双截尾 Tobit 模型中的随机加权逼近方法

王占锋*, 姚双双, 吴耀华

中国科学技术大学管理学院统计与金融系, 合肥 230026

E-mail: zfw@ustc.edu.cn, yshsh7@mail.ustc.edu.cn, wuyh@ustc.edu.cn

收稿日期: 2017-01-09; 接受日期: 2017-10-18; 网络出版日期: 2018-03-13; * 通信作者
国家自然科学基金 (批准号: 11231010 和 11471302) 资助项目

摘要 本文研究双截尾删失回归模型中参数的随机加权估计 (RWE), 获得了 RWE 的统计渐近性质, 如相合性和渐近分布. 本文证明了 RWE 在给定样本下的条件渐近分布与参数的最小绝对偏差 (LAD) 估计的渐近分布是一样的, 则可以利用 RWE 的条件分布去逼近回归参数的 LAD 估计的分布, 从而避免冗余参数的估计, 如误差项的密度函数. 另外, 本文也提出了一个 M 检验统计量和随机加权 M 检验统计量 (RWM) 来检验参数的线性假设问题, 建立了该检验的统计性质. 数值模拟和实际数据分析结果表明所提方法是可行的.

关键词 双截尾删失回归模型 随机加权方法 渐近性 冗余参数

MSC (2010) 主题分类 62F10, 62F03

1 引言

删失回归 (“Tobit”) 模型常常用来分析受限制的响应变量数据. 本文考虑如下模型:

$$Y_i^* = (x_i^\top \beta_0 + e_i)I(d \leq x_i^\top \beta_0 + e_i \leq M) + MI(x_i^\top \beta_0 + e_i > M), \quad i = 1, \dots, n, \quad (1.1)$$

其中 x_i 是 p 维协变量, β_0 是未知的回归系数, e_i 是误差项, $I(\cdot)$ 是示性函数, d 和 M 是两个给定的常数. 在模型 (1.1) 中, 潜在随机变量 Y_i 受到下界 d 和上界 M 的限制, 导致观测到的变量为 Y_i^* , 所以, 模型 (1.1) 称为双截尾 Tobit 回归模型. 不失一般性, 取 $d = 0$ 和 $M > 0$.

当 $M = \infty$ 时, 模型 (1.1) 退化成常见的 Tobit 模型, 该模型的估计和检验问题已经被广泛研究, 参见文献 [1–8] 及其所引用的参考文献. 例如, 文献 [2] 提出了参数 β_0 的最小绝对偏差 (LAD) 估计; 文献 [7] 利用随机加权方法去逼近参数估计的分布. 最近, 文献 [8] 研究了在协变量的维度随样本量变化下的 Tobit 模型, 并且证明了参数估计的统计性质.

英文引用格式: Wang Z F, Yao S S, Wu Y H. A randomly weighting approximation approach for two-tailed censored regression model (in Chinese). Sci Sin Math, 2018, 48: 955–968, doi: 10.1360/SCM-2017-0016

然而, 在实际应用中, 我们经常会遇到 $M < \infty$ 的情形, 例如, 受到测量仪器和工具的上界精度限制. 文献 [9] 提出了 β_0 的一个 LAD 估计

$$\hat{\beta}_n = \operatorname{argmin}_{\beta \in \bar{B}} \sum_{i=1}^n |Y_i^* - [x_i^\top \beta I(0 \leq x_i^\top \beta \leq M) + M(x_i^\top \beta > M)]|, \quad (1.2)$$

其中参数空间 B 是 $\mathbb{R}^p$ 中的有界开集 (闭包为 $\bar{B}$). 他们证明了 $\hat{\beta}_n$ 是相合的, 且具有渐近正态分布

$$2f(0)S_n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, I_p), \quad (1.3)$$

其中 $S_n = \sum_{i=1}^n I(0 \leq \mu_i \leq M)x_i x_i^\top$, $\mu_i = x_i^\top \beta_0$, $f(0)$ 是误差项的密度函数在 0 处的值. 由此可见, $\hat{\beta}_n$ 的渐近分布中包含冗余参数, 如密度函数 $f(0)$. 众所周知, 由于数据点在 0 点处左删失, $f(0)$ 比较难估计, 尤其在样本量较小的情形下.

本文应用随机加权方法给出了模型 (1.1) 中回归系数的随机加权估计 (RWE). 我们证明了 RWE 是 β_0 的相合估计, 且给出了 RWE 的 Bahadur 表示. 另外, 我们指出, 在给定观测样本下, RWE 的条件渐近分布与 $\hat{\beta}_n$ 的渐近分布相同, 因此, RWE 的条件分布可以用来逼近 $\hat{\beta}_n$ 的分布, 从而避免估计冗余参数. 据我们所知, 很少文献研究当 $M < \infty$ 时回归参数的线性假设检验问题, 我们构造了一个 M 检验统计量去检验线性假设. 为了避免估计密度函数, 我们也提出了随机加权 M 检验统计量 (RWM), 用来逼近 M 检验统计量的分布, 并且研究了 RWM 的渐近性质.

本文的主要内容如下: 第 2 节介绍 RWE、M 检验统计量和 RWM, 研究它们的统计性质; 第 3 节为数值模拟和真实数据分析; 第 4 节为结论; 所有定理的证明见附录 A.

2 随机加权估计和检验

假设模型 (1.1) 中的潜在变量 (Y_i, x_i) ($i = 1, \dots, n$) 独立同分布. 由于响应变量受限制, 我们只能观测到 $Y_i^* = Y_i I(0 \leq Y_i \leq M) + M I(Y_i > M)$, 因此, 观测样本为 (Y_i^*, x_i) , $i = 1, \dots, n$. 记 $\lambda(A)$ 是对称矩阵 A 的最小特征值.

在给出定理之前, 我们还需要下述假设条件:

(A₁) $e_1, e_2, \dots$ 是独立同分布 (i.i.d.) 随机变量, 且其分布函数的中位数为 0, 它的密度函数在 0 处的值 $f(0)$ 大于 0;

(A₂) β_0 的参数空间 B 是 $\mathbb{R}^p$ 上的有界开集 (闭包为 $\bar{B}$);

(A₃) 对 $\forall \varepsilon > 0$, 存在一个有界 $N > 0$, 使得

$$\sum_{i=1}^n \|x_i\|^2 I(\|x_i\| > N) < \varepsilon \lambda(S_n), \quad \text{当 } n \text{ 足够大};$$

(A₄) 对 $\forall \varepsilon > 0$, 存在一个有界 $\delta > 0$, 使得

$$\sum_{i=1}^n \|x_i\|^2 I(|\mu_i| \leq \delta) < \varepsilon \lambda(S_n), \quad \text{当 } n \text{ 足够大};$$

(A₅) 当 $n \rightarrow \infty$ 时, $\lambda(S_n)/\log^2 n \rightarrow \infty$;

(A₆) $w_1, w_2, \dots$ 是 i.i.d. 的有界随机变量, $w_1 \geq 0$ a.s., $E(w_1) = 1$, $\operatorname{Var}(w_1) = \sigma^2$, 且 $\{(Y_i, x_i)\}$ 与 $\{w_i\}$ 独立.

