

M Estimators and Influence Diagnostics in Linear Models

Rong Jiang, Weimin Qian

Department of Mathematics, Tongji University, Shanghai
Email: jrtrying@126.com, wmqian2003@yahoo.com.cn

Received: Nov. 12th, 2012; revised: Nov. 17th, 2012; accepted: Nov. 29th, 2012

Abstract: In order to overcome the large deviation between the actual observed data and established models. There are two common methods: the robust estimation and statistical diagnostics. M estimator is a robust estimation, it is the method which got the most attention and research results in linear regression. Therefore, influence diagnostics and M estimator are used to judge the impact of outliers or strong influence points in this paper. Moreover, the expression of parameter estimation deviate and diagnostic statistics are given. Finally, the proposed methods are applied to two data sets.

Keywords: Case-Deletion Model; Mean-Shift Outlier Model; Local Influence; M Estimate; Cook Distance

线性模型基于 M 估计的统计诊断与影响分析

姜 荣, 钱伟民

同济大学应用数学系, 上海
Email: jrtrying@126.com, wmqian2003@yahoo.com.cn

收稿日期: 2012 年 11 月 12 日; 修回日期: 2012 年 11 月 17 日; 录用日期: 2012 年 11 月 29 日

摘 要: 为了克服实际观测数据与既定模型之间可能存在的较大偏离, 目前有两种常用的处理方法: 稳健统计与统计诊断。M 方法是最重要的稳健统计之一, 也是线性回归分析中是最受重视和研究成果最多的方法之一。所以本文结合 M 估计方法分析数据的统计诊断, 从而得到异常点或强影响点。本文给出了参数估计偏离的表达式及几个诊断统计量, 最后通过两个实例验证了本文所提方法的可行性。

关键词: 数据删除模型; 均值漂移模型; 局部影响; M 估计; Cook 距离

1. 引言

统计诊断从 20 世纪 70 年代中期受到统计学家的广泛关注, 经过近 40 年的发展, 异常点识别、残差分析、影响分析和数据变换等内容现已成为统计诊断的主要课题。特别地, 基于数据删除模型和局部影响的诊断分析方法现已成为统计诊断的通用方法, 它们可广泛地应用于各种统计模型的影响分析。例如, 线性模型(Cook and Weisberg^[1]), 非线性回归模型(Seber and Wild^[2]), 半参数非线性模型(姜荣, 邵明江, 钱伟民^[3]), 指数族非线性模型(Wei^[4]), 线性混合效应

模型(Beckman et al.^[5])。M 估计是目前应用最广泛的稳健估计方法。吴小燕, 赵林城, 杨亚宁^[6]研究了线性模型的随机加权 M 估计。Jiang et al.^[7]对测量误差模型提出了随机加权 M 估计。最近, 孙慧慧, 林金官^[8]利用 M 方法提出了稳健极大似然估计, 并基于此估计方法研究了线性混合模型的随机误差的方差加权扰动模型的影响分析。

本文基于 M 估计, 考虑删除模型、均值漂移模型、因变量扰动模型以及自变量扰动模型的影响分析。

考虑如下线性回归模型:

$$y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

其中 x_1, \dots, x_p 为已知的 p 维向量, $\varepsilon_1, \dots, \varepsilon_p$ 为独立同分布随机误差, β 为未知的 p 维回归参数向量。设 β 的 M 估计 $\hat{\beta}$ 是求下述极小问题的解:

$$S(\beta) = \sum_{i=1}^n \rho(y_i - x_i^T \beta) \quad (2)$$

其中 ρ 为 R^1 上的非单调凸函数, 且以 ψ_- 和 ψ_+ 分别记 ρ 的左右导数, ψ 是介于 ψ_- 和 ψ_+ 之间的一个函数。

本文结合统计诊断与 M 方法来诊断实际观测数据与既定模型之间的偏离。这样检测出来的点相比最小二乘法更有理由相信它是异常点或强影响点。本文给出了参数偏离的表达式及几个诊断统计量, 并通过实例验证本文所提出的方法是可行的。

本文的安排如下: 第二节介绍基于删除模型的影响分析, 给出了参数估计偏离的表达式及诊断统计量; 第三节介绍了均值漂移模型, 并证明了数据删除模型与均值漂移模型基于 M 估计的等价性; 第四节介绍了局部影响分析, 给出了两个不同的扰动下的诊断统计量; 第五节通过实例验证了本文所提出的诊断方法的有效性。附件中是定理的证明。

