

基于均值漂移模型的异常值检测方法

张探探, 樊亚莉, 钟先乐

(上海理工大学 理学院, 上海 200093)

摘要: 由于异常值的存在对统计推断有很大影响, 因此异常值检测是数据分析中的一个重要步骤。对于横截面数据的线性模型, 改写模型的设计矩阵后, 基于均值漂移模型, 利用系数压缩估计方法来进行异常值检测。由于系数压缩估计中调节参数的选择对检测效果有很大影响, 基于两种调节方法的加权, 提出了一种新的调节方法。数值模拟结果表明, 使用这种基于均值漂移模型的异常值检测调节方法, 可以显著降低犯两种错误的概率。

关键词: 异常值检测; 参数调节; 均值漂移模型; 系数压缩估计

中图分类号: O 212.1 **文献标志码:** A

Outlier Detection Method Based on the Mean Shift Model

ZHANG Tantan, FAN Yali, ZHONG Xianle

(College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Outlier detection is an important part for data analysis, since outliers can infect the statistical inference evidently. The design matrix in a linear model for cross sectional data was rewritten, and an outlier detection method was proposed based on the mean shift model by using the coefficient shrink estimation. Because the selection of tuning model parameters is very important for outlier detection, a new tuning method based on a weighted tuning process was proposed. The numerical simulation results show that when the new tuning method is applied in the outlier detection procedure, two false identification probability can be decreased observably.

Keywords: outlier detection; parameter tuning; mean shift model; coefficient shrink estimation

1 问题的提出

近些年, 随着人们对数据重视程度的增加, 统计分析的应用也越来越广泛。但在大量的数据中, 异常值的存在一直是一个问题。简单地说, 异常值就是指那些少数的与数据主体不属于同一

分布的数据值。在数据分析中, 对异常值的检测是统计分析工作的重要部分。这是因为: 一方面, 如果无视异常值的存在而直接对其进行建模分析, 必定会影响到分析效果, 甚至会导致得出错误的结论; 而另一方面, 异常值又经常含有某些值得注意的信息。有时候它的出现说明了新情

收稿日期: 2017-06-24

基金项目: 国家自然科学基金资助项目(11401383)

第一作者: 张探探(1993-), 男, 硕士研究生. 研究方向: 异常值检测. E-mail: 792295478@qq.com

通信作者: 樊亚莉(1978-), 女, 讲师. 研究方向: 概率论与数理统计. E-mail: yalifan@usst.edu.cn

况或新事物的出现, 比如经济模型中某个经济政策的颁布, 疾病监控模型中某个激素含量的变化。这些情形下, 异常值本身可能更值得关注。

异常值的概念从 1777 年被 Bernoulli 第一次提出之后, 人们对它的研究就一直没有停止过。二十世纪以来, Abraham 等^[1-2], Cook^[3-4]也在这个方面作了大量的研究工作。总体来说, 传统的异常值检测方法可以分为两大类。第一类是基于残差分析。由于残差往往揭示了模型基本假设的大部分信息, 因此残差及其各种变体可以作为很重要的检测统计量, 参见文献[5-7]。第二类是基于影响分析。例如常见的点删除方法, 是通过检测一个观测点被删除前后对模型的影响力, 从而判断它是否为异常值, 参见文献[3-4]。文献[8]则是构造了观测值的影响矩阵来进行异常值诊断。近年来, 有一种新的异常值检测方法, 该方法利用基于系数压缩估计的变量选择方法来进行异常值检测。例如, She 等^[9]在均值漂移模型中引进系数压缩估计方法对漂移项进行惩罚, 从而达到异常值检测的效果。

围绕异常值检测的各种方法, 人们往往关心这些方法犯两种错误的概率。第一种错误称之为掩盖现象, 即只检测出数据中的部分异常值, 而遗漏了其他的异常值; 第二种错误称之为淹没现象, 即一些正常的观测值被认作异常值。文献[4]认为点删除方法容易犯第一种错误。文献[10]研究多个异常值的检测, 并认为他们的方法可以有效降低犯这两种错误的概率。文献[8]的方法致力于减少掩盖现象的发生。文献[9]比较了基于压缩估计的异常值检测方法与传统方法的优劣, 研究认为, 前者比传统方法更能有效地降低犯两种错误的概率。

