

纵向数据下变系数线性模型的惩罚秩回归

张雅斐

西南大学数学与统计学院, 重庆

收稿日期: 2025年6月24日; 录用日期: 2025年7月14日; 发布日期: 2025年7月29日

摘要

针对纵向数据变系数线性模型中数据具有组内相关性和组间独立性, 以及组内协方差矩阵估计效率低的特点, 本文基于修正的Cholesky分解和B样条函数处理非参数部分, 对组内协方差矩阵提出更有效的估计, 改善了不平衡纵向数据的估计效率。同时, 文章结合了两个非凸惩罚函数SCAD和MCP实现稳健的变量选择, 模拟研究和实证分析表明新方法能够获得更有效的估计, 并且相同参数准则下, 对高相关性的样本数据, MCP惩罚函数的显著性更强。

关键词

纵向数据, 修正的Cholesky分解, 秩回归, B样条, 变量选择

Penalized Rank Regression for Varying-Coefficient Linear Models with Longitudinal Data

Yafei Zhang

School of Mathematics and Statistics, Southwest University, Chongqing

Received: Jun. 24th, 2025; accepted: Jul. 14th, 2025; published: Jul. 29th, 2025

Abstract

For longitudinal data in varying coefficient linear models characterized by within-group correlation and between-group independence, as well as the low estimation efficiency of within-group covariance matrices, this paper proposes a more efficient estimation method for the within-group covariance matrix based on modified Cholesky decomposition and B-spline functions to handle the non-parametric components, thereby improving the estimation efficiency for unbalanced longitudinal data. Additionally, the study incorporates two non-convex penalty functions, SCAD and MCP, to

achieve robust variable selection. Simulation studies and empirical analyses demonstrate that the proposed method yields more efficient estimates. Under the same parameter criteria, the MCP penalty exhibits stronger significance for highly correlated sample data.

Keywords

Longitudinal Data, Modified Cholesky Decomposition, Rank Regression, B-Spline, Variable Selection

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

纵向数据是指通过对观测个体在不同时间节点上进行多次重复测量而获得的数据。因此，纵向数据分析可以对同一个体内部的变化规律以及协变量对响应变量的影响进行动态分析。在独立数据的回归分析中引入数据的组内相关性，提高统计推断的效率的同时，纵向数据分析也对统计方法提出了巨大的挑战[1]-[3]。常见的适用于纵向数据的模型，如边际回归模型、转移回归模型和混合效应模型。传统的参数回归模型只考虑参数部分，在处理实际问题，尤其是分析长期或动态环境下的数据时，模型解释能力不足。因此，考虑系数参数随时间变化是提升模型适应性的重要方向。变系数线性模型的首次提出是由 Shumway [4]，Cleveland [5]和 Hastie 以及 Tibshirani [6]给出的。其他基于该模型的讨论和扩展可从 Fan 和 Zhang [7]以及 Chiang 等人[8]的论文中得到。

传统的变系数线性模型通常利用最小二乘法估计系数参数，然而其对异常值敏感且依赖误差分布假设。但秩回归作为一种能有效降低异常值影响的非参数回归方法，它的引入恰好弥补了这一缺陷。该方法也在纵向数据下得到了广泛研究，如将基于秩回归的方法运用到纵向数据中[9]-[11]，并扩展到变量选择研究中。Fu 等[12]基于独立性假设提出了基于秩的变量选择方法；Fu 等[13]基于加权秩提出了带有 SCAD (平滑裁剪绝对偏差)惩罚的 Wilcoxon 离散函数。但是在估计纵向数据的协方差矩阵时，大部分文献选择提前指定具体结构，如 Jung 等人[9]在独立工作结构下构造了广义的 Wilcoxon-Mann-Whitney 秩统计量，但得到的估计不是渐进有效的。Fu 等人[14]在可交换结构下考虑了基于秩的估计函数，但是当组内相关结构不是可交换时估计效率会降低。因此，建立协方差模型，在更一般的相关结构下研究变系数模型是值得关注的课题。吕晶等[15]利用修正的 Cholesky 分解处理了纵向秩回归的组内相关性，提出了更加有效的无偏性秩回归得分函数。

以上提到的文献模型虽然在变系数估计方面取得了显著进展，但大多未系统解决变量选择问题。往往在实际建模中，研究者倾向于初始纳入过多变量以避免遗漏重要预测因子，但这一做法不可避免地导致模型过拟合、预测精度下降及计算效率损失。对此，Wang 等人[16]结合了基差函数扩展和 SCAD 惩罚函数提出了正则化变量选择方法，成功识别并选出具有时变效应的重要变量。Zhang [17]提出了 MCP (极小极大凹惩罚)惩罚函数，在保留 SCAD 惩罚优点的同时对回归系数进行有差别惩罚，从而得到更加精确的估计，同时证明了基于 MCP 惩罚的回归模型具有 Oracle 性质。但是相较 SCAD 惩罚函数，基于 MCP 惩罚的函数形式更为简洁、参数估计偏差更小、模型结果也更加精简。

综合上述讨论，本文结合 B 样条函数近似和修正的 Cholesky 分解提出一种纵向数据下变系数线性模型的秩回归方法。利用 B 样条光滑函数近似变系数参数，并在修正的 Cholesky 分解下估计更具一般性结

v 构的协方差工作矩阵。然而，基于秩回归的估计函数非凸，不连续且不可微。本文将利用 Brown 等人 [18] 提出的诱导光滑的方法获得更精确的估计。在光滑估计函数的基础上，文章分别添加了非凸惩罚函数 SCAD 和 MCP 进行对比分析，以针对不同相关性数据，实现更有效的变量选择。