2. 基于删除模型的影响分析

2.1. 数据删除模型

考虑模型(1)式中删除第 i 个数据点 $(y_i, x_i^T)^T$ 以后的模型及参数估计。这个模型可表示为

$$y_j = x_j^T \beta + \varepsilon_j, \quad j \neq i \quad (3)$$

模型(3)称为数据删除模型, 或简称删除模型(CDM), 记删除第 i 个数据点的数据删除模型(3)式中 β 的 M 估计记为 $\hat{\beta}_{(i)}$ 。为研究数据点 $(y_i, x_i^T)^T$ 对估计 $\hat{\beta}$ 的影响, 最简单直接的方法就是比较两组估计 $\hat{\beta}$ 和 $\hat{\beta}_{(i)}$ 的差异。为此有以下重要定理。

首先给出定理所需要的条件

A1: 当 $|u|$ 充分小时, $G(u) \equiv E\{\psi(\varepsilon_1 + u)\}$ 存在, 在 $u=0$ 有正的导数 $\lambda = G'(0)$, 且当 $u \rightarrow \infty$ 时, $|G'(u) - \lambda| = O(|u|^{1/2})$ 。

A2: $0 < \sigma^2 = E\{\psi^2(\varepsilon_1)\} < \infty$, 且 $E|\psi(\varepsilon_1)|^3 < \infty$ 。

A3: 对充分小的 $t > 0$, 有 $\sup_{|u| < \infty} |\psi(u+t) - \psi(u)| < \infty$

注: 条件 A1~A3 是 M 估计线性表示和参数相合

性最基本的条件。

定理 1: 若模型(1)满足条件 A1~A3, 则模型(3)中 β 的 M 估计 $\hat{\beta}_{(i)}$ 的一阶近似可表示为:

$$\hat{\beta}_{(i)} = \hat{\beta} - \left[(X^T X)^{-1} x_i / (1 - P_{ii}) \right] \left(h^{-1} \psi(\varepsilon_i) + \hat{\varepsilon}_i \right) + o_p(1)$$

其中: $h = E\{\psi'(\varepsilon_1)\} \neq 0$, $P_{ii} = x_i^T (X^T X)^{-1} x_i$, $\hat{\varepsilon}_i = y_i - x_i^T \hat{\beta}$ 。

考虑去掉第 i 个点 $(y_i, x_i^T)^T$ 的删除模型, 即模型(3)式。差值 $\hat{\beta}_{(i)} - \hat{\beta}$ 就是数据点 $(y_i, x_i^T)^T$ 的影响大小的一种度量, 差值越大, 影响越大。但是由于 $\hat{\beta}_{(i)} - \hat{\beta}$ 是一个向量, 不便于比较, 必须选择一个适合的数量或距离, 以便于定量地比较影响的大小。为了解决以上问题, 我们研究在影响诊断中最重要的距离之一: Cook 距离。

2.2. 广义 Cook 距离

Cook and Weisberg^[1] 基于参数置信域的观点提出了 Cook 距离。取以下非负定矩阵 M, 及实数 C 作为尺度因子。定义 $\hat{\beta}_{(i)} - \hat{\beta}$ 的距离为

$$D_i = (\hat{\beta}_{(i)} - \hat{\beta})^T M (\hat{\beta}_{(i)} - \hat{\beta}) / C$$

D_i 称为广义 Cook 距离, 它是度量 $\hat{\beta}_{(i)} - \hat{\beta}$ 大小的有效度量。根据陈希孺和赵林城^[2] 的定理 4.1 和定理 5.2, 我们有:

$$(\hat{\beta} - \beta_0)^T X^T X (\hat{\beta} - \beta_0) / (\hat{\sigma}^2 \hat{\lambda}^{-2}) \rightarrow^L \chi^2(p)$$

其中:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \psi^2(y_i - x_i^T \hat{\beta}),$$

$$\hat{\lambda} = \frac{1}{2nh} \sum_{i=1}^n \left\{ \psi(y_i - x_i^T \hat{\beta} + h) - \psi(y_i - x_i^T \hat{\beta} - h) \right\}.$$