本文对文献[9]中异常值的检测方法进行了改进和推广。具体来说是在线性回归模型中, 基于均值漂移模型和系数压缩估计方法, 对漂移项进行惩罚, 从而达到对异常值进行检测的效果。通过大量模拟发现, 惩罚项中调节参数的选取对异常值检测的效果至关重要。因此, 提出了一种新的调节方式, 对 BIC(Bayesian information criterion)准则和 CV(cross validation)准则选出来的调节参数进行加权平均, 作为最终的调节参数。模拟计算证明, 该调节方式既保持了原方式的优良之处, 又显著改善了原方式失效的情况。随后, 通过大量的模拟比较了上述方法应用于最小绝对压缩选

择算子方法(LASSO)^[11]、带有光滑截断绝对差的惩罚似然方法(SCAD)^[12]、硬截断惩罚方法(HARD)^[13]等惩罚方式的结果, 并与文献[7]的结果相比较, 证实了本方法的有效性及其显著优势。

2 模型和方法

2.1 均值漂移模型

多元线性回归模型的一般形式为

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, i = 1, 2, \cdots, n$$

式中: $\beta_i (i = 0, 1, \cdots, p)$ 是未知的回归系数; 假设 $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \cdots, n$ 。将此线性模型以矩阵的形式简写成

$$Y = X\beta + \varepsilon \quad (1)$$

式中: $Y = (y_1, y_2, \cdots, y_n)^T$; $X = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p)$; $\mathbf{1}$ 为元素为 1 的 n 维列向量; $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})^T$; $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^T$ 。若对每个观测值引入一个新的变量, 即异常值示性变量 γ_i 。若 $\gamma_i = 0$, 则说明第 i 个观测值不是异常值, 否则是异常值。将模型(1)扩展为

$$Y = X\beta + \gamma \cdot \mathbf{1} + \varepsilon \quad (2)$$

式中, $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_n)^T$ 。模型(2)就是均值漂移模型, 也是本文所使用的模型。

2.2 估计方法

针对模型(2), 需要最小化的目标函数是

$$f(\beta, \gamma, \lambda) = \frac{1}{2} \|Y - X\beta - \gamma\|_2^2 + \sum_{i=1}^n P_\lambda(\gamma_i) \quad (3)$$

式中, $\sum_{i=1}^n P_\lambda(\gamma_i)$ 是惩罚函数项, $\lambda > 0$ 是惩罚函数中的调节参数。与文献[9]相似, 这里假设 γ 是稀疏的, 由于 β 和 γ 对于式(3)给出的目标函数关于 β 和 γ 都是凸的, 因此可以交替对 β 和 γ 进行优化, 具体优化步骤详见第 3 节。根据一定的调节标准, 选定一个大小适中的 λ 值。对于给定的 λ 值, 式(3)的最小值点记为 $\hat{\beta}$ 与 $\hat{\gamma}$ 。 $\hat{\beta}$ 是回归系数的估计, $\hat{\gamma}$ 则是异常值示性函数的估计。若 $\hat{\gamma}_i \neq 0$, 则意味着第 i 个观测值为异常值。

在本文中, 考虑了 3 种惩罚函数, 与之相对应的截断函数如下:

$$\text{LASSO截断函数 } \theta_{\text{lasso}}(x; \lambda) = \begin{cases} 0, & |x| \leq \lambda \\ x - \text{sign}(x)\lambda, & |x| > \lambda \end{cases}$$

HARD截断函数 $\Theta_{\text{hard}}(x; \lambda) = \begin{cases} 0, & |x| \leq \lambda \\ x, & |x| > \lambda \end{cases}$

SCAD截断函数 $\Theta_{\text{scad}}(x; \lambda; a) =$

$$\begin{cases} (|x| - \lambda)_+ \text{sign}(x), & |x| \leq 2\lambda \\ a - 2^{-1} \{(a-1)x - \text{sign}(x)a\lambda\}, & 2\lambda < |x| \leq a\lambda \\ x, & |x| > a\lambda \end{cases}$$

其中, $(\cdot)_+$ 表示函数的正部。这里 SCAD 截断函数的参数除 λ 外还有参数 a 。根据文献[13]的建议, a 的最佳取值为 3.7, 故本文使用此值。

3 算法

3.1 算法步骤

使用迭代算法进行计算, 步骤分为以下几步。

第一步, 用一种稳健的方法对 β 进行估计, 得到一个初始值 $\hat{\beta}$, 之后计算 $\gamma^{(0)} = Y - X\hat{\beta}$ 。在本文以及文献[9]中, 目标函数关于 β 和 γ 都是凸的, 因此初始点的选择并不是很关键, 甚至可以让 $\gamma^{(0)} = 0$, 所以同文献[9]一致, 使用初始值 $\gamma^{(0)} = Y - X\hat{\beta}$ 。随后, 将 X 进行 QR 分解, 得 $X = QR$ 。模拟发现, 通过不同的方法对 β 进行初始估计时, 最终结果差别不大。在本文中, 使用的方法是 MM 估计。

