第44卷第2期　　中国科学技术大学学报　　Vol. 44, No. 2
2014年2月　　JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA　　Feb. 2014

文章编号:0253-2778(2014)02-0101-5

删失回归模型的分位数变量选择和压缩估计

叶仁玉[1,2]

(1.安庆师范学院数学与计算科学学院，安徽安庆 246133;2.中国科学技术大学管理学院统计与金融系，安徽合肥 230026)

摘要:删失回归模型是一种响应变量受限制的模型,广泛应用于计量经济学中. 针对删失回归模型,借助于分位数估计方法和SCAD型惩罚函数,提出了一种变量选择和压缩估计方法. 该方法可选出对模型有贡献的回归变量,即非0回归系数,同时给出非零参数的一个相合估计. 另外,获得了变量选择方法的oracle性质. 最后,利用数值模拟计算说明所提出方法的效果.

关键词:删失回归模型;分位数估计; SCAD;变量选择; oracle性质

中图分类号:O212.1　　**文献标识码**:A　　doi:10.3969/j.issn.0253-2778.2014.02.004

AMS Subject Classification (2000): Primary 62F12; Secondary 62E20

引用格式: Ye Renyu. Variable selection and shrinkage quantile estimation for censored regression model[J]. Journal of University of Science and Technology of China, 2014,44(2):101-105.
叶仁玉. 删失回归模型的分位数变量选择和压缩估计[J]. 中国科学技术大学学报,2014,44(2):101-105.

Variable selection and shrinkage quantile estimation for censored regression model

YE Renyu[1,2]

(1. *School of Mathematics and Computation Science, Anqing Teachers College, Anqing* 246133, *China*;
2. *Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei* 230026, *China*)

Abstract: Censored regression ("Tobit") model is a kind of limited dependent variable model widely used in econometrics research. Based on the quantiles estimation and the smoothly clipped absolute deviation (SCAD), a method for variable selection and shrinking estimation was presented, which selects the non-zero coefficients corresponding to the significant variables and simultaneously gives a consistent estimate of the parameters. In addition, the variable selection possesses the oracle properties. Finally, numerical studies were conducted to evaluate the performance of the proposed method for censored regression model.

Key words: censored regression model; quantiles estimation; SCAD; variable selection; oracle property

0　引言

考虑删失回归模型

$$Y_i^+ = (x_i^{\mathrm{T}}\beta_0 + e_i)^+,\ i = 1, \cdots, n \tag{1}$$

式中,$Y_i^+ = \max(Y_i, 0)$,$\{x_i\}$为p维的设计向量,β_0为p维的未知参数,且e_i为不可观测的随机误差序

收稿日期:2012-07-09;**修回日期**:2012-12-10

基金项目:安徽省教育厅高等学校自然科学基金(KJ2011A197,KJ2011B084)资助.

作者简介:叶仁玉,女,1977年生,硕士/副教授. 研究方向:数理统计. E-mail:yereny@163.com

列，它的 τ 条件分位数为 0. 显而易见，删失回归模型是一种特殊的响应变量非负限制模型，广泛应用于计量经济学中.

当 τ 取特殊值 0.5 时，Powell[1] 引进了 β_0 的一个最小绝对偏差估计(LAD)，并研究了它的渐近性质，后来关于 β_0 的估计和变量选择，可以参看文献[7-13]. 当 τ 取一般的值($\tau\in(0,1)$)时，Powell[2] 提出了删失回归分位数估计(CRQ) $\hat{\beta}^{(0)}$，即使得准则函数

$$\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i^+-(x_i^{\mathrm{T}}\beta)^+) \tag{2}$$

达到最小值. 式中，

$$\rho_\tau(x)=2(\tau xI(x\geqslant 0)-(1-\tau)xI(x<0)),$$

$0<\tau<1$；$\hat{\beta}^{(0)}$ 的渐近性质也得到了研究. Li[6] 借助于极大值不等式，也获得了删失回归分位数估计的渐近正态性.