对模型(1)的 M 估计问题, 我们选取

$$M = X^T X; \quad C = \hat{\sigma}^2 \hat{\lambda}^{-2}$$

所以广义 Cook 距离的公式:

$$D_i = (\hat{\beta}_{(i)} - \hat{\beta})^T X^T X (\hat{\beta}_{(i)} - \hat{\beta}) / (\hat{\sigma}^2 \hat{\lambda}^{-2})$$

3. 均值漂移模型与异常点检验

数据删除模型, 由于它非常直观和容易计算, 所以它是构造有效诊断统计量的基础, 且在实践中也是

一种最重要的诊断模型。另一种经常使用的诊断模型是均值漂移模型(MSOM)。均值漂移模型可表示为

$$\begin{cases} y_j = x_j^T \beta + \varepsilon_j, j \neq i \\ y_i = x_i^T \beta + \gamma + \varepsilon_i \end{cases} \quad (4)$$

其中 γ 为扰动值, 是一个新参数。该模型表明对第 i 个数据点增加一个扰动值。记该模型相应参数的 M 估计为 $\hat{\beta}_{mi}, \hat{\gamma}$ 。

对于最小二乘估计, 韦博成, 林金官, 解锋昌^[9]证明了数据删除模型和均值漂移模型的等价性, 即这两个模型的估计量相等。但对于 M 估计, 类似的结论文献中还未见到。因此我们将这一结果推广到 M 估计。有以下定理。

定理 2: 若模型(4)满足条件 A1~A3, 则有

$$\hat{\beta}_{(i)} = \hat{\beta}_{mi}$$

定理 2 说明, 虽然数据删除模型(3)和均值漂移模型(4)表面形式不一样, 但有关估计量的统计性质完全相同。因而用它们研究第 i 个数据点 $(y_i, x_i^T)^T$ 对估计量的影响, 其效果一样。比较模型(3)与模型(1), 若 $\hat{\beta}$ 与 $\hat{\beta}_{(i)}$ 有显著差异, 则说明 $(y_i, x_i^T)^T$ 可能是异常点; 比较模型(4)与模型(1), 若 γ 显著不为零, 则说明 $(y_i, x_i^T)^T$ 可能是异常点。而定理 2 说明两者是一致的。在实践中, 往往根据需要选择其中之一加以研究。通常, 数据删除模型更为直观, 均值漂移模型更便于分析。

4. 局部影响分析

Cook^[10]从微分几何观点出发提出了局部影响分析方法, 随后, 经过许多作者(例如, Escobar and Mecker^[11]和 Wei^[12])的研究, 这一方法得到了充分发展, 本节研究线性模型基于 M 估计的局部影响分析。

首先介绍一般的局部影响分析方法。令 φ 是未知的 k 维参数, 其定义域为 R^k 中的某一开集。 $l(\varphi)$ 是为估计 φ 设计的某一目标函数(例如似然函数, 又如(2)式的 $S(\beta)$ 函数)。 ω 是定义在 $\Omega \in R^n$ 上的 n 维向量, 表示对模型的扰动因素, 记受扰动的模型为 $M(\omega)$, 其相应的目标函数为 $l(\varphi|\omega)$, 由 $M(\omega)$ 得到的相应估计记为 $\hat{\varphi}_\omega$ 。假定存在一个 ω_0 使 $l(\varphi|\omega_0) = l(\varphi)$ 且 $\hat{\varphi} = \hat{\varphi}_{\omega_0}$, $l(\varphi|\omega)$ 关于 φ 和 ω 存在二阶以上连续导数。 $l(\hat{\varphi}_\omega)$ 是 ω 的函数。从几何上看函数 $l(\hat{\varphi}_\omega)$ 表示 $n+1$ 维空间中的一个 n 维曲面

$$\eta(\omega) = (\omega^T, l(\hat{\varphi}_\omega))^T \quad (5)$$

这个图称为影响图。影响图随 ω 变化的情况更全面地反映了扰动的影响。由于 ω_0 对应于无扰动模型, 因此影响图在 ω_0 处的变化率反映了原模型对于扰动的敏感程度, 现称为局部影响。Cook^[11]提出利用影响曲率度量影响图在 ω_0 附近的变化情况。