第二步, 确定调节参数 λ 的取值范围。将 $\|(I - H)Y / \sqrt{\text{diag}(I - H)}\|_\infty$ 设为最大值, 以步长 -0.1 进行取值至 0。其中, I 是 n 维单位矩阵, H 是帽子矩阵, $H = H(x) = X(X^T X)^{-1} X^T$ 。

第三步, 计算 $Y^* = Y - \gamma^{(i)}$, $\beta^{(i)} = R^{-1} Q^T Y^*$, 以及 $r = Y - X\beta^{(i)}$ 。

第四步, 通过所选用惩罚函数产生的截断函数计算 $\gamma^{(i+1)} = \Theta(r, \lambda)$, 由此得到一个新的 γ 。

第五步, 通过 $\|\gamma^{(i+1)} - \gamma^{(i)}\|_\infty < \varepsilon$ 判断 γ 是否收敛。其中, ε 是一个极小的值, 文中取值为 0.000 1。如果未收敛, 继续运行第三步和第四步进行迭代。若收敛, 则可结束迭代, 得到最终的 β 的估计以及 γ 的估计。

3.2 调节方式

在 λ 的取值范围内, 对一个给定的 λ , 迭代后得到了相对应的 $\hat{\beta}_{(\lambda)}$ 以及 $\hat{\gamma}_{(\lambda)}$ 。此时需要通过一些准则来确定哪个 λ 是最优的, 其对应的 $\hat{\beta}_{(\lambda)}$ 以及 $\hat{\gamma}_{(\lambda)}$ 也将被最终采用。目前已有 AIC(Akaike information criterion), BIC, CV 等调节方式。文献[9]采用了一种经过优化的 BIC 方式。本文模拟过程中发现使用这种方式选择 λ 时, 会导致第二种错误概率较高。故本文对其进一步优化, 记作

BIC_M 。

$$BIC_M(\lambda) = m \log \left(\frac{RSS_M}{m} \right) + k(\lambda)(\log m + 1)$$

式中: $m = n - p$; $RSS_M = \|(I - H)(Y - \hat{\gamma})^2\|_2$; $k(\lambda) = DF(\lambda) + 1$; $(Y - \hat{\gamma})^2$ 是指对 $Y - \hat{\gamma}$ 所得矩阵的每一元素都进行平方; $DF(\lambda)$ 是 $\hat{\gamma}_{(\lambda)}$ 中非零值的数目。优化之后的效果有所提升。但这两种基于 BIC 的调节方式都存在一个共同的问题, 即当异常值相对于整体的偏差较小时, 选择的效果很不理想。考虑到这种情况, 本文又考虑了 CV 方法, 其具体步骤如下所述。首先, 先给定一个 λ , 然后依次选取出数据中的一组, 并使用剩下的数据对 β 进行估计得到 $\hat{\beta}_{(i)}$, 然后使用该 $\hat{\beta}_{(i)}$ 估计相应被选取的一组数据, 并得到一个残差。对所有的数据组依次进行选取计算, 最后得到一系列残差, 将这些残差相加得到总的残差和。这个残差和就是与该给定的 λ 相对应的残差和。在 λ 的取值范围内所得到的最小的残差和对应的 λ 值就是应使用的 λ 值。但经大量模拟发现, CV 方式选择的 λ 值会偏小, 导致惩罚力度偏小。同时, BIC_M 方式选择的 λ 又偏大, 惩罚力度太强。故综合考虑这两种方式, 对 BIC_M 和 CV 方式所选的 λ 进行加权平均。

$$\lambda = (1 - M) \lambda_{BIC_M} + M \lambda_{CV}$$

式中: λ_{BIC_M} 是由 BIC_M 选择得到的 λ ; λ_{CV} 是由 CV 选择得到的 λ 。通过模拟发现, 调节结果最好时的系数 M 为

$$M = \left(\frac{\lambda_{BIC_M} - \lambda_{CV}}{\lambda_{BIC_M}} \right) * \left(\frac{\lambda_{BIC_M}}{\lambda_{\max}} \right)^2$$

