

基于 L_1 -范数拟合的 $AR(p)$ 模型的 稳健的拟合优度检验*

蒋建成

(北京大学概率统计系, 北京 100871)

许溢宏

(香港城市大学管理科学系, 香港)

郑忠国

(北京大学概率统计系, 北京 100871)

摘要 基于 L_1 -回归定义了一个稳健的残差自相关函数. 在非常一般的条件下, 获得了这个稳健的残差自相关的渐近分布. 然后, 构造了一个稳健的多用途(port-manteau)统计量, 它能用于 L_1 -范数拟合的 $AR(p)$ 模型的拟合优度检验. 经验结果表明, 对一给定的容量有限样本, L_1 -范数估计和所提出的多用途统计量对异常值、误差分布和精度是稳健的.

关键词 L_1 -回归 模型诊断检查 稳健的残差自相关 时间序列

在时间序列的分析中, 标准的基于最大似然、最小二乘和矩方法的估计和检验已得到广泛的研究. 众所周知, 这些方法对强影响点和误差分布的偏离很敏感. 进一步, 异常值对这些方法有负面影响. 为了克服这些缺点, 可考虑稳健的时间序列建模方法. 文献[1]考虑了自回归过程的广义 M-估计; 文献[2]使用估计函数研究了广义 M-估计并构造了稳健的得分(score)统计量和 Wald 统计量; 文献[3]提出了所谓的残差自协方差估计, 这包括了经典的 Gauss 估计和 M-估计. 细节可以参考综述文章文献[4].

作为最小二乘的备选, 最小绝对偏离(LAD)估计, 即 L_1 -估计, 是一个自然的能达到一定稳健性的候选方法. 然而, 在时间序列的建模中, 它受到远较一般线性模型少的关注. 在线性模型中, 对非对称误差分布和重尾对称误差分布, L_1 -回归比最小二乘回归有较大的功效; 同时, LAD 估计对关于因变量的异常值有稳健性^[5]. 当误差是独立同分布的双指数分布时, LAD 估计是中值无偏的最大似然估计^[6]. 使用 S-plus 统计软件, LAD 估计的计算在工作站上非常容易执行. 因此, LAD 标准对时间序列的建模很有吸引力. 沿着这一方向, 文献[7, 8]等已做了一些工作. 当使用这一标准时, 自然碰到两个重要问题: 模型阶的选择和诊断检查. 然而, 正如文献[9]指出的, LAD 回归的广泛使用受到对备选模型的可接受的假设检验方法的限制. 虽然可得到其渐近检验方法, 但是这要求估计未知误差分布密度函数在中值点的值 $f(0)^{[10]}$. 文献[11]就线性模型提出了对 LAD 估计的分布执行自助(bootstrap)法逼近, 这似

乎提供了一个检验回归系数的合理方法. 然而, 与最小二乘回归不一样, LAD 回归估计在模型诊断检查中会遇到困难. 虽然, 标准的基于残差的模型诊断检查技术能对模型拟合提供一个定量的评估方法, 但是一个更令人满意的拟合优度的评估方法是定义一个合适的统计量来度量模型的显著性. 文献[12]提出的多用途统计量是检查模型合适性的一个广泛使用的诊断统计量, 然而, 在用 L_1 -回归拟合模型时它显得不合适, 因为缺乏稳健性而与 L_1 -回归的目的相矛盾. 因此, 在此情况下, 构造一个稳健的拟合优度检验是很有价值的. 本文定义了一个稳健的统计量, 当模型成立时, 对一大类误差分布而言, 其渐近分布是已知的, 这将提供一个易于使用的模型诊断工具. 由于自回归模型在计量经济及预测中广泛使用, 将集中考虑这一模型. 内在的思想也适用于自回归滑动平均模型.