2. 基于修正的 Cholesky 分解的秩回归

2.1. B 样条近似

假定纵向数据 $\{(t_{ik}, X_{ik}, Y_{ik}), 1 \leq i \leq n, 1 \leq k \leq n_i\}$ 满足如下形式的变系数模型：

$$Y_{ik} = X_{ik}^T \beta(t_{ik}) + \varepsilon_{ik}, \quad (1)$$

其中， Y_{ik} 为响应变量， $X_{ik} = (X_{ik}^1, X_{ik}^2, \dots, X_{ik}^p)^T$ 是一个 p 维协变量， $\beta(t_{ik}) = (\beta_1(t_{ik}), \beta_2(t_{ik}), \dots, \beta_p(t_{ik}))^T$ ， $\beta_r(t)$ ($r=1, 2, \dots, p$) 是未知的非参数函数，并且 ε_{ik} 是第 i 个个体在 t_{ik} 时刻测得的零均值随机过程。

对于模型的非参数函数，利用 B 样条函数近似未知变系数 $\beta_r(t)$ ，可得到 $\beta_r(t) \approx B_r(t) \gamma_r = \sum_{h=1}^H \gamma_{rh} B_{rh}(t)$ 。

其中， $B_{rh}(t) \geq 0$ ， $\sum_{h=1}^H B_{rh}(t) = 1$ ，且控制点总数 $H = h + m + 1$ ， m 为样条曲线的阶数， h 为内节点个数。经过上述样条近似，可得到如下模型：

$$Y_{ik} \approx \sum_{r=1}^p \sum_{h=1}^H X_{ik}^{(r)} \gamma_{rh} B_{rh}(t_{ik}) + \varepsilon_{ik}, \quad (2)$$

现定义 $\pi^{(r)}(\cdot) = (B_{r1}(\cdot), \dots, B_{rH}(\cdot))^T$ 为标准化 B 样条基函数，定义 $\Pi(T, X) = (X^{(1)} \pi^{(1)}(T)^T, \dots, X^{(p)} \pi^{(p)}(T)^T)^T$ ， $\Pi_{ik} = \Pi(T_{ik}, X_{ik})$ ， $\pi_{ik}^{(r)} = \pi^{(r)}(t_{ik})$ 。因此，(2) 可以写为 $Y_{ik} \approx \Pi_{ik}^T \gamma + \varepsilon_{ik}$ ，其中 $\gamma = (\gamma_1^T, \dots, \gamma_p^T)^T$ ， $\gamma_r = (\gamma_{r1}, \dots, \gamma_{rH})^T$ 。

基于上面的近似，我们可以得到在 t_{ik} 处的残差 $\hat{\varepsilon}_{ik}(\hat{\gamma}) = Y_{ik} - \Pi_{ik}^T \hat{\gamma}$ 。对于模型(2)，定义 S_{ik} 为 $\varepsilon_{ik}(\gamma)$ 的平均秩， $S_{ik} = M^{-1} \sum_{j=1}^n \sum_{l=1}^{n_j} \left[I\{\varepsilon_{jl}(\gamma) < \varepsilon_{ik}(\gamma)\} - \frac{1}{2} \right]$ ，其中 $M = \sum_{i=1}^n n_i$ ，记 $S_i(\gamma) = (S_{i1}, \dots, S_{in_i})^T$ ，

$$S(\gamma) = \{S_1(\gamma)^T, \dots, S_n(\gamma)^T\}^T。$$

对 $S(\gamma)$ 有秩估计函数：

$$U_{OPT}(\gamma) = \bar{D}^T V^{-1} S(\gamma), \quad (3)$$

其中 $V = \text{Cov}(S(\gamma))$ ， $\bar{D} = \{\bar{D}_{ik}, i=1, \dots, n; k=1, \dots, n_i\}$ 为 $S(\gamma)$ 的期望关于 γ 的导数，且

$\bar{D}_{ik}(\gamma_0) = M^{-1} \sum_{j=1}^n \sum_{l=1}^{n_j} (\Pi_{jl}^T - \Pi_{ik}^T) f_{ikjl}(\gamma_0)$ ， $f_{ikjl}(\cdot)$ 为 $\varepsilon_{ik} - \varepsilon_{jl}$ 的密度函数， \bar{D}_{ik} 涉及未知的密度函数 $f_{ikjl}(\cdot)$ 。假设 $f_{ikjl}(\gamma_0)$ 为常数，并在模型(3)中用 $D_{ik}^T = \bar{\Pi} - \Pi_{ik}$ 代替 \bar{D}_{ik}^T 。

假定使用一个块对角矩阵 $\text{diag}(V_1, \dots, V_n)$ 作为 V 的工作协方差矩阵，可得到 γ 的估计函数为

$$U_B(\gamma) = \sum_{i=1}^n D_i^T V_i^{-1} S_i(\gamma), \quad \text{其中，} V_i = \text{var}\{S_i(\gamma)\}。$$

2.2. 一般工作结构的秩估计函数

在以往的秩估计方法中，Jung 和 Ying [9] 假定误差具有独立工作结构用于估计 β ，但估计结果 $\hat{\beta}_i$ 在其他工作结构情况下会严重损失估计效率。基于此，Fu 和 Wang [14] 在可交换结构下提出了最佳秩估计函数，并利用矩方法估计协方差矩阵的逆中的参数 σ_i^2 和 σ_{ii} 。在该模型下，只有 ε_i 具有可交换结构时，

S_i 和 ε_i 才具有相同的相关结构。因此, 当 ε_i 不具有可交换相关结构时, $\hat{\beta}_F$ 会损失估计效率。

已知上述两个模型只有在给定结构下的估计才是有效的, 当误差具有其它结构时, 估计效率大大降低。此时, 修正的 Cholesky 分解可以保证协方差矩阵正定, 同时提高数值计算的稳定性。故本文将基于修正的 Cholesky 分解建模, 基于 Yao 和 Li [19], $V_i = Cov\{S_i(\gamma_0)\}$ 可进行如下分解,

$$Cov\{\Phi_i S_i\} = \Phi_i^T V_i \Phi_i = \Lambda_i \quad (4)$$

则 $V_i = (\Phi_i^{-1})^T \Lambda_i \Phi_i^{-1}$ 。其中, Φ_i 是主对角元素均为 1 的下三角矩阵, $\Lambda_i = diag(d_{i1}^2, \dots, d_{in_i}^2)$ 为 $n_i \times n_i$ 的对角矩阵。记 $e_i = (e_{i1}, \dots, e_{in_i})^T = \Phi_i S_i$, Φ_i 的第 (k, l) 个元素为 $-\phi_{k,l}^{(i)}$ ($k = 1, \dots, n_i; l = 1, \dots, k-1$), 则有如下关系式:

$$\begin{cases} S_{i1} = e_{i1} \\ S_{ik} = \phi_{k,1}^{(i)} S_{i1} + \dots + \phi_{k,k-1}^{(i)} S_{i,k-1} + e_{ik}, i = 1, 2, \dots, n; k = 2, \dots, n_i \end{cases}$$