条件 (A₁)–(A₅) 是研究 Tobit 模型所需的常用条件, 参见文献 [2, 5]. 条件 (A₆) 是研究随机加权估计的一般条件, 参见文献 [5, 7].

2.1 随机加权估计

类似于文献 [5], β_0 的随机加权估计 (RWE) 可定义为

$$\beta_n^* = \operatorname{argmin}_{\beta \in \bar{B}} \sum_{i=1}^n w_i |Y_i^* - [x_i^\top \beta I(0 \leq x_i^\top \beta \leq M) + M(x_i^\top \beta > M)]|. \quad (2.1)$$

下面给出 β_n^* 的相合性和 Bahadur 强表示定理, 其证明参见附录 A.

定理 2.1 假设在模型 (1.1) 中, 条件 (A₁)–(A₄) 和 (A₆) 成立, 且

$$\frac{\lambda(S_n)}{\log n} \rightarrow \infty, \quad n \rightarrow \infty, \quad (2.2)$$

则

$$\beta_n^* \xrightarrow{P} \beta_0, \quad n \rightarrow \infty.$$

定理 2.2 假设在模型 (1.1) 中, 条件 (A₁)–(A₆) 成立, 则

$$2f(0)S_n^{\frac{1}{2}}(\beta_n^* - \beta_0) = \sum_{i=1}^n w_i I(0 < \mu_i < M) \operatorname{sgn}(e_i) x_{in} + o_p(1),$$

其中 $x_{in} = S_n^{-\frac{1}{2}} x_i$.

当 $w_i = 1$ 时, 由 (2.1) 和定理 2.2 可得到 $\beta_n^* = \hat{\beta}_n$ 且

$$2f(0)S_n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0) = \sum_{i=1}^n I(0 < \mu_i < M) \operatorname{sgn}(e_i) x_{in} + o_p(1) \xrightarrow{d} N(0, I_p), \quad (2.3)$$

这与文献 [9] 的结论是一致的. 由定理 2.1 和 2.2 及 (2.3), 可以得到

$$2f(0)S_n^{\frac{1}{2}}(\beta_n^* - \hat{\beta}_n) = \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \operatorname{sgn}(e_i) x_{in} + o_p(1). \quad (2.4)$$

令 P^* 和 $\mathcal{L}^*$ 表示在样本 $\{(Y_i^*, x_i), i = 1, \dots, n\}$ 给定条件下的条件概率和条件分布. 当 $E(w_i) = 1$ 且 $\operatorname{Var}(w_i) = \sigma^2 < \infty$ 时, 由 (2.4) 可以得到下述定理.

定理 2.3 假设条件 (A₁)–(A₆) 成立, 则

$$\mathcal{L}^*\left(\frac{2f(0)S_n^{\frac{1}{2}}(\beta_n^* - \hat{\beta}_n)}{\sigma}\right) \xrightarrow{d} N(0, I_p).$$

进一步地,

$$\sup_u \left| P^*\left(\frac{S_n^{\frac{1}{2}}(\beta_n^* - \hat{\beta}_n)}{\sigma} \leq u\right) - P(S_n^{\frac{1}{2}}(\hat{\beta}_n - \beta_0) \leq u) \right| \xrightarrow{P} 0,$$

其中 u 遍历所有 p 维向量.

特别地, 当 $\sigma^2 = 1$ 时, 在给定样本 $\{(Y_i^*, x_i) : i = 1, \dots, n\}$ 下, $\sqrt{n}(\beta_n^* - \hat{\beta}_n)$ 的条件分布可以用来估计 $\sqrt{n}(\hat{\beta}_n - \beta_0)$ 的分布, 从而避免估计冗余参数 $f(0)$. 过程如下: 对于一个足够大的正数 N , 取随机变量 $\{w_1, \dots, w_n\}$ 的 N 个独立复制, 从而, 由 (2.1) 得到 N 个 β_n^* , 记 $D = \{\beta_{n1}^*, \dots, \beta_{nN}^*\}$. 由定理 2.3 可知, D 的经验分布和方差可以用来估计 $\hat{\beta}_n$ 的分布和方差.

2.2 随机加权检验

现在来讨论回归参数的假设检验问题. 尽管前面章节的渐近理论可以用来构造 Wald 型检验统计量, 这里将集中讨论检验参数线性假设问题, 即考虑原假设

$$H_0 : H^\top (\beta - b_0) = 0, \quad (2.5)$$

其中 H 是一个已知的 $p \times q$ 维、秩为 q 的矩阵, b_0 是一个已知的 p 维向量 ($0 < q \leq p$). 不失一般性, 本文只考虑一列备择假设

$$H_1 : H^\top (\beta - b_0) = \omega_n, \quad (2.6)$$

其中 ω_n 是一个已知 p 维向量, 满足

$$\|S_n^{1/2}\omega_n\| = O(1), \quad n \rightarrow \infty. \quad (2.7)$$

在原假设 H_0 下, 定义 β_0 的估计为 $\hat{\beta}_c^*$, 满足

$$\sum_{i=1}^n |Y_i^* - (x_i^\top \hat{\beta}_c^*)^*| = \inf_{H^\top (\beta - b_0) = 0} \sum_{i=1}^n |Y_i^* - (x_i^\top \beta)^*|.$$

为了检验假设 (2.5), 构造一个 M 检验统计量

$$M_n = \sum_{i=1}^n |Y_i^* - (x_i^\top \hat{\beta}_c^*)^*| - \sum_{i=1}^n |Y_i^* - (x_i^\top \hat{\beta}_n)^*|.$$

在原假设下, 可以证明 $4f(0)M_n$ 的渐近分布是自由度为 q 的卡方分布. 同样地, 为了避免估计 $f(0)$, 类似文献 [10], 定义一个随机加权 M 检验统计量 (RWM)

$$\begin{aligned} M_n^* = & \frac{1}{\sigma^2} \left\{ \sum_{i=1}^n w_i |Y_i^* - (x_i^\top \beta_c^*)^*| - \sum_{i=1}^n w_i |Y_i^* - (x_i^\top \beta_n^*)^*| \right\} \\ & - \frac{1}{\sigma^2} \left\{ \sum_{i=1}^n w_i |Y_i^* - (x_i^\top \hat{\beta}_c^*)^*| - \sum_{i=1}^n w_i |Y_i^* - (x_i^\top \hat{\beta}_n)^*| \right\}, \end{aligned} \quad (2.8)$$

其中 β_c^* 满足

$$\sum_{i=1}^n w_i |Y_i^* - (x_i^\top \beta_c^*)^*| = \inf_{H^\top (\beta - b_0) = 0} \sum_{i=1}^n w_i |Y_i^* - (x_i^\top \beta)^*|.$$

下面的定理给出 RWM 的渐近分布, 证明见附录 A.

