

$$Y_i=\boldsymbol{X}_i'\boldsymbol{\beta}_0+\sin\left(2\pi Z_{i1}\right)+Z_{i2}^3+\left(0.5\left|X_{i9}\right|+0.5\right)\left\{\varepsilon_i-e_\tau\left(\varepsilon_i\right)\right\} \tag{10}$$

其中$\boldsymbol{\beta}_0=\left(\beta_1,\cdots,\beta_{10}\right)'=\left(2,1.6,1.3,1,0.7,0.4,0.1,-0,1,-0.4,-0.7\right)'$，$\beta_1$为截距，$\varepsilon_i$可分两种情形，一是来自正态分布$N\left(0,2\right)$，二是来自自由度为2.1的$t$分布，样本量$n$分为200和400两种情形，$\tau$水平分为0.5和0.8两种情况。

表4和表5分别展示了随机噪声来$N\left(0,2\right)$和$t\left(2.1\right)$的模型(13)在$\tau=\left\{0.5,0.8\right\}$时四种不同方法的拟合表现，可以看出随着样本量增加，所有模型的估计效果都有所改进，另外$\tau=0.5$时retire与huber估计结果相同，因为$\tau=0.5$时二者为同一个模型。表5可见随机噪声来自正态分布时，retire的估计误差是所有模型中表现最好的；表4显示随机误差来自t分布时，retire的估计误差比huber和sales小，比quantile的估计误稍大，原因在于quantile遇到异常值时是最稳健的，但是quantile计算速度比retire慢。总而言之，数值模拟验证retire-PLAMs在拥有稳健性的同时还具有计算效率高的优势。

Table 4. Comparison of the fitting performance of the four methods for model (13) with random noise from $t(2.1)$ at τ levels of 0.5 and 0.8

表4. 四种方法在随机噪声来自$t(2.1)$的模型(13)在τ水平为0.5和0.8的拟合表现比较

		n = 200	p = 10		n = 400	p = 10	
τ	method	AE	MSE	AADE	AE	MSE	AADE
0.5	retire	1.413	0.310	0.303	0.974	0.152	0.205
	huber	1.413	0.310	0.303	0.974	0.152	0.205
	sales	1.784	0.545	0.469	1.437	0.488	0.337
	qr	0.974	0.150	0.270	0.618	0.061	0.192
0.8	retire	1.584	0.402	0.390	1.080	0.192	0.260
	huber	1.431	0.320	0.841	1.002	0.160	0.820
	sales	2.309	1.072	0.623	1.957	1.384	0.485
	qr	1.343	0.290	0.403	0.889	0.130	0.275

Table 5. Comparison of the fitting performance of the four methods for model (13) with random noise from $N(0,2)$ at τ levels of 0.5 and 0.8
表 5. 四种方法在随机噪声来自 $N(0,2)$的模型(13)在 τ 水平为 0.5 和 0.8 的拟合表现比较

		n = 200	p = 10		n = 400	p = 10	
τ	method	AE	MSE	AADE	AE	MSE	AADE
0.5	retire	1.007	0.163	0.265	0.725	0.082	0.184
	huber	1.007	0.163	0.265	0.725	0.082	0.184
	sales	1.010	0.163	0.272	0.726	0.082	0.189
	qr	1.066	0.181	0.283	0.737	0.087	0.197
0.8	retire	1.088	0.187	0.327	0.822	0.103	0.260
	huber	1.035	0.172	0.612	0.728	0.083	0.608
	sales	1.073	0.182	0.280	0.805	0.099	0.193
	qr	1.348	0.287	0.359	0.870	0.117	0.283

4.2. 带惩罚的稀疏模型

生成来自 $N_{p+2}\left(\mathbf{0}_{p+2},\Sigma\right)$ 的 $\tilde{X}=\left(\tilde{X}_1,\cdots,\tilde{X}_{p+2}\right)'$，其中 $\Sigma=\left(\sigma_{jk}\right)_{(p+2)\times(p+2)}$，$\sigma_{jk}=0.5^{|j-k|}$。让 $Z_1=\Phi\left(\tilde{X}_{10}\right)$，$Z_2=\Phi\left(\tilde{X}_{11}\right)$，$\Phi(\cdot)$ 是标准正态分布的分布函数；对于 $l=1,\cdots,9$，$X_l=\tilde{X}_l$；对于 $l=13,\cdots,p+2$，$X_l=\tilde{X}_{l-2}$。响应变量 Y_i 有以下回归模型生成：

$$Y_i=\boldsymbol{X}_i'\boldsymbol{\beta}_0+\sin\left(2\pi Z_{i1}\right)+Z_{i2}^3+\left(0.5\left|X_{i9}\right|+0.5\right)\left\{\varepsilon_i-e_\tau\left(\varepsilon_i\right)\right\} \tag{11}$$