已知 Λ_i 为对角矩阵, 故 e_{ik} 互不相关, 且 $var(e_{ik}) = d_{ik}^2$ 。根据 Liu 和 Li [20] 的思想, 在此构造线性模型和非参数模型, 引入 $s \times 1$ 维系数向量 θ 和一维未知光滑函数估计自回归系数 $\phi_{k,l}^{(i)} = z_{k,l}^{(i)T} \theta$ 和更新方差 $d_{ik}^2 = d_{ik}^2(t_{ik}) = g(t_{ik})$ 。此处取 $z_{k,l}^{(i)} = (1, t_{ik} - t_{il}, \dots, (t_{ik} - t_{il})^{s-1})^T$ 。

利用最小二乘法对参数向量 θ 进行估计, 求解下面的最小二乘问题即可得到估计值 $\hat{\theta}$:

$$\begin{aligned} \hat{\theta} &= \arg \min \sum_{i=1}^n \sum_{k=2}^{n_i} (\hat{S}_{ik} - \hat{S}_{i1} z_{k,1}^{(i)T} \theta - \dots - \hat{S}_{i,k-1} z_{k,k-1}^{(i)T} \theta)^2 \\ &= \arg \min \sum_{i=1}^n \sum_{k=2}^{n_i} (\hat{S}_{ik} - \hat{v}_{ik} \theta)^2 \\ &= \left(\sum_{i=1}^n \sum_{k=2}^{n_i} (\hat{v}_{ik} \hat{v}_{ik}^T) \right)^{-1} \left(\sum_{i=1}^n \sum_{k=2}^{n_i} (\hat{v}_{ik} \hat{S}_{ik}) \right) \end{aligned}$$

$$\text{其中, } \hat{v}_{ik} = \left(\sum_{l=1}^{k-1} \hat{S}_{il} z_{k,l,1}^{(i)T}, \dots, \sum_{l=1}^{k-1} \hat{S}_{il} z_{k,l,s}^{(i)T} \right)^T。$$

根据 Fan and Yao [21] 的思想, 对光滑函数 $g(t_{ik})$ 的估计可通过求解如下最小二乘问题得到:

$$(\hat{\eta}_1, \hat{\eta}_2) = \arg \min \sum_{i=1}^n \sum_{k=2}^{n_i} [\hat{e}_{ik}^2 - \eta_1 - \eta_2 (t_{ik} - t)]^2 K_h(t_{ik} - t)$$

其中, $K_h(t_{ik} - t) = K\left(\frac{t_{ik} - t}{h}\right) / h$, h 为带宽, $K(\cdot)$ 为核函数, 通过上述问题的求解可知 $d_{ik}^2(t_{ik}) = \hat{\eta}_1$ 。

结合上述讨论, 可得到估计函数 $U_C(\gamma) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} S_i(\gamma)$ 。估计函数的解记为 γ_C , 求出结果需要对估计函数求导, 但 $S_{ik} = M^{-1} \sum_{j=1}^n \sum_{l=1}^{n_j} \left[I\{\varepsilon_{jl}(\gamma) < \varepsilon_{ik}(\gamma)\} - \frac{1}{2} \right]$ 不可微, 则采用 Brown B 和 Wang [18] 提出的诱导光滑方法得到 S_{ik} 的近似式, 即

$$\bar{S}_{ik}(\gamma) = M^{-1} \sum_{j=1}^n \sum_{l=1}^{n_j} \left[\Phi(\varepsilon_{ikjl} / r_{ikjl}) - 0.5 \right], (i, k) \neq (j, l) \quad (5)$$

得到诱导光滑秩估计函数为:

$$\tilde{U}_C(\gamma) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \bar{S}_i(\gamma) \quad (6)$$

其中, $\Phi(\cdot)$ 表示标准正态的卷积分布函数, $\varepsilon_{ikjl} = \varepsilon_{ik} - \varepsilon_{jl}$, $r_{ikjl}^2 = d_{ikjl}^T \Omega d_{ikjl}$, $d_{ikjl} = \Pi_{ik} - \Pi_{jl}$, Ω 是 $\tilde{\gamma}_C$ 的

$p \times p$ 维协方差矩阵, $\tilde{\gamma}_C$ 为诱导光滑秩估计函数的解, $\bar{S}_i(\gamma) = (\bar{S}_{i1}(\gamma), \dots, \bar{S}_{in_i}(\gamma))^T$ 。此时 $\tilde{U}_C(\gamma)$ 函数的特性满足了我们的需求。

2.3. 迭代算法

已知上述求导结果, 算法具体迭代步骤如下:

步骤 1 给定初始值 $\tilde{\gamma}^{(0)} = (0, 0, \dots, 0)^T$ 和 $\Omega^{(0)} = n_{-1}I_p$;

步骤 2 利用第 m 次迭代得到的 $\tilde{\gamma}^{(m)}$ 和 $\tilde{\Omega}^{(m)}$, 在下面公式的基础上更新 $\tilde{\gamma}^{(m+1)}$ 和 $\tilde{\Omega}^{(m+1)}$;

$$\tilde{\gamma}^{(m+1)} = \tilde{\gamma}^{(m)} - \left(\frac{\partial \tilde{U}_C(\gamma)}{\partial \gamma} \Big|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} \tilde{U}_C(\gamma) \quad (7)$$

$$\tilde{\Omega}^{(m+1)} = \left(\frac{\partial \tilde{U}_C(\gamma)}{\partial \gamma} \Big|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} \text{Cov} \left(\tilde{U}_C(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}) \right) \left(\left(\frac{\partial \tilde{U}_C(\gamma)}{\partial \gamma} \Big|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} \right)^T \quad (8)$$