根据 Cook^[10]可知, 由(5)式定义的影响图的影响曲率可表示为

$$l_d = d^T \Delta^T [-\ddot{l}]^{-1} \Delta d \quad (6)$$

其中 \ddot{l} 为 l 关于 φ 的二阶导数, 在 $\hat{\varphi}$ 处的值;

$$\Delta = \partial^2 l(\varphi|\omega) / \partial \varphi \partial \omega^T \Big|_{\hat{\varphi}, \omega_0}, \quad d = \omega - \omega_0。$$

所以下面我们应用(6)式, 研究线性模型基于 M 估计的局部影响分析。

4.1. 因变量扰动模型

对因变量加权扰动, 即在 $Y_\alpha = Y + \omega$ 下找强影响点。假定 $\omega = (\omega_1, \dots, \omega_n)^T$ 为扰动向量, $\omega_0 = (0, \dots, 0)^T$ 对应无扰动模型, 则扰动模型可表示为:

$$S(\beta|\omega) = \sum_{i=1}^n \rho(y_i + \omega_i - x_i^T \beta)$$

经简单的计算可得:

$$\ddot{S}(\hat{\beta}|\omega_0) = \psi'(\hat{\varepsilon}_i) X^T X \quad (7)$$

$$\Delta_{\omega_i} = -x_i \psi'(\hat{\varepsilon}_i), \Delta = (\Delta_{\omega_1}, \dots, \Delta_{\omega_n}) \quad (8)$$

把(7)式和(8)代入(6)式由此可得相应的影响曲率。

4.2. 自变量扰动模型

设第 k 个自变量有扰动, 这时

$X_k \rightarrow X_{k\omega} = X_k + \omega$ 。假定 ω 为加权扰动向量, $\omega_0 = (0, \dots, 0)^T$ 对应无扰动模型, 则扰动模型可表示为:

$$S(\beta|\omega) = \sum_{i=1}^n \rho(y_i - (x_i + \omega_i h_k)^T \beta)$$

其中 h_k 为第 k 个分量为 1 其他为 0 的 p 维向量, 再经简单的计算可得:

$$\Delta_{\omega_i} = \psi(\hat{\varepsilon}_i) h_k^T \hat{\beta} x_i - \psi(\hat{\varepsilon}_i) h_k^T d_a$$

$$\Delta = (\Delta_{\omega_1}, \dots, \Delta_{\omega_n}) \quad (9)$$

其中：当 $k = a$ 时 $h_k^T d_a = 1$ ；当 $k \neq a$ ， $h_k^T d_a = 0, a = 1, \dots, p$ 。

把(7)式和(9)代入(6)式由此可得相应的影响曲率。

5. 实际例子分析

5.1. 镀锌数据(Cook^[13])

在钢铁上镀锌的传统工艺代价较高且污染严重，今考虑一种比较节约且无害的新工艺。为比较新旧工艺的性能，以镀层厚度(单位为 10^{-5} 英寸)来度量， y 表示按新工艺操作的镀层厚度， x 表示按旧工艺操作的镀层厚度，共有 11 块测试样品。考虑 y 关于 x 的线性回归

$$y_i = \beta_0 + x_i^T \beta_1 + \varepsilon_i, i = 1, \dots, 11 \quad (10)$$

这里我们取惩罚函数为：

$$\rho(u) = \begin{cases} u^2, & |u| \leq 1 \\ 2u - 1, & u > 1 \\ -2u - 1, & u < -1 \end{cases} \quad (11)$$

运用本文的方法得到参数 β 的 M 估计 $\hat{\beta}_0 = 4.4974, \hat{\beta}_1 = 0.8695$ 。

表 1 给出了有关的统计量，其中 β_0, β_1 是参数的

估计， ε 是相应删除数据下的残差， ri, ti 分别是 M 估计得学生化残差和学生化外残差， Di 为广义 Cook 距离的值。因为第 7 号点的广义 Cook 距离 Di 的值非常的大，所以第 7 号点为异常点或强影响点。分析原因是在原始数据中第 7 号点远离其它的点，还可以从参数 β 的估计里看出，7 号点与其他点的差异非常明显，而其他点之间的差异很小。

图 1 和图 2 分别是因变量扰动模型和自变量扰动模型下影响曲率的散点图，从图 1 和图 2 可看出，第 7 号点是强影响点与前面分析的一致。所以，本文的方法是有效的。

5.2. 雪雁数据(Weisberg^[14])