1 $AR(p)$ 模型的 L_1 -估计

令 $\{Y_{-(p-1)}, \dots, Y_{-1}, Y_0, Y_1, \dots, Y_n\}$ 是来自如下 p 阶自回归模型的观测:

$$\Phi(B)Y_t = \varepsilon_t, \quad (1.1)$$

其中 $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$, B 表示后移算子; $\varphi_i (i = 1, \dots, p)$ 是回归参数; ε_t 是中值为零的独立噪声. 为了平稳性, 要求 $\Phi(B)$ 的根都在单位圆外.

模型(1.1)可改写为

$$Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} = \varepsilon_t.$$

记

$$\begin{aligned} X_t &= (Y_{t-1}, \dots, Y_{t-p})^T, \\ \varphi &= (\varphi_1, \dots, \varphi_p)^T, \end{aligned}$$

则模型可表示成

$$Y_t = X_t^T \varphi + \varepsilon_t. \quad (1.2)$$

φ 的 LAD 回归估计 $\hat{\varphi}$ 是下式的最小化解:

$$\sum_{t=1}^n |Y_t - X_t^T \varphi|. \quad (1.3)$$

本文需要下述符号和假设:

(A₁) 误差 ε_t 是独立同分布(i. i. d.)的, 其分布函数为 $F(x)$, 密度函数 $f(x)$ 是有界的;

(A₂) 密度函数 $f(x)$ 在其中值 0 处连续且 $f(0) > 0$;

(A₃) 存在 $\delta > 0$, 使得 $E|Y_t|^{2+\delta} < +\infty$;

(A₄) 令 F_t 是由 $\{Y_t, Y_{t-1}, \dots\}$ 生成的 σ -域, 则 ε_t 独立于 F_{t-1} .

上述假设是很一般的且在很多应用中都满足. 这里, 甚至不要求 $f(x)$ 的对称性. $f(x)$ 在 0 点处的连续性是需要的, 它似乎是对 $f(x)$ 的极小光滑性假设. 令 $Q = EX_t X_t^T$, 则 Q 是正定的.

定理 1 在假设(A₁) ~ (A₄)下, 有关于 φ 的 Bahadur 表示

$$\sqrt{n}(\hat{\varphi} - \varphi) = \frac{Q^{-1}}{f(0)} n^{-\frac{1}{2}} \sum_1^n X_t \psi(\varepsilon_t) + o_p(1),$$

其中 $\psi(x) = \frac{1}{2} - I(x < 0)$, $I(x < 0)$ 是集合 $\{x: x < 0\}$ 的示性函数. 进一步, 有

$$\sqrt{n}(\varphi - \hat{\varphi}) \xrightarrow{D} N\left(0, \frac{Q^{-1}}{4f^2(0)}\right).$$

文献[13]得到了自回归模型的广义 M-估计的 Bahadur-Kiefer 表示, 它包括了 LAD 估计, 但是要求八阶以上的矩. 这在具有重尾误差分布的实际问题中是非常有限的. 与此相比较, 定理 1 不需要 $f(x)$ 在 $x=0$ 处的 Lipschitz 连续性, 同时要求的条件弱得多.

注意到 LAD 估计是 \sqrt{n} -相合并渐近正态的, 人们可能对如下假设检验感兴趣: $(H_0) A\varphi=0$, 其中 A 是一个 $m \times p$ 矩阵. 由定理 1, 在 (H_0) 和条件 $(A_1) \sim (A_4)$ 下, 有

$$T_n = 4nf^2(0)\varphi^T A^T Q A \varphi \xrightarrow{D} \chi^2(m).$$

利用文献[14], 可构造自回归系数的一个稳健检验(在此不详述). 将集中考虑上述 L_1 -回归拟合模型的稳健拟合优度检验. 上述 Bahadur 表示将用于导出下节引入的稳健残差自相关的渐近表示.