其中, $\text{Cov} \left(\tilde{U}_C(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}) \right) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \tilde{S}_i(\gamma) \tilde{S}_i^T(\gamma) \hat{V}_i^{-1} D_i$ 。

步骤 3 重复上述步骤 2 直至收敛。

3. 变量选择与惩罚秩估计函数

在实际情况下, 真实的模型通常是未知的, 并且我们不知道哪些系数为 0, 但提前假定系数为 0 往往会造成损失, 因此为避免错误假设带来的损失, 考虑到参数的稀疏性问题, 我们在上述诱导光滑秩估计函数的基础上引入满足无偏性的 SCAD 惩罚函数和 MCP 惩罚函数改进模型, 通过构建惩罚秩估计函数得到系数参数更一致的估计。

3.1. 非凸惩罚函数

SCAD 惩罚函数 $p_\lambda^{SCAD}(|\gamma_k|)$ 形式如下:

$$p_\lambda^{SCAD}(|\gamma|) = \begin{cases} \lambda|\gamma|, & |\gamma| \leq \lambda \\ (a^2 - 1)\lambda^2 - (|\gamma| - a\lambda)^2 / 2(a - 1), & \lambda < |\gamma| \leq a\lambda \\ (a + 1)\lambda^2 / 2, & |\gamma| > a\lambda \end{cases} \quad (9)$$

其中 SCAD 惩罚函数的一阶导为

$$p_\lambda^{SCAD'}(|\gamma|) = \begin{cases} \lambda \text{sign}(|\gamma|), & |\gamma| \leq \lambda \\ (a\lambda \text{sign}(|\gamma|) - |\gamma|) / (a - 1), & \lambda < |\gamma| \leq a\lambda \\ 0, & |\gamma| > a\lambda \end{cases}$$

已知 SCAD 惩罚函数没有连续的二阶导数, 但是对于该惩罚函数可以通过一个二次函数近似。现假定有初始值 γ_0 , 惩罚函数的近似函数如下所示:

$$\sum_{k=1}^p p_\lambda^{SCAD}(|\gamma_k|) \approx \sum_{k=1}^p \frac{1}{2} \frac{p_\lambda^{SCAD'}(|\tilde{\gamma}_k|)}{|\tilde{\gamma}_k|} |\gamma_k|^2 = \gamma^T M_{SCAD} \gamma \quad (10)$$

在近似式下, 得到近似的秩估计函数:

$$\tilde{U}_{SCAD}(\gamma) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \bar{S}_i(\gamma) + n\gamma^T M_{SCAD} \quad (11)$$

其中, $\gamma = (\gamma_1^T, \dots, \gamma_p^T)^T$, $M_{SCAD} = \text{diag} \left(\frac{p_\lambda^{SCAD'}(|\tilde{\gamma}_1|)}{2|\tilde{\gamma}_1|}, \dots, \frac{p_\lambda^{SCAD'}(|\tilde{\gamma}_p|)}{2|\tilde{\gamma}_p|} \right)$.

可得到近似的估计等式为: $\sum_{i=1}^n \sum_{k=1}^{n_i} D_{ik}^T \hat{V}_{ik}^{-1} \Delta_{ik}(\gamma) + n \sum_{k=1}^p p_\lambda^{SCAD'}(|\gamma_k^0|) |\gamma_k| = 0$, 其中, γ_k^0 为一个初始估计值, 即非惩罚加权 Wilcoxon 估计量。

计算得到算法迭代式如下:

$$\tilde{\gamma}^{(m+1)} = \tilde{\gamma}^{(m)} - \left(\frac{\partial \tilde{U}_{SCAD}(\gamma)}{\partial \gamma} \right)^{-1} \tilde{U}_{SCAD}(\gamma) \quad (12)$$

$$\tilde{\Omega}^{(m+1)} = \left(\frac{\partial \tilde{U}_{SCAD}(\gamma)}{\partial \gamma} \Big|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} \text{Cov} \left(\tilde{U}_{SCAD}(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}) \right) \left(\left(\frac{\partial \tilde{U}_{SCAD}(\gamma)}{\partial \gamma} \Big|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} \right)^T \quad (13)$$

其中, $\frac{\partial \tilde{U}_{SCAD}(\gamma)}{\partial \gamma} = \sum_{i=1}^n \sum_{k=1}^{n_i} D_{ik}^T \hat{V}_{ik}^{-1} \Delta_{ik}(\gamma) + nM_{SCAD}$,

$$\text{Cov} \left(\tilde{U}_{SCAD}(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}) \right) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \tilde{S}_i(\tilde{\gamma}^{(m)}) \tilde{S}_i^T(\tilde{\gamma}^{(m)}) \hat{V}_i^{-1} D_i + nM_{SCAD}^T \tilde{\gamma}^{(m)} \tilde{\gamma}^{(m)T} M_{SCAD}.$$

MCP 惩罚函数也同样满足 Oracle 性质, 且在处理特征之间有很高相关性数据时, 表现要比 SCAD 惩罚函数更好。MCP 惩罚函数的惩罚项 $p_\lambda^{MCP}(|\gamma_k|)$ 为:

$$p_\lambda^{MCP}(|\gamma|) = \begin{cases} \lambda|\gamma| - \frac{|\gamma|^2}{2a}, & |\gamma| \leq a\lambda \\ \frac{1}{2}a\lambda^2, & |\gamma| > a\lambda \end{cases} \quad (14)$$

MCP 惩罚函数的一阶导为

$$p_\lambda^{MCP'}(|\gamma|) = \begin{cases} \lambda \text{sign}(|\gamma|) - \frac{|\gamma|}{a}, & |\gamma| \leq a\lambda \\ 0, & |\gamma| > a\lambda \end{cases}$$

同样利用二次函数对 MCP 惩罚近似, 得到 MCP 惩罚下的秩估计函数为

$$\tilde{U}_{MCP}(\gamma) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \bar{S}_i(\gamma) + n\gamma^T M_{MCP}, \quad (15)$$