为调查加拿大某地雪雁的数据，可由有经验的人在小飞机上通过目测估计出每群雪雁的数目。为研究目测方法的可靠性，进行了一项实验：在目测每群雪雁数目(记为 x)的同时，也用通过拍照测出每群雪雁的准确数目(记为 y)；共测量了 45 群。

本例考虑的模型和惩罚函数与例 5.1 中的一样。(即模型为(10)式，惩罚函数为(11)式)。

运用本文的方法得到参数 β 的 M 估计 $\hat{\beta}_0 = 5.1785, \hat{\beta}_1 = 0.6004$ 。图 3 是雪雁数据广义 Cook 距离的散点图，从图中可以发现第 28 号点为异常点或强影响点。图 4 和图 5 分别是因变量扰动模型和自变量扰动模型下影响曲率的散点图，从图 4 和图 5 可看出，第 28 号点是强影响点并且第 41 号点的影响也较大。

Table 1. Statistics for galvanized data
表 1. 镀锌数据的相关统计量

NO	β_0	β_1	ε	ri	ti	Di
1	4.7491	0.8692	-0.3636	-0.2468	-0.1828	0.1294
2	4.0316	0.8702	0.7238	0.4883	0.4766	0.4598
3	5.0195	0.8687	-9.9292	-6.7737	-6.8032	0.5549
4	4.972	0.8689	-4.3629	-2.9365	-2.9184	0.4925
5	4.1500	0.8701	11.3755	7.7274	8.6404	0.2424
6	4.1557	0.8700	10.3324	6.9784	7.8663	0.2451
7	-3.3953	0.9286	-0.5635	-0.6174	-0.4728	736.9482
8	4.9243	0.8690	-0.7967	-0.5311	-0.4458	0.4374
9	4.7363	0.8696	-27.3927	-18.2146	-17.0120	0.2955
10	4.3065	0.8693	11.5994	7.7574	9.6184	0.2636
11	4.4495	0.8688	34.1683	24.2697	30.4102	0.2965

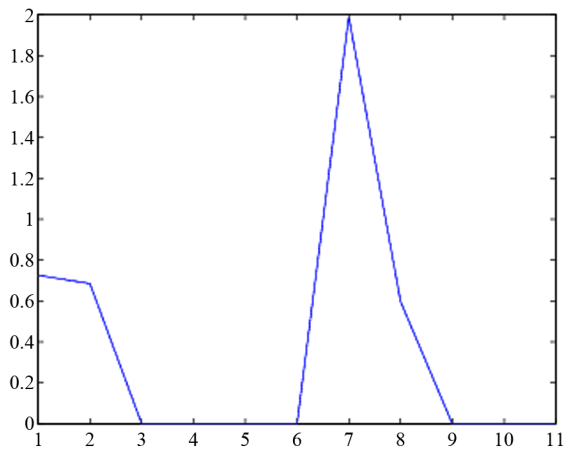


Figure 1. The scatter diagram of the variable perturbation influence for galvanized data
图 1. 镀锌数据在因变量扰动下影响曲率的散点图

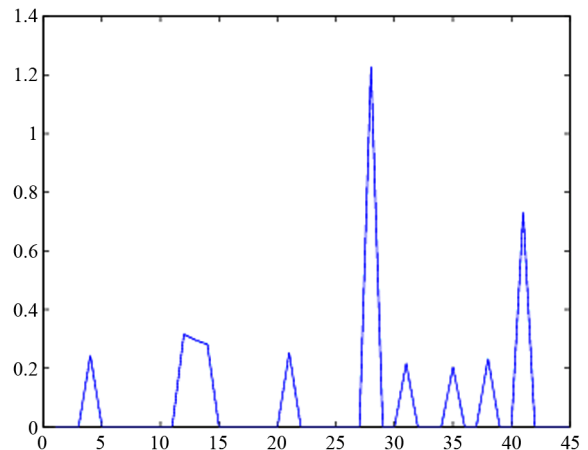


Figure 4. The scatter diagram of the variable perturbation influence for snow geese data
图 4. 雪雁数据在因变量扰动下影响曲率的散点图