其中 $\boldsymbol{\beta}_0=\left(\beta_1,\cdots,\beta_{10},\mathbf{0}_{p-8}\right)'=\left(2,1.6,1.3,1,0.7,0.4,0.1,-0,1,-0.4,-0.7,\mathbf{0}'_{p-8}\right)'$，$\varepsilon_i$ 可分两种情形，一是来自正态分布 $N(0,2)$，二是来自自由度为 2.1 的 t 分布，样本量 $n=400$，$p=300,500$ 两种情形，τ 水平为 0.5 和 0.8 两种情况。表 6 展示了带 SCAD 惩罚的 retire 与带 Lasso 惩罚的另外三种方法的拟合结果表现比较，可见 retire-SCAD 的线性参数估计误差最小，TP 最大，FP 最小，是变量筛选最一致的。

Table 6. Comparison of the fitting performance of the four methods for model (14) with random noise from $t(2.1)$ at τ levels of 0.5 and 0.8
表 6. 四种方法在随机噪声来自 $t(2.1)$的模型(14)在 τ 水平为 0.5 和 0.8 的拟合表现比较

		n = 400	p = 300			n = 400	p = 500		
τ	method	MSE	AADE	TP	FP	MSE	AADE	TP	FP
0.5	retire-SCAD	0.107	0.693	1	0.015	0.132	0.7123	1	0.010
	roer-L1	0.443	0.400	1	0.095	0.510	0.439	1	0.056
	huber-L1	0.443	0.400	1	0.095	0.510	0.439	1	0.056
	sales-L1	0.991	0.519	0.995	0.093	1.092	0.546	0.993	0.577
	qr-L1	0.006	0.018	1	0.017	0.072	0.019	1	0.129

续表

	retire-SCAD	0.157	0.656	1	0.010	0.161	0.607	1	0.007
	roer-L1	0.609	0.463	1	0.099	0.621	0.461	1	0.062
0.8	huber-L1	0.471	0.560	1	0.095	0.497	0.497	1	0.057
	salse-L1	1.002	0.615	0.994	0.091	1.207	0.780	0.996	0.072
	qr-L1	0.263	0.021	1	0.169	0.257	0.019	1	0.135

4.3. 时间对比

笔者增加此部分模拟以展示所提方法 retire-L1 与 qr-L1 相比在计算效率上的优势。分别借助 R 语言的 adaHuber 和 rqPen 两个包，稀疏正则参数 λ 由十则交叉验证选择，retire 的稳健参数 γ 如表 1 所示由数据自适应调节。100 次模拟的数据由模型(15)生成：

$$Y_i = \boldsymbol{X}_i'\boldsymbol{\beta}_0 + \sin\left(2\pi Z_{i1}\right) + Z_{i2}^3 + \varepsilon_i \tag{12}$$

其中 $\boldsymbol{X}_i$， Z_{i1} 和 Z_{i2} 同 4.2 节的设定， ε_i 来自自由度为 2.1 的 t 分布， $\tau = 0.5$，样本数量 $n = p/2$，解释变量维数 $p = \{100, 200, 300, 400, 500\}$。

模拟结果如表 7 所示，运行时间的单位为秒，随着 n 和 p 的增加估计误差都在减少。虽然在 $t_{2.1}$ 的随机噪声下 qr-L1 比 retire-L1 较稳健，但计算效率不如 retire-L1 快，特别是 p 很大的时候，例 $p = 500$ 时，qr-L1 的计算时间是 retire-L1 的 16 倍。

Table 7. Comparison of time and estimation results between retire-L1 and qr-L1 at $\tau = 0.5$
表 7. Retire-L1 与 qr-L1 在 $\tau = 0.5$ 的时间和估计结果对比

	p = 100, n = 50	p = 200, n = 100	p = 300, n = 150	p = 400, n = 200	p = 500, n = 250
qr-L1-time/qr-L1-time	1	1	1	1	1
qr-L1-AADE/qr-L1-AADE	1	1	1	1	1
qr-L1-MSE/qr-L1-MSE	1	1	1	1	1
retire-L1-time/qr-L1-time	0.176	0.121	0.092	0.065	0.060
retire-L1-AADE/qr-L1-AADE	1.496	1.351	1.240545	1.261	1.081
retire-L1-MSE/qr-L1-MSE	1.139	1.207	1.340393	1.406	1.432