在建模时，变量选择是一个重要的环节. 对于线性回归模型，Fan 等[3] 提出了 SCAD(smoothly clipped absolute deviation)方法，且获得了它的最优性质(oracle property). 基于删失回归模型，当 $\tau=1/2$ 时，Wang 等[4] 提出了一个 LASSO 型估计；近来，Liu 等[5] 提出了模型(1)的 SCAD 型变量选择和估计方法，得到了该估计的相合性、稀疏性和渐近正态性.

而当 τ 取一般的值($\tau\in(0,1)$)时，还没有文献研究删失回归模型(1)的变量选择和压缩估计方法。所以，本文主要针对删失回归模型，借助于分位数估计方法和 SCAD 型惩罚函数，提出了一种变量选择和压缩估计方法，并研究了该估计的相合性、稀疏性和渐近正态性. 该方法可选出重要的回归变量，即真参数中非零系数，同时给出非零参数相合估计. 还证明了变量选择方法是相合的，具有 oracle 性质.

1 主要结果

定义 1.1 准则函数

$$Q_n(\beta)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i^+-(x_i^{\mathrm{T}}\beta)^+)+\sum_{j=1}^{p}p_{\lambda_n}(|\beta_j|) \tag{3}$$

式中，$p_{\lambda_n}(\cdot)$ 为依赖于正则参数 λ_n 的 SCAD 型惩罚函数，对 $p_\lambda(\cdot)$ 满足

$$p'_\lambda(t)=\lambda\{I(t\leqslant\lambda)+\frac{(a\lambda-t)^+}{(a-1)\lambda}I(t>\lambda)\},$$

$$a>2,\ t>0.$$

使得 $Q_n(\beta)$ 达到最小的 $\hat{\beta}_n$ 为 β 的删失回归模型分位数估计的 SACD 型压缩估计(CRSSE)，即 $\hat{\beta}_n=\operatorname{argmin} Q_n(\beta)$. 该定义包括了文献[4-5]中的估计，$p_{\lambda_n}(\cdot)$ 良好的性质参见文献[3].

这里研究的主要内容包括该估计的相合性、稀疏性和渐近性. 设 $\beta_0=(\beta_0^1,\beta_0^2)^{\mathrm{T}}$，其中，$\beta_0^1$ 为 $1\times s$ 向量，β_0^2 为 $1\times(p-s)$ 向量. 不失一般性，假设 β_0^1 为 β_0 中非 0 元素，$\beta_0^2=0$. 为了简单起见，记

$$\mu_i=x_i^{\mathrm{T}}\beta_0,\ S_n=\sum_{i=1}^{n}I(\mu_i>0)x_ix_i^{\mathrm{T}},$$

$$a_n=\max\{p'_{\lambda_n}(|\beta_{0i}|):\beta_{0i}\neq 0\},$$

$$b_n=\max\{p''_{\lambda_n}(|\beta_{0i}|):\beta_{0i}\neq 0\}.$$

阐述定理之前先进行一些条件的假设：

(A1) $e_1,e_2,\cdots$ 为 i. i. d. 随机变量，e_i 的 τ 条件分位数为 0，其密度函数 $f(\cdot)$ 在 0 附近连续，且严格大于 0.

(A2) β_0 的参数空间 B 是 R^p 中的一个有界开凸集合(它的闭包为 $\overline{B}$)，且 $0\in B$.

(A3) $S_n/n\to V^2$，$n\to\infty$，其中，$V>0$.

(A4) 对任意 $\gamma>0$，存在一个有限正数 $\alpha>0$ 使得

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2I(\|x_i\|>\alpha)<\gamma,\ \text{当 } n \text{ 充分大}.$$

(A5) 对任意 $\gamma>0$，存在一个有限正数 $\delta>0$ 使得

$$\frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2I(|\mu_i|\leqslant\delta)<\gamma,\ \text{当 } n \text{ 充分大}.$$

(A6) $\max\{\|x_i\|:i=1,\cdots,n\}=o(\sqrt{n})$.

(A7) $\liminf\limits_{n\to\infty}\liminf\limits_{\theta\to 0+}p'_{\lambda_n}(\theta)/\lambda_n>0$.