2 稳健残差自相关和稳健多用途统计量的渐近性质

在拟合一个即时模型后, 需要评估拟合模型的合适性. 残差自相关是评估模型的一个有用工具, 因为它能用于发现指定模型误差偏离白噪声扰动这一关键假设. 文献[15]使用文献[3]提出的残差自协方差估计获得了 ARMA 模型稳健的拟合优度检验. 受这一检验的启发, 考虑 L_1 -拟合的拟合优度检验. 将残差表示为 $\varepsilon_t = Y_t - X_t^T \varphi$, 并定义时延为 k 的样本 L_1 -残差自相关为

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \psi(\varepsilon_t) \psi(\varepsilon_{t-k})}{\sum_{t=k+1}^n \psi^2(\varepsilon_t)},$$

其中 $\psi(x)$ 与定理 1 相同; 表示时延为 k 的样本 L_1 -噪声自相关为

$$r_k = \frac{\sum_{t=k+1}^n \psi(\varepsilon_t) \psi(\varepsilon_{t-k})}{\sum_{t=k+1}^n \psi^2(\varepsilon_t)},$$

那么, 有下列渐近表示:

定理 2 在条件 $(A_1) \sim (A_4)$ 下, 有

$$\hat{r}_k = r_k - 4f(0) W_k^T (\varphi - \hat{\varphi}) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

其中 $W_k = E \psi(\varepsilon_{t-k}) X_t$. 进一步, 对任一给定的正整数 M , 如果令

$$W = (W_1, \dots, W_M)^T, r = (r_1, \dots, r_M)^T, \hat{r} = (\hat{r}_1, \dots, \hat{r}_M)^T,$$

则

$$\hat{r} = r - 4f(0) W (\varphi - \hat{\varphi}) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

现在研究 L_1 -残差自相关的渐近分布. 注意到 $\sqrt{n}(\varphi - \hat{\varphi})$ 是渐近正态的. 由定理 1 和 2, 使用 Mann-Wald 工具及鞅中心极限定理, 得

定理 3 在假设 $(A_1) \sim (A_4)$ 下, \sqrt{nr} 是渐近零均值协方差 $V = I - 4WQ^{-1}W^T$ 的正态分布. 因此, $G_M = nr^T V^{-1}r$ 是渐近 M 个自由度的 χ^2 分布.

注 1 注意到 \sqrt{nr} 的渐近协方差 V 能被任一相合估计取代而不改变 G_M 的渐近分布, 因此, 定义稳健的多用途统计量为 $G_M = nr^T V^{-1}r$, 其中 V 是 V 的相合估计, 则 G_M 也是渐近 M 个自由度的 χ^2 分布.

注 2 在应用中, V 的一个方便选择是相应的样本平均, 即 W 和 Q 的元素由样本平均取代. 更精确地, 令

$$Q = n^{-1} \sum_{t=1}^n X_t X_t^T,$$

$$W_k = n^{-1} \sum_{t=k+1}^n X_t \psi(\epsilon_{t-k}),$$

$$W = (W_1, \dots, W_M)^T,$$

则易证 $V = I - 4WQ^{-1}W^T$ 是 V 的相合估计. 显然, 为了达到稳健性, 可以使用切尾均值代替上述平均值.

稳健的多用途统计量 G_M 为检验 $AR(p)$ 模型的拟合优度提供了一个有用的工具. 下一节中, 将使用经验结果来说明其具体应用.

3 经验结果

3.1 对 RESEX 数据的应用

RESEX 序列由 89 个观测组成, 是加拿大某一区域居民电话分机 1966 年 1 月至 1973 年 5 月的内部移动次数, 有两个观测是异常值^[16].

文献[16]在对 RESEX 序列进行季节性差分后, 采用 $AR(2)$ 模型拟合. 表 1 给出了该模型使用 L_1 -回归和文献[16]提出的 ACM-型稳健滤波方法的参数估计. 可以看出, 相对于 Box-Jenkins 最小二乘来说, L_1 -范数拟合大大地校正了参数 φ_1 , 而很小改变参数 φ_2 . 使用注 2 中提到的 5% 的切尾均值, 相应的稳健多用途统计量为 $G_5 = 1.770 1$, $G_6 = 2.467 0$, $G_7 = 2.472 0$, 参考其渐近分布, 它们都不显著. 表 1 也给出了残差和绝对残差的中值及残差的经验标准误, 可见 L_1 -范数估计对异常值是稳健的.