定理 2.4 假设在模型 (1.1) 中, 条件 (A₁)–(A₆) 成立, 则在原假设或备择假设下, 有

$$4f(0)M_n^* = \left\| \sum_{i=1}^n \frac{w_i - 1}{\sigma} I(0 < \mu_i < M) \text{sgn}(e_i) H_n^\top x_{in} \right\|^2 + o_p(1), \quad (2.9)$$

其中 $H_n = S_n^{-1/2} H (H^\top S_n^{-1} H)^{-1/2}$. 进一步地, 有

$$\sup_{u>0} |\mathbf{P}^*(M_n^* \leq u) - \mathbf{P}(K_q \leq u)| \rightarrow 0, \quad \text{依概率} \quad n \rightarrow \infty, \quad (2.10)$$

其中 $4f(0)K_q$ 是一个自由度为 q 的中心卡方变量.

由定理 2.4 可知, 在给定数据 $\{(Y_i^*, x_i) : i = 1, \dots, n\}$ 下, M_n^* 的渐近分布在原假设和备择假设下都是 $\chi_q^2/(4f(0))$. 因此, 在原假设下, 可以用 M_n^* 去近似 M_n 的分布, 从而避免估计 $f(0)$. 我们需要一个子列 $c_n(\alpha)$, 在假设 H_0 下, 使得 $\lim_{n \rightarrow \infty} P(M_n \leq c_n(\alpha)) = 1 - \alpha$, 其中 $\alpha \in (0, 1)$. 下面用 M_n^* 的 $1 - \alpha$ 分位数 $c_n^*(\alpha)$ 去估计 $c_n(\alpha)$. 取足够大的 N , 产生 N 个独立复制的随机权重, 并且得到 N 个 M_n 的随机加权估计 $D_N = \{M_{nj}^*, j = 1, 2, \dots, N\}$. D_N 的 $1 - \alpha$ 分位数记作 $c_n^*(\alpha)$. 当 $M_n > c_n^*(\alpha)$ 时, 拒绝原假设 H_0 , 否则接受原假设.

3 数值研究

3.1 模拟研究

本节用模拟研究来评估上述方法的效果. 模拟数据由下述模型生成:

$$Y_i^* = (\beta_{00} + \beta_{01}x_{1i} + \beta_{02}x_{2i} + e_i)^*, \quad i = 1, 2, \dots, n, \quad (3.1)$$

其中 $M = 8$. 解释变量 x_{1i} 和 x_{2i} 由标准正态分布 $N(0, 1)$ 生成, 且 e_i 有 3 种不同的误差分布: $N(0, 1)$ (Normal)、自由度为 2 的 t_2 分布 (t_2) 和 Cauchy 分布. 真实参数 $\beta_0 = (\beta_{00}, \beta_{01}, \beta_{02}) = (3, 1, 2)$. 样本量分别为 $n = 100$ 、 $n = 150$ 和 $n = 200$. 随机加权变量 w_i 是参数为 1 的指数分布. 随机加权次数 $N = 200$, 所有的模拟都重复 1,000 次.

表 1 列出了 RWE、 $\hat{\beta}_n$ 的标准偏差 (Sd) 和它的估计值 (Esd). 可以看出所有的 RWE 都接近它对应的真实值, 且 Esd 接近 Sd. 这表明通过随机加权方法得到的 $\hat{\beta}_n$ 的方差估计效果是可行的.

根据模型 (1.1), 比较了 3 种估计方法的效果: (1) 不考虑响应变量的约束 (下界为 $-\infty$, 且 $M = \infty$), 记作 NOLP; (2) 不考虑上界约束 ($M = \infty$), 记作 NOP; (3) 考虑上下界约束 (LP). 事实上, NOLP 假设数据来自一个无响应约束的线性回归模型; NOP 是有下界约束的 Tobit 模型; LP 是双尾约束 Tobit 模型. 表 2 列出了在模型 (1.1) 下, 3 种参数估计偏离真实参数的情况. 由结果可以看出 LP 偏差最小, 因此当数据来自双截尾删失回归模型中时, 忽略响应约束会使参数估计产生很大的偏差.

现在通过模拟研究来研究随机加权检验的效果. 根据模型 (3.1) 生成数据, 其中 $\beta_{00} = 1.0$, $n = 100$, $M = 10$, 原假设 $H_0 : \beta_{10} = \beta_{20} = 0$. 表 3 表明了在不同值时的经验显著水平和检验的效

表 1 回归参数的随机加权估计

分布	n	β_{00}			β_{01}			β_{02}		
		RWE	Sd	Esd	RWE	Sd	Esd	RWE	Sd	Esd
Normal	100	2.9924	0.1330	0.1426	1.0031	0.1401	0.1488	2.0207	0.1654	0.1676
	150	2.9895	0.1089	0.1144	1.0055	0.1158	0.1205	2.0159	0.1265	0.1336
	200	2.9962	0.0952	0.0984	1.0074	0.1014	0.1022	2.0119	0.1133	0.1156
t_2	100	2.9890	0.1547	0.1700	1.0198	0.1705	0.1812	2.0274	0.1900	0.2039
	150	2.9919	0.1269	0.1358	1.0103	0.1388	0.1438	2.0246	0.1555	0.1611
	200	2.9944	0.1102	0.1137	1.0072	0.1127	0.1209	2.0151	0.1302	0.1350
Cauchy	100	2.9793	0.1802	0.2041	1.0114	0.2000	0.2243	2.0335	0.2249	0.2489
	150	2.9850	0.1467	0.1583	1.0164	0.1539	0.1704	2.0325	0.1742	0.1921
	200	2.9944	0.1217	0.1330	1.0108	0.1330	0.1430	2.0292	0.1521	0.1598