其中, $M_{MCP} = \text{diag} \left(\frac{p_\lambda^{MCP'}(|\tilde{\gamma}_1|)}{2|\tilde{\gamma}_1|}, \dots, \frac{p_\lambda^{MCP'}(|\tilde{\gamma}_p|)}{2|\tilde{\gamma}_p|} \right)$ 。则有估计等式 $\sum_{i=1}^n \sum_{k=1}^{n_i} D_{ik}^T \hat{V}_{ik}^{-1} \Delta_{ik}(\gamma) + n \sum_{k=1}^p p_\lambda^{MCP'}(|\gamma_k^0|) |\gamma_k| = 0$ 。

同样可得算法迭代式:

$$\tilde{\gamma}^{(m+1)} = \tilde{\gamma}^{(m)} - \left(\frac{\partial \tilde{U}_{MCP}(\gamma)}{\partial \gamma} \right)^{-1} \tilde{U}_{MCP}(\gamma) \quad (16)$$

$$\tilde{\Omega}^{(m+1)} = \left(\left. \frac{\partial \tilde{U}_{MCP}(\gamma)}{\partial \gamma} \right|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1} Cov\left(\tilde{U}_{MCP}\left(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}\right)\right) \left(\left. \frac{\partial \tilde{U}_{MCP}(\gamma)}{\partial \gamma} \right|_{\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}} \right)^{-1T} \quad (17)$$

其中, $\frac{\partial \tilde{U}_{MCP}(\gamma)}{\partial \gamma} = \sum_{i=1}^n \sum_{k=1}^{n_i} D_{ik}^T \hat{V}_{ik}^{-1} \Delta_{ik}(\gamma) + nM_{MCP}$,

$$Cov\left(\tilde{U}_{MCP}\left(\tilde{\gamma}^{(m)}, \tilde{\Omega}^{(m)}\right)\right) = \sum_{i=1}^n D_i^T \hat{V}_i^{-1} \tilde{S}_i\left(\tilde{\gamma}^{(m)}\right) \tilde{S}_i^T\left(\tilde{\gamma}^{(m)}\right) \hat{V}_i^{-1} D_i + nM_{MCP}^T \tilde{\gamma}^{(m)} \tilde{\gamma}^{(m)T} M_{MCP}.$$

3.2. 最佳调节参数选择

在(9)和(14)式中, 还有参数 a 需要估计, a 的不同取值会直接影响估计性能。对于(9)式中的参数 a , Fan 和 Li [22]通过蒙特卡洛模拟得出 a 的最优值约等于 3.7, (14)式的 a 在实际使用中通常默认为 3。对于参数 λ , 本文使用 BIC (贝叶斯信息准则)进行参数选择。具体计算式如下:

$$BIC_\lambda = \log\left(\sum_{i=1}^n D_i^T \hat{V}_i^{-1} \bar{S}_i(\gamma)\right) + \frac{df(\lambda)\log(n)}{n} \quad (18)$$

4. 模拟研究

数据来源于如下变系数数模型

$$Y_{ik} = \sum_{k=1}^{n_i} X_{ik}^T \beta(t_{ik}) + \varepsilon_{ik}, i = 1, \dots, n; j = 1, \dots, n_i \quad (19)$$

其中 $\beta_1(t) = \frac{3}{2} - \cos((t-2)\pi)$, $\beta_2(t) = \left(\frac{1}{4} - 2t\right)^2$, $\beta_3(t) = -\frac{2}{3} + \frac{5.5-4t}{300}$, $\beta_{p-3}(t) = 0$ 。 $t_{ik} \sim U(-1,1)$, X_{ik}^1 服从 $[t_{ik}, 3+t_{ik}]$ 上的均匀分布, 在 X_{ik}^1 的条件下, X_{ik}^2 产生于正态分布 $N(0, \Sigma_2)$ 生成的随机数加 1, 其中 $\Sigma_2 = \frac{X_{ik}^1 + 1}{X_{ik}^1 + 2}$; X_{ik}^3 服从参数为 0.8 的伯努利分布, 其余协变量 $(X_{ik}^4, \dots, X_{ik}^{p-3})$ 服从均值为 0, 协方差矩阵为 Υ 的多元正态分布。

为对比不同相关性数据下惩罚函数变量选择的结果, 考虑两种生成协变量的协方差矩阵, 具体如下:

$$\text{CASE I 协方差矩阵 } \Upsilon = Cov\left(X_{ik}^r, X_{jl}^r\right) = \begin{cases} 15e^{-|t_{ik}-t_{jl}|}, & i = j; \\ 0, & i \neq j; \end{cases}$$

$$\text{CASE II 协方差矩阵 } \Upsilon = Cov\left(X_{ik}^r, X_{jl}^r\right) = \begin{cases} 30e^{-0.3|t_{ik}-t_{jl}|}, & i = j; \\ 0, & i \neq j. \end{cases}$$

随机误差 ε_i 服从多元正态分布 $N(0, \Xi_i)$ 。这里协方差矩阵 Ξ_i 采用定义如下的矩阵 $\Xi_i = \Delta_i^{-1} B_i (\Delta_i^T)^{-1}$,

其中 B_i 是 $m_i \times m_i$ 的对角矩阵且第 k 个对角元素为 $\frac{\sin(\pi \zeta_{ik})}{3} + 0.5$, $\zeta_{ik} \sim U(0,2)$, 以及 Δ_i 是对角元素全为 1 的下三角矩阵且第 (k,l) 个元素为 $-\delta_{k,l}^i (k \neq l)$, $\delta_{k,l}^i = W_{k,l}^{(i)T} \theta^*$, 其中 $W_{k,l}^{(i)} = (1, \omega_{ik} - \omega_{il})^T$, $\theta^* = (0.2, 0.5)^T$ 和 $\omega_{ik} \sim U(0,1)$ 。

对于重复观测次数 n_i , 此处我们考虑不平衡的情形, 令 n_i 取 2 到 10 之间的随机数。取 $n = 30, 50, 100$ 为小中大样本, 模拟次数为 100 次。

