Econometric Theory, **23**, 2007, 852–879. Printed in the United States of America.
DOI: 10.1017/S0266466607070363

WEIGHTED LEAST ABSOLUTE DEVIATIONS ESTIMATION FOR ARMA MODELS WITH INFINITE VARIANCE

JIAZHU PAN
Peking University
and
London School of Economics

HUI WANG
Peking University

QIWEI YAO
Peking University
and
London School of Economics

For autoregressive moving average (ARMA) models with infinite variance innovations, quasi-likelihood-based estimators (such as Whittle estimators) suffer from complex asymptotic distributions depending on unknown tail indices. This makes statistical inference for such models difficult. In contrast, the least absolute deviations estimators (LADE) are more appealing in dealing with heavy tailed processes. In this paper, we propose a weighted least absolute deviations estimator (WLADE) for ARMA models. We show that the proposed WLADE is asymptotically normal, is unbiased, and has the standard root-n convergence rate even when the variance of innovations is infinity. This paves the way for statistical inference based on asymptotic normality for heavy-tailed ARMA processes. For relatively small samples numerical results illustrate that the WLADE with appropriate weight is more accurate than the Whittle estimator, the quasi-maximum-likelihood estimator (QMLE), and the Gauss–Newton estimator when the innovation variance is infinite and that the efficiency loss due to the use of weights in estimation is not substantial.

1. INTRODUCTION

Let $\{y_t\}$ be a stationary autoregressive moving average (ARMA) time series generated by the equation

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{1.1}$$

The authors thank the two referees for their valuable suggestions. The work was partially supported by an EPSRC research grant (UK) and the Natural Science Foundation of China (grant 10471005). Address correspondence to Jiazhu Pan, School of Mathematical Sciences, Peking University, Beijing 100871, China; e-mail: jzpan@math.pku.edu.cn.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

where the innovation process $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables and $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$ is an unknown parameter vector. When $E\varepsilon_t^2 < \infty$, it is well known that various estimators such as the maximum likelihood estimator (MLE), Whittle estimators, and least squares estimators (LSE) for β are all asymptotically normal and unbiased (Brockwell and Davis, 1991). Furthermore, the least absolute deviations estimator (LADE) is also asymptotically normal; see Dunsmuir and Spencer (1991) and Davis and Dunsmuir (1997). When $E\varepsilon_t^2 = \infty$, model (1.1) is called the infinite variance ARMA (IVARMA) model, which defines a heavy-tailed process $\{y_t\}$. The IVARMA models are pertinent in modeling heavy-tailed time series data often encountered in, for example, economics and finance (Koedijk, Schafgans, and De Vries, 1990; Jansen and de Vries, 1991). For further references on statistical modeling for heavy-tailed phenomena, we refer to Resnick (1997) and Adler, Feldaman, and Taqqu (1997).

Statistical inference for IVARMA models has not been well explored yet. Most available results concern infinite variance autoregressive (IVAR) models (i.e., $q = 0$ in (1.1)). Gross and Steiger (1979) and An and Chen (1982) obtained the strong consistency and the convergence rates for LADE for IVAR models. Davis and Resnick (1985, 1986) derived the limiting distributions of the LSE for IVAR models. A more comprehensive asymptotic theory of *M*-estimators for autoregressive (AR) models was derived by Davis, Knight, and Liu (1992). For IVARMA models, the asymptotic properties for several estimators have been derived when the innovation distribution is in the domain of attraction of a stable law distribution with index between 0 and 2. For example, both the Whittle estimator proposed by Mikosch, Gadrich, and Adler (1995) and the Gauss–Newton estimator proposed by Davis (1996) converge in distribution to some functions of a sequence of stable random variables. Furthermore, it has been proved that the *M*-estimator converges in distribution to the minimizer of a stochastic process; see Davis (1996) and Calder and Davis (1998). Kokoszka and Taqqu (1996) extended the results of Mikosch et al. (1995) to fractional ARMA models with infinite variance. However, all the preceding limiting distributions are complicated and depend intimately on the unknown tail indices of the underlying processes. This makes it difficult to develop asymptotic approximations for the purpose of statistical inference. This paper provides a remedy for this problem.

The difficulty of the asymptotic theory for LADE for IVARMA processes may at least partially result from the fact that the residual of a linear prediction for y_t based on its lagged values depends on β *nonlinearly*, whereas such a dependence is completely linear for pure AR processes. Note that this linearity implies that the objective function for least absolute deviations estimation is convex and therefore the asymptotic normality of LADE may be readily derived from the convex lemma (Hjort and Pollard, 1993). One way to deal with a nonconvex objective function is to adopt a local

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

linearization around the true value of the parameter, which enables one to establish asymptotic properties of a local estimator defined as a local minimizer around the true value of the parameter. This is the line taken in the paper by Davis and Dunsmuir (1997), which dealt with LADE for ARMA models with $E\varepsilon_t^2 < \infty$. On the other hand, Ling (2005) proposed a weighted least absolute deviations estimator (WLADE) for IVAR models. The key idea of the WLADE is to weigh down the observations that are excessively large, either positively or negatively. Ling (2005) showed that the WLADE is asymptotically normal. The idea of weighing down the large observations has also been used in estimation for autoregressive conditionally heteroskedastic (ARCH) models with heavy tailed innovations by Horvath and Liese (2004).

In this paper, we deal with the WLADE for IVARMA models. By adopting the idea of local linearization mentioned previously, we show that a local WLADE is asymptotic normal and unbiased under the condition that $E|\varepsilon_t|^\delta < +\infty$ for some $\delta > 0$ and the density function of ε_t and its derivative are bounded. This facilitates statistical inference for IVARMA models (even when $E|\varepsilon_t| = \infty$) in a conventional fashion. For example, a Wald test for a linear hypothesis can be constructed; see Section 2. For relatively small samples a simulation study indicates that the proposed WLADE is more accurate than the Whittle estimator, the quasi-maximum-likelihood estimator (QMLE), and the Gauss–Newton estimator when $\mathrm{Var}(\varepsilon_t) = \infty$. Furthermore, the efficiency loss of the WLADE with respect to the (unweighted) LADE is not significant with appropriately selected weights. Because the WLADE converges at a slower rate than the Whittle estimator and the Gauss–Newton estimator, we also studied the large-sample properties of the WLADE numerically.

Although we only deal with IVARMA models in this paper, the basic idea of combining a weighted objective function with local linearization of the residuals may apply to other infinite variance time series models, such as the infinite variance autoregressive integrated moving average (ARIMA) and the integrated generalized autoregressive conditionally heteroskedastic (IGARCH), which are popular in financial econometrics. Another open problem is to develop appropriate methods for choosing weight functions; see Remark 3 in Section 2.2.

The rest of paper is organized as follows. The WLADE and the associated asymptotic properties are presented in Section 2. In addition to showing the asymptotic normality of the local WLADE, we also show that a (global) estimator sharing the same asymptotic property could be obtained by minimizing a convex objective function if there is available an initial estimator within root-n distance from the true value; see Theorem 2 in Section 2.2. Section 3 gives all the theoretical proofs of the results in Section 2. Section 4 reports some numerical results from a simulation study.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

2. WLADE AND ITS ASYMPTOTIC PROPERTIES

2.1. Weighted Least Absolute Deviations Estimators

Denote by $\Theta \subset R^{p+q}$ the parameter space that contains the true value $\beta_0 = (\phi_1^0, \ldots, \phi_p^0, \theta_1^0, \ldots, \theta_q^0)'$ of the parameter β as an inner point. For $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$, put

$$\varepsilon_t(\beta) = \begin{cases} y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1}(\beta) - \cdots - \theta_q \varepsilon_{t-q}(\beta), \\ \qquad \text{if } t > 0, \\ 0, \quad \text{otherwise}, \end{cases} \tag{2.1}$$

where $y_t \equiv 0$ for all $t \leq 0$. Note that $\varepsilon_t \neq \varepsilon_t(\beta_0)$ because of this truncation.

We define the objective function as

$$W_n(\beta) = \sum_{t=u+1}^{n} \widetilde{w}_t |\varepsilon_t(\beta)| \tag{2.2}$$

and the WLADE as

$$\hat{\beta} = \arg\min_{\beta} W_n(\beta), \tag{2.3}$$

where $u = u(n)$ is a positive integer and the weight function, depending on a constant $\alpha > 2$, is defined as

$$\widetilde{w}_t = \left(1 + \sum_{k=1}^{t-1} k^{-\alpha} |y_{t-k}|\right)^{-2}. \tag{2.4}$$

2.2. Asymptotic Normality of WLADE

To state the asymptotic normality of $\hat{\beta}$, we introduce some notation first. Let

$$v = (v_1, \ldots, v_{p+q})' = \sqrt{n}(\beta - \beta_0).$$

It is easy to see that $\hat{\beta} = \beta_0 + \hat{v}/\sqrt{n}$, where $\hat{v}$ is a minimizer of

$$T_n(v) = \sum_{t=u+1}^{n} \widetilde{w}_t (|\varepsilon_t(\beta_0 + n^{-1/2} v)| - |\varepsilon_t(\beta_0)|).$$

Denote $A_t(\beta) = (A_{t,1}(\beta), \ldots, A_{t,p+q}(\beta))'$, where $A_{t,i}(\beta) = -\partial \varepsilon_t(\beta)/\partial \beta_i$. By (8.11.9) of Brockwell and Davis (1991), it holds for $t \geq \max(p,q)$ that

$$\begin{cases} \theta(B) A_{t,i}(\beta) = y_{t-i}, & i = 1, \ldots, p \\ \theta(B) A_{t,i+p}(\beta) = \varepsilon_{t-i}(\beta), & i = 1, \ldots, q, \end{cases} \tag{2.5}$$

where B is the backshift operator.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

For $t = 0, \pm 1, \pm 2, \ldots$, define

$$U_t - \sum_{i=1}^{p} \phi_i^0 U_{t-i} = \varepsilon_t, \qquad V_t + \sum_{j=1}^{q} \theta_j^0 V_{t-j} = \varepsilon_t. \tag{2.6}$$

Put $Q_t = (U_{t-1}, \ldots, U_{t-p}, V_{t-1}, \ldots, V_{t-q})'$, $w_t \equiv (1 + \sum_{k=1}^{\infty} k^{-\alpha} |y_{t-k}|)^{-2}$, and

$$\Sigma = E(w_t Q_t Q_t'), \qquad \Omega = E(w_t^2 Q_t Q_t'). \tag{2.7}$$

We denote by $\|v\|$ the euclidean norm for a vector v.