表 1 对 RESEX 数据的相关结果^{a)}

	L_1 -回归	Box-Jenkins	Martin 等
φ_1	0.511	0.537	0.522
φ_2	-0.104	-0.106	0.350
MR	19.942 4	31.69	73.922
$MAR \times 10^3$	1.038 6	1.053 5	0.904 4
$StdR \times 10^3$	6.313	6.327	7.316

a) 表中 Box-Jenkins 为无条件后向预测最小二乘估计, Martin 等为 ACM-型稳健滤波方法, MR 为残差的中值, MAR 为绝对残差的中值, StdR 为残差的样本标准误

3.2 模拟

为了评估本文的方法, 这里做一个小的模拟实验. 具体地, 考虑 $AR(2)$ 模型

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad t = 1, \dots, 100, \quad (3.1)$$

其中 $\varphi = (-1/6, 1/3)^T$.

考虑三类噪声变量集, 即自由度为 3 的学生 t 分布, 混合正态分布 $(0.95N(0, 1^2) + 0.05N(0, 3^2))$ 及标准正态分布. 对每一误差分布和样本容量, 抽取 100 个独立样本. 使用 L_1 -范数估计参数 φ . 表 2 给出了 100 次模拟的参数估计的中值. 可以看出, L_1 -范数估计是精确的, 且对误差分布是稳健的. 表 3 给出了 $G_M(M=6)$ 与 χ^2 分布的分位数相比较的频数.

表 2 100 次模拟中 AR(2)模型($\varphi_1 = -1/6, \varphi_2 = 1/3$)参数的 L_1 -估计

误差分布	φ_1 的中值	φ_2 的中值
标准正态	-0.183 380 6	0.313 989 2
混合正态	-0.167 608 6	0.305 046 2
学生 $t(3)$	-0.163 346 1	0.321 043 1

表 3 100 次模拟中经验多用途统计量的频数

误差分布	$G_M \geq \chi_{6,0.05}^2$	$G_M \geq \chi_{6,0.1}^2$
标准正态	9	13
混合正态	6	9
学生 $t(3)$	7	13

图 1~3 描述了在 100 次模拟中 G_M 统计量的 $Q-Q$ 图. 这些概率图表明 G_M 渐近服从自由度为 6 的 χ^2 分布. 从而, 多用途检验支持了 L_1 -范数拟合的模型的合适性.