式中, $\lambda_{\max}$ 是上文确定 λ 取值范围时所采用的最大值。系数 M 的第一部分考虑的是 λ_{BIC_M} 和 λ_{CV} 差值的影响。因为异常值偏差的大小会影响到 $\lambda_{\max}$, 所以第二部分考虑了这个因素。之所以这样定义 M 的表达式, 是因为本文在大量模拟中发现, BIC 准则更侧重于模型的选择, 因而总是选出较大的 λ ; 而 CV 准则更侧重于模型的拟合优度, 因而总选出较小的 λ 。所以, 总有 $\lambda_{CV} < \lambda_{BIC_M}$ 。这种发现与文献[14]和[15]的观点一致, 故 $0 < \frac{\lambda_{BIC_M} - \lambda_{CV}}{\lambda_{BIC_M}} < 1$ 。另外, 当异常值偏差较小时, λ_{BIC_M} 接近于 $\lambda_{\max}$, 而 λ_{CV} 很小。当异常值偏差较大时, λ_{BIC_M} 慢慢接近于 λ_{CV} 。故, $0 < M < 1$, 且 M 随着异常程度的增大而减小。

4 模拟研究

通过模型(2)产生模拟数据, 其中, x_{ij} 和误差 ε_i 都来自于标准正态分布。令 $\beta = (1, 1, -1, 0, \dots, 0)^T$ 。为全面考察本方法的稳健性, 讨论了 $n = 100$, $p = 10$ 以及 $n = 100$, $p = 5$ 这两种情形。对于污染方式, 通过污染 Y 的前十行来对数据进行污染, 具体方式是令模型(2)中的 $\gamma_i (i = 1, 2, \dots, 10)$ 分别取 $(1, 2, \dots, 12)$, 这里 γ_i 的取值逐渐增大, 代表异常程度从低到高。为排除随机性对本方法的影响, 对每种方式都重复模拟了 200 次。

为比较结果的优劣, 使用了两种指标。第一种指标是不犯第一种错误的比率, 记作 NM, 其计

算的是检测出的异常值占有所有异常值的比例, 也就是未被掩盖的异常值的比例。第二种指标是不犯第二种错误的比率, 记作 NS, 其计算的是检测出的正常值占有所有正常值的比例, 也就是未被淹没的正常值的比例。这两种指标越接近于 1, 说明方法越有效。

表 1 给出了当 $n=100$, $p=10$ 时, 两指标的模拟结果。其中, 每种方式对应的第一行是 NM 指标, 第二行是 NS 指标。LASSO_M, HARD_M, SCAD_M 是使用本文调节方式所得到的结果, LASSO, HARD, SCAD 是使用文献[9]的调节方式得到的结果, SH 是使用文献[7]的调节方式得到的结果。

表 1 各种方法关于正确检测比例的模拟结果 ($n=100$, $p=10$)

Tab.1 Simulation results about correctly detecting proportion with different methods ($n=100$, $p=10$)

γ_i	1	2	3	4	5	6	7	8	9	10	11	12
LASSO _M	0.90	0.60	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.13	0.98	0.92	0.79	0.97	0.79	0.98	0.98	0.98	0.98	0.98	0.99
HARD _M	0.89	0.90	0.90	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.36	0.77	0.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SCAD _M	0.00	0.20	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	0.98	0.91	0.94	0.98	0.98	0.99	0.99	0.99	1.00	1.00
LASSO	0.90	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.08	0.16	0.10	0.07	0.13	0.07	0.09	0.09	0.12	0.13	0.13	0.13
HARD	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.01	0.08	0.03	0.00	0.07	0.00	0.01	0.06	0.06	0.06	0.08	0.17
SCAD	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.02	0.22	0.13	0.00	0.19	0.03	0.09	0.17	0.20	0.18	0.24	0.27
SH	0.00	0.00	0.10	0.40	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

由表 1 可以看出, LASSO_M, HARD_M, SCAD_M 得到的结果中, 两指标都相对较高。虽然在异常程度较低时, 本文方法会发生一些淹没现象, 但随着异常程度的变大, 其两指标均为 1 或接近于 1。而使用文献[9]的调节方式时, 其 NM 指标很高, 但其 NS 指标太低, 即使异常程度较大时也很低, 甚至全部未能超过 0.3。这也说明文献[9]的调节方式侧重于模型的选择, 使用了较大的 λ 值, 但其淹没现象太严重, 大量正常值被误判为异常值。至于 SH 方法, 其两指标相对较高。但在异常的程度较低时, 其 NM 指标为 0 或者接近于 0, 明显低于本文结果。这说明其对异常值不敏感, 容易发生掩盖现象。而且 SH 方法还涉及统计量阈值的取值问题, 这也是该方法的一个难点。本文对 SH 方法进行模拟时, 使用的是文献