Some regularity conditions are now in order.

A1. For $\beta \in \Theta$, the polynomials $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ have no common zeros, and all roots of $\phi(z)$ and $\theta(z)$ are outside the unit circle.

A2. Innovation ε_t has zero median and a differentiable density function $f(x)$ satisfying the conditions $f(0) > 0$, $\sup_{x \in R} |f(x)| < B_1 < \infty$, and $\sup_{x \in R} |f'(x)| < B_2 < \infty$. Furthermore, $E|\varepsilon_t|^{\delta} < +\infty$ for some $\delta > 0$, and $\alpha > \max\{2, 2/\delta\}$.

A3. As $n \to \infty$, $u \to \infty$ and $u/n \to 0$.

The following proposition indicates that model (1.1) has a unique strictly stationary and ergodic solution under conditions A1 and A2.

Proposition 1. *Suppose that condition A1 holds and* $E|\varepsilon_t|^{\delta} < +\infty$ *for some* $\delta > 0$. *Then model (1.1) defines a unique strictly stationary and ergodic process* $\{y_t\}$.

Proof. Condition A1 implies that

$$\phi^{-1}(z)\theta(z) = \sum_{j=0}^{\infty} \psi_j z^j,$$

where $|\psi_j| \le Cr^j$ for some constants $C > 0$ and $0 < r < 1$. Let $\tilde{\delta} = \min\{\delta, 1\}$. Then

$$\sum_{j=0}^{\infty} |\psi_j|^{\tilde{\delta}} E|\varepsilon_{t-j}|^{\tilde{\delta}} < \infty.$$

The same argument as for Proposition 13.3.2 of Brockwell and Davis (1991) yields the result. ■

Remark 1. Condition A3 eliminates asymptotically the bias in the estimation resulting from the lack of observations y_t for $t \le 0$.

Remark 2. Condition A2 does not rule out the possibility that $E|\varepsilon_t| = \infty$. The purpose of introducing weights $\tilde{w}_t$ is to weigh down excessively large observations that reflect the heavy-tailed innovation distribution. Therefore the asymptotic covariance matrix of the normalized WLADE, depending on Σ and Ω given in (2.7), is a well-defined (finite) matrix. Note that $\tilde{w}_t \in (0, 1]$. Conditions A1 and A2 imply that for $\tilde{\delta} = \min\{\delta, 1\}$,

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

$$E\left(\sum_{k=1}^{\infty} k^{-\alpha}|y_{t-k}|\right)^{\bar{\delta}} \le \sum_{k=1}^{\infty} k^{-\alpha\bar{\delta}} E|y_{t-k}|^{\bar{\delta}} < +\infty.$$

Hence $\sum_{k=1}^{\infty} k^{-\alpha}|y_{t-k}| < \infty$ with probability one, which ensures that w_t is well defined. Note that w_t is stationary and ergodic under condition A1 and it is asymptotically equivalent to $\tilde{w}_t$ for $t > u$ (see A3).

We are now ready to state our main results.

THEOREM 1. *Let conditions A1–A3 hold. For any given positive random variable M with* $P(0 < M < \infty) = 1$, *there exists a local minimizer* $\hat{\beta}$ *of* $W_n(\beta)$ *that lies in the random region* $\{\beta : \|\beta - \beta_0 - \xi/\sqrt{n}\| \le M/\sqrt{n}\}$ *for which*

$$\sqrt{n}(\hat{\beta} - \beta_0) \to_{\mathcal{L}} N\left(0, \frac{1}{4f^2(0)} \Sigma^{-1}\Omega\Sigma^{-1}\right),$$

where ξ *is a normal random vector with mean 0 and covariance matrix* $(1/4f^2(0))\Sigma^{-1}\Omega\Sigma^{-1}$.

Notice that the lack of convexity for the objective function $W_n(\beta)$ complicates the search for its minimizer. As in Davis and Dunsmuir (1997), $W_n(\beta)$ may be linearized in a neighborhood of a *good* initial estimate $\hat{\beta}^0$ as follows:

$$\widetilde{W}_n(\beta) = \sum_{t=u+1}^{n} \tilde{w}_t |\varepsilon_t(\hat{\beta}^0) - A_t'(\hat{\beta}^0)(\beta - \hat{\beta}^0)|.$$

The resulting estimator $\tilde{\beta} = \arg\min_{\beta} \widetilde{W}_n(\beta)$ shares the same asymptotic property as the local WLADE. See Theorem 2, which follows.

THEOREM 2. *Let conditions A1–A3 hold. Then*

$$\sqrt{n}(\tilde{\beta} - \beta_0) \to_{\mathcal{L}} N\left(0, \frac{1}{4f^2(0)} \Sigma^{-1}\Omega\Sigma^{-1}\right),$$

provided that $\hat{\beta}^0 = \beta_0 + O_p(n^{-1/2})$.

Remark 3. Although we only deal with the weight function defined in (2.4) explicitly, the preceding theorem holds for general weight function $g_t \equiv g(y_{t-1}, y_{t-2}, \ldots)$ provided

$$E\{(g_t + g_t^2)(\xi^2 + \xi^3)\} < \infty, \qquad E\{(g_t + g_t^2)(\xi_t^2 + \xi_t^3 + \xi_t^4)\} < \infty,$$

where $\xi = \sum_{i=0}^{\infty} r^i |y_{-i}|$, $\xi_t = C_0 \sum_{i=t}^{\infty} r^i |y_{t-i}|$, $0 < r < 1$, and $C_0 > 0$ are constants. For example, we may use weights of more general form:

$$g_t = \left(1 + \sum_{k=1}^{t-1} k^{-\alpha} (\log k)^d |y_{t-k}|\right)^{-\gamma}, \qquad \alpha > 2, \qquad \gamma \ge 2, \qquad d \ge 0. \qquad \textbf{(2.8)}$$

The numerical examples in Section 3 indicate that the accuracy of the WLADE is not sensitive with respect to the value of α. However the choice of $\gamma = 2$ typically leads to a better estimator than those with $\gamma > 2$, at least for model

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

(4.1). Furthermore, it seems that $d = 2,3,4$ behave almost equally well. However how to choose a weight function in general such that the resulting estimator is of certain optimality remains an open question.

2.3. A Wald Test for Linear Hypotheses

The asymptotic normality of the estimator $\hat{\beta}$ stated in Theorem 1 facilitates inference for model (1.1). For example, we may consider a general form of linear null hypothesis:

$$H_0 : \Gamma\beta_0 = \kappa,$$

where Γ is an $s \times (p+q)$ constant matrix with rank s and κ is an $s \times 1$ constant vector. A Wald test statistic may be defined as

$$Z_n = (\Gamma\hat{\beta} - \kappa)' \left\{ \Gamma \frac{1}{4n\tilde{f}^2(0)} \hat{\Sigma}^{-1}\hat{\Omega}\hat{\Sigma}^{-1}\Gamma' \right\}^{-1} (\Gamma\hat{\beta} - \kappa),$$

and we reject H_0 for large values of Z_n. In the preceding expression,

$$\hat{\Sigma} = \frac{1}{n-u} \sum_{t=u+1}^{n} \widetilde{w}_t \hat{Q}_t \hat{Q}_t', \qquad \hat{\Omega} = \frac{1}{n-u} \sum_{t=u+1}^{n} (\widetilde{w}_t^2 \hat{Q}_t \hat{Q}_t'). \tag{2.9}$$

Here $\hat{Q}_t$ is defined in the same manner as Q_t but with β_0 replaced by $\hat{\beta}$, ε_t replaced by $\varepsilon_t(\hat{\beta})$, and $y_t = 0$ for all $t \le 0$ (see (2.6) and (2.7)), and $\tilde{f}(0)$ is an estimate for $f(0)$ defined as

$$\tilde{f}(0) = \frac{1}{b_n} \sum_{t=u+1}^{n} \widetilde{w}_t K\left(\frac{\varepsilon_t(\hat{\beta})}{b_n} \right) \Big/ \sum_{t=u+1}^{n} \widetilde{w}_t, \tag{2.10}$$

where $K(\cdot)$ is a kernel function on R and $b_n > 0$ is a bandwidth. Theorem 3, which follows, shows that the asymptotic null distribution of Z_n is χ_s^2. It in fact still holds if $\hat{\beta}$ in the definition of Z_n is replaced by $\tilde{\beta}$.

THEOREM 3. *Suppose conditions A1–A3 hold. Let kernel function K be bounded, Lipschitz continuous, and of finite first moment. Let $b_n \to 0$ and $nb_n^4 \to \infty$ as $n \to \infty$. Then $Z_n \to_{\mathcal{L}} \chi_s^2$ under H_0.*

3. PROOFS

We use the same notation as in Section 2. For any fixed $v \in R^{p+q}$, put

$$S_n(v) = \sum_{t=u+1}^{n} w_t(|\varepsilon_t(\beta_0 + n^{-1/2}v)| - |\varepsilon_t(\beta_0)|),$$

$$S_n^+(v) = \sum_{t=u+1}^{n} w_t(|\varepsilon_t - n^{-1/2}Q_t'v| - |\varepsilon_t|),$$

$$S_n^*(v) = \sum_{t=u+1}^{n} w_t(|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v| - |\varepsilon_t(\beta_0)|).$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

We linearize $\varepsilon_t(\beta)$ around β_0; that is, $\varepsilon_t(\beta)$ is approximated by

$$\varepsilon_t(\beta_0) - A_t'(\beta_0)(\beta - \beta_0),$$

where $A_t(\beta)$ is defined in (2.5).

We denote by $\to_{\mathcal{L}}$ the convergence in distribution and by $\xrightarrow{P}$ the convergence in probability. Let $C(R^s)$ be the space of the real-valued continuous functions on R^s (Rudin, 1991). For probability measures P_n and P on $C(R^s)$, we say that P_n converges weakly to P in $C(R^s)$ if $\int f dP_n \to \int f dP$ for any bounded and continuous function f defined on $C(R^s)$. For random functions S_n, S defined on $C(R^s)$, $S_n \to_{\mathcal{L}} S$ if the distribution of S_n converges weakly to that of S in $C(R^s)$ (Billingsley, 1999). The term C denotes a positive constant that may be different at different places.

3.1. Proof of Theorem 1

Before we prove Theorem 1, we first introduce a proposition that is of independent interest. Its proof is divided into several lemmas. We always assume that conditions A1–A3 hold.