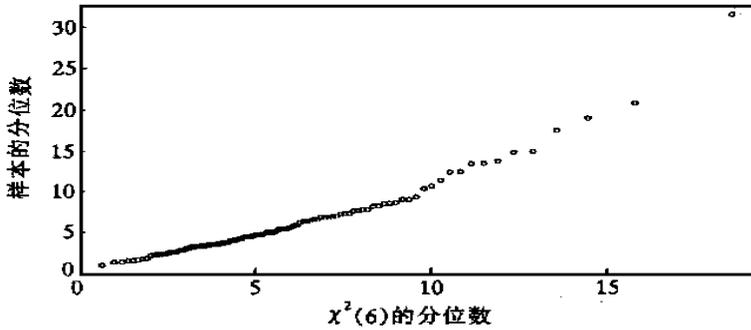


图 1 对正态误差, 统计量 $G_M(M=6)$ 的 $Q-Q$ 图

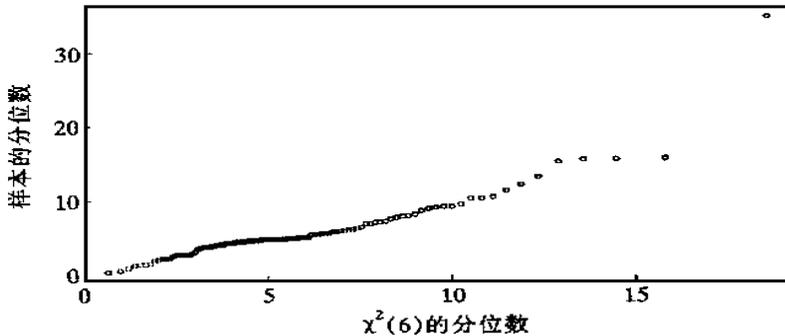


图 2 对混合正态误差, 统计量 $G_M(M=6)$ 的 $Q-Q$ 图

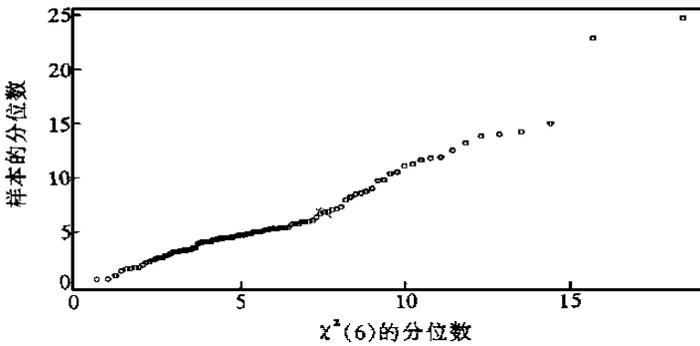


图 3 对 $t(3)$ 误差, 统计量 $G_M (M=6)$ 的 $Q-Q$ 图

参 考 文 献

- 1 Martin R D. Robust estimation of autoregressive models (with discussion). In: Brillinger D R, Tiao G C, eds. Directions in Time Series. Hayward: Inst Math Statist Pub, 1980. 228 ~ 262
- 2 Basawa I V, Huggins R M, Staudte R G. Robust tests for time series with an application to first-order autoregressive processes. Biometrika 1985, 72: 559 ~ 571
- 3 Bustos P H, Yohai V J. Robust estimates for ARMA models. J Amer Statist Assoc. 1986, 81: 155 ~ 168
- 4 Martin R D, Yohai V J. Robustness in time series and estimating ARMA models. In: Brillinger D R, Krishnaiah P R, eds. Handbook of Statistics, 4. New York; Elsevier, 1985
- 5 Birkes D, Dodge Y. Alternative Methods of Regression. New York: Wiley, 1993. 191 ~ 192
- 6 Cade B S, Richards J D. Permutation tests for least absolute deviation regression. Biometrics 1996, 52: 886 ~ 902
- 7 An H Z, Chen Z G. On convergence of LAD estimates in autoregression with infinite variance. J Multivariate Anal. 1982, 12: 335 ~ 345
- 8 Dunsmuir W T, Spencer N M. Strong consistency and asymptotic normality L_1 estimates of the autoregressive moving-average models. J Time Series Anal 1991, 12: 95 ~ 104
- 9 Rao C R. Methodology based on the L_1 -norm in statistical inference. Sankhya A, 1988, 50: 289 ~ 313
- 10 陈希孺, 白志东, 赵林城, 等. 线性模型中最小一乘估计的渐近正态性. 中国科学, A 辑, 1990, (5): 449 ~ 463
- 11 De Angelis D, Hall P, Young G A. Analytical and bootstrap approximations to estimator distributions in l_1 regression. J Amer Statist Assoc. 1993, 88: 1 310 ~ 1 316
- 12 Box G E P, Pierce D A. Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. J Amer Statist Assoc, 1970, 65: 1 509 ~ 1 526
- 13 Koul H L, Zhu Z. Bahadur-Kiefer representations for GM-estimators in autoregression models. Stochastic Processes and Their Applications. 1995, 57: 167 ~ 189
- 14 Koenker R, Bassett G. Tests of linear hypotheses and l_1 estimation. Econometrica 1982, 50: 1 577 ~ 1 583
- 15 Li W K. A goodness-of-fit test in robust time series modelling. Biometrika 1988, 75: 355 ~ 361
- 16 Martin R D, Samarov A, Vandaele W. Robust methods for ARIMA models. In: Zellner A, ed. Applied Time Series Analysis of Economic Data. Washington: US Bureau of Census, 1983. 153 ~ 169
- 17 Koenker R, Zhao Q. Conditional quantile estimation and inference for ARCH models. Econometric Theory, 1996, 12: 793 ~ 813
- 18 Ruppert D, Carroll R J. Trimmed least square estimation in the linear model. J Amer Statist Assoc. 1980, 75: 828 ~ 838
- 19 Stout W F. Almost Sure Convergence. New York; Wiley, 1974