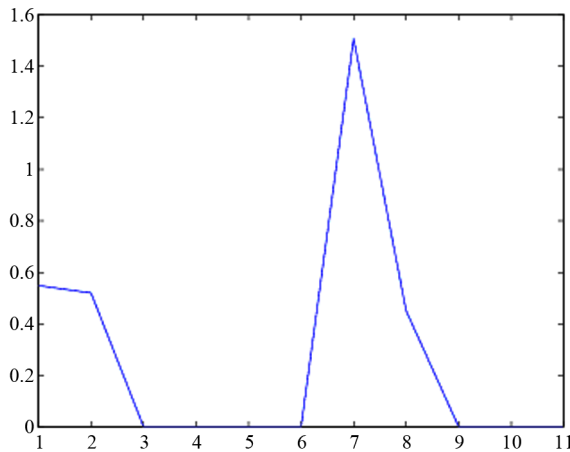


Figure 2. The scatter diagram of the independent variable perturbation influence for galvanized data
图 2. 镀锌数据在自变量扰动下影响曲率的散点图

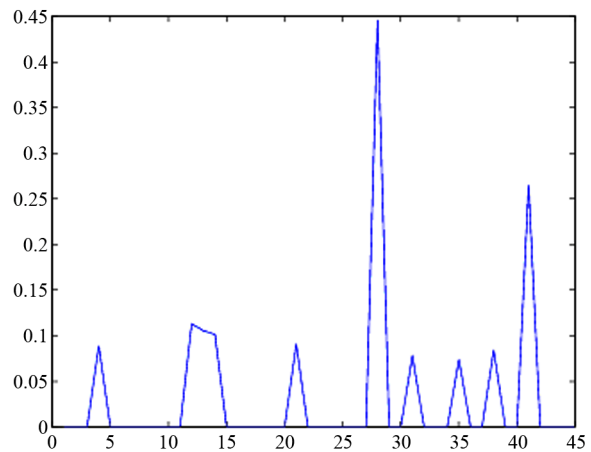


Figure 5. The scatter diagram of the independent variable perturbation influence for snow geese data
图 5. 雪雁数据在自变量扰动下影响曲率的散点图

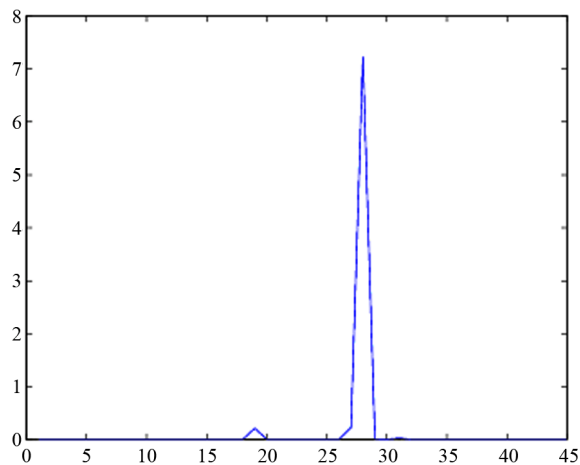


Figure 3. The scatter diagram of generalized Cook distance for snow geese data
图 3. 雪雁数据广义 Cook 距离的散点图

6. 致谢

感谢编辑及审稿人的宝贵意见。

参考文献 (References)

- [1] R. D. Cook, S. Weisberg. Residual and influence in regression. New York: Chapman and Hall, 1982.
- [2] G. Seber, C. J. Wild. Nonlinear Regression. New York: Wiley, 1989.
- [3] 姜荣, 邵明江, 钱伟民. 半参数非线性模型中的 t-型估计和影响分析[J]. 华东师范大学学报(自然科学版), 2011, 3: 1-11.
- [4] B. C. Wei. Exponential family nonlinear models. Singapore: Springer-Verlag, 1994.
- [5] R. J. Beckman, C. J. Nachtshiem and R. D. Cook. Diagnostics for mixed-model analysis of variance. Technometrics, 1987, 29(4): 413-426.
- [6] 吴小燕, 赵林城, 杨亚宁. 线性模型中 M 估计分布的随机加权