PROPOSITION 2. *As $n \to \infty$, it holds that $T_n(v) \to_{\mathcal{L}} T(v)$ on $C(R^{p+q})$, where $T(v) = f(0)v'\Sigma v + v'N$ and N denotes an $N(0, \Omega)$ random vector.*

LEMMA 1. *It holds that $|\varepsilon_t - \varepsilon_t(\beta_0)| \le \xi_t$ and $|A_{t,i}(\beta_0) - Q_{t,i}| \le \xi_t$ for $i = 1, \ldots, p+q$, where $\xi_t = C_0 \sum_{j=t}^{\infty} r^j |y_{t-j}|$, $0 < r < 1$, C_0 is a positive constant, and $Q_{t,i}$ is the ith component of Q_t.*

Proof. See Brockwell and Davis (1991, pp. 265–268).

LEMMA 2. *$S_n^+(v) \to_{\mathcal{L}} T(v)$ on $C(R^{p+q})$.*

Proof. We first prove the convergence for any fixed v. Using the identity

$$|z - y| - |z| = -y sgn(z) + 2(y - z)\{I(0 < z < y) - I(y < z < 0)\},$$

which holds for $z \neq 0$, we have

$$\begin{aligned} S_n^+(v) = & -n^{-1/2} \sum_{t=u+1}^{n} w_t Q_t' v sgn(\varepsilon_t) \\ & + 2 \sum_{t=u+1}^{n} w_t (n^{-1/2} Q_t' v - \varepsilon_t) \\ & \times [I(0 < \varepsilon_t < n^{-1/2} Q_t' v) - I(n^{-1/2} Q_t' v < \varepsilon_t < 0)] \\ =: & A_n + B_n. \end{aligned}$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

Notice that, by Lemma 1, we have $|Q_{t,i}| \le C\sum_{j=1}^{\infty} r^j |y_{t-j}|$ and

$$|w_t^{1/2} Q_{t,i}| \le \frac{C\sum_{j=1}^{\infty} r^j |y_{t-j}|}{1 + \sum_{k=1}^{\infty} k^{-\alpha} |y_{t-k}|} \le C\sum_{k=1}^{\infty} k^{\alpha} r^k. \tag{3.1}$$

Then, $Ew_t^2 (Q_t' v)^2 < +\infty$. But, from conditions A1 and A2, $\{w_t Q_t' v sgn(\varepsilon_t)\}$ is a stationary martingale difference sequence. Therefore, applying a martingale central limit theorem (Hall and Heyde, 1980), we obtain $A_n \to_{\mathcal{L}} v'N$.

For B_n, let

$$W_{nt} = w_t (n^{-1/2} Q_t' v - \varepsilon_t) I(0 < \varepsilon_t < n^{-1/2} Q_t' v)$$

and $\mathcal{F}_{t-1} = \sigma(\varepsilon_j, j \le t-1)$. Then

$$\begin{aligned} nEW_{nt}^2 &= nE(E(W_{nt}^2 | \mathcal{F}_{t-1})) \\ &= nE\left(w_t^2 \left[\int_0^{n^{-1/2} Q_t' v} (n^{-1/2} Q_t' v - z)^2 (f(z) - f(0))\, dz \right.\right. \\ &\qquad \left.\left. + \int_0^{n^{-1/2} Q_t' v} (n^{-1/2} Q_t' v - z)^2 f(0)\, dz \right] \right) \\ &\le nE(w_t^2 B_2 n^{-2} (Q_t' v)^4 + w_t^2 B_1 n^{-3/2} (Q_t' v)^3). \end{aligned}$$

Similarly to (3.1), we can obtain

$$Ew_t^2 (Q_t' v)^4 < +\infty, \qquad Ew_t^2 (Q_t' v)^3 < +\infty.$$

Therefore, we have proved that

$$\limsup_{n \to \infty} nEW_{nt}^2 = 0. \tag{3.2}$$

On the other hand, on the set $\{Q_t' v > 0\}$, we may show that

$$\sum_{t=u+1}^{n} E(W_{nt} | \mathcal{F}_{t-1}) \to \frac{f(0)}{2} E[w_t (Q_t' v)^2 I(Q_t' v > 0)]$$

and

$$\mathrm{Var}\left(\sum_{t=u+1}^{n} (W_{nt} - E(W_{nt} | \mathcal{F}_{t-1})) \right) \to 0;$$

see Davis and Dunsmuir (1997). Therefore,

$$\sum_{t=u+1}^{n} W_{nt} \to \frac{f(0)}{2} E[w_t (Q_t' v)^2 I(Q_t' v > 0)]. \tag{3.3}$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

Using the same argument for the second indicator in the summands of B_n, we obtain that

$$B_n \xrightarrow{P} f(0)v'\Sigma v, \tag{3.4}$$

so that the finite-dimensional distributions of S_n^+ converge to those of T. But because S_n^+ has convex sample paths, this implies that the convergence is in fact on $C(R^{p+q})$ (see Davis and Dunsmuir, 1997, proof of Prop. 1). ■

LEMMA 3. $S_n^*(v) - S_n^+(v) \xrightarrow{P} 0$ *uniformly on compact sets.*

Proof. Notice that

$$\begin{aligned}
S_n^*(v) - S_n^+(v) &= \sum_{t=u+1}^{n} w_t[(|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v| - |\varepsilon_t(\beta_0)|) - (|\varepsilon_t - n^{-1/2}Q_t'v| - |\varepsilon_t|)] \\
&= -n^{-1/2}\sum_{t=u+1}^{n} w_t A_t'(\beta_0)v\,sgn(\varepsilon_t(\beta_0)) + n^{-1/2}\sum_{t=u+1}^{n} w_t Q_t'v\,sgn(\varepsilon_t) \\
&\quad + 2\sum_{t=u+1}^{n} w_t(n^{-1/2}A_t'(\beta_0)v - \varepsilon_t(\beta_0))[I(0 < \varepsilon_t(\beta_0) < n^{-1/2}A_t'(\beta_0)v) \\
&\qquad\qquad - I(n^{-1/2}A_t'(\beta_0)v < \varepsilon_t(\beta_0) < 0)] \\
&\quad - 2\sum_{t=u+1}^{n} \omega_t(n^{-1/2}Q_t'v - \varepsilon_t) \\
&\quad \times [I(0 < \varepsilon_t < n^{-1/2}Q_t'v) - I(n^{-1/2}Q_t'v < \varepsilon_t < 0)].
\end{aligned}$$

First, we consider

$$\begin{aligned}
\Lambda_1 &= n^{-1/2}\sum_{t=u+1}^{n} w_t[Q_t'v\,sgn(\varepsilon_t) - A_t'(\beta_0)v\,sgn(\varepsilon_t(\beta_0))] \\
&= n^{-1/2}\sum_{t=u+1}^{n} w_t Q_t'v(sgn(\varepsilon_t) - sgn(\varepsilon_t(\beta_0))) \\
&\quad + n^{-1/2}\sum_{t=u+1}^{n} w_t(Q_t'v - A_t'(\beta_0)v)sgn(\varepsilon_t(\beta_0)) \\
&=: K_1 + K_2.
\end{aligned}$$

By Lemma 1 and the proof of (3.1), we know that

$$|w_t(Q_t' - A_t'(\beta_0))v| \le C\|v\|\xi_t.$$

Then

$$K_2 \le n^{-1/2}C\|v\|\sum_{t=u+1}^{n}\sum_{j=t}^{\infty} r^j|y_{t-j}| = n^{-1/2}C\|v\|\sum_{t=u+1}^{n} r^t\sum_{h=0}^{\infty} r^h|y_{-h}| \xrightarrow{P} 0$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

uniformly on compact sets. For K_1, we have

$$
\begin{aligned}
E|K_1| &\le n^{-1/2}\sum_{t=u+1}^{n} E(E(w_t|Q_t'v|\,|sgn(\varepsilon_t)-sgn(\varepsilon_t(\beta_0))||\mathcal{F}_{t-1})) \\
&= 2n^{-1/2}\sum_{t=u+1}^{n} E(w_t|Q_t'v|P(\{\varepsilon_t>0,\varepsilon_t(\beta_0)<0\}\cup\{\varepsilon_t<0,\varepsilon_t(\beta_0)>0\}|\mathcal{F}_{t-1})) \\
&\le 2n^{-1/2}\sum_{t=u+1}^{n} E(w_t|Q_t'v|P(\{0<\varepsilon_t<\xi_t\}\cup\{-\xi_t<\varepsilon_t<0\}|\mathcal{F}_{t-1})) \\
&= 2n^{-1/2}\sum_{t=u+1}^{n}\left[E\left(w_t|Q_t'v|\int_{-\xi_t}^{\xi_t} f(x)\,dx\right)\right] \\
&\le 4B_1 n^{-1/2}\sum_{t=u+1}^{n} E(w_t^{1/2}|Q_t'v|\cdot w_t^{1/2}\xi_t) \\
&\le 4CB_1\|v\|n^{-1/2}\sum_{t=u+1}^{n}\sum_{h=0}^{\infty}(t+h)^{\alpha}r^{t+h} \\
&\le 4CB_1\|v\|n^{-1/2}\left[\sum_{t=u+1}^{n} t^{\alpha}r^{t}\sum_{h=0}^{\infty} r^{h}+\sum_{t=u+1}^{n} r^{t}\sum_{h=0}^{\infty} h^{\alpha}r^{h}\right]\to 0
\end{aligned}
$$

uniformly on compact sets, because we have the facts

$$
w_t^{1/2}|Q_t'v|\le C\|v\|,\qquad w_t^{1/2}\xi_t\le\sum_{j=t}^{\infty} r^{j}j^{\alpha},\qquad \text{and}\quad (t+h)^{\alpha}\le 2^{\alpha}(t^{\alpha}+h^{\alpha}).
$$

Therefore, $\Lambda_1 \xrightarrow{P} 0$ uniformly on compact sets as $n\to\infty$.