表 2 三种不同估计方法的参数估计偏差结果

分布	n	β_{00}			β_{01}			β_{02}		
		LP	NOP	NOLP	LP	NOP	NOLP	LP	NOP	NOLP
Normal	100	-0.0017	-0.0087	0.0959	-0.0021	-0.0308	-0.1193	0.0101	-0.0423	-0.2300
	150	-0.0055	-0.0136	0.0932	0.0019	-0.0269	-0.1181	0.0083	-0.0420	-0.2285
	200	-0.0021	-0.0091	0.0976	0.0065	-0.0200	-0.1121	0.0095	-0.0426	-0.2302
t_2	100	-0.0061	-0.0137	0.1021	0.0133	-0.0154	-0.1196	0.0133	-0.0479	-0.2532
	150	-0.0034	-0.0111	0.1081	0.0045	-0.0253	-0.1269	0.0154	-0.0454	-0.2492
	200	-0.0029	-0.0118	0.1056	0.0031	-0.0258	-0.1259	0.0079	-0.0492	-0.2503
Cauchy	100	-0.0128	-0.0229	0.1061	0.0040	-0.0331	-0.1481	0.0161	-0.0557	-0.2817
	150	-0.0100	-0.0175	0.1130	0.0104	-0.0254	-0.1361	0.0180	-0.0516	-0.2698
	200	-0.0016	-0.0091	0.1193	0.0065	-0.0293	-0.1390	0.0200	-0.0502	-0.2639

表 3 经验显著性水平和随机加权检验的功效

(β_{10}, β_{20})	0.01 ^{a)}	0.05 ^{a)}
(0.00, 0.00)	0.007	0.042
(0.25, 0.00)	0.161	0.362
(0.25, 0.25)	0.360	0.630
(0.50, 0.00)	0.757	0.905

a) 显著性水平

果. 当 $(\beta_{10}, \beta_{20}) = (0, 0)$ (原假设) 时, 经验显著性水平近似名义显著性水平 α . (β_{10}, β_{20}) 与原假设偏离越大, 检验的效用越大.

3.2 妇女劳动力供应的例子

我们用随机加权方法去分析女性劳动力付出水平数据集^[11], 这个数据集调查了女性工作的平均年工时 (WHRS) 与其他解释变量之间的联系: 小于 6 岁孩子的数量 (KL6)、6–18 岁孩子的数量 (K618)、妻子的年龄 (WA)、妻子的教育水平 (WE)、妻子的工资 (WW)、妻子的记录工资 (RPWG)、丈夫的工作时间 (HHRS)、丈夫的教育水平 (HE)、丈夫的工资 (HW)、妻子的边际税率 (MTR)、妻子父亲的教育水平 (WFED)、失业率 (UN)、大城市指数 (CIT) 和妻子的实际工作经验 (AX). 在建模前, 先对数据进行以下处理: WHRS 除以它的标准差; 所有的解释变量都标准化. 样本总量为 753, 随机选取 500 个作为训练集, 剩下的作为检验集. 在这个例子中, 取上界 $M = 2.295$ ($2000/\text{Sd}(\text{WHRS})$) 来说明所提的方法.

表 4 给出了 NOP 和 LP 方法中参数估计和标准差估计; 其中 p -value 是 $1 - \Phi(|\text{Est}/\text{Esd}|)$, 且 $\Phi(\cdot)$ 是正态分布 $N(0, 1)$ 的累积分布函数. 可以看到在名义显著性水平为 0.05 时, LP 和 NOP 得到相同的显著的解释变量, 如 KL6、WA、RPWG、HW、MTR 和 AX, 但 NOP 低估了这些显著变量, 如 RPWG 的系数, LP 和 NOP 的估计值分别为 0.9757 和 0.5307. 另外, 对变量 K618 系数估计, LP 方法估计值是正的, NOP 方法下是负的. 因为 6–18 岁的孩子一般都待在学校里, 母亲会有更多的时间去工作, 所以正的参数估计更加合理.

下面记检验数据集为 $\{(x_i^*, y_i^*), i = 1, \dots, 253\}$. 两个指标: 预测值与真实值的绝对偏差 Median =

表 4 女性劳动力支付数据的参数估计

参数	LP			NOP		
	Est ^{a)}	Esd ^{b)}	p-value	Est ^{a)}	Esd ^{b)}	p-value
Intercept	0.3102	0.1297	0.0084	0.5002	0.0629	<0.0001
KL6	-0.5565	0.1522	0.0001	-0.3579	0.1106	0.0006
K618	0.0785	0.0869	0.1834	-0.0173	0.0471	0.3567
WA	-0.6628	0.1683	<0.0001	-0.4059	0.0822	<0.0001
WE	0.2181	0.1365	0.0551	-0.1568	0.0732	0.0160
WW	0.1645	0.2206	0.2279	0.0646	0.0835	0.2197
RPWG	0.9757	0.2041	<0.0001	0.5307	0.1047	<0.0001
HRSS	-0.4717	0.1133	<0.0001	-0.2906	0.0815	0.0002
HE	-0.2897	0.1136	0.0054	-0.1430	0.0621	0.0107
HW	-1.4502	0.2589	<0.0001	-0.6099	0.1588	0.0001
MTR	-1.5760	0.2626	<0.0001	-0.6723	0.1540	<0.0001
WFED	0.0146	0.1155	0.4496	0.0420	0.0578	0.2334
UN	-0.1245	0.0972	0.1001	-0.0310	0.0495	0.2655
CIT	0.0142	0.0846	0.4332	-0.0009	0.0512	0.4927
AX	0.4893	0.1144	<0.0001	0.3704	0.0562	<0.0001

a) 参数估计; b) 标准差估计

$\text{median}\{|y_i^* - (\hat{y}_i^* I(0 < \hat{y}_i^* < M) + MI(\hat{y}_i^* > M))|, i = 1, \dots, m\}$ 的中位数 (Median) 和均值 (Mean). 可以计算出, LP 给出 (Median, Mean) 是 (0.0459, 0.3344), NOP 的 (Median, Mean) 是 (0.0918, 0.3485), 则 LP 的中位数和均值都比 NOP 小.

4 结论

对于双截尾删失回归模型, 本文提出了一个随机加权逼近方法来避免估计冗余参数, 在模型中给出了随机加权估计 (RWE) 并且构造了随机加权 M 检验统计量 (RWM), 证明了在观测数据给定的条件下, RWE 和 RWM 的条件分布可以用来近似参数估计和 M 检验统计量的分布. 数值模拟也验证了本文的理论. 然而, 本文假设协变量的维度是一个固定常数, 当维度随样本量增加时, 函数的非凸和非平滑性使得证明存在一些技术上的困难, 作者将在未来的工作中研究这个问题.