(A8) $a_n=O(n^{-1/2})$，$b_n=o(1)$.

条件(A1)～(A5)用于讨论 $\hat{\beta}_n$ 的相合性和渐近正态性[4]，(A7)和(A8)保证了 $\hat{\beta}_n$ 是 β_0 的稀疏估计[3]，(A6)为证明该稀疏性提供了支持，一般情况下，解释变量 x 是有限的，显然满足条件(A6).

定理 1.1 假设在模型(1)下，条件(A1)～(A5)成立；若 $b_n\to 0$，当 $n\to\infty$，则 $Q_n(\beta)$ 存在局部最小值 $\hat{\beta}_n$ 使得

$$\|\hat{\beta}_n-\beta_0\|=O_p(n^{-1/2}+a_n).$$

从定理 1.1 可知通过选择一个合适的 λ_n，可以得到一个关于 β_0 的 n 次方根的一致处罚估计.

定理 1.2 在模型(1)下，假设定理 1.1 中条件

和条件(A6)～(A8)成立，若 $\lambda_n \to 0$ 和 $\sqrt{n}\lambda_n \to \infty$，当 $n \to \infty$ 时，则以概率 1 有

① (稀疏性) $\hat{\beta}_{n0}^2 = 0$.

② (渐近正态性)

$$\frac{f(0)}{\sqrt{\tau(1-\tau)}}\sqrt{n}V_1(\hat{\beta}_{n0}^1 - \beta_0^1) \xrightarrow{d} N(0, V_1) \quad (4)$$

式中，$\hat{\beta}_n = (\hat{\beta}_{n0}^1, \hat{\beta}_{n0}^2)^T$，$V_1$ 为 V^2 的前 $s \times s$ 的矩阵.

2　定理证明

定理 1.1 证明

设 $r_n = n^{-1/2}$，$u = (u_1, u_2, \cdots, u_p)^T$ 且 $\| u \| = c$，其中，c 为充分大的一个常数，我们要证明对任意给定的 $\varepsilon > 0$，使得对充分大的 n 有

$$P\{\inf_{\| u \| = c} Q_n(\beta_0 + r_n u) > Q_n(\beta_0)\} \geqslant 1 - \varepsilon \quad (5)$$

式(5)说明在空间球体区域 $\{\beta_0 + r_n u: \| u \| \leqslant c\}$ 存在最小值的概率至少为 $1-\varepsilon$，所以存在局部最小值 $\hat{\beta}_n$ 使得 $\| \hat{\beta}_n - \beta_0 \| = O_p(r_n)$. 由 $p_{\lambda_n}(0) = 0$，可得

$$\begin{aligned} &Q_n(\beta_0 + r_n u) - Q_n(\beta_0) = \\ &\frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i^+ - (x_i^T\beta_0 + r_n x_i^T u)^+) - \\ &\frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i^+ - (x_i^T\beta_0)^+) + \\ &\sum_{j=1}^s p_{\lambda_n}(|\beta_{0j} + r_n u_j|) - \sum_{j=1}^s p_{\lambda_n}(|\beta_{0j}|) = \\ &\mathrm{I} + \mathrm{II} \end{aligned} \quad (6)$$

式中，s 为 β_0 中非 0 元素的个数，由 Taylor 展开，可得

$$\begin{aligned} \mathrm{II} &= r_n\sum_{j=1}^s p'_{\lambda_n}(|\beta_{0j}|)u_j + \\ &\frac{1}{2}r_n^2\sum_{j=1}^s p''_{\lambda_n}(|\beta_{0j}|)u_j^2(1 + o(1)) \leqslant \\ &r_n a_n\sqrt{s}\| u \| + r_n^2 b_n \| u \|^2 \end{aligned} \quad (7)$$

由文献[4,6]，得

$$\begin{aligned} \mathrm{I} &= f(0)r_n^2 u^T V^2 u + o(r_n^2 u^T V^2 u) + \\ &W_n^T r_n u + o_p\left(\frac{\sqrt{r_n^2 u^T V^2 u}}{\sqrt{n}}\right) \end{aligned} \quad (8)$$