为了体现模型评估的准确性, 我们采用均方根平均误差(RASE)来对模型结果进行评价, RASE 的具体计算表达式如下:

$$\text{RASE} = \left\{ n^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} (\hat{\beta}_j(U_i) - \beta_j(U_i))^2 \right\}^{\frac{1}{2}} \quad (20)$$

将本文方法下的估计结果记为 β_{PB}^{SCAD} 、 β_{PB}^{MCP} ，与三种结构下的 PGEE 方法进行对比分析，估计结果记为 β_{PGEE}^{indep} 、 β_{PGEE}^{AR} 、 β_{PGEE}^{CS} 。具体见表 1。

Table 1. RASE results for CASE I and CASE II (β_{PB}^{SCAD} 、 β_{PB}^{MCP} ($\times 10^{-1}$))

表 1. CASE I 和 CASE II 的 RASE 结果(其中 β_{PB}^{SCAD} 、 β_{PB}^{MCP} ($\times 10^{-1}$))

n	p_n		CASE I	CASE II
			RASE	
30	12	β_{PGEE}^{indep}	0.7663	0.2021
		β_{PGEE}^{AR}	0.7637	0.1908
		β_{PGEE}^{CS}	0.7025	0.2174
		β_{PB}^{SCAD}	0.0946	0.0870
		β_{PB}^{MCP}	0.0944	0.0870
	15	β_{PGEE}^{indep}	0.6214	0.1936
		β_{PGEE}^{AR}	0.7213	0.1523
		β_{PGEE}^{CS}	0.6547	0.1237
		β_{PB}^{SCAD}	0.0088	0.0073
		β_{PB}^{MCP}	0.0077	0.0072
50	12	β_{PGEE}^{indep}	0.6688	0.1939
		β_{PGEE}^{AR}	0.7633	0.1417
		β_{PGEE}^{CS}	0.6981	0.0871
		β_{PB}^{SCAD}	0.0855	0.0850
		β_{PB}^{MCP}	0.0795	0.0747
	15	β_{PGEE}^{indep}	0.5652	0.1543
		β_{PGEE}^{AR}	0.5012	0.1540
		β_{PGEE}^{CS}	0.4980	0.1246
		β_{PB}^{SCAD}	0.0654	0.0531
		β_{PB}^{MCP}	0.0554	0.0526
100	12	β_{PGEE}^{indep}	0.5165	0.2016
		β_{PGEE}^{AR}	0.5275	0.2058
		β_{PGEE}^{CS}	0.4932	0.2091
		β_{PB}^{SCAD}	0.0393	0.0300
		β_{PB}^{MCP}	0.0393	0.0290
	15	β_{PGEE}^{indep}	0.3960	0.1180
		β_{PGEE}^{AR}	0.3910	0.1324
		β_{PGEE}^{CS}	0.2862	0.1461
		β_{PB}^{SCAD}	0.0386	0.0319
		β_{PB}^{MCP}	0.0376	0.0291

根据上面的表格，可以得到如下结论：

1) 不同维度下处理效果

为了体现本文提出的方法对不同维度样本数据的处理效果差异, 逐步增加模型维数, 分别取 $p_n = 12$ 和 $p_n = 15$ 。对比上述结果, 在小中大三种样本类型下, 模型维数越大, 估计的均方根平均误差越小, 说明模型的估计效率越高。

2) 不同协方差工作结构对估计结果的影响

基于文章研究, 在相同惩罚函数和样本数据结构下, PGEE 方法对应的模型平均误差略高于本文提出方法下的模型误差, 说明利用修正的 Cholesky 分解处理变系数纵向秩回归模型的组内协方差矩阵是有效的, 明显降低了估计效率。同时, 在模拟研究中, 会出现两种方法的模型误差非常接近的情况, 这说明此时对协方差矩阵结构的两种估计都是正确的。总之, 与现有方法相比, 本文提出的方法具有更强的优势, 主要原因在于它允许协方差结构具有更一般的形式, 不需提前指定矩阵结构, 从而可以达到更强的稳健性和有效性。

3) 不同协方差相关性对惩罚函数变量选择的影响

对比相同维度下 CASE I 和 CASE II 的模型结果, 发现当样本协方差矩阵相关性变强时, CASE II 的模型平均误差更小。这是因为当变量高度相关时, MCP 惩罚的硬阈值特性和无偏性使其能更有效地筛选关键变量并减少估计偏差, 而 SCAD 的渐进惩罚可能导致冗余变量残留。结合两个惩罚的表达式分析, $|\gamma| > a\lambda$ 时, MCP 惩罚完全为 0, 而 SCAD 仍然存在轻微惩罚, 为非零常数。

5. 模拟研究

我们将本文提出的方法应用到纵向孕酮数据中, 该数据的具体表述见文章[23]。这项研究收集了 34 名女性的尿样, 每个妇女的重复观测次数从 11 到 28 变化且这组数据共有 492 个观测值, 每位女性的观测次数覆盖了完整的月经周期。但由于不同女性的周期长度不同, 研究采用时间标准化, 将每位女性的月经周期按比例缩放到 28 天的参考周期。同时, 考虑到激素数据通常呈右偏分布, 研究对原始孕酮值进行对数转换, 使其更接近正态分布, 便于后续统计分析。已经有一些学者利用线性模型分析这组数据, 前面的文献旨在研究血清孕酮水平的对数与年龄(Age)、身体质量指数(BMI)和时间(t_{ij})的关系。Fan 等人[24]利用 Huber 得分函数分析这组数据, 尽管有界的 Huber 得分函数可以减弱响应变量中异常值的影响从而得到稳健估计, 但是在非稳健估计情形下会损失一些效率。吕晶等人[15]利用修正的 Cholesky 分解对这组数据进行了有效的秩推断分析, 但是文章并没有考虑变量选择时造成的模型损失。考虑到孕酮数据随时间变化的特殊性, 本文建立了 B 样条近似下的变系数秩回归模型, 叠加非凸惩罚函数进行变量选择, 研究血清孕酮水平的对数与 Age、BMI、 t_{ij} 以及它们交互项之间的关系, 具体模型如下:

$$Y_{ij} = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{BMI}_i + \beta_3 t_{ij} + \beta_4 \text{Age}_i * t_{ij} + \beta_5 \text{BMI}_i * t_{ij} + \beta_6 t_{ij}^2 + \beta_7 \text{Age}_i * \text{BMI}_i + \varepsilon_{ij} \quad (21)$$