参考文献

- 1 Powell J L. Least absolute deviations estimation for the censored regression model. J Econometrics, 1984, 25: 303–325
- 2 Pollard D. Empirical Process: Theory and Application. NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2. Hayward: Institute of Mathematical Statistics, 1990
- 3 Rao C, Zhao L. Asymptotic normality of LAD estimators in censored regression models. Math Methods Statist, 1993, 2: 228–239
- 4 Fang Y, Jin M, Zhao L C. Strong convergence of LAD estimates in a censored regression model. Sci China Ser A, 2005, 48: 155–168
- 5 Wang Z F, Wu Y H, Zhao L C. Approximation by randomly weighting method in censored regression model. Sci China Ser A, 2009, 52: 561–576

- 6 Liu X, Wang Z, Wu Y. Group variable selection and estimation in the Tobit censored response model. *Comput Statist Data Anal*, 2013, 60: 80–89
- 7 Xiao L Q, Hou B, Wang Z F, et al. Random weighting approximation for Tobit regression models with longitudinal data. *Comput Statist Data Anal*, 2014, 79: 235–247
- 8 Ding H, Wang A F, Wu Y H. Tobit regression model with parameters of increasing dimensions. *Statist Probab Lett*, 2017, 120: 1–7
- 9 Wang X, Wang Z, Wu Y. Asymptotic normality of LAD estimates in two-tailed Tobit regression models. *J Univ Sci Technol China*, 2014, 44: 709–717
- 10 Chen K, Ying Z, Zhang H, et al. Analysis of least absolute deviation. *Biometrika*, 2008, 95: 107–122
- 11 Mroz T A. The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 1987, 55: 765–799
- 12 Bennett G. Probability inequalities for the sum of independent random variables. *J Amer Statist Assoc*, 1962, 57: 33–45

附录 A 定理的证明

不妨假设 c 是独立于 n 的正常数, 它在不同的章节, 甚至在同一个不等式中都可能是不同的. 首先介绍一个引理:

引理 A.1 ^[12] 假设 $\xi_1, \xi_2, \dots$ 是独立随机变量, 任意 i 满足 $E\xi_i = 0$, $|\xi_i| \leq b < \infty$; $\nu = \sum_{i=1}^n \text{Var}(\xi_i) < \infty$, 对 $\forall \varepsilon > 0$, 有

$$P\left(\left|\sum_{i=1}^n \xi_i\right| \geq \varepsilon\right) \leq 2 \exp\left\{-\frac{\varepsilon^2}{2\nu + 2b\varepsilon}\right\}.$$

定理 2.1 的证明类似文献 [3, 引理 2.2] 中的证明, 在此省略.

定理 2.2 的证明 为了简单, 令 $\gamma = \beta - \beta_0$, $\gamma_n^* = \beta_n^* - \beta_0$, 则 $\gamma \in B^* = B - \beta_0 = \{\beta - \beta_0 : \beta \in B\}$. 定义

$$\begin{aligned} G_n(\gamma) &= \sum_{i=1}^n g_{ni}(\gamma) = \sum_{i=1}^n w_i (|Y_i^* - (\mu_i + x_i^\top \gamma)^*| - |Y_i^* - \mu_i^*|) \\ &= \sum_{i=1}^n w_i \int_{\mu_i^*}^{(\mu_i + x_i^\top \gamma)^*} \text{sgn}(v - Y_i) dv. \end{aligned}$$

令

$$W_n = \sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) S_n^{-1/2} x_i, \quad \bar{\gamma}_n = \frac{S_n^{-1/2} W_n}{2f(0)}. \quad (\text{A.1})$$

由 (A₃)、(A₅) 和 $\max_{1 \leq i \leq n} x_i^\top S_n^{-1} x_i I(0 < \mu_i < M) \rightarrow 0$ 知, Lindeberg 条件满足, 则可以得到

$$2f(0) S_n^{1/2} \bar{\gamma}_n = W_n \xrightarrow{d} N(0, (\sigma^2 + 1) I_p). \quad (\text{A.2})$$

为了证明这个定理, 只需证明

$$S_n^{1/2} (\gamma_n^* - \bar{\gamma}_n) \xrightarrow{P} 0, \quad n \rightarrow \infty. \quad (\text{A.3})$$

由引理 A.1 和定理 2.1 知, 存在一列 $\varepsilon_n \in (0, 1)$ 且 $\varepsilon_n \rightarrow 0$, 使得

$$P(\|\gamma_n^*\| + \|\bar{\gamma}_n\| \leq \varepsilon_n) \rightarrow 1, \quad n \rightarrow \infty.$$

由 (A₅) 可知, 存在一列 $\alpha_n \rightarrow \infty$, 使得

$$\alpha_n^2 \varepsilon_n \rightarrow 0, \quad \frac{\alpha_n^{-2} \lambda(S_n)}{\log^2 n} \rightarrow \infty. \quad (\text{A.4})$$

记

$$\begin{aligned} S_{n0} &= \sum_{i=1}^n I(0 < \mu_i < M, \|x_i\| \leq \alpha_n) x_i x_i^\top, \\ S_{n1} &= \sum_{i=1}^n I(0 < \mu_i < M, \|x_i\| > \alpha_n) x_i x_i^\top. \end{aligned} \quad (\text{A.5})$$

由 (A₃) 和 (A₄) 可得, 当 $n \rightarrow \infty$ 时,

$$\begin{aligned} \sum_{i=1}^n I(\|x_i\| > \alpha_n) \|x_i\|^2 &= o(\lambda(S_n)), \\ \sum_{i=1}^n I(|\mu_i| \leq \alpha_n \varepsilon_n) \|x_i\|^2 &= o(\lambda(S_n)), \end{aligned} \quad (\text{A.6})$$

且

$$\frac{\lambda(S_{n0})}{\lambda(S_n)} \rightarrow 1. \quad (\text{A.7})$$

定义

$$\begin{aligned} H_n(\gamma) &= \sum_{i=1}^n h_{ni}(\gamma) = \sum_{i=1}^n w_i \int_{(\mu_i)^*}^{(\mu_i + x_i^\top \gamma)^*} (\text{sgn}(v - Y_i) + \text{sgn}(e_i)) dv \\ &= \sum_{i=1}^n w_i (|Y_i^* - (\mu_i + x_i^\top \gamma)^*| - |Y_i^* - \mu_i^*|) + \sum_{i=1}^n w_i ((\mu_i + x_i^\top \gamma)^* - \mu_i)^* \text{sgn}(e_i) \\ &\triangleq \sum_{i=1}^n h_{ni}^{(1)}(\gamma) + \sum_{i=1}^n h_{ni}^{(2)}(\gamma), \\ H_{n0}(\gamma) &= \sum_{i=1}^n I(\|x_i\| \leq \alpha_n) h_{ni}(\gamma), \\ H_{n1}(\gamma) &= \sum_{i=1}^n I(\|x_i\| > \alpha_n) h_{ni}(\gamma). \end{aligned} \quad (\text{A.8})$$

我们知道集合

$$\bar{h}_{n,w}^{(k)} = \left\{ \left(\frac{h_{n1}^{(k)}}{\|\gamma\|}, \dots, \frac{h_{nn}^{(k)}}{\|\gamma\|} \right)^\top \in \mathbb{R}^n, 0 \neq \gamma \in \bar{B}_T \right\}, \quad k = 1, 2$$