5. 实证研究

文献[24]指出 β 胡萝卜素与肺癌、乳腺癌等癌症之间存在直接关系，一些流行病学研究表明其抗氧化特性能清除可能导致癌症的自由基，充足的 β 胡萝卜素能提高人体免疫力有效地对抗癌症等疾病。我们分析 Nierenberg 等人[25]收集的血浆 β 胡萝卜素水平数据集并建立部分线性可加模型，寻找血浆 β 胡萝卜素水平与个人特征之间的关系，包括年龄、性别、体重指数和其他因素，响应变量 Y 是以 ng/ml 为单位的血浆 β 胡萝卜素，命名为 BETAPLASMA，各变量的说明见表 8，使用以下部分线性加性模型进行建模：

$$\begin{aligned} Y &= u + \beta_1 \mathrm{SEX} + \beta_2 \mathrm{SMOK1} + \beta_3 \mathrm{SMOK2} + \beta_4 \mathrm{BMI} + \beta_5 \mathrm{VIT1} + \beta_6 \mathrm{VIT2} + \beta_7 \mathrm{CAL} + \beta_8 \mathrm{FAT} \\ &\quad + \beta_9 \mathrm{ALCOHOL} + \beta_{10} \mathrm{BETADIET} + \eta_1(\mathrm{AGE}) + \eta_2(\mathrm{CHOL}) + \eta_2(\mathrm{FIBER}) + \sigma\varepsilon \\ &= u + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X} + \sum_{j=1}^{3} \eta_1(z_j) + \sigma\varepsilon \end{aligned} \tag{16}$$

Table 8. Explanation of the variables for the Plasma β Carotene Levels dataset
表 8. 血浆 β 胡萝卜素水平数据集变量的解释说明

变量名	含义
SEX	1 = 男，0 = 女
SMOK1	1 = 以前吸烟，0 = 其他
SMOK2	1 = 现在吸烟，0 = 其他
BMI	身体质量指数
VIT1	1 = 经常补充维生素，0 = 其他
VIT2	1 = 偶尔补充维生素，0 = 其他
CAL	每天消耗的卡路里数
AGE	年龄
FAT	每天消耗的脂肪克数
ALCOHOL	每周饮用的酒精饮料数量
BETADIET	膳食 β-胡萝卜素消耗量(微克/天)
CHOL	每日摄入的胆固醇(毫克/天)
FIBER	每日摄入的纤维克数(克/天)
BETAPLASMA	血浆 β-胡萝卜素(纳克/毫升)

在建模前用 Min-Max 标准化法消除变量量纲不同可能造成的影响，AGE、CHOL 和 FIBER 放入非线性部分，其余解释变量放入线性部分。用带 SCAD 惩罚的稳健 expectile 回归筛选影响人体 β 胡萝卜素含量的重要变量，研究血浆 β 胡萝卜素在不同 expectile 水平 $\tau = \{0.2, 0.5, 0.9\}$ 的条件分布，回归结果见表 9。可以看到三个不同 τ 水平的回归系数有差异，说明数据存在异质性，从十个变量里筛选出 BMI、胡萝卜消耗、吸烟、维他命摄入和性别对 β 胡萝卜素含量占主要影响的五个因素。进而推出合理的 BMI 范围、不吸烟、多摄入维生素能保证人体 β 胡萝卜素含量充足，能预防癌症等疾病的结论。

Table 9. Regression coefficients of important variables at three different τ levels
表 9. 在三个不同 τ 水平下的重要变量的回归系数

τ	BMI	BETADIET	SMOK1	SMOK2	VIT1	VIT2	SEX
0.2	−143	84	−9	−39	0	−30	−29
0.5	−187	135	0	−42	23	−39	−23
0.8	−222	140	−11	−59	33	−38	−31

6. 总结

本文在 PLAMs 框架下研究高维稳健 expectile 回归模型，数值模拟和实证研究显示该法与 quantile regression 一样能通过取不同的 τ 水平来分析数据异质性，但比 quantile regression 的计算效率高，而估计效果和 quantile regression 同样具有稳健性。

参考文献

[1] Rigby, R.A. and Stasinopoulos, D.M. (1996) A Semi-Parametric Additive Model for Variance Heterogeneity. *Statistics and Computing*, **6**, 57-65. https://doi.org/10.1007/bf00161574

[2] Horowitz, J.L. (1999) Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity. *Econometrica*, **67**, 1001-1028. https://doi.org/10.1111/1468-0262.00068

[3] Hastie, T. and Tibshirani, R. (1990) Exploring the Nature of Covariate Effects in the Proportional Hazards Model. *Biometrics*, **46**, 1005-1016. https://doi.org/10.2307/2532444