式中，

$$\begin{aligned} W_n &= \frac{1}{n}\sum_{i=1}^n [2(1-\tau)I(e_i(\omega) < 0) - \\ &2\tau I(e_i(\omega) \geqslant 0)]I(\mu_i > 0)x_i, \end{aligned}$$

$I(\cdot)$ 为示性函数，因为 $W_n \xrightarrow{d} N(0, 4\tau(1-\tau)V^2)$，所以

$$\mathrm{I} = f(0)r_n^2 u^T V^2 u(1 + o(1) + O_p(1) + o_p(1)) \quad (9)$$

由式(6)、(7)、(9)知，选择充分大的 c，使得 $\| u \| = c$ 以及 $f(0)r_n^2 u^T V^2 u > 0$，定理 1.1 得证.　□

定理 1.2 证明

由定理 1.1 和(A8)，对充分大的 c，在球体区域 $\{\beta_0 + r_n u: \| u \| \leqslant c\}$ 中 $\hat{\beta}_n$ 以概率 1 存在，其中 $r_n = n^{-1/2}$. 设 $\hat{\beta}_{n1} = (\hat{\beta}_{n0}^1, 0)^T$，$f_k = (0, \cdots, 1, \cdots, 0)^T$，其中，1 位于第 k 个坐标，$k = s+1, \cdots, p$. 定义

$$\begin{aligned} &H(t) = Q_n(\hat{\beta}_{n1} + tf_k) - Q_n(\hat{\beta}_{n1}) = \\ &\frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i^+ - (x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+) - \\ &\frac{1}{n}\sum_{i=1}^n \rho_\tau(y_i^+ - (x_i^T\hat{\beta}_{n1})^+) + p_{\lambda_n}(|t|) = \\ &\frac{1}{n}\sum_{i=1}^n \int_{(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+}^{(x_i^T\hat{\beta}_{n1})^+} \varphi_\tau(y_i - v)\mathrm{d}v + p_{\lambda_n}(|t|). \end{aligned}$$

被积函数 $\varphi_\tau(x) = 2\tau - 2I(x < 0) + (4\tau - 2)xD(x)$，其中 $D(\cdot)$ 为 Dirac δ 函数. 对 t 求导得

$$\begin{aligned} &\frac{\mathrm{d}H(t)}{\mathrm{d}t} = \\ &\frac{1}{n}\sum_{i=1}^n \varphi_\tau(y_i - (x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+) \cdot \\ &\frac{\mathrm{d}(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+}{\mathrm{d}t} + p'_{\lambda_n}(|t|)\mathrm{sgn}(t) = \\ &\frac{1}{n}\sum_{i=1}^n \varphi_\tau(y_i - (x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+) \cdot \\ &I(x_i^T\hat{\beta}_{n1} + tx_i^T f_k > 0)x_i^T f_k + \\ &\frac{1}{n}\sum_{i=1}^n \varphi_\tau(y_i - (x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+) \cdot \\ &(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)D(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)x_i^T f_k + \\ &p'_{\lambda_n}(|t|)\mathrm{sgn}(t) = \\ &\frac{1}{n}\sum_{i \in M} \varphi_\tau(y_i - x_i^T\hat{\beta}_{n1} - tx_i^T f_k)x_i^T f_k + \\ &\frac{1}{n}\sum_{i=1}^n \varphi_\tau(y_i - (x_i^T\hat{\beta}_{n1} + tx_i^T f_k)^+) \cdot \\ &(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)D(x_i^T\hat{\beta}_{n1} + tx_i^T f_k)x_i^T f_k + \\ &p'_{\lambda_n}(|t|)\mathrm{sgn}(t) = \\ &\frac{1}{n}\sum_{i \in M} \varphi_\tau(x_i^T\beta_0 + e_i - x_i^T\hat{\beta}_{n1} - tx_i^T f_k)x_i^T f_k + \\ &p'_{\lambda_n}(|t|)\mathrm{sgn}(t), \end{aligned}$$