的伪维度是有界的 (由 p 决定). 注意到 $\frac{h_{n1}^{(k)}}{\|\gamma\|} \leq c\|x_i\|$ 和 (A.6), 使用极大值不等式, 得到

$$\mathbb{E} \sup_{\|\gamma\| \leq \varepsilon_n} (\|\gamma\|^{-1} |H_{n1}(\gamma) - \mathbb{E} H_{n1}(\gamma)|)^2 \leq c \sum_{i=1}^n I(\|x_i\| > \alpha_n) \|x_i\|^2 = o(\lambda(S_n)). \quad (\text{A.9})$$

由 (A.5), $\exists \Delta > 0$, 当 $|v| < \Delta$ 时, $|F(v) - 1/2| \leq cv$, 这表明, 对 $\forall \gamma$, 有

$$0 \leq Eh_{ni}(\gamma) \leq c(x_i^\top \gamma)^2. \quad (\text{A.10})$$

为了证明这一点, 以 $0 < \mu_i < M$ 和 $0 < \mu_i + x_i^\top \gamma < M$ 作为例子, 其他情形类似. 在这种情形下,

$$Eh_{ni}(\gamma) = 2 \int_{\mu_i}^{\mu_i + x_i^\top \gamma} \left(F(v - \mu_i) - \frac{1}{2} \right) dv = 2 \int_0^{x_i^\top \gamma} \left(F(v) - \frac{1}{2} \right) dv.$$

对 $F(v)$ 进行 Taylor 展开, 如果 $|x_i^\top \gamma| < \Delta$, 那么 $Eh_{ni}(\gamma) \leq c(x_i^\top \gamma)^2$; 如果 $|x_i^\top \gamma| \geq \Delta$, 那么 $Eh_{ni}(\gamma) \leq (c/\Delta)(x_i^\top \gamma)^2$. 综上, 有

$$0 \leq \frac{EH_{n1}(\gamma)}{\|\gamma\|^2} \leq c \sum_{i=1}^n I(\|x_i\| > \alpha_n) \|x_i\|^2 = o(\lambda(S_n)), \quad (\text{A.11})$$

$$\frac{\gamma^\top S_{n1} \gamma}{\|\gamma\|^2} \leq c \sum_{i=1}^n I(\|x_i\| > \alpha_n) \|x_i\|^2 = o(\lambda(S_n)). \quad (\text{A.12})$$

由 (A.9)、(A.11) 和 (A.12), 对 $\forall \|\gamma\| \leq \varepsilon_n$, 一致地有

$$H_{n1}(\gamma) - f(0)\gamma^\top S_{n1}\gamma = o_p(\|S_n^{1/2}\gamma\| + \|S_n^{1/2}\gamma\|^2), \quad n \rightarrow \infty. \quad (\text{A.13})$$

定义

$$M_n(\gamma) = \sum_{i=1}^n m_{ni} = \sum_{i=1}^n w_i I(\|x_i\| \leq \alpha_n) ((\mu_i + x_i^\top \gamma)^* - (\mu_i)^*) - I(0 < \mu_i < M) x_i^\top \gamma \operatorname{sgn}(e_i).$$

与之前一样, 对可处理过程使用极大值不等式, 注意到 $EM_n(\gamma) = 0$, 且

$$\|\gamma\|^{-1} |m_{ni}(\gamma)| \leq c \|x_i\| I(|\mu_i| \leq \alpha_n \varepsilon_n), \quad \forall \gamma \leq \varepsilon_n.$$

我们得到

$$E \sup_{\|\gamma\| \leq \varepsilon_n} (\|\gamma\|^{-1} |M_n(\gamma)|)^2 \leq c \sum_{i=1}^n I(|\mu_i| \leq \alpha_n \varepsilon_n) \|x_i\|^2. \quad (\text{A.14})$$

由 (A.6)、(A.14) 和 $\alpha_n \varepsilon_n \rightarrow 0$, $\forall \|\gamma\| \leq \varepsilon_n$, 一致地有

$$|M_n(\gamma)| = o_p(\|S_n^{\frac{1}{2}} \gamma\|), \quad n \rightarrow \infty. \quad (\text{A.15})$$

类似于文献 [2, 第 62 和 63 页], 我们证明了当 $\|x_i\| \leq \alpha_n$ 时, 对 $\forall \|\gamma\| \leq \varepsilon_n$, 一致地有

$$EH_{n0}(\gamma) = f(0)\gamma^\top S_n \gamma (1 + o(1)) \quad (\text{A.16})$$

和

$$\operatorname{Var} H_{n0}(\gamma) \leq c \alpha_n \|\gamma\| \gamma^\top S_n \gamma. \quad (\text{A.17})$$

考虑集合 $D = \{(j_1 n^{-2}, \dots, j_p n^{-2})^\top : j_1, \dots, j_p = 0, \pm 1, \pm 2, \dots\}$, 集合

$$B_{\varepsilon_n} \cap D - \{0\} = \{(\gamma_1, \gamma_2, \dots, \gamma_L) : L \leq (2n^2 + 1)^p\}$$

是有界的. 对 $\forall \varepsilon > 0$, 定义

$$E_{nl} = \{|H_{n0}(\gamma_l) - EH_{n0}(\gamma_l)| \geq \varepsilon(\|S_n^{\frac{1}{2}}\gamma_l\| + \|S_n^{\frac{1}{2}}\gamma_l\|^2) : l \leq L\}, \quad E_n = \bigcup_{l \leq L} E_{nl}.$$

因此, $\|x_i\| \leq \alpha_n$, $|x_i^\top \gamma| \leq \alpha_n \|\gamma_l\|$, 由引理 A.1 和 (A.5), 可以得到

$$P(E_{nl}) \leq 2\exp\left\{\frac{-2\varepsilon^2(\|S_n^{\frac{1}{2}}\gamma_l\|^2 + \|S_n^{\frac{1}{2}}\gamma_l\|^4)}{c\alpha_n\|\gamma_l\|\|S_n^{\frac{1}{2}}\gamma_l\|^2 + c\varepsilon\alpha_n\|\gamma_l\|(\|S_n^{\frac{1}{2}}\gamma_l\|^2 + \|S_n^{1/2}\gamma_l\|)}\right\}.$$

如果 $\|S_n^{1/2}\gamma_l\| \geq 1$, 那么,

$$P(E_{nl}) \leq 2\exp\left\{\frac{-c\|S_n^{\frac{1}{2}}\gamma_l\|^3}{\alpha_n\|\gamma_l\|\|S_n^{\frac{1}{2}}\gamma_l\|^2}\right\} \leq 2\exp\{-c(\lambda(S_n)\alpha_n^{-2})^{\frac{1}{2}}\};$$