Now we consider

$$
\begin{aligned}
\Lambda_2 &= \sum_{t=u+1}^{n} w_t(n^{-1/2}A_t'(\beta_0)v-\varepsilon_t(\beta_0))I(0<\varepsilon_t(\beta_0)<n^{-1/2}A_t'(\beta_0)v) \\
&\quad -\sum_{t=u+1}^{n} w_t(n^{-1/2}Q_t'v-\varepsilon_t)I(0<\varepsilon_t<n^{-1/2}Q_t'v) \\
&= \sum_{t=u+1}^{n} w_t(n^{-1/2}A_t'(\beta_0)v-\varepsilon_t(\beta_0)-n^{-1/2}Q_t'v+\varepsilon_t) \\
&\quad \times I(0<\varepsilon_t(\beta_0)<n^{-1/2}A_t'(\beta_0)v) \\
&\quad +\sum_{t=u+1}^{n} w_t(n^{-1/2}Q_t'v-\varepsilon_t) \\
&\quad \times[I(0<\varepsilon_t(\beta_0)<n^{-1/2}Q_t'v)-I(0<\varepsilon_t<n^{-1/2}Q_t'v)] \\
&\quad +\sum_{t=u+1}^{n} w_t(n^{-1/2}Q_t'v-\varepsilon_t) \\
&\quad \times[I(0<\varepsilon_t(\beta_0)<n^{-1/2}A_t'(\beta_0)v)-I(0<\varepsilon_t(\beta_0)<n^{-1/2}Q_t'v)] \\
&= D_1+D_2+D_3.
\end{aligned}
$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

By a method similar to the proof for K_2, we can obtain that $D_1 \xrightarrow{P} 0$ uniformly on compact sets. On the other hand, denote $\eta_t = \varepsilon_t(\beta_0) - \varepsilon_t$. It can be verified that

$$\begin{aligned}
&|I(0 < \varepsilon_t(\beta_0) < n^{-1/2}Q_t'v) - I(0 < \varepsilon_t < n^{-1/2}Q_t'v)| \\
&\quad \le (I(-\eta_t < \varepsilon_t < 0) + I(n^{-1/2}Q_t'v - \eta_t < \varepsilon_t < n^{-1/2}Q_t'v))I(\eta_t > 0) \\
&\qquad + (I(0 < \varepsilon_t < -\eta_t) + I(n^{-1/2}Q_t'v < \varepsilon_t < n^{-1/2}Q_t'v - \eta_t))I(\eta_t < 0).
\end{aligned}$$

Then from condition A3 and by a method similar to the proof of Lemma 2, we have

$$\begin{aligned}
E|D_2| &\le \sum_{t=u+1}^{n} E\Bigg\{ w_t I(Q_t'v > 0) \\
&\qquad \times \Bigg[I(\eta_t > 0)\left(\int_{-\eta_t}^{0} + \int_{n^{-1/2}Q_t'v-\eta_t}^{n^{-1/2}Q_t'v} \right) |n^{-1/2}Q_t'v - x| f(x)\,dx \\
&\qquad\quad + I(\eta_t < 0)\left(\int_{0}^{-\eta_t} + \int_{n^{-1/2}Q_t'v}^{n^{-1/2}Q_t'v-\eta_t} \right) \\
&\qquad\quad \times |n^{-1/2}Q_t'v - x| f(x)\,dx \Bigg] \Bigg\} \\
&\le C \sum_{t=u+1}^{n} E\{w_t(|n^{-1/2}Q_t'v|\xi_t + \xi_t^2)\} \\
&= Cn^{-1/2} \sum_{t=u+1}^{n} E\{w_t^{1/2}|Q_t'v| \cdot w_t^{1/2}\xi_t\} + C \sum_{t=u+1}^{n} E\{w_t^{1/2}\xi_t\}^2 \\
&= D_{21} + D_{22}.
\end{aligned}$$

By the same method as for K_1, we can obtain that $D_{21} \to 0$. On the other hand,

$$\begin{aligned}
D_{22} &\le C \sum_{t=u+1}^{n} \left(\sum_{h=0}^{\infty} r^{h+t}(h+t)^{\alpha} \right)^2 \le C \sum_{t=u+1}^{n} r^{2t} \left(\sum_{h=0}^{\infty} r^h h^{\alpha} + \sum_{h=0}^{\infty} r^h t^{\alpha} \right)^2 \\
&\le C \sum_{t=u+1}^{n} r^{2t} \left(\sum_{h=0}^{\infty} r^h h^{\alpha} \right)^2 + C \sum_{t=u+1}^{n} r^{2t} \left(\sum_{h=0}^{\infty} r^h t^{\alpha} \right)^2 \to 0.
\end{aligned}$$

Therefore, we have

$$E|D_2| \to 0. \tag{3.5}$$

And we can obtain $E|D_3| \to 0$ similarly to (3.5). Hence $\Lambda_2 \xrightarrow{P} 0$ uniformly on compact sets.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

By the same method as for Λ_2, we have

$$\Lambda_3 = \sum_{t=u+1}^{n} w_t(n^{-1/2}A_t'(\beta_0)v - \varepsilon_t(\beta_0))I(n^{-1/2}A_t'(\beta_0)v < \varepsilon_t(\beta_0) < 0)$$

$$- \sum_{t=u+1}^{n} w_t(n^{-1/2}Q_t'v - \varepsilon_t)I(n^{-1/2}Q_t'v < \varepsilon_t < 0) \xrightarrow{P} 0$$

uniformly on compact sets. This completes the proof of this lemma. ■

LEMMA 4. $S_n(v) - T_n(v) \xrightarrow{P} 0$ *uniformly on compact sets.*

Proof. Notice that we have the fact that under conditions A1 and A2, for any given positive number M, there exist positive constants $C > 0$ and $0 < r < 1$ such that for sufficiently large n

$$\sup_{\|v\| \le M} |\varepsilon_t(\beta_0 + n^{-1/2}v)| \le C \sum_{j=0}^{\infty} r^j |y_{t-j}|.$$

Now we have

$$\sup_{\|v\| \le M} |S_n(v) - T_n(v)| \le \sup_{\|v\| \le M} \sum_{t=u+1}^{n} |w_t - \tilde{w}_t|[|\varepsilon_t(\beta_0 + n^{-1/2}v)| + |\varepsilon_t(\beta_0)|].$$

$$\le C \sum_{t=u+1}^{n} \frac{\sum_{k=t}^{\infty} k^{-\alpha}|y_{t-k}| \sum_{j=0}^{\infty} r^j |y_{t-j}|}{1 + \sum_{k=1}^{\infty} k^{-\alpha}|y_{t-k}|}$$

$$\le C \sum_{t=u+1}^{n} \sum_{k=t}^{\infty} k^{-\alpha}|y_{t-k}| \left(\sum_{j=1}^{\infty} r^j j^{\alpha} + |y_t| \right)$$

$$\le C \sum_{t=u+1}^{n} |y_t| \sum_{h=0}^{\infty} (h+t)^{-\alpha}|y_{-h}| + C \sum_{t=u+1}^{n} \sum_{h=0}^{\infty} (h+t)^{-\alpha}|y_{-h}|$$

$$\le C \left(\sum_{t=u+1}^{n} t^{-\alpha/2}(|y_t| + 1) \right) \sum_{h=0}^{\infty} h^{-\alpha/2}|y_{-h}| \xrightarrow{P} 0$$

as $n \to \infty$, by conditions A2 and A3 and the fact that $(h+t)^{-\alpha} \le 2^{-\alpha}(ht)^{-\alpha/2}$. This completes the proof of Lemma 4. ■

Proof of Proposition 2. For $v \in R^{p+q}$, define

$$S_n^{**}(v) = \sum_{t=u+1}^{n} w_t(|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v - n^{-1}v'H_t(\beta_0)v| - |\varepsilon_t(\beta_0)|),$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

where

$$H_t(\beta) = -\frac{1}{2}\frac{\partial^2 \varepsilon_t(\beta)}{\partial\beta\partial\beta'} = (h_{t(i,j)})_{(p+q)\times(p+q)}.$$

From the definition of $\varepsilon_t(\beta)$ in expression (2.1), we have

$$\theta(B)\frac{\partial^2 \varepsilon_t(\beta)}{\partial\phi_i\partial\phi_j} = 0, \qquad i,j = 1,\ldots,p,$$

$$\theta(B)\frac{\partial^2 \varepsilon_t(\beta)}{\partial\phi_i\partial\theta_j} = A_{t-j,i}(\beta), \qquad i = 1,\ldots,p, \qquad j = 1,\ldots,q,$$

$$\theta(B)\frac{\partial^2 \varepsilon_t(\beta)}{\partial\theta_i\partial\theta_j} = A_{t-j,i}(\beta) + A_{t-i,j}(\beta), \qquad i,j = 1,\ldots,q.$$

Similarly, replacing $A_{t-i,j}$ (or $A_{t-j,i}$) by $Q_{t-i,j}$ (or $Q_{t-j,i}$) in the preceding three equalities, we can define $X_t = (X_{t(i,j)})$. Then, $(h_{t(i,j)}(\beta_0))$ can be well approximated by $(X_{t(i,j)})$ such that

$$|X_{t(i,j)} - h_{t(i,j)}(\beta_0)| \le \xi_t, \qquad i,j = 1,\ldots,p+q, \tag{3.6}$$

which can be proved in the same way as Lemma 1.

Notice that

$$\begin{aligned}
S_n^{**}(v) - S_n^{*}(v)
&= \sum_{t=u+1}^{n} w_t(|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v - n^{-1}v'H_t(\beta_0)v| - |\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v|)\\
&= -\frac{1}{n}\sum_{t=u+1}^{n} w_t v'H_t(\beta_0)v\,sgn(\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v)\\
&\quad + 2\sum_{t=u+1}^{n} w_t(n^{-1}v'H_t(\beta_0)v - \varepsilon_t(\beta_0) + n^{-1/2}A_t'(\beta_0)v)\\
&\quad \times [I(0 < \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v < n^{-1}v'H_t(\beta_0)v)\\
&\quad\quad - I(n^{-1}v'H_t(\beta_0)v < \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v < 0)]\\
&= T_1 + T_2.
\end{aligned}$$

Notice that $-(1/n)\sum_{t=u+1}^{n} w_t v'X_t v\,sgn(\varepsilon_t) \to 0$; we obtain

$$T_1 = -\frac{1}{n}\sum_{t=u+1}^{n} w_t v'X_t v[sgn(\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v) - sgn(\varepsilon_t)] + o_p(1)$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

by (3.6). Using the same argument of $E|K_1| \to 0$ in Lemma 3, we have $T_1 \xrightarrow{P} 0$. Similarly, we have

$$\begin{aligned} T_{21} &= \sum_{t=u+1}^{n} w_t(n^{-1}v'H_t(\beta_0)v - \varepsilon_t(\beta_0) + n^{-1/2}A_t'(\beta_0)v) \\ &\quad \times I(0 < \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v < n^{-1}v'H_t(\beta_0)v) \\ &= \sum_{t=u+1}^{n} w_t(n^{-1}v'X_tv - \varepsilon_t + n^{-1/2}Q_t'v) \\ &\quad \times I(0 < \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v < n^{-1}v'H_t(\beta_0)v) \\ &\quad + o_p(1). \end{aligned}$$

Let

$$T_{21}^{(1)} = \sum_{t=u+1}^{n} w_t(n^{-1}v'X_tv - \varepsilon_t + n^{-1/2}Q_t'v)I(0 < \varepsilon_t - n^{-1/2}Q_t'v < n^{-1}v'X_tv)$$

and

$$\begin{aligned} T_{21}^{(2)} &= \sum_{t=u+1}^{n} w_t(n^{-1}v'X_tv - \varepsilon_t + n^{-1/2}Q_t'v) \\ &\quad \times [I(0 < \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v < n^{-1}v'H_t(\beta_0)v) \\ &\qquad - I(0 < \varepsilon_t - n^{-1/2}Q_t'v < n^{-1}v'X_tv)]. \end{aligned}$$

In the same way as the proofs of (3.3) and (3.5) for $T_{21}^{(1)}$ and $T_{21}^{(2)}$, respectively, we can obtain that $T_{21} \xrightarrow{P} 0$. The same result holds for the rest term of T_2. Thus, it follows that $S_n^{**}(v) - S_n^{*}(v) \xrightarrow{P} 0$ uniformly on compact sets.