[7]的推荐值, 但该值会导致较严重的掩盖问题。而本文提出的调节方式则不需要考虑这些问题, 对各种情况均易达到理想的效果。从表 1 也可看出, 对于不同的惩罚方式, 本文的调节方式都能对其进行有效调节, 证明了该方法的普适性, 而且相对于其他人的异常值检测方法而言, 本方法也能取得更好的结果。由此可说明本方法能明显提升异常值检测的效果。

表 2 给出了当 $n = 100$, $p = 5$ 时, 两指标的模拟结果, 其结果排布方式同表 1。

由表 2 可发现, 使用本文的调节方式仍能得到两指标都很高的结果。对比表 1 和表 2 的结果可发现, 在这两种模拟情形中, 各方式所得的结果基本不变, 可以得出相同的结论。另外, 当 p 较小时, 各方式对异常程度较低的数据能更好地

进行异常值监测,尤其是本文的调节方式,效果改善更为明显。

表2 各种方法关于正确检测比例的模拟结果 ($n=100, p=5$)

Tab.2 Simulation results about correctly detecting proportion with different methods ($n=100, p=5$)

γ_i	1	2	3	4	5	6	7	8	9	10	11	12
LASSO _M	0.01	0.20	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.99	0.99	1.00	1.00	1.00
HARD _M	0.80	0.20	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.27	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SCAD _M	0.00	0.20	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
LASSO	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.04	0.03	0.00	0.12	0.12	0.12	0.11	0.07	0.07	0.07	0.07	0.06
HARD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.01	0.01	0.01	0.05	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.01
SCAD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.02	0.01	0.01	0.16	0.16	0.15	0.11	0.08	0.07	0.06	0.04	0.01
SH	0.00	0.20	0.20	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

5 总结与展望

统计分析研究中有关异常值检测的问题一直是我们的热点,且也在这个方向取得了不少成果。使用系数压缩方法对线性模型进行异常值检测时,会面临一个调节问题。本文研究这个问题时,发现了一种新的调节方式。随后,通过模拟不同的情形并进行大量的模拟计算来检验这种新方式。结果表明,该方法相对于其他方法能更准确地解决线性模型数据中异常值的检测问题。

目前,本文只讨论了异常值的检测问题,没有考虑协变量的选择问题。试想,在目标函数(3)中再增加一个对 β 的惩罚项,则可以将异常值检测与变量选择同时进行。至于如何限制异常值对变量选择效果的影响,这是值得进一步探讨的问题。

参考文献:

- [1] ABRAHAM B, CHUANG A. Outlier detection and time series modeling[J]. Technometrics, 1989, 31(2): 241–248.
- [2] ABRAHAM B, YATAWARA N. A score test for detection of time series outliers[J]. Journal of the Time Series Analysis, 1988, 9(2): 109–119.
- [3] COOK R D. Detection of influential observation in linear regression[J]. Technometrics, 1977, 19(1): 15–18.
- [4] COOK R D. Influential observations in linear regression[J]. Journal of the American Statistical Association, 1979, 74(365): 169–174.
- [5] 陈希孺, 王松桂. 近代回归分析——原理方法及应用[M]. 合肥: 安徽教育出版社, 1987.
- [6] DRAPER N R, SMITH H. Applied regression analysis[M]. New York: John Wiley, 1981.
- [7] 尚华, 张贝贝, 纪宏. 一种新的基于回归分析的异常值检测[J]. 河南大学学报(自然科学版), 2015, 45(6): 635–639.
- [8] PEÑA D, YOHAI V J. The detection of influential subsets in linear regression by using an influence matrix[J]. Journal of the Royal Statistical Society, 1995, 57(1): 145–156.
- [9] SHE Y Y, OWEN A B. Outlier detection using nonconvex penalized regression[J]. Journal of the American Statistical Association, 2011, 106(494): 626–639.
- [10] MARASINGHE M G. A multistage procedure for detecting several outliers in linear regression[J]. Technometrics, 1985, 27(4): 395–399.
- [11] TIBSHIRANI R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1996, 58(1): 267–288.
- [12] FAN J Q, LI R Z. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American Statistical Association, 2001, 96(456): 1348–1360.
- [13] DONOHO D L, JOHNSTONE J M. Ideal spatial adaptation by wavelet shrinkage[J]. Biometrika, 1994, 81(3): 425–455.
- [14] SHAO J. Linear model selection by cross-validation[J]. Journal of the American Statistical Association, 1993, 88(422): 486–494.
- [15] SHAO J. An asymptotic theory for linear model selection[J]. Statistica Sinica, 1997, 7(2): 221–242.

(编辑: 董伟)