其中，

$$M = \{1 \leqslant i \leqslant n: x_i^{\mathrm{T}}\beta_{n1} + t x_i^{\mathrm{T}} f_k > 0\},$$

由条件(A4)和(A6),得 $\frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2 = O(1)$ 和 $\max\{|n^{-1/2}x_i^{\mathrm{T}} f_k|: i=1,\cdots,n\} = o(1)$. 利用中心极限定理可得

$$\frac{1}{\sqrt{n}}\sum_{i\in M}\varphi_\tau(x_i^{\mathrm{T}}\beta_0 + e_i - x_i^{\mathrm{T}}\hat{\beta}_{n1} - t x_i^{\mathrm{T}} f_k) x_i^{\mathrm{T}} f_k\} = O_p(1).$$

因为 $|t|\leqslant cn^{-1/2}$,利用 $\sqrt{n}\lambda_n \to \infty$ 和条件(A7),

$$\frac{\mathrm{d}H(t)}{\mathrm{d}t} = \frac{1}{\sqrt{n}}\Big(\frac{1}{\sqrt{n}}\sum_{i\in M}\varphi_\tau(x_i^{\mathrm{T}}\beta_0 + e_i - x_i^{\mathrm{T}}\hat{\beta}_{n1} - t x_i^{\mathrm{T}} f_k) x_i^{\mathrm{T}} f_k + \sqrt{n} p'_{\lambda_n}(|t|)\mathrm{sgn}(t)\Big) = \frac{1}{\sqrt{n}}\Big(O_p(1) + \sqrt{n}\lambda_n \frac{p'_{\lambda_n}(|t|)}{\lambda_n}\mathrm{sgn}(t)\Big).$$

显然 t 决定了 $\mathrm{d}H(t)/\mathrm{d}t$ 符号,①证毕,下证②.

为简化,设 $\gamma_n = \hat{\beta}_{n0}^1 - \beta_0^1$. 由 $Q_n(\beta)$ 的定义,得

$$Q_n((\hat{\beta}_{n0}^1, 0)^{\mathrm{T}}) - Q_n((\beta_0^1, 0)^{\mathrm{T}}) = f(0)\gamma_n^{\mathrm{T}} V_1\gamma_n + W_n^{*\mathrm{T}}\gamma_n + o(\gamma_n^{\mathrm{T}} V_1\gamma_n) + o_p\left(\frac{\sqrt{\gamma_n^{\mathrm{T}} V_1\gamma_n}}{\sqrt{n}}\right) + \sum_{j=1}^{s} p'_{\lambda_n}(|\beta_{0j}|)\mathrm{sgn}(\beta_{0j})\gamma_{nj} + \frac{1}{2}\sum_{j=1}^{s} p''_{\lambda_n}(|\beta_{0j}|)\gamma_{nj}^2(1+o(1)).$$

其中,V_1 为 V^2 的前 $s\times s$ 的矩阵,W_n^* 为式(8)中 W_n 相对应的前 s 个元素组成的向量.

设

$$u_n = \sqrt{n}\gamma_n,$$

$$\Sigma = \mathrm{diag}\{p''_{\lambda_n}(|\beta_{01}|),\cdots,p''_{\lambda_n}(|\beta_{0s}|)\},$$

$$b = (p'_{\lambda_n}(|\beta_{01}|)\mathrm{sgn}(\beta_{01}),\cdots, p'_{\lambda_n}(|\beta_{0s}|)\mathrm{sgn}(\beta_{0s}))^{\mathrm{T}},$$

令 $B_n(u) = f(0)u^{\mathrm{T}} V_1 u + W_n^{*\mathrm{T}} u + n^{1/2} b^{\mathrm{T}} u + \frac{1}{2}u^{\mathrm{T}}\Sigma u$,对 $B_n(u)$ 关于 u 求一阶导数,得 $B_n(u)$ 最小值

$$\tilde{u}_n = (2f(0)V_1 + \Sigma)^{-1}(-W_n^{*\mathrm{T}} - n^{1/2}b).$$

因为

$$W_n^* \xrightarrow{d} N(0, 4\tau(1-\tau)V_1),$$

所以

$$(2f(0)V_1 + \Sigma)\tilde{u}_n + n^{1/2}b \xrightarrow{d} N(0, 4\tau(1-\tau)V_1) \tag{10}$$