But,

$$\begin{aligned} &|S_n^{**}(v) - S_n(v)| \\ &\quad = \left| \sum_{t=u+1}^{n} w_t(|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v - n^{-1}v'H_t(\beta_0)v| - |\varepsilon_t(\beta_0 + n^{-1/2}v)|) \right| \\ &\quad \le \frac{1}{n}\sum_{t=1}^{n} w_t|v'(H_t(\beta_n^*) - H_t(\beta_0))v|, \end{aligned}$$

where β_n^* is between β_0 and $\beta_0 + n^{-1/2}v$. Then, $S_n^{**}(v) - S_n(v) \xrightarrow{P} 0$ uniformly on compact sets. Therefore, combining Lemmas 3 and 4, we have $T_n(v) - S_n^{+}(v) \xrightarrow{P} 0$ uniformly on compact sets.

By Lemma 2, we obtain that

$$T_n(v) \to_{\mathcal{L}} T(v) \quad \text{on } C(R^{p+q})$$

as $n \to \infty$. This completes the proof. ■

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

Proof of Theorem 1. Note that the limit process in Proposition 2 has convex sample paths and a unique minimizer:

$$\xi = -(1/[2f(0)])\Sigma^{-1}N.$$

Let P_n and P be the probability measures on $C(R^{p+q})$ induced by T_n and T, respectively. By Skorokhod's representation theorem (cf. Pollard, 1984, pp. 71–73), there exists a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ with processes T^* and T_n^* having the distributions P and P_n, respectively, such that $T_n^* \xrightarrow{a.s.} T^*$. Hence, there exists a subset Ω' of Ω^* with $P^*(\Omega') = 1$ such that for any $\omega \in \Omega'$,

$$\sup_{v \in K} |T_n^*(v, \omega) - T^*(v, \omega)| \to 0 \tag{3.7}$$

holds for any compact set K. Denote the minimizer of T^* by ξ^*. Then ξ^* has the same distribution as ξ. For any given positive random variable M, let

$$\xi_n^*(\omega) = \underset{\|v - \xi^*(\omega)\| \le M(\omega)}{\arg\min} T_n^*(v, \omega).$$

Now we show that $\xi_n^*(\omega) \to \xi^*(\omega)$. Suppose $\xi_n^*(\omega) \to \xi^*(\omega)$ does not hold; then there is a subsequence n' such that $\xi_{n'}^*(\omega) \to \xi'(\omega) \neq \xi^*(\omega)$ and we know $\|\xi'(\omega) - \xi^*(\omega)\| \le M(\omega)$. From the definition of $\xi_n^*(\omega)$, we have

$$T_n^*(\xi^*(\omega)) - T_n^*(\xi_n^*(\omega)) \ge 0.$$

But, on the other hand,

$$\begin{aligned} &T_{n'}^*(\xi^*(\omega)) - T_{n'}^*(\xi_{n'}^*(\omega)) \\ &\quad = [T_{n'}^*(\xi^*(\omega)) - T^*(\xi^*(\omega))] + [T^*(\xi^*(\omega)) - T^*(\xi_{n'}^*(\omega))] \\ &\qquad + [T^*(\xi_{n'}^*(\omega)) - T_{n'}^*(\xi_{n'}^*(\omega))] \\ &\quad = G_1 + G_2 + G_3. \end{aligned}$$

From (3.7), we can obtain that $G_1 \to 0$ and $G_3 \to 0$. Noticing that $T^*(v, \omega)$ is continuous, we have

$$G_2 \to T^*(\xi^*(\omega)) - T^*(\xi'(\omega)) < 0.$$

This is a contradiction. Therefore, $\xi_n^* \xrightarrow{a.s.} \xi^*$.

Define a sequence of local minimizers $\{\xi_n\}$ of T_n by

$$\xi_n(\omega) = \underset{\|v - \xi(\omega)\| \le M(\omega)}{\arg\min} T_n(v, \omega).$$

Then, ξ_n converges in distribution to ξ. The proof is completed. ■

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

3.2. Proof of Theorem 2

We also use the substitution $v = \sqrt{n}(\beta - \beta_0)$. Then, minimizing $\widetilde{W}_n(\beta)$ is equivalent to minimizing

$$\tilde{T}_n(v) = \sum_{t=u+1}^{n} \widetilde{w}_t(|\varepsilon_t(\hat{\beta}^0) - A_t'(\hat{\beta}^0)(\beta_0 - \hat{\beta}^0 + n^{-1/2}v)| - |\varepsilon_t|).$$

Because $\tilde{T}_n(v)$ has convex sample paths, we only need to prove that $\tilde{T}_n(v) \to_{\mathcal{L}} T(v)$ on $C(R^{p+q})$ by Lemma 2.2 in Davis et al. (1992).

Denote

$$\tilde{H}_t(\beta) = \frac{1}{2}\sum_{i,j=1}^{p+q} \frac{\partial^2 \varepsilon_t(\beta)}{\partial\beta_i \partial\beta_j}.$$

By Taylor expansion near $\hat{\beta}^0$, for any fixed $v \in R^{p+q}$, we have

$$\varepsilon_t(\beta_0 + n^{-1/2}v) = \varepsilon_t(\hat{\beta}^0) - A_t'(\hat{\beta}^0)(\beta_0 - \hat{\beta}^0 + n^{-1/2}v) + \tilde{H}_t(\beta_{1t})O_p(n^{-1}),$$

where β_{1t} lies between $\hat{\beta}^0$ and $\beta_0 + n^{-1/2}v$. We now have

$$\begin{aligned}
&\varepsilon_t(\hat{\beta}^0) - A_t'(\hat{\beta}^0)(\beta_0 - \hat{\beta}^0 + n^{-1/2}v) \\
&\quad = \varepsilon_t(\beta_0 + n^{-1/2}v) - \tilde{H}_t(\beta_{1t})O_p(n^{-1}) \\
&\quad = \varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v + \tilde{H}_t(\beta_{2t})O_p(n^{-1}) - \tilde{H}_t(\beta_{1t})O_p(n^{-1}),
\end{aligned}$$

where β_{2t} lies between β_0 and $\beta_0 + n^{-1/2}v$. Hence, by a similar way to the proof of Lemma 2 but replacing $n^{-1/2}Q_t'v$ by $n^{-1/2}Q_t'v - \tilde{H}_t(\beta_0)O_p(n^{-1})$, we have

$$\begin{aligned}
\tilde{T}_n(v) &= \sum_{t=u+1}^{n} \widetilde{w}_t[|\varepsilon_t(\beta_0) - n^{-1/2}A_t'(\beta_0)v + \tilde{H}_t(\beta_{2t})O_p(n^{-1}) \\
&\qquad - \tilde{H}_t(\beta_{1t})O_p(n^{-1})| - |\varepsilon_t|] \\
&= \sum_{t=u+1}^{n} \widetilde{w}_t(|\varepsilon_t - n^{-1/2}Q_t'v + \tilde{H}_t(\beta_0)O_p(n^{-1})| - |\varepsilon_t|) + o_p(1) \to_{\mathcal{L}} T(v).
\end{aligned}$$

■

3.3. Proof of Theorem 3

Based on Theorem 1, Theorem 3 follows immediately from Lemmas 5 and 6 in this section. We assume that the conditions of Theorem 3 hold.

LEMMA 5. $\hat{\Sigma} \xrightarrow{P} \Sigma$ *and* $\hat{\Omega} \xrightarrow{P} \Omega$.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

Proof. Denote

$$\hat{\theta}(z) = 1 + \hat{\theta}_1 z + \cdots + \hat{\theta}_q z^q, \qquad \hat{\phi}(z) = 1 - \hat{\phi}_1 z - \cdots - \hat{\phi}_p z^p.$$

Then

$$\hat{Q}_t = (\hat{U}_{t-1}, \ldots, \hat{U}_{t-p}, \hat{V}_{t-1}, \ldots, \hat{V}_{t-q})',$$

where $\hat{U}_t$ and $\hat{V}_t$ are determined by

$$\hat{\phi}(B)\hat{U}_t = \varepsilon_t(\hat{\beta}), \qquad \hat{\theta}(B)\hat{V}_t = \varepsilon_t(\hat{\beta}).$$

Note that, by the definition of $\varepsilon_t(\beta)$, we have $\hat{\phi}(B)y_t = \hat{\theta}(B)\varepsilon_t(\hat{\beta})$ where $y_t = 0$ for $t \leq 0$. Hence,

$$\hat{U}_t = \hat{\theta}^{-1}(B)y_t, \qquad \hat{V}_t = \hat{\theta}^{-2}(B)\hat{\phi}(B)y_t, \qquad (y_t = 0 \text{ for } t \leq 0).$$