20 Bickel P J. One step Huber estimates in linear models. J Amer Statist Assoc. 1975, 70: 428 ~ 433

附录 A

这里给出定理 1 和 2 的证明.

令 L 是任一给定的正整数, E_{t-1} 表示关于 F_{t-1} 的条件期望. 记 $\Delta_n = \sqrt{n}(\varphi - \varphi)$, $\varphi(\Delta) = \varphi + n^{-\frac{1}{2}}\Delta$, $\epsilon_t(\Delta) = Y_t - X_t^T\varphi(\Delta) = \epsilon_t - n^{-\frac{1}{2}}X_t^T\Delta$, $\xi_t(\Delta) = I(\epsilon_t < 0) - I(\epsilon_t(\Delta) < 0)$, $V_n(\Delta) = 4n^{-\frac{1}{2}} \sum_{k=1}^n \psi(\epsilon_t(\Delta)) \psi(\epsilon_{t-k}(\Delta))$, 则 $\varphi(\Delta_n) = \varphi$, $\epsilon_t(\Delta_n) = \epsilon_t$, $V_n(0) = \sqrt{nr_k}$, $V_n(\Delta_n) = \sqrt{nr_k}$.

引理 假定条件(A1) ~ (A4)成立, 则

$$J_n \doteq \sup_{\|\Delta\| \leq L} \left| n^{-\frac{1}{2}} \sum_{k=1}^n \psi(\epsilon_{t-k}(\Delta)) E_{t-1} \xi_t(\Delta) + n^{-1} f(0) \sum_{k=1}^n \psi(\epsilon_{t-k}) X_t^T \Delta \right| = o_p(1).$$

证 取 $\xi_t(\Delta)$ 关于 F_{t-1} 的条件期望, 得

$$E_{t-1} \xi_t(\Delta) = F(0) - F(n^{-\frac{1}{2}} X_t^T \Delta), \tag{A1}$$

利用中值定理, 有

$$E_{t-1} \xi_t(\Delta) = - f(-\theta n^{-\frac{1}{2}} X_t^T \Delta) n^{-\frac{1}{2}} X_t^T \Delta, \tag{A2}$$

其中 $\theta \in (0, 1)$. 注意到 $n^{-\frac{1}{2}} \max_{t \leq n} \|X_t\| = o_p(1)$ 及 $f(x)$ 在零点的连续性, 有

$$f(-\theta n^{-\frac{1}{2}} X_t^T \Delta) \xrightarrow{P} f(0)$$

关于 $\|\Delta\| \leq L$ 和 $t \leq n$ 一致成立. 再由条件(A3)和(A2)式, 可得

$$\sup_{\|\Delta\| \leq L} \left| n^{-\frac{1}{2}} \sum_{k=1}^n \psi(\epsilon_{t-k}(\Delta)) [E_{t-1} \xi_t(\Delta) + f(0) n^{-\frac{1}{2}} X_t^T \Delta] \right| = o_p(1). \tag{A3}$$