类似于文献[4]的定理2.2证明,

$$Q_n((\hat{\beta}_{n0}^1, 0)^{\mathrm{T}}) - Q_n((\beta_0^1, 0)^{\mathrm{T}}) = \frac{1}{n}B_n(u_n) + o_p\left(\frac{1}{n}\right),$$

$$B_n(u_n) - B_n(\tilde{u}_n) \geqslant (u_n - \tilde{u}_n)^{\mathrm{T}}\left(f(0)V_1 + \frac{1}{2}\Sigma\right)(u_n - \tilde{u}_n).$$

所以,

$$Q_n((\hat{\beta}_{n0}^1, 0)^{\mathrm{T}}) - Q_n((\beta_0^1 + n^{-1/2}\tilde{u}_n, 0)^{\mathrm{T}}) = \frac{1}{n}(B_n(u_n) - B_n(\tilde{u}_n)) + o_p\left(\frac{1}{n}\right) \geqslant \frac{1}{n}f(0)(u_n - \tilde{u}_n)^{\mathrm{T}} V_1(u_n - \tilde{u}_n) + o_p\left(\frac{1}{n}\right) \geqslant \frac{c}{n}f(0)(u_n - \tilde{u}_n)^{\mathrm{T}}(u_n - \tilde{u}_n) + o_p\left(\frac{1}{n}\right),$$

其中,c 为独立于样本容量 n 的一个正常数. $\hat{\beta}_n$ 为 $(\hat{\beta}_{n0}^1, 0)$ 形式,由式(6)知

$$Q_n((\hat{\beta}_{n0}^1, 0)^{\mathrm{T}}) - Q_n((\beta_0^1 + n^{-1/2}\tilde{u}_n, 0)^{\mathrm{T}}) \leqslant 0,$$

所以,$u_n \overset{d}{=\!=} \tilde{u}_n$.

若 $a_n = o(n^{-1/2})$,则当 n 趋于无穷时,$n^{1/2}b = o(1)$,$\Sigma \to 0$. 此时式(10)变为

$$\frac{f(0)}{\sqrt{\tau(1-\tau)}}V_1 u_n \xrightarrow{d} N(0, V_1) \tag{11}$$

定理1.2证毕. □

3 模拟计算

我们利用数值模拟来研究所提方法的表现能力. 由于规则化参数 λ_n 未知,BIC 准则被用来选择合适的规则化参数,BIC 准则如下:

$$\mathrm{BIC}(\lambda) = \frac{nL_n(\hat{\beta}_n(\lambda))}{L_n(\hat{\beta}^{(0)})} + \frac{\ln n}{2}\#\{j: \hat{\beta}_{nj} \neq 0\},$$

其中,

$$L_n(\beta) = \sum_{i=1}^{n}\rho_\tau(y_i^+ - (x_i^{\mathrm{T}}\beta)^+),$$

且 $\hat{\beta}^{(0)}$ 是模型(1)的 LAD 估计.

模拟数据从模型(1)产生,其中,解释变量 x 除了截距项外各个元素从标准正态中产生,真参数 $\beta_0 = (1.5, 1, -1, 1, 0, 0, 0, 0, 0)^{\mathrm{T}}$,取 $n=50$,100 和 200 不同的样本量. 每个模拟过程重复1 000次. 模型 M1 为我们提出的分位数 SCAD 型变量选择和估计,模型 M2 为文献[5]提出的对模型(1)SCAD 型变量选择和估计,对这两种变量选择方法和估计进行比较,结果如表1所示.