Let

$$II_1 = \frac{1}{n-u}\sum_{t=u+1}^{n} \tilde{w}_t \hat{Q}_t \hat{Q}_t' - \frac{1}{n-u}\sum_{t=u+1}^{n} w_t Q_t Q_t',$$

$$II_2 = \frac{1}{n-u}\sum_{t=u+1}^{n} w_t Q_t Q_t'.$$

Then, $\hat{\Sigma} = II_1 + II_2$. Obviously, $II_2 \to \Sigma$ by the ergodic theorem. For II_1, we first define a vector-valued function $Q_t(\beta) = (U_{t-1}(\beta), \ldots, U_{t-p}(\beta), V_{t-1}(\beta), \ldots, V_{t-q}(\beta))'$, and its components are determined as follows:

$$\phi(B)U_t(\beta) = \varepsilon_t, \qquad \theta(B)V_t(\beta) = \varepsilon_t.$$

We denote $\tilde{Q}_t = Q_t(\hat{\beta})$ and divide II_1 into three terms as follows:

$$\begin{aligned} II_1 &= \frac{1}{n-u}\sum_{t=u+1}^{n} (\tilde{w}_t - w_t)\hat{Q}_t \hat{Q}_t' + \frac{1}{n-u}\sum_{t=u+1}^{n} w_t(\hat{Q}_t \hat{Q}_t' - \tilde{Q}_t \tilde{Q}_t') \\ &\quad + \frac{1}{n-u}\sum_{t=u+1}^{n} w_t(\tilde{Q}_t \tilde{Q}_t' - Q_t Q_t') \\ &= J_1 + J_2 + J_3. \end{aligned}$$

It is easy to obtain that $J_1 \xrightarrow{P} 0$ by Lemma 1 and the definition of $\hat{Q}_t$. Notice that the (1,1)th element of J_2 is

$$J_2^{(1,1)} = \frac{1}{n-u}\sum_{t=u+1}^{n} w_t(\hat{U}_{t-1} - U_{t-1}(\hat{\beta}))(\hat{U}_{t-1} + U_{t-1}(\hat{\beta}))$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

and

$$w_t^{1/2}|\hat{U}_{t-1} - U_{t-1}(\hat{\beta})| \le w_t^{1/2}\xi_t \le C, \qquad w_t^{1/2}|\hat{U}_{t-1} + U_{t-1}(\hat{\beta})| \le C$$

for some positive constant C and ξ_t defined in Lemma 1. It follows that $J_2^{(1,1)} \xrightarrow{P} 0$, and similarly we can prove that the rest elements of J_2 converge to zero in probability. Hence $J_2 \xrightarrow{P} 0$. For J_3, noticing that $Q_t = Q_t(\beta_0)$, we have

$$\tilde{Q}_t - Q_t = Q_t(\hat{\beta}) - Q_t(\beta_0) = \frac{\partial Q_t(\beta_t^*)}{\partial \beta'}(\hat{\beta} - \beta_0),$$

where β_t^* lies between $\hat{\beta}$ and β_0. Noticing that $w_t^{1/2}\|(\partial Q_t(\beta^*))/\partial\beta'\| \le C$, $w_t^{1/2}\|Q_t\| \le C$, and $w_t^{1/2}\|\tilde{Q}_t\| \le C$, we obtain that $J_3 \xrightarrow{P} 0$. Now it has been proved that $\hat{\Sigma} \xrightarrow{P} \Sigma$.

In the same way we can prove $\hat{\Omega} \xrightarrow{P} \Omega$. ■

LEMMA 6. *It holds that* $\tilde{f}(0) \xrightarrow{P} f(0)$.

Proof. Define

$$\hat{f}(0) = \frac{1}{\hat{\sigma}_w b_n(n-u)} \sum_{t=u+1}^{n} w_t K\left(\frac{\varepsilon_t(\hat{\beta})}{b_n}\right),$$

where $\hat{\sigma}_w = (n-u)^{-1}\sum_{t=u+1}^{n} w_t$; then

$$\begin{aligned}
|\hat{f}(0) - \tilde{f}(0)| &\le \left| \frac{1}{\hat{\sigma}_{\tilde{w}} b_n(n-u)} \sum_{t=u+1}^{n} (\tilde{w}_t - w_t) K\left(\frac{\varepsilon_t(\hat{\beta})}{b_n}\right) \right| \\
&\quad + \left| \frac{\hat{\sigma}_{\tilde{w}} - \hat{\sigma}_w}{\hat{\sigma}_{\tilde{w}}\hat{\sigma}_w b_n(n-u)} \right| \sum_{t=u+1}^{n} w_t K\left(\frac{\varepsilon_t(\hat{\beta})}{b_n}\right) \\
&\le \frac{C}{\hat{\sigma}_{\tilde{w}} b_n(n-u)} \sum_{t=u+1}^{n} \sum_{k=t}^{\infty} k^{-\alpha}|y_{t-k}| + O_p(1)\left|\frac{\hat{\sigma}_{\tilde{w}} - \hat{\sigma}_w}{\hat{\sigma}_{\tilde{w}}\hat{\sigma}_w b_n}\right| \\
&\le o_p(1) + O_p(1)\frac{1}{(n-u)b_n}\sum_{t=u+1}^{n}(\tilde{w}_t - w_t) \xrightarrow{P} 0,
\end{aligned}$$

where $\hat{\sigma}_{\tilde{w}} = (n-u)^{-1}\sum_{t=u+1}^{n}\tilde{w}_t$. So we need to prove that

$$\hat{f}(0) \xrightarrow{P} f(0).$$

Notice that

$$|\hat{f}(0) - f(0)| \le P_1 + P_2,$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

where

$$P_1 = \left| \frac{1}{\hat{\sigma}_w b_n (n-u)} \sum_{t=u+1}^{n} w_t \left(K\left(\frac{\varepsilon_t(\hat{\beta})}{b_n}\right) - K\left(\frac{\varepsilon_t}{b_n}\right)\right)\right|,$$

$$P_2 = \left| \frac{1}{\hat{\sigma}_w b_n (n-u)} \sum_{t=u+1}^{n} w_t K\left(\frac{\varepsilon_t}{b_n}\right) - f(0)\right|.$$

But, Lipschitz continuity of $K(x)$ insures that there exists a positive number L such that $|K(x) - K(y)| \le L|x - y|$ for any x, y, and then, from Theorem 1 and the assumptions on b_n, we have

$$\begin{aligned} P_1 &\le \frac{L}{\hat{\sigma}_w b_n^2 (n-u)} \sum_{t=u+1}^{n} w_t |\varepsilon_t(\hat{\beta}) - \varepsilon_t| \\ &\le \frac{L}{\hat{\sigma}_w b_n^2 (n-u)} \sum_{t=u+1}^{n} w_t (|\varepsilon_t(\hat{\beta}) - \varepsilon_t(\beta_0)| + |\varepsilon_t(\beta_0) - \varepsilon_t|) \\ &\le \frac{L}{\hat{\sigma}_w b_n^2 (n-u)} \sum_{t=u+1}^{n} w_t \|\hat{\beta} - \beta_0\| \|A_t(\beta_t^{**})\| \\ &\quad + \frac{C}{\hat{\sigma}_w b_n^2 (n-u)} \sum_{t=u+1}^{n} \sum_{k=t}^{\infty} r^k |y_{t-k}| \\ &\le \frac{C}{\hat{\sigma}_w b_n^2 \sqrt{n}} \sqrt{n} \|\hat{\beta} - \beta_0\| + \frac{C}{\hat{\sigma}_w b_n^2 (n-u)} \sum_{t=u+1}^{n} r^t \sum_{h=0}^{\infty} r^h |y_{-h}| \xrightarrow{P} 0, \end{aligned}$$

where β^* lies between $\hat{\beta}$ and β_0. For P_2, equivalently we need to prove that

$$I = \frac{1}{\hat{\sigma}_w b_n (n-u)} \sum_{t=u+1}^{n} w_t K\left(\frac{\varepsilon_t}{b_n}\right) \xrightarrow{P} f(0). \tag{3.8}$$

In fact, $I = (1/\hat{\sigma}_w) I_1 + f(0)$ where

$$I_1 = \frac{1}{b_n (n-u)} \sum_{t=1}^{n} w_t K\left(\frac{\varepsilon_t}{b_n}\right) - \frac{1}{n-u} \sum_{t=u+1}^{n} w_t f(0).$$

Notice that

$$\begin{aligned} EI_1^2 &= \frac{1}{(n-u)^2} \sum_{t=u+1}^{n} E\left[w_t \left(\frac{1}{b_n} K\left(\frac{\varepsilon_t}{b_n}\right) - f(0)\right)\right]^2 \\ &\quad + \frac{2}{(n-u)^2} \sum_{u+1 \le i < j \le n} E\left[w_i w_j \left(\frac{1}{b_n} K\left(\frac{\varepsilon_i}{b_n}\right) - f(0)\right)\left(\frac{1}{b_n} K\left(\frac{\varepsilon_j}{b_n}\right) - f(0)\right)\right] \\ &= I_{11} + I_{12}. \end{aligned}$$

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

But

$$I_{11} \le \frac{2}{n-u} E\left[w_t^2 \left(K^2\left(\frac{\varepsilon_t}{b_n}\right)\frac{1}{b_n^2} + f^2(0)\right)\right]$$

$$= \frac{2}{b_n(n-u)} \int_{-\infty}^{\infty} K^2(x) f(b_n x)\, dx E(w_t^2) + \frac{2f^2(0)}{n-u} E w_t^2 \to 0$$

by the assumptions on $K(x)$, and

$$|I_{12}| = \frac{2}{(n-u)^2} \left| \sum_{u+1\le i<j\le n} E\left[w_i w_j \left(K\left(\frac{\varepsilon_i}{b_n}\right)\frac{1}{b_n} - f(0)\right)\right] \right.$$

$$\left. \times \int_{-\infty}^{\infty} K(x)(f(b_n x) - f(0))\, dx \right|$$

$$\le \int_{-\infty}^{\infty} K(x)|f(b_n x) - f(0)|\, dx\, E\left[\left(K\left(\frac{\varepsilon_i}{b_n}\right)\frac{1}{b_n} + f(0)\right)\right] \le C b_n \to 0$$

by condition A3 and the assumptions on $K(x)$. Noticing that $\hat{\sigma}_w \to E w_t$, we have (3.8). The proof of Lemma 6 is completed. ■

4. NUMERICAL PROPERTIES

We conducted a simulation study to illustrate the finite-sample properties of the proposed WLADE in five aspects: namely, (i) its accuracy; (ii) its sampling distribution; (iii) comparison with the unweighted LADE, the Whittle estimator (Mikosch et al., 1995), QMLE, and the Gauss–Newton estimator (Davis, 1996); (iv) the selection of α, γ, and d in (2.8); and (v) the performance of the Wald test statistic Z_n.