[4] Stone, C.J. (1985) Additive Regression and Other Nonparametric Models. *The Annals of Statistics*, **13**, 689-705. https://doi.org/10.1214/aos/1176349548

[5] Opsomer, J.D. and Ruppert, D. (1997) Fitting a Bivariate Additive Model by Local Polynomial Regression. *The Annals of Statistics*, **25**, 186-211. https://doi.org/10.1214/aos/1034276626

[6] Opsomer, J.D. and Ruppert, D. (1999) A Root-Nconsistent Backfitting Estimator for Semiparametric Additive Modeling. *Journal of Computational and Graphical Statistics*, **8**, 715-732. https://doi.org/10.1080/10618600.1999.10474845

[7] Liu, X., Wang, L. and Liang, H. (2011) Estimation and Variable Selection for Semiparametric Additive Partial Linear Models. *Statistica Sinica*, **21**, 1225-1248. https://doi.org/10.5705/ss.2009.140

[8] Hoshino, T. (2014) Quantile Regression Estimation of Partially Linear Additive Models. *Journal of Nonparametric Statistics*, **26**, 509-536. https://doi.org/10.1080/10485252.2014.929675

[9] Sherwood, B. and Wang, L. (2016) Partially Linear Additive Quantile Regression in Ultra-High Dimension. *The Annals of Statistics*, **44**, 288-317. https://doi.org/10.1214/15-aos1367

[10] Newey, W.K. and Powell, J.L. (1987) Asymmetric Least Squares Estimation and Testing. *Econometrica*, **55**, 819-847. https://doi.org/10.2307/1911031

[11] Kuan, C., Yeh, J. and Hsu, Y. (2009) Assessing Value at Risk with CARE, the Conditional Autoregressive Expectile Models. *Journal of Econometrics*, **150**, 261-270. https://doi.org/10.1016/j.jeconom.2008.12.002

[12] Daouia, A., Girard, S. and Stupfler, G. (2017) Estimation of Tail Risk Based on Extreme Expectiles. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, **80**, 263-292. https://doi.org/10.1111/rssb.12254

[13] Sobotka, F., Kauermann, G., Schulze Waltrup, L. and Kneib, T. (2011) On Confidence Intervals for Semiparametric Expectile Regression. *Statistics and Computing*, **23**, 135-148. https://doi.org/10.1007/s11222-011-9297-1

[14] Zhao, J., Yan, G. and Zhang, Y. (2019) Semiparametric Expectile Regression for High-Dimensional Heavy-Tailed and Heterogeneous Data.

[15] Man, R., Tan, K.M., Wang, Z. and Zhou, W. (2024) Retire: Robust Expectile Regression in High Dimensions. *Journal of Econometrics*, **239**, Article 105459. https://doi.org/10.1016/j.jeconom.2023.04.004

[16] Sun, Q., Zhou, W. and Fan, J. (2019) Adaptive Huber Regression. *Journal of the American Statistical Association*, **115**, 254-265. https://doi.org/10.1080/01621459.2018.1543124

[17] Abadie, A., Angrist, J. and Imbens, G. (2002) Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. *Econometrica*, **70**, 91-117. https://doi.org/10.1111/1468-0262.00270

[18] Schumaker, L. (2007) Spline Functions: Basic Theory. 3rd Edition, Cambridge University Press. https://doi.org/10.1017/cbo9780511618994

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

[20] Fan, J. and Li, R. (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

[21] Zou, H. and Li, R. (2008) One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models. *The Annals of Statistics*, **36**, 1509-1533. https://doi.org/10.1214/009053607000000802

[22] Barzilai, J. and Borwein, J.M. (1988) Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, **8**, 141-148. https://doi.org/10.1093/imanum/8.1.141

[23] Fan, J., Liu, H., Sun, Q. and Zhang, T. (2018) I-LAMM for Sparse Learning: Simultaneous Control of Algorithmic Complexity and Statistical Error. *The Annals of Statistics*, **46**, 814-841. https://doi.org/10.1214/17-aos1568

[24] Fairfield, K.M. and Fletcher, R.H. (2002) Vitamins for Chronic Disease Prevention in Adults. *Journal of the American Medical Association*, **287**, 3116-3126. https://doi.org/10.1001/jama.287.23.3116

[25] Nierenberg, D.W., Stukel, T.A., Baron, J.A., Dain, B.J. and Greenberg, E.R. (1989) Determinants of Plasma Levels of Beta-Carotene and Retinol. *American Journal of Epidemiology*, **130**, 511-521. https://doi.org/10.1093/oxfordjournals.aje.a115365