表 1　真参数 $\beta_0=(1.5,1,-1,1,0,0,0,0,0)^T$ 时非 0 参数估计和变量选择结果

Tab. 1　Non-zero parameters estimation and variable selection results for true parameter $\beta_0=(1.5,1,-1,1,0,0,0,0,0)^T$

model	n	1.5	1	−1	1	γ	γ_1
M1	50	1.427(0.496)	1.041(0.246)	−1.026(0.293)	1.043(0.257)	0.015(0.197)	3.619(1.643)
	100	1.478(0.170)	1.015(0.165)	−1.016(0.170)	1.017(0.170)	0(0)	4.115(1.375)
	200	1.486(0.120)	1.008(0.110)	−1.009(0.122)	1.008(0.114)	0(0)	4.566(0.966)
M2	50	1.286(0.454)	0.979(0.404)	−0.963(0.385)	0.936(0.420)	0.34(1.099)	2.440(2.187)
	100	1.443(0.157)	1.03(0.185)	−1.059(0.230)	0.997(0.168)	0(0)	4.040(1.616)
	200	1.478(0.117)	1.026(0.133)	−1.013(0.124)	1.019(0.101)	0(0)	4.520(0.931)

【注】 ① 括号内数值为标准差.

② γ 为非 0 参数错误地估为 0 的平均个数，γ_1 为 0 参数被估为 0 的平均个数.

从表 1 可以看出，随着样本量的增大，两种模型方法中的非 0 参数的估计值逐渐逼近真参数，且真参数为 0 的错估平均个数趋于 0，即真参数为 0 被正确地估为 0 的平均个数趋于真值 5，说明这两种方法得到的估计都具有相合性以及 oracle 性质. 同时，可以看出我们提出的方法得到的估计在精度上略优于文献[5]的估计.

本文在删失回归模型中，把分位数估计方法和 SCAD 型惩罚函数相结合，提出了对于一般的概率值的一种变量选择和压缩估计方法. 该估计具有相合性、稀疏性、渐近正态性和变量选择的 oracle 性质.

致谢　在本文形成的过程中，得到王占锋、赵林城老师的指导帮助，特致衷心的感谢！

参考文献(References)

[1] Powell J L. Least absolute deviations estimates for the censored regression model [J]. Journal of Econometrics, 1984, 25: 303-325.

[2] Powell J L. Censored regression quantiles[J]. Journal of Econometrics, 1986, 32(1): 43-155.

[3] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. Journal of the American Statistical Association, 2001, 96: 1 348-1 360.

[4] Wang Zhanfeng, Wu Yaohua, Zhao Lincheng. A LASSO-type approach to variable selection and estimation for censored regression model [J]. China Journal of Applied Probability and Statistic, 2010, 26(1):66-80.
王占锋，吴耀华，赵林城. 删失回归模型中一个 LASSO 型变量选择和估计方法[J]. 应用概率统计，2010，26(1)：66-80.

[5] Liu Xianhui, Wang Zhanfeng, Wu Yaohua. Variable selection and estimation via SCAD-type for censored regression model[J]. Journal of University of Science and Technology of China, 2013,43(3):182-189.
刘显慧，王占锋，吴耀华. 删失回归模型中 SCAD 型变量选择与估计[J]. 中国科学技术大学学报，2013，43(3):182-189.

[6] 李硕. 删失回归模型中的分位数统计[D]. 合肥：中国科学技术大学，2012.

[7] Pollard D. Empirical Process: Theory and Application [M]. Hayward: Institute of Mathematical Statistics, 1990.

[8] Chen X, Wu Y. Consistency of L_1 estimates in censored linear regression models[J]. Commu Statist Theor Meth, 1993,23:1 847-1 858.

[9] Rao C, Zhao L. Asymptotic normality of LAD estimators in censored regression models [J]. Math Meth Statist, 1993, 2:228-239.

[10] Zhao L, Fang Y. Random weighting method for censored regression model [J]. J Systems Science and Complexity, 2004, 17: 262-270.

[11] Jin M, Fang Y, Zhao L. Variable selection for censored regression models[J]. Chin J Applied Probab Statist, 2005,21:141-149.

[12] Fang Y, Jin M,Zhao L. Strong convergence of LAD estimates in a censored regression model [J]. Science in China, Ser A, Math, 2005, 48:155-168.

[13] Wang Z, Wu Y,Zhao L. Change-point estimation for censored regression model [J]. Science in China, Ser A, Math, 2007, 50:63-72.