We generated data from a simple ARMA(1,1) model:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \tag{4.1}$$

with t_2, Cauchy, or $N(0,1)$ innovation distribution. Unless specified otherwise, we always set $u = 20$, $\alpha = 3$, and $d = 0$.

Tables 1 and 2 list the means and the standard deviations (SD) of the WLADE for ϕ_1 and θ_1 from the 1,000 samples from model (4.1) with sample size $n =$ 200 or 400 and the true value $(\phi_1, \theta_1) = (0.4, 0.7)$, $(0.3, 0.5)$, or $(-0.5, -0.5)$. The estimates are very accurate in general, and the accuracy increases when the sample size increases from 200 to 400. We also included in the tables asymptotic standard deviations (AD) derived from Theorem 1 with (Σ, Ω) replaced by their estimators in (2.9) and with $f(0)$ replaced by (2.10) with the kernel $K(x) = e^{-x}/(1 + e^{-x})^2$ and the rule-of-thumb bandwidth $b_n = 1.06 \times n^{-1/5}$ (Silverman 1986, p. 40). The values of SD and AD are quite close to one another.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

TABLE 1. Means and standard deviations of WLADE for model (4.1) with $t(2)$ innovations

			$n = 200$		$n = 400$	
ϕ_1	θ_1		ϕ_1	θ_1	ϕ_1	θ_1
0.4	0.7	β	0.4092	0.6982	0.4078	0.6959
		SD	0.1013	0.0831	0.0716	0.0537
		AD	0.1156	0.0846	0.0779	0.0574
0.3	0.5	β	0.3019	0.5038	0.3009	0.4970
		SD	0.1153	0.1101	0.0916	0.0842
		AD	0.1165	0.1144	0.0926	0.0848
−0.5	−0.5	β	−0.4928	−0.4952	−0.4992	−0.4959
		SD	0.1075	0.1044	0.0705	0.0799
		AD	0.1164	0.1201	0.0779	0.0807

To investigate the sampling distributions of the WLADE, we drew 16,000 samples of size $n = 400$ from (4.1) with $(\phi_1, \theta_1) = (0.3, 0.5)$ and t_2 or Cauchy innovations. For each sample, the WLADEs for both ϕ_1 and θ_1 were obtained. We divided $[-6,6]$ into small intervals with equal length 0.2. For each small interval, we computed the (normalized) relative frequency for the occurrence of the event that the normalized WLADE falls into the interval. Those relative frequencies are plotted against the center of intervals in Figures 1 and 2. We superimposed the $N(0,1)$ density function in the figures, which is the limiting density of the normalized WLADE. Even with sample size $n = 400$, the

TABLE 2. Means and standard deviations of WLADE for model (4.1) with Cauchy innovations

			$n = 200$		$n = 400$	
ϕ_1	θ_1		ϕ_1	θ_1	ϕ_1	θ_1
0.4	0.7	β	0.4087	0.6918	0.4095	0.6924
		SD	0.0787	0.0535	0.0550	0.0382
		AD	0.0748	0.0530	0.0498	0.0379
0.3	0.5	β	0.3095	0.4917	0.3014	0.4936
		SD	0.0850	0.0776	0.0405	0.0360
		AD	0.0806	0.0749	0.0441	0.0371
−0.5	−0.5	β	−0.4994	−0.4925	−0.5009	−0.4974
		SD	0.0774	0.0849	0.0440	0.0498
		AD	0.0754	0.0831	0.0438	0.0478

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

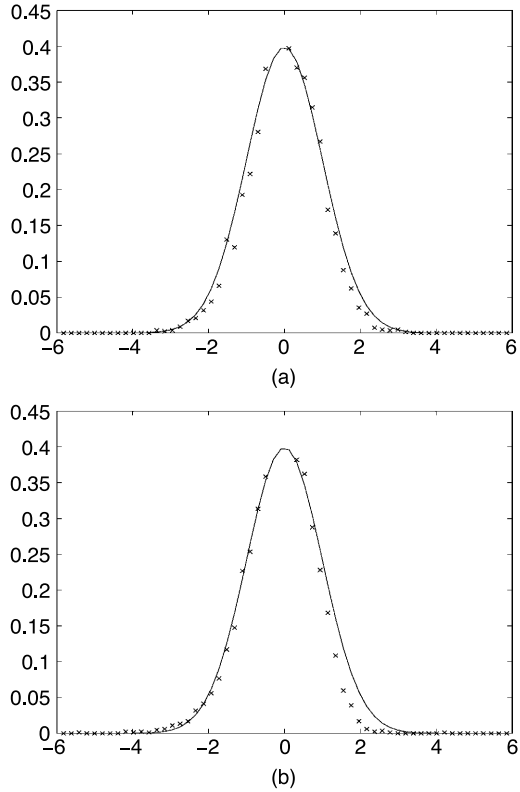


FIGURE 1. Estimated values ($\times$) for the density functions of the normalized $\hat{\phi}_1$ (shown in (a)) and $\hat{\theta}_1$ (shown in (b)), together with their asymptotic limit—the $N(0,1)$ density (solid curves) for model (4.1) with t_2 innovations.

estimated values of the density functions match their asymptotic limits very well.

Figure 3 presents the box plots of the average absolute error (AAE) $(|\hat{\phi}_1 - 0.3| + |\hat{\theta}_1 - 0.5|)/2$ for the unweighted LADE, the WLADE, the Whittle estimator, the QMLE, and the Gauss–Newton estimator from 1,000 samples with sample size $n = 400$ drawn from (4.1). Here, we set $d = 2$. For the samples with heavy-tailed innovations, that is, t_2 and Cauchy, the WLADE performed better than the Whittle estimator, QMLE, and the Gauss–Newton estimator. In fact, the improvement from using the WLADE over the preceding three estimators is more pronounced when the tails are heavier (i.e., with Cauchy distribution). The Gauss–Newton estimator performed the best with Gaussian innovations. However it is noticeable that there was an efficiency loss due to the introduction of weights in the estimation, although such a loss was not significant at least in the setting used in our simulation.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

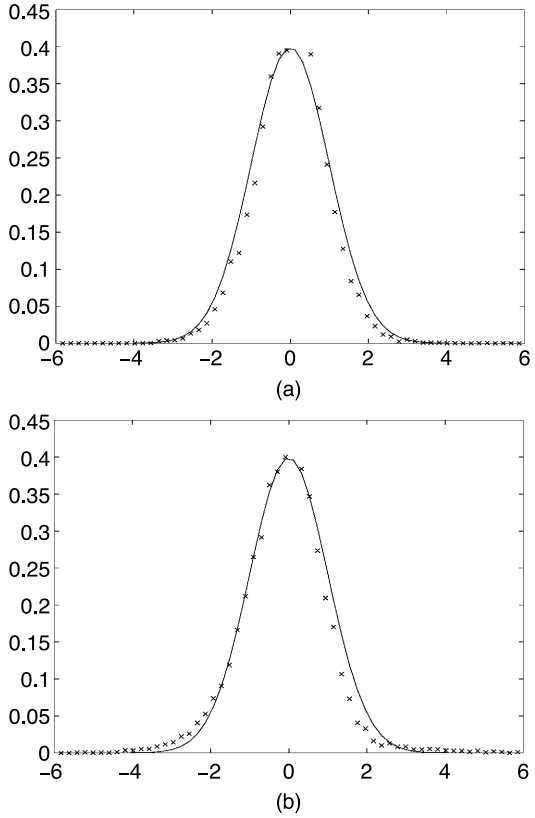


FIGURE 2. Estimated values ($\times$) for the density functions of the normalized $\hat{\phi}_1$ (shown in (a)) and $\hat{\theta}_1$ (shown in (b)), together with their asymptotic limit—the $N(0,1)$ density (solid curves) for model (4.1) with Cauchy innovations.

Because the convergence rate of the WLADE is slower than the Gauss–Newton estimator (Mikosch et al., 1995; Davis, 1996), we compared the two estimators with large sample size n between 2,000 and 8,000. For each setting, 1,000 samples were drawn from model (4.1) with Cauchy innovations. The parameters (ϕ_1, θ_1) were set at $(-0.3,0.2)$, $(0.3,-0.5)$, $(-0.3,0.5)$, $(0.1,-0.6)$, $(0.3,0.5)$, $(0.6,0.7)$, $(0.4,0.6)$, $(-0.4,-0.6)$, $(-0.3,-0.4)$, and $u = \sqrt{n}$. It turned out that when $\phi_1\theta_1 > 0$, the WLADE performed better than the Gauss–Newton estimator for all $n \leq 8{,}000$. However, when $\phi_1\theta_1 < 0$, the WLADE was better than the Gauss–Newton estimator only for $n \leq 3{,}000$. The box plots of the AAE for parameters $(\phi_1, \theta_1) = (-0.3, 0.2), (-0.3, -0.4), (0.6, 0.7)$ are displayed in Figure 4.

We also compared the WLADE using a general form of weights (2.8) with different α, γ, and d. To this end, we drew 1,000 samples from model (4.1) with

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

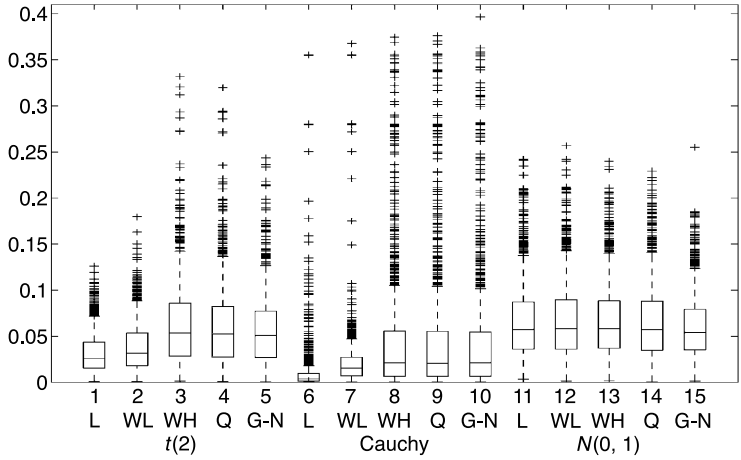


FIGURE 3. Box plots of the AAE of LADE (L), WLADE (WL), Whittle estimator (WH), QMLE (Q) and Gauss–Newton estimator (G–N) for model (4.1).