如果 $0 \leq \|S_n^{\frac{1}{2}}\gamma_l\| \leq 1$, 那么,

$$P(E_{nl}) \leq 2\exp\left\{\frac{-c\|S_n^{\frac{1}{2}}\gamma_l\|^2}{\alpha_n\|\gamma_l\|\|S_n^{\frac{1}{2}}\gamma_l\|}\right\} \leq 2\exp\{-c(\lambda(S_n)\alpha_n^{-2})^{\frac{1}{2}}\}.$$

因此, 由 (A.5), 有

$$\sum_n P(E_n) \leq c \sum_n n^{2p} \exp\{-c(\lambda(S_n)\alpha_n^{-2})^{\frac{1}{2}}\} < \infty.$$

由 Borel-Cantelli 引理, 对 $\forall \varepsilon > 0$, 有

$$P(E_n, \text{i.o.}) = 0, \quad (\text{A.18})$$

其中 i.o. 表示无穷次发生. 由 (A.16) 和 (A.18), 当 n 足够大时, 对 $l \leq L$, 一致地依概率 1, 有

$$H_{n0}(\gamma_l) = f(0)\gamma_l^\top S_{n0}\gamma_l + o(\|S_n^{\frac{1}{2}}\gamma_l\| + \|S_n^{\frac{1}{2}}\gamma_l\|^2). \quad (\text{A.19})$$

对 $\forall \gamma \in \overline{B}_{\varepsilon_n}$, $\exists \gamma_l \in D$ 使得 $\|\gamma - \gamma_l\| \leq p/n^2$ 且

$$|H_{n0}(\gamma_l) - H_{n0}(\gamma)| < \frac{c\alpha_n\varepsilon_n}{n} \leq \frac{c}{n}, \quad (\text{A.20})$$

$$|f(0)\gamma_l^\top S_{n0}\gamma_l - f(0)\gamma^\top S_{n0}\gamma| \leq \frac{c\alpha_n^2\varepsilon_n}{n} \leq \frac{c}{n}. \quad (\text{A.21})$$

由 (A.13) 和 (A.19)–(A.21), 当 n 足够大时, 对 $l \leq L$, 一致地依概率 1, 有

$$H_{n0}(\gamma) = f(0)\gamma^\top S_{n0}\gamma + o(\|S_n^{\frac{1}{2}}\gamma\| + \|S_n^{\frac{1}{2}}\gamma\|^2) + O\left(\frac{1}{n}\right). \quad (\text{A.22})$$

由 (A.1)、(A.2)、(A.8)、(A.13)、(A.15) 和 (A.22), $\exists \|\gamma\| \leq \varepsilon_n$, 一致地有

$$G_n(\gamma) = -W_n^\top S_n^{\frac{1}{2}}\gamma + f(0)\gamma^\top S_n\gamma + o_p(1 + \gamma^\top S_n\gamma). \quad (\text{A.23})$$

由 (A.2)、(A.4) 和 (A.23), 得到

$$G_n(\gamma_n^*) = -2f(0)\gamma_n^{\top} S_n\gamma_n^* + f(0)(\gamma_n^*)^\top S_n\gamma_n^*(1 + o_p(1)) + o_p(1),$$

$$G_n(\bar{\gamma}_n) = -f(0)(1 + o_p(1))\bar{\gamma}_n S_n \bar{\gamma}_n + 2f(0)\bar{\gamma}_n S_n \bar{\gamma}_n(o_p(1)) + o_p(1).$$

因为 $S_n^{\frac{1}{2}}\bar{\gamma}_n(o_p(1)) = o_p(1)$, 并且注意到 $G_n(\gamma_n^*) - G_n(\bar{\gamma}_n) \leq 0$, 所以有

$$0 \geq G_n(\gamma_n^*) - G_n(\bar{\gamma}_n) = f(0)(1 + o_p(1))\|S_n^{\frac{1}{2}}(\gamma_n^* - \bar{\gamma}_n) + o_p(1)\|^2 + o_p(1) \geq 0,$$

这表明 $S_n^{\frac{1}{2}}(\gamma_n^* - \bar{\gamma}_n) \xrightarrow{P} 0, n \rightarrow \infty$. 定理 2.2 证明完毕. $\square$

定理 2.4 的证明 假设 $0 < q < p$, K 是一个秩为 $p - q$ 的 $p \times (p - q)$ 维矩阵, 使得 $H^\top K = 0$ 且 $K^\top \omega_n = 0$. 记 $K_n = S_n^{\frac{1}{2}}K(K^\top S_n K)^{-\frac{1}{2}}$, 有 $K_n^\top K_n = I_{p-q}, H_n^\top H_n = I_q$ 且 $H_n^\top K_n = 0$. 由定理 2.2 知, 对 $\|\gamma\| \leq \varepsilon_n$, 一致地有

$$\begin{aligned} 2f(0)S_n^{\frac{1}{2}}\gamma_n^* &= \sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) x_{in} + o_p(1), \\ G_n(\gamma) &= -W_n^\top S_n^{\frac{1}{2}}\gamma + f(0)\gamma^\top S_n \gamma(1 + o_p(1)) + o_p(1) \\ &= -\left(\sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) x_{in}\right) S_n^{\frac{1}{2}}\gamma + f(0)\gamma^\top S_n \gamma(1 + o_p(1)) + o_p(1). \end{aligned}$$

令 $H^\top(b - b_0) = 0$ 和 $H^\top(\beta_0 - b_0) = 0$. 存在一个唯一的 $\gamma_c \in \mathbb{R}^{p-q}$ 使得 $b - \beta_0 = K\gamma_c$. 所以, 对某个 $\gamma_c \in \mathbb{R}^{p-q}$, H_0 和 H_1 可以表示为

$$H_0 : b - \beta_0 = K\gamma_c \leftrightarrow H_1 : b - \beta_0 = K\gamma_c + \omega_n.$$

定义

$$\begin{aligned} G_{nc}(\gamma_c) &= \sum_{i=1}^n w_i (|Y_i^* - (x_i^\top b)^*| - |\mu_i^* - Y_i^*|) \\ &= \sum_{i=1}^n w_i (|Y_i^* - (\mu_i + x_i^\top K\gamma_c)^*| - |\mu_i^* - Y_i^*|), \quad \gamma_c \in \mathbb{R}^{p-q}. \end{aligned}$$