下面比较同一样本数据下, 本文所提方法($\hat{\beta}_{PB}^{SCAD}$ 、 $\hat{\beta}_{PB}^{MCP}$)、PGEE 方法($\hat{\beta}_{PGEE}^{CS}$ 、 $\hat{\beta}_{PGEE}^{indep}$ 、 $\hat{\beta}_{PGEE}^{AR}$)以及吕晶等人[15]提出的方法($\hat{\beta}_C$)的结果。表 2 列举了 200 次 Bootstrap 抽样得到的估计系数(Estimate)和对应的标准差(SE)。

由表 2 可得如下结论:

- 1) 在上述所有方法中, 本文提出的方法具有最小标准差, 这说明本文提出的方法更有效;
- 2) 对比 $\hat{\beta}_{PGEE}^{CS}$ 、 $\hat{\beta}_{PGEE}^{indep}$ 、 $\hat{\beta}_{PGEE}^{AR}$ 以及 $\hat{\beta}_{PB}^{SCAD}$ 、 $\hat{\beta}_{PB}^{MCP}$ 可知, 利用修正的 Cholesky 分解处理纵向秩回归协方差矩阵的结构是有效的, 明显降低了模型估计过程中损失的效率;
- 3) 在本文提出的方法下, $\hat{\beta}_{PB}^{MCP}$ 较 $\hat{\beta}_{PB}^{SCAD}$ 的标准差更小, 说明同样作为非凸惩罚函数, 在修正的 Cholesky 分解下的变系数纵向秩回归模型中, MCP 惩罚函数对模型有效性的提高更显著。图 1 是对 200

次 Bootstrap 抽样的一次随机样本数据绘制得到的相关性热图。可知由于交互项的存在，变量 X_3 、 X_4 、 X_5 、 X_6 之间出现了强相关性，比较 $\hat{\beta}_{PB}^{MCP}$ 和 $\hat{\beta}_{PB}^{SCAD}$ 对应的两种估计方法下得到的四个变量的标准差结果，可以看出，强相关下，MCP 惩罚函数表现出更明显的优势；

Table 2. Estimated coefficients ($\times 10^{-2}$) and standard deviations ($\times 10^{-1}$) derived from the progesterone data
表 2. 孕酮数据中所得的估计系数($\times 10^{-2}$)以及标准差($\times 10^{-1}$)

		Age	BMI	Time	Age*Time	BMI* Time	Time* Time	Age* BMI
$\hat{\beta}_C$	Estimate	-5.0348	-5.3870	-15.481	0.2718	0.2670	-0.0201	0.1434
	SE	5.6838	8.2926	1.8319	0.0344	0.0442	0.0243	0.2211
$\hat{\beta}_{PGEE}^{CS}$	Estimate	7.3184	11.5133	-4.2705	0.1122	0.0557	-0.0121	-0.2473
	SE	1.4396	2.2622	1.1004	0.0244	0.0313	0.0173	0.0658
$\hat{\beta}_{PGEE}^{indep}$	Estimate	-7.2348	-12.870	-3.2181	0.1218	0.3670	-0.0211	0.1134
	SE	5.1661	7.7001	1.4750	0.0321	0.0362	0.0339	0.1058
$\hat{\beta}_{PGEE}^{AR}$	Estimate	7.1611	11.6469	-3.8152	0.1128	0.0395	-0.0138	-0.2435
	SE	1.4663	2.2829	1.1037	0.0244	0.0327	0.0176	0.0669
$\hat{\beta}_{PB}^{SCAD}$	Estimate	24.9950	39.2499	-3.5169	0.1956	-0.0744	-0.0484	-0.9459
	SE	1.9106	3.0075	6.9158	0.1284	0.1664	0.0711	0.1113
$\hat{\beta}_{PB}^{MCP}$	Estimate	0.4673	0.7336	-2.6240	0.0725	0.0848	0.0049	0.1542
	SE	0.6155	1.0663	2.9786	0.0671	0.0650	0.0382	0.0498

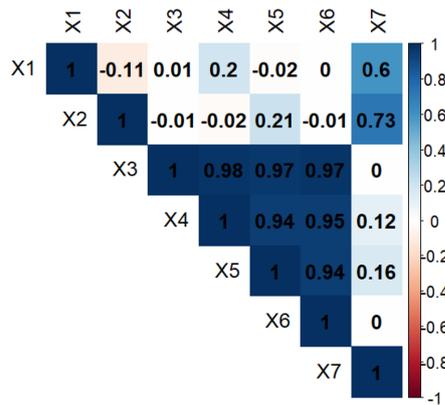


Figure 1. Bootstrap resampling-based correlation heatmap

图 1. Bootstrap 随机抽样样本相关性热图

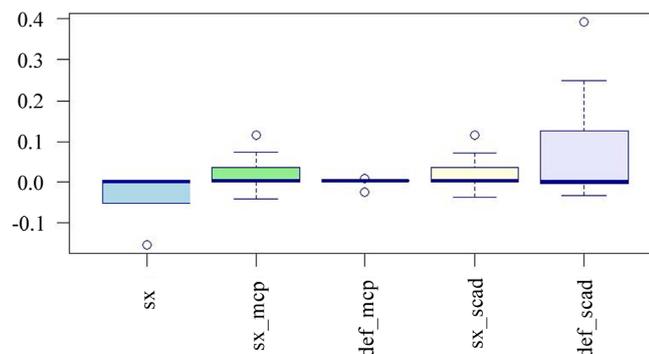


Figure 2. Boxplot of estimation results from 200 bootstrap resampling trials

图 2. 200 次 Bootstrap 随机抽样估计结果的箱线图

4) 绘制上述五种方法的参数估计结果对应箱线图如下, 结合图 2 的特点分析, MCP 惩罚函数下的箱线图相对集中, 说明估计结果的稳定性和一致性更强, 而 SCAD 惩罚下的箱体较宽, 这是因为 SCAD 惩罚对高相关性变量的处理不稳定, 会导致系数被过度压缩或方向不一致。总之, 根据箱线图分析可知, 对于维度相同的样本数据, MCP 惩罚函数对变量的选择更果断, 能有效降低估计方差。