- 方法逼近[J]. 系统科学与数学, 2008, 28(9): 1092-1100.
- [7] R. Jiang, X. Yang and W. Qian. Random weighting M-estimation for linear errors-in-variables models. Journal of the Korean Statistical Society, 2012, 41(4): 505-514.
- [8] 孙慧慧, 林金官. 基于 M 估计的线性混合模型的局部影响分析[J]. 应用概率统计, 2012, 28(2): 217-223.
- [9] 韦博成, 林金官, 解锋昌. 统计诊断[M]. 北京: 高等教育出版社, 2009.
- [10] R. D. Cook. Assessment of local influence (with discussion). Journal of the Royal Statistical Society Series B, 1986, 48(2): 133-169.
- [11] L. A. Escobar, W. Q. Meeker. Assessing influence in regression analysis with censored data. Biometrics, 1992, 48(2): 507-508.
- [12] B. C. We. Exponential family nonlinear models. Singapore: Springer-Verlag, 1994.
- [13] R. D. Cook, N. Holschuh, and S. Weisberg. A notes on alternative outlier model. Journal of the Royal Statistical Society Series B, 1982, 44(3): 370-376.
- [14] S. Weisberg, Applied linear regression. New York: Wiley, 1985.
- [15] 陈希孺, 赵林城. 线性模型中的M方法[M]. 上海: 上海科学技术出版社, 1996.

附录

定理 1 的证明: 根据陈希孺和赵林城^[15]中的(6.2)式, 设 β_0 为真值, 则我们有:

$$\hat{\beta} - \beta_0 = h^{-1} S^{-1} \sum_{i=1}^n \psi(\varepsilon_i) x_i + o_p(1) \quad (12)$$

$$\begin{aligned} \hat{\beta}_{(i)} - \beta_0 &= h^{-1} S_{(i)}^{-1} \sum_{j \neq i} \psi(\varepsilon_j) x_j + o_p(1) \\ &= h^{-1} (X^T X - x_i^T x_i)^{-1} \left\{ \sum_{i=1}^n \psi(\varepsilon_i) x_i - \psi(\varepsilon_i) x_i \right\} + o_p(1) \\ &= h^{-1} \left[(X^T X)^{-1} + \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - P_{ii}} \right] \left\{ \sum_{i=1}^n \psi(\varepsilon_i) x_i - \psi(\varepsilon_i) x_i \right\} + o_p(1) \\ &= h^{-1} (X^T X)^{-1} \sum_{i=1}^n \psi(\varepsilon_i) x_i - h^{-1} (X^T X)^{-1} \psi(\varepsilon_i) x_i \\ &\quad + h^{-1} \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - P_{ii}} \sum_{i=1}^n \psi(\varepsilon_i) x_i - h^{-1} \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - P_{ii}} \psi(\varepsilon_i) x_i + o_p(1) \\ &= \hat{\beta} - \beta_0 - \frac{(X^T X)^{-1} x_i}{1 - P_{ii}} (h^{-1} \psi(\varepsilon_i) + \hat{\varepsilon}_i) + o_p(1) \end{aligned}$$

所以, 定理 1 得证。

定理 2 的证明: 由 $\hat{\beta}_{(i)}$ 的定义, $\hat{\beta}_{(i)}$ 是以下最小问题的解

$$S_{(i)}(\beta) = \sum_{j \neq i} \rho(y_j - x_j^T \beta)$$

对模型(4)式, $\hat{\beta}_{mi}$ 和 $\hat{\gamma}$ 是以下最小问题的解

$$S_{mi}(\beta, \gamma) = \sum_{j \neq i} \rho(y_j - x_j^T \beta) + \rho(y_i - x_i^T \beta - \gamma)$$

因此 $\hat{\gamma}$ 满足以下方程

$$\frac{\partial S_{mi}}{\partial \gamma} = -\psi(y_i - x_i^T \beta - \gamma) = 0 \quad (13)$$

其中: $S = X^T X$ 。记

$$S_{(i)} = \sum_{j \neq i} x_j^T x_j = X_{(i)}^T X_{(i)}$$

则类似(12), 我们有

而 $\hat{\beta}_{mi}$ 满足以下方程

$$\frac{\partial S_{mi}}{\partial \beta} = \frac{\partial S_{(i)}}{\partial \beta} - x_i \psi(y_i - x_i^T \beta - \gamma) = 0 \quad (14)$$

由(13)和(14), 我们有

$$\frac{\partial S_{mi}}{\partial \beta} = \frac{\partial S_{(i)}}{\partial \beta} = 0$$

因此 $\hat{\beta}_{(i)} = \hat{\beta}_{mi}$, 定理得证。