令 $K\gamma_{nc}^* = \beta_n^* - \beta_0$, $G_{nc}(\gamma_{nc}^*) = \min_{\gamma_{nc}} G_{nc}(\gamma_{nc})$. 用 $K^\top x_i$ 替换 x_i , 由定理 2.2 知, 对 $\|\gamma_c\| \leq \varepsilon_n$, 一致地有

$$\begin{aligned} 2f(0)(K^\top S_n K)^{\frac{1}{2}}\gamma_{nc}^* &= \sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) K_n^\top x_{in} + o_p(1), \\ G_{nc}(\gamma_c) &= -\left(\sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) K_n^\top x_{in}\right) (K^\top S_n K)^{\frac{1}{2}}\gamma_c \\ &\quad + f(0)\gamma_c^\top K^\top S_n K \gamma_c(1 + o_p(1)) + o_p(1). \end{aligned}$$

由 $\gamma_n^{*\top} S_n \gamma_n^*(o_p(1)) = o_p(1)$, $\hat{\gamma}_n^\top S_n \hat{\gamma}_n(o_p(1)) = o_p(1)$, 有

$$\begin{aligned} G_n(\gamma_n^*) - G_n(\hat{\gamma}_n) &= -\left(\sum_{i=1}^n w_i I(0 < \mu_i < M) \text{sgn}(e_i) x_{in}\right) S_n^{\frac{1}{2}}(\gamma_n^* - \hat{\gamma}_n) \\ &\quad + f(0)(\gamma_n^{*\top} S_n \gamma_n^* - \hat{\gamma}_n^\top S_n \hat{\gamma}_n) + o_p(1) \\ &= -(4f(0))^{-1} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \text{sgn}(e_i) x_{in} \right\|^2 + o_p(1). \end{aligned}$$

类似地,

$$G_{nc}(\gamma_{nc}^*) - G_{nc}(\hat{\gamma}_{nc}) = -(4f(0))^{-1} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \operatorname{sgn}(e_i) K_n^\top x_{in} \right\|^2 + o_p(1).$$

所以, 在原假设下, 有

$$\begin{aligned} 4f(0)M_n^* &= \frac{4f(0)}{\sigma^2} \{G_{nc}(\gamma_{nc}^*) - G_{nc}(\hat{\gamma}_{nc}) - (G_n(\gamma_n^*) - G_n(\hat{\gamma}_n))\} \\ &= \frac{1}{\sigma^2} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \operatorname{sgn}(e_i) H_n^\top x_{in} \right\|^2 + o_p(1). \end{aligned} \quad (\text{A.24})$$

在备择假设下, $b - \beta_0 = K\gamma_c + \omega_n$, $\gamma_c \in \mathbb{R}^{p-q}$, 令

$$\begin{aligned} G'_{nc}(\gamma_c) &= \sum_{i=1}^n w_i (|Y_i^* - (\mu_i + x_i^\top \omega_n + x_i^\top K\gamma_c)^*|) \\ &\quad - \sum_{i=1}^n w_i |(\mu_i + x_i^\top \omega_n)^* - Y_i^*|, \quad \text{对 } \gamma_c \in \mathbb{R}^{p-q}. \end{aligned}$$

在定理 2.2 中, 用 $K^\top x_i$ 替代 x_i , 用 $\beta_0 + \omega_n$ 替代 β_0 , 对 $\|\gamma_c\| \leq \varepsilon_n$, 一致地有

$$\begin{aligned} 2f(0)(K^\top S_n K)^{\frac{1}{2}} \gamma_{nc}^* &= \sum_{i=1}^n w_i I(0 < \mu_i + x_i^\top \omega_n < M) \operatorname{sgn}(e_i) K_n^\top x_{in} + o_p(1), \\ G'_{nc}(\gamma_c) &= - \left(\sum_{i=1}^n w_i I(0 < \mu_i + x_i^\top \omega_n < M) \operatorname{sgn}(e_i) K_n^\top x_{in} \right) (K^\top S_n K)^{\frac{1}{2}} \gamma_c \\ &\quad + f(0) \gamma_c^\top K^\top S_n K \gamma_c (1 + o_p(1)) + o_p(1), \end{aligned}$$

在备择假设下, 有

$$G_{nc}(\gamma_{nc}^*) - G_{nc}(\hat{\gamma}_{nc}) = G'_{nc}(\gamma_{nc}^*) - G'_{nc}(\hat{\gamma}_{nc}).$$

因此,

$$\begin{aligned} G_{nc}(\gamma_{nc}^*) - G_n(\hat{\gamma}_{nc}) &= - \left(\sum_{i=1}^n w_i I(0 < \mu_i + x_i^\top \omega_n < M) \operatorname{sgn}(e_i) K_n^\top x_{in} \right) (K^\top S_n K)^{\frac{1}{2}} (\gamma_{nc}^* - \hat{\gamma}_{nc}) \\ &\quad + f(0) (\gamma_{nc}^{*\top} K^\top S_n K \gamma_{nc}^* - \hat{\gamma}_{nc}^\top K^\top S_n K \hat{\gamma}_{nc}) + o_p(1) \\ &= -(4f(0))^{-1} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i + x_i^\top \omega_n < M) \operatorname{sgn}(e_i) K_n^\top x_{in} \right\|^2 + o_p(1) \\ &= -(4f(0))^{-1} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \operatorname{sgn}(e_i) K_n^\top x_{in} \right\|^2 + o_p(1). \end{aligned}$$

从而有

$$\begin{aligned} 4f(0)M_n^* &= \frac{4f(0)}{\sigma^2} \{G_{nc}(\gamma_{nc}^*) - G_{nc}(\hat{\gamma}_{nc}) - (G_n(\gamma_n^*) - G_n(\hat{\gamma}_n))\} \\ &= \frac{1}{\sigma^2} \left\| \sum_{i=1}^n (w_i - 1) I(0 < \mu_i < M) \operatorname{sgn}(e_i) H_n^\top x_{in} \right\|^2 + o_p(1). \end{aligned} \quad (\text{A.25})$$

由 (A.24) 和 (A.25) 知, 定理 2.4 成立. $\square$

A randomly weighting approximation approach for two-tailed censored regression model

Zhanfeng Wang, Shuangshuang Yao & Yaohua Wu

Abstract A random weight estimator (RWE) of the parameter is studied in a two-tailed censored regression model. To test linear hypotheses of the parameters, we propose an M-statistic and a randomly weighting M-statistic (RWM). Conditional distributions of RWE and RWM are used to approximate distributions of the least absolute deviation (LAD) estimator of the regression parameter and the M-statistic without estimating nuisance parameters, such as density function. Statistical properties, such as consistency and asymptotic distribution, of RWE and RWM are developed. Numerical studies including simulation and a real example show that the proposed method performs well.

Keywords two-tailed censored regression model, randomly weighting method, asymptotic properties, nuisance parameter

MSC(2010) 62F10, 62F03

doi: 10.1360/SCM-2017-0016