因为

$$\sup_{\|\Delta\| \leq L} |\psi(\epsilon_{t-k}(\Delta)) - \psi(\epsilon_{t-k})| = \sup_{\|\Delta\| \leq L} |I(\epsilon_{t-k} < n^{-\frac{1}{2}} X_{t-k}^T \Delta) - I(\epsilon_{t-k} < 0)| \leq I\left(|\epsilon_{t-k}| < n^{-\frac{1}{2}} L \max_{t \leq n} \|X_t\|\right),$$

则由(A3)式, 有

$$\begin{aligned} J_n &\leq \sup_{\|\Delta\| \leq L} \left| n^{-1} \sum_{k=1}^n [\psi(\epsilon_{t-k}) - \psi(\epsilon_{t-k}(\Delta))] f(0) X_t^T \Delta \right| + \\ &\sup_{\|\Delta\| \leq L} \left| n^{-\frac{1}{2}} \sum_{k=1}^n \psi(\epsilon_{t-k}(\Delta)) [E_{t-1} \xi_t(\Delta) + f(0) n^{-\frac{1}{2}} X_t^T \Delta] \right| \leq \\ &f(0) L n^{-1} \sum_{k=1}^n \|X_t\| I\left(|\epsilon_{t-k}| < L n^{-\frac{1}{2}} \max_{t \leq n} \|X_t\|\right) + o_p(1) = o_p(1). \end{aligned}$$

定理 1 的证 该证明是文献[17] 的方法的修正. 记

$$V^*(\Delta) = n^{-\frac{1}{2}} \sum_1^n X_t \psi(\epsilon_t(\Delta)). \tag{A4}$$

注意到 φ 是下述函数的极小化解:

$$\sum_1^n |Y_t - X_t^T \varphi|, \tag{A5}$$

其中 X_t 是 F_{t-1} -可测的. 因此, 利用文献[18] 的引理 A. 2 可得

$$\|V^*(\Delta_n)\| = \left\| n^{-\frac{1}{2}} \sum_1^n X_t \psi(Y_t - X_t^T \varphi) \right\| \leq \dim(X_t) n^{-\frac{1}{2}} \max_{t \leq n} \|X_t\| = o_p(1). \tag{A6}$$

其次, 由文献[19]定理 3.5.8 可知, X_t 是平稳遍历的且有 $2+\delta$ 矩. 因此

$$n^{-1} \sum_1^n X_t X_t^T \xrightarrow{P} E X_t X_t^T = Q. \tag{A7}$$

然后, 由文献[17]引理 A.2 可得

$$\sup_{\|\Delta\| \leq L} \|V^*(\Delta) - V^*(0) + f(0)Q\Delta\| = o_p(1). \tag{A8}$$

因为 $\sum_1^n |\epsilon_t - \lambda n^{-\frac{1}{2}} X_t^T \Delta|$ 是关于 λ 的凸函数, 所以它关于 λ 的梯度

$$-\Delta^T V^*(\lambda\Delta) = n^{-\frac{1}{2}} \sum_1^n (-X_t^T \Delta) \psi(\epsilon_t - \lambda n^{-\frac{1}{2}} X_t^T \Delta)$$

关于 λ 是非降的. 应用文献[17]引理 A.4 可得, $\|\Delta_n\| = O_p(1)$ 和

$$\Delta_n = \frac{Q^{-1}}{f(0)} V^*(0) + o_p(1). \tag{A9}$$

由鞅中心极限定理和 Slutsky 定理, 定理 1 成立.