$(\phi_1, \theta_1) = (0.3, 0.5)$ and $n = 400$. Figure 5 presents the box plots of the AAEs with t_2, Cauchy, and normal innovations. They suggest that the WLADE is fairly robust with respect to the value of α. However, the WLADE is more sensitive to γ. It is evident that we should choose γ as small as possible, that

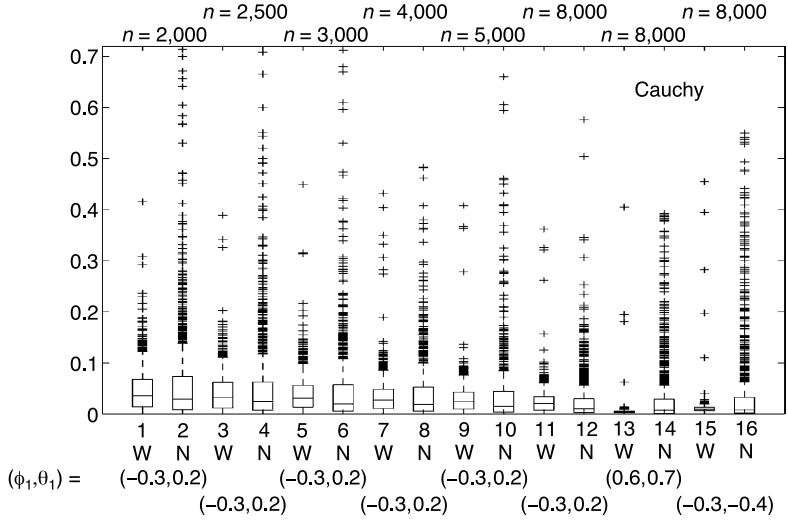


FIGURE 4. Box plots of the AAE of WLADE (W) and Gauss–Newton (N) estimator for model (4.1) for the large-sample case.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

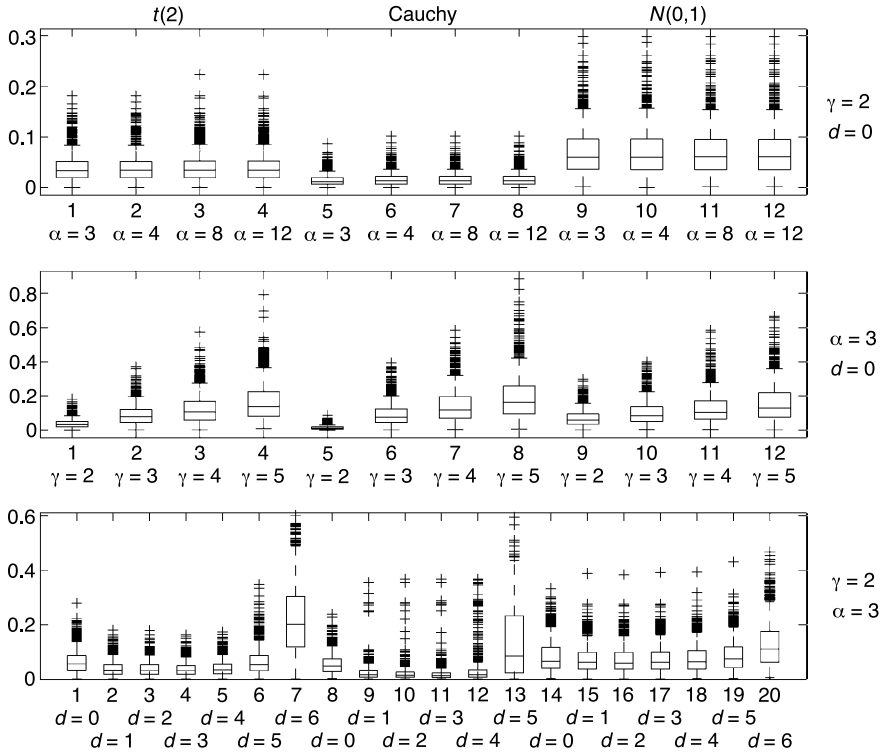


FIGURE 5. Box plots of AAE of WLADE with different α, γ, and d for model (4.1).

is, $\gamma = 2$, which corresponds to the default weight function used in this paper. On the other hand, the performances with $d = 1, 2$, and 3 do not differ significantly.

Finally, we approximated the P-value of the Wald test proposed in Section 2.3 for testing AR(1) against ARMA(1,1) models by the relative frequency $\hat{P}$ for the occurrence of the event that the AR(1) null hypothesis was rejected in a simulation with 1,000 replications. The data were generated from (4.1) with $(\phi_1, \theta_1) = (0.3, 0)$ or $(\phi_1, \theta_1) = (0.3, 0.5)$ and innovation to be $t(2)$, Cauchy, or $N(0, 1)$. We repeated the experiment with sample size equal to, respectively, 200, 400, and 600 for the nominal significance level between 0 and 0.1. Figure 6 plots the difference between $\hat{P}$ and the nominal significance level against the nominal level with data generated from (4.1). With $(\phi_1, \theta_1) = (0.3, 0)$, Figure 6(a)–(c) indicate that the χ^2-asymptotic approximation for the significance level is accurate, especially for $n = 400$ and 600. With $(\phi_1, \theta_1) = (0.3, 0.5)$, Figure 6(d)–(f) illustrate that the test is powerful in detecting the departure from the AR(1) hypothesis.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

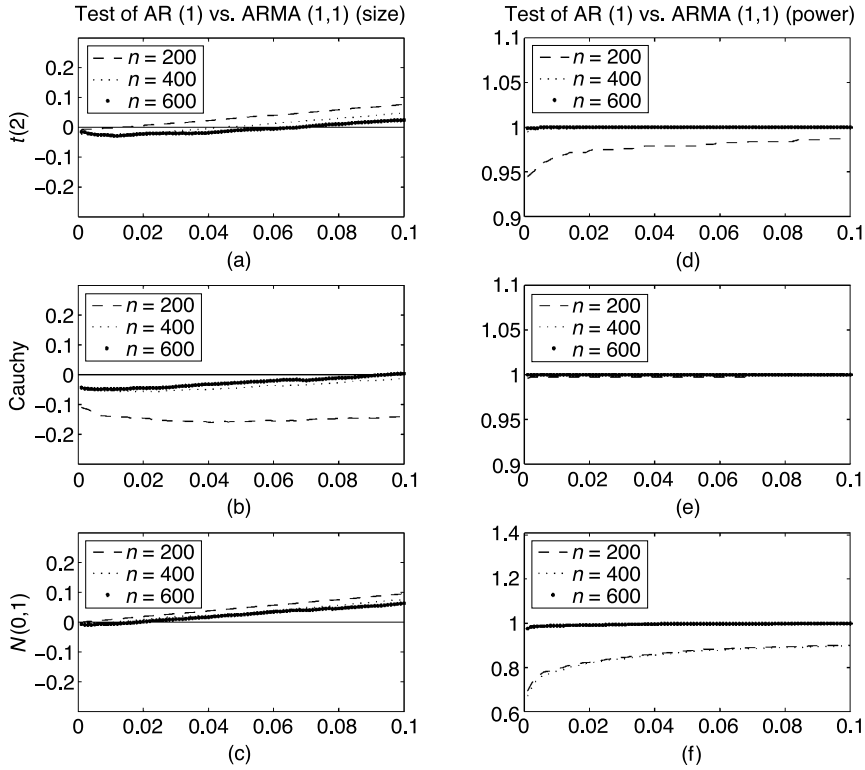


FIGURE 6. Simulation results for the Wald test. Differences between approximated *P*-value and the nominal significance level are plotted against the nominal level. The data were generated from (4.1) with $(\phi_1, \theta_1) = (0.3, 0)$ for (a)–(c) and $(\phi_1, \theta_1) = (0.3, 0.5)$ for (d)–(f).

REFERENCES

Adler, R.J., R. Feldaman, & M. Taqqu (1997) *A User's Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distribution and Processes*. Birkhauser.

An, H.Z. & Z.G. Chen (1982) On convergence of LAD estimates in autoregression with infinite variance. *Journal of Multivariate Analysis* 12, 335–345.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods.* Springer-Verlag.

Calder, M. & R.A. Davis (1998) Inference for linear processes with stable noise. In R.J. Adler, R. Feldaman, & M. Taqqu (eds.), *A Practical Guide to Heavy Tails*, pp. 159–176. Birkhauser.

Davis, R.A. (1996) Gauss-Newton and M–estimation for ARMA processes with infinite variance. *Stochastic Processes and Their Applications* 63, 75–95.

Davis, R.A. & W.T.M. Dunsmuir (1997) Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497.

Davis, R.A., K. Knight, & J. Liu (1992) M–estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications* 40, 145–180.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press

Davis, R.A. & S.I. Resnick (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* 13, 179–195.

Davis, R.A. & S.I. Resnick (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558.

Dunsmuir, W.T.M. & N.M. Spencer (1991) Strong consistency and asymptotic normality of L_1 estimates of the autoregressive moving-average model. *Journal of Time Series Analysis* 12, 95–104.

Gross, S. & W.L. Steiger (1979) Least absolute deviation estimates in autoregressions with infinite variance. *Journal of Applied Probability* 16, 104–116.

Hall, P.C. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. Academic Press.

Hjort, N.L. & D. Pollard (1993) Asymptotics for Minimisers of Convex Processes. Yale University. Available at http://www.stat.yale.edu/~pollard/Papers/convex.pdf.

Horvath, L. & F. Liese (2004) L_p-estimators in ARCH models. *Journal of Statistical Planning and Inference* 119, 277–309.

Jansen, D. & C. de Vries (1991) On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics* 73, 18–28.

Koedijk, K.G., M.M.A. Schafgans, & C.G. De Vries (1990) The tail index of exchange rate returns. *Journal of International Economics* 29, 93–108.

Kokoszka, P.S. & M.S. Taqqu (1996) Parameter estimation for infinite variance fractional ARIMA. *Annals of Statistics* 24, 1880–1913.

Ling, S. (2005) Self-weighted LAD estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society, Series B* 67, 381–393.

Mikosch, T., T. Gadrich, & R.J. Adler (1995) Parameter estimation for ARMA models with infinite variance innovations. *Annals of Statistics* 23, 305–326.

Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag.

Resnick, S.I. (1997) Heavy tail modeling and teletraffic data (with discussions). *Annals of Statistics* 25, 1805–1869.

Rudin, W. (1991) *Functional Analysis*. McGraw-Hill.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

https://doi.org/10.1017/S0266466607070363 Published online by Cambridge University Press