6. 结论

本文针对纵向数据中存在的时变效应与异质性结构, 提出了一种基于变系数线性模型的稳健估计方法。该方法让每个变量的影响大小可以随时间变化, 比如在研究药物疗效时, 药物对病人的作用效果可能会随着治疗时间的推移逐渐增强或减弱。相比之下, 传统方法只能假设药物的效果在整个治疗过程中保持不变, 这往往不符合实际情况。同时, 为进一步提升估计效率, 本文设计了基于修正 Cholesky 分解的协方差矩阵估计方法, 避免了预设参数结构的局限性, 并通过惩罚秩回归技术实现了对离群值与厚尾分布的高鲁棒性。理论证明与数值模拟均表明, 所提方法在参数估计的一致性、效率性及稳健性方面均优于现有方法。除了纵向数据, 本文提出的方法还可扩展至非平稳时间序列分析与时变比例风险模型等场景, 为复杂动态系统的统计建模提供了通用工具。

参考文献

- [1] Diggle, P., Heagerty, P. and Liang, K.Y. (2002) *Analysis of Longitudinal Data*. Oxford University Press.
- [2] Hsiao, C. (2003) *Analysis of Panel Data*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511754203>
- [3] Song, X.K. (2007) *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer.
- [4] Shumway, R.H. (1988) *Applied Statistical Time Series Analysis*. Prentice Hall.
- [5] Cleveland, W.S., Grosse, E. and Shyu, W.M. (2017) Local Regression Models. In: *Statistical Models in S*, Routledge, 309-376. <https://doi.org/10.1201/9780203738535-8>
- [6] Hastie, T. and Tibshirani, R. (1993) Varying-Coefficient Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **55**, 757-779. <https://doi.org/10.1111/j.2517-6161.1993.tb01939.x>
- [7] Fan, J. and Zhang, W. (1999) Statistical Estimation in Varying Coefficient Models. *The Annals of Statistics*, **27**, 1491-1518. <https://doi.org/10.1214/aos/1017939139>
- [8] Chiang, C., Rice, J.A. and Wu, C.O. (2001) Smoothing Spline Estimation for Varying Coefficient Models with Repeatedly Measured Dependent Variables. *Journal of the American Statistical Association*, **96**, 605-619. <https://doi.org/10.1198/016214501753168280>
- [9] Jung, S.H. and Ying, Z.L. (2003) Rank-Based Regression with Repeated Measurements Data. *Biometrika*, **90**, 732-740. <https://doi.org/10.1093/biomet/90.3.732>
- [10] Wang, Y. and Zhu, M. (2006) Rank-Based Regression for Analysis of Repeated Measures. *Biometrika*, **93**, 459-464. <https://doi.org/10.1093/biomet/93.2.459>
- [11] Wang, Y. and Zhao, Y. (2008) Weighted Rank Regression for Clustered Data Analysis. *Biometrics*, **64**, 39-45. <https://doi.org/10.1111/j.1541-0420.2007.00842.x>
- [12] Fu, L. and Wang, Y. (2016) Variable Selection in Rank Regression for Analyzing Longitudinal Data. *Statistical Methods in Medical Research*, **27**, 2447-2458. <https://doi.org/10.1177/0962280216681347>
- [13] Fu, L., Yang, Z., Cai, F. and Wang, Y. (2020) Efficient and Doubly-Robust Methods for Variable Selection and Parameter Estimation in Longitudinal Data Analysis. *Computational Statistics*, **36**, 781-804. <https://doi.org/10.1007/s00180-020-01038-3>
- [14] Fu, L. and Wang, Y. (2012) Efficient Estimation for Rank-Based Regression with Clustered Data. *Biometrics*, **68**, 1074-1082. <https://doi.org/10.1111/j.1541-0420.2012.01760.x>
- [15] 吕晶, 郭朝会, 杨虎, 等. 纵向数据的有效秩推断基于修正的 Cholesky 分解[J]. *数学学报(中文版)*, 2018, 61(4): 549-568.
- [16] Wang, L., Li, H. and Huang, J.Z. (2008) Variable Selection in Nonparametric Varying-Coefficient Models for Analysis of Repeated Measurements. *Journal of the American Statistical Association*, **103**, 1556-1569. <https://doi.org/10.1198/016214508000000788>

-
- [17] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-aos729>
- [18] Brown, B.M. and Wang, Y. (2005) Standard Errors and Covariance Matrices for Smoothed Rank Estimators. *Biometrika*, **92**, 149-158. <https://doi.org/10.1093/biomet/92.1.149>
- [19] Yao, W. and Li, R. (2012) New Local Estimation Procedure for a Non-Parametric Regression Function for Longitudinal Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **75**, 123-138. <https://doi.org/10.1111/j.1467-9868.2012.01038.x>
- [20] Liu, S. and Li, G. (2015) Varying-Coefficient Mean-Covariance Regression Analysis for Longitudinal Data. *Journal of Statistical Planning and Inference*, **160**, 89-106. <https://doi.org/10.1016/j.jspi.2014.11.008>
- [21] Fan, J. and Yao, Q. (1998) Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika*, **85**, 645-660. <https://doi.org/10.1093/biomet/85.3.645>
- [22] 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>
- [23] Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998) Semiparametric Stochastic Mixed Models for Longitudinal Data. *Journal of the American Statistical Association*, **93**, 710-719. <https://doi.org/10.1080/01621459.1998.10473723>
- [24] Fan, Y., Qin, G. and Zhu, Z. (2012) Variable Selection in Robust Regression Models for Longitudinal Data. *Journal of Multivariate Analysis*, **109**, 156-167. <https://doi.org/10.1016/j.jmva.2012.03.007>