定理 2 的证 注意到

$$V_n(\Delta) - V_n(0) = 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) \xi_t(\Delta) + 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_t) \xi_{t-k}(\Delta) = : I_{n1}(\Delta) + I_{n2}(\Delta). \tag{A10}$$

取期望可得

$$E I_{n2}(\Delta) = 0 \tag{A11}$$

$$\begin{aligned} \text{var} I_{n2}(\Delta) &= 16n^{-1} E \left\{ \sum_{k+1}^n \psi(\epsilon_t) \xi_{t-k}(\Delta) \right\}^2 = 4n^{-1} \sum_{k+1}^n E \xi_{t-k}^2(\Delta) \leq \\ &4n^{-1} \sum_{k+1}^n E I \left(|\epsilon_{t-k}| \leq n^{-\frac{1}{2}} \mid X_{t-k}^T \Delta \right) \leq \\ &4n^{-1} \sum_{k+1}^n E [\|f\|_{\infty} 2Ln^{-\frac{1}{2}} \|X_{t-k}\|] = o(1), \end{aligned} \tag{A12}$$

其中 $\|f\|_{\infty} = \sup_x |f(x)|$. 由(A11)和(A12)式得 $I_{n2}(\Delta) = o_p(1)$.

其次, 使用文献[20]的链方法, 可证明

$$\sup_{\|\Delta\| \leq L} |I_{n2}(\Delta)| = o_p(1). \tag{A13}$$

由(A10)和(A13)式有

$$\sup_{\|\Delta\| \leq L} |V_n(\Delta) - V_n(0) - I_{n1}(\Delta)| = o_p(1). \tag{A14}$$

注意到

$$\begin{aligned} I_{n1}(\Delta) &= 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) [\xi_t(\Delta) - E_{t-1} \xi_t(\Delta)] + 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) E_{t-1} \xi_t(\Delta) = : \\ &T(\Delta) + 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) E_{t-1} \xi_t(\Delta). \end{aligned} \tag{A15}$$

取期望可得

$$\begin{aligned} E T^2(\Delta) &= 16n^{-1} E \left[\sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) (\xi_t(\Delta) - E_{t-1} \xi_t(\Delta)) \right]^2 = \\ &4n^{-1} \sum_{k+1}^n E (\xi_t(\Delta) - E_{t-1} \xi_t(\Delta))^2 \leq 4n^{-1} \sum_{k+1}^n E \xi_t^2(\Delta) = o(1), \end{aligned}$$

其中最后的等式由和(A12)式一样的方法可得. 因此 $T(\Delta) = o_p(1)$.

利用类似(A13)式的链方法, 可证

$$\sup_{\|\Delta\| \leq L} |T(\Delta)| = o_p(1). \quad (\text{A16})$$

结合(A14)和(A15)式可得

$$\sup_{\|\Delta\| \leq L} \left| V_n(\Delta) - V_n(0) - 4n^{-\frac{1}{2}} \sum_{k+1}^n \psi(\epsilon_{t-k}(\Delta)) E_{t-1} \xi_t(\Delta) \right| = o_p(1). \quad (\text{A17})$$

由引理 1, 得

$$\sup_{\|\Delta\| \leq L} \left| V_n(\Delta) - V_n(0) + 4n^{-1} f(0) \sum_{k+1}^n \psi(\epsilon_{t-k}) X_t^T \Delta \right| = o_p(1). \quad (\text{A18})$$

注意到定理 1 蕴涵 $\|\Delta_n\| = O_p(1)$, 有

$$V_n(\Delta_n) = V_n(0) - 4f(0) n^{-1} \sum_{k+1}^n \psi(\epsilon_{t-k}) X_t^T \Delta_n + o_p(1). \quad (\text{A19})$$

因为 $\{\epsilon_t\}$ 是一独立同分布过程, 由文献[19]的定理 3.5.8 可知 $\psi(\epsilon_{t-k}) X_t$ 是平稳遍历的, 故由条件(A₂)和遍历定理, 得到 $n^{-1} \sum_{k+1}^n \psi(\epsilon_{t-k}) X_t = E \psi(\epsilon_{t-k}) X_t + o_p(1) = W_k + o_p(1)$. 这与(A19)式给出定理的结论.