

Least tail-trimmed absolute deviation estimation for autoregressions with infinite/finite variance

Rongning Wu and Yunwei Cui

*Zicklin School of Business, The City University of New York, New York, NY 10010, U.S.A.
Department of Mathematics, Towson University, Towson, MD 21252, U.S.A.
e-mail: rongning.wu@baruch.cuny.edu; ycui@towson.edu*

Abstract: We propose least tail-trimmed absolute deviation estimation for autoregressive processes with infinite/finite variance. We explore the large sample properties of the resulting estimator and establish its asymptotic normality. Moreover, we study convergence rates of the estimator under different moment settings and show that it attains a super- \sqrt{n} convergence rate when the innovation variance is infinite. Simulation studies are carried out to examine the finite-sample performance of the proposed method and that of relevant statistical inferences. A real example is also presented.

MSC 2010 subject classifications: Primary 62M10; secondary 62F12.

Keywords and phrases: Asymptotics, autoregressive process, estimation, financial data, infinite variance.

Received December 2016.

Contents

1	Introduction	941
2	Least tail-trimmed absolute deviation estimation	943
3	Convergence rate	945
4	Numerical studies	946
4.1	Simulation studies	946
4.2	Empirical example	949
A	Technical proofs	954
	Acknowledgements	959
	References	959

1. Introduction

Heavy-tailed time series is often encountered in the real world and particularly in economics and finance; see [5, 6] for some of the examples. Autoregressive (AR) processes with infinite variance have been widely used to model such a series. Let $\{X_t\}$ be an AR process of order p (AR(p)); that is, $\{X_t\}$ satisfies the recursions

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t, \quad (1.1)$$

where $\{Z_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables. Heavy-tailed time series $\{X_t\}$ is induced when the innovation Z_t has an infinite variance. A number of papers have been devoted to estimation of $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_p)^T$, the vector of unknown AR coefficients; see [10, 11] for a review. Specifically, [9] established the limiting distribution of the least squares (LS) and Yule–Walker estimators. [8] made available the respective asymptotic theory of the M-estimator and least absolute deviation (LAD) estimator when the innovation Z_t is in the domain of attraction of a stable law. However, neither these estimators are asymptotically normal nor their limiting distributions have a closed form. As a result, it is difficult to apply the asymptotic theory to statistical inferences due to the lack of effective inference techniques. Typically one needs to resort to nonparametric methods for an approximation of limiting distribution. Along another strand of research, [11] proposed self-weighted LAD (SLAD) estimation and showed asymptotic normality of the resulting estimator with a standard- \sqrt{n} convergence rate when the innovation density and its derivative are uniformly bounded. Ling’s result admits statistical inferences in a conventional fashion. More recently, aiming at robust parameter estimation, [10] developed a least tail-trimmed squares (LTTS) procedure. It is shown that the LTTS estimator is asymptotically normal and can achieve a super- \sqrt{n} convergence rate when the innovation variance is infinite. Also, Hill’s simulation study indicated that the LTTS procedure is robust to outlying observations in both small and large samples.

In the present paper we explore least tail-trimmed absolute deviation (LTTAD) estimation, the LAD version of Hill’s LTTS estimation, for AR processes. Our study is motivated by the well known fact that the LAD method in general is superior to the LS method when modeling heavy-tailed data due to its robustness to outliers. For infinite variance AR processes, it has been shown that the LAD estimator has a higher convergence rate than the LS estimator. Our study suggests that the merits of the LAD method carry over to LTTAD estimation when compared to LTTS estimation; for example, the LTTAD estimator converges faster than the LTTS estimator. Meanwhile, asymptotic normality of the LTTAD estimator is alike ensured. Moreover, without exerting tail-trimming on the extreme residual values, the LTTAD estimation on the one hand is much simpler to implement numerically and on the other hand requires weaker model assumptions.

Our study is also related to [11], but major difference exists. Ling employed a constant trimming threshold as an example of weight selection. However, when the constant is replaced by quantile order statistic of observations (as in Ling’s simulation studies), the corresponding weight no longer meets the requirement (i.e. Ling’s Assumption 2 on page 383). Our study develops theoretical treatments under such a situation. It turns out that, while the resulting estimator retains asymptotic normality, it actually achieves a super- \sqrt{n} (instead of standard- \sqrt{n}) convergence rate in the case of infinite variance.

The remainder of the paper is organized as follows. Section 2 sets up the study and defines the proper LTTAD estimator. We develop the asymptotic properties of the estimator. Statistical inferences are also explored. In Section 3 we

discuss convergence rate of the LTTAD estimator under different cases, showing its super- \sqrt{n} convergence when the innovation variance is infinite. Simulation studies are conducted in Section 4 to assess the finite sample performance of the proposed procedures and a real example of financial time series is presented as well. Technical proofs are relegated to the Appendix.

2. Least tail-trimmed absolute deviation estimation

In this paper we assume that

- (A1) the innovation Z_t has zero median and a probability density function $f(\cdot)$ that is continuous in a neighborhood of zero, $f(0) > 0$, and $\sup_{z \in \mathbb{R}} f(z) < B < \infty$; and
- (A2) the distribution of Z_t has Pareto-like tails:

$$P(|Z_t| > z) = Cz^{-\alpha}(1 + o(1)), \quad (2.1)$$

where $C > 0$ is a constant and $\alpha > 0$ is the tail index.

Assumption (A2) is satisfied by the Pareto distribution, stable distribution with index $\alpha \in (0, 2)$, Student's t distribution, Fréchet distribution, etc. Under this assumption, $E|Z_t|^\vartheta < \infty$ for $\vartheta \in (0, \alpha)$, and therefore there exists a unique strictly stationary and ergodic process $\{X_t\}$. When $\alpha \leq 2$, Z_t has no finite variance and whence $\{X_t\}$ is a heavy-tailed series, which is the case of particular interest.

For the model (1.1), $\phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$ is called the AR polynomial. Under the assumption

- (A3) the polynomial $\phi(z)$ has all its roots outside the unit circle in the complex plane,

X_t can be represented as $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, where ψ_j , $j = 0, 1, \dots$, is the coefficient of z^j in the power series expansion of $1/\phi(z)$. According to the results of [4],

$$\lim_{z \rightarrow \infty} \frac{P(|X_t| > z)}{P(|Z_t| > z)} = \sum_{j=0}^{\infty} |\psi_j|^\alpha,$$

from which it follows that the distribution of X_t also has Pareto-like tails with the same decay rate as in (2.1).

Let $\phi_0 = (\phi_{00}, \phi_{01}, \dots, \phi_{0p})^T$ denote the true value of ϕ . Suppose X_1, \dots, X_n are the observations from the true model. At $t = p + 1, \dots, n$, the residual corresponding to ϕ is given by

$$z_t(\phi) = X_t - \phi_0 - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}.$$

To introduce the LTTAD estimator, we will trim the extreme values of X_{t-i} , $t = p + 1, \dots, n, i = 1, \dots, p$. Define $X_t^a = |X_t|$, $t = 1, \dots, n$ and let $X_{(1)}^a \geq \dots \geq X_{(n)}^a$ denote the order statistics of X_1^a, \dots, X_n^a . Also, let k_n be an integer

number such that $1 \leq k_n < n$ and $1/k_n + k_n/n \rightarrow 0$ as $n \rightarrow \infty$. The LTTAD estimator $\hat{\phi}_n$ of ϕ_0 is defined as the minimizer of the objective function

$$l_n(\phi) = \sum_{t=p+1}^n |z_t(\phi)| I_{nt}, \tag{2.2}$$

where $I_{nt} = \prod_{i=1}^p I(|X_{t-i}| < X_{(k_n)}^a)$ with $I(\cdot)$ being the indicator function. Unlike Hill's LTT procedure, there is no trimming in (2.2) based on the magnitude of $|z_t(\phi)|$. There are several advantages of doing so. First, $l_n(\phi)$ is a convex function of ϕ , which on the one hand guarantees that the established theoretical results hold for the global minimizer and on the other hand considerably reduces the burden of numerical computations. Second, the "fractiles" assumption of [10] can be dropped. Third, the symmetry is no longer required for the distribution of Z_t , that is, skewed distributions are allowed for Z_t .

For $t > p$, noting that $z_t(\phi_0) = Z_t$, we have

$$\begin{aligned} z_t(\phi) &= Z_t - (\phi_0 - \phi_{00}) - (\phi_1 - \phi_{01}) X_{t-1} - \cdots - (\phi_p - \phi_{0p}) X_{t-p} \\ &=: Z_t - (\phi - \phi_0)^T \mathbf{Y}_t, \end{aligned}$$

where $\mathbf{Y}_t = (1, X_{t-1}, \dots, X_{t-p})^T$. Let $\mathbf{V}_n = E(\mathbf{Y}_t \mathbf{Y}_t^T I_{nt}^c)$, where I_{nt}^c is the model version of I_{nt} ; i.e. $I_{nt}^c = \prod_{i=1}^p I(|X_{t-i}| < c_n)$, where $\{c_n\}$ is a non-random sequence such that $P(|X_t| \geq c_n) = k_n/n$. By similar arguments of [10], it can be shown that \mathbf{V}_n is positive definite for sufficiently large n .

The asymptotic distribution of the LTTAD estimator is given in the following theorem.

Theorem 2.1. *Under the Assumptions (A1)–(A3), the LTTAD estimator satisfies*

$$n^{1/2} \mathbf{V}_n^{1/2} (\hat{\phi}_n - \phi_0) \xrightarrow{d} \frac{1}{2f(0)} N(\mathbf{0}, \mathbf{I}_{p+1}). \tag{2.3}$$

Proof. See Appendix. □

Theorem 2.1 admits statistical inferences with respect to ϕ_0 in a conventional fashion. For example, consider a linear null hypothesis of the form

$$H_0 : \mathbf{\Gamma} \phi_0 = \boldsymbol{\gamma},$$

where $\mathbf{\Gamma}$ and $\boldsymbol{\gamma}$ are $s \times (p + 1)$ constant matrix and $s \times 1$ constant vector, respectively. A Wald test statistic can be constructed once \mathbf{V}_n and $f(0)$ are consistently estimated. An estimator of \mathbf{V}_n is given by

$$\hat{\mathbf{V}}_n := \frac{1}{n-p} \sum_{t=p+1}^n \mathbf{Y}_t \mathbf{Y}_t^T I_{nt}.$$

On the other hand, a kernel density estimator of $f(0)$ is

$$\hat{f}(0) = \frac{1}{\sum_{t=p+1}^n I_{nt}} \sum_{t=p+1}^n I_{nt} K_h(X_t - \hat{\phi}_n^T \mathbf{Y}_t),$$

where $K_h(\cdot) = K(\cdot/h_n)/h_n$ with $K(\cdot)$ being a symmetric kernel function and $h_n > 0$ being a bandwidth. It can be shown that $\widehat{\mathbf{V}}_n = \mathbf{V}_n(1 + o_p(1))$. Moreover, $\widehat{f}(0) = f(0) + o_p(1)$ under some regularity conditions on the kernel function and bandwidth. Then, a Wald test statistic defined as

$$\widehat{W} := (\mathbf{\Gamma}\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\gamma})^\top \left(\frac{1}{4n\widehat{f}^2(0)} \mathbf{\Gamma}\widehat{\mathbf{V}}_n^{-1}\mathbf{\Gamma}^\top \right)^{-1} (\mathbf{\Gamma}\widehat{\boldsymbol{\phi}}_n - \boldsymbol{\gamma})$$

has a null limiting chi-squared distribution with s degrees of freedom.

3. Convergence rate

It follows from Theorem 2.1 that, for $i = 0, 1, \dots, p$,

$$n^{1/2}V_{i+1,i+1,n}^{1/2}(\widehat{\phi}_{i,n} - \phi_{0i}) \xrightarrow{d} \frac{1}{2f(0)}N(0, 1) \quad (3.1)$$

where $\widehat{\phi}_{i,n}$ is the LTTAD estimator of ϕ_{0i} and $V_{i+1,i+1,n}$ is the $(i+1)^{th}$ element on the diagonal of the matrix \mathbf{V}_n . As we specify the same distribution tails of Z_t as in [10], the convergence rate of the LTTAD estimator can be straightforwardly derived. First, by the results of [10] that

$$E(I_{nt}^c) = 1 + o(1) \quad \text{and} \quad E(X_{t-i}^2 I_{nt}^c) = E(X_t^2 I(|X_t| < c_n))(1 + o(1)), \quad i = 1, \dots, p,$$

we have

$$V_{1,1,n} \sim 1 \quad \text{and} \quad V_{i,i,n} \sim E(X_t^2 I(|X_t| < c_n)), \quad i = 2, \dots, p + 1.$$

On the other hand, [10] showed that

$$E(X_t^2 I(|X_t| < c_n)) \sim \begin{cases} d \ln(n), & \text{if } \alpha = 2, \\ \frac{\alpha}{2-\alpha} d^{2/\alpha} \left(\frac{n}{k_n}\right)^{2/\alpha-1}, & \text{if } \alpha \in (0, 2), \end{cases}$$

where d is some scale parameter. So, the convergence rate of our LTTAD estimator is in order.

Theorem 3.1. *Suppose Assumptions (A1)–(A3) are satisfied.*

- (a) *If $\alpha > 2$, then the convergence rate of $\widehat{\phi}_{i,n}$ is $n^{1/2}$, $i = 0, 1, \dots, p$.*
- (b) *If $\alpha = 2$, then the convergence rates of $\widehat{\phi}_{0,n}$ and $\widehat{\phi}_{i,n}$, $i = 1, \dots, p$, are $n^{1/2}$ and $(n \ln(n))^{1/2}$, respectively.*
- (c) *If $\alpha \in (0, 2)$, then the convergence rates of $\widehat{\phi}_{0,n}$ and $\widehat{\phi}_{i,n}$, $i = 1, \dots, p$, are $n^{1/2}$ and $n^{1/2}(n/k_n)^{1/\alpha-1/2}$, respectively.*

Remark 1. When the innovation variance is infinite (i.e. $\alpha \in (0, 2]$), the LTTAD estimator has a higher convergence rate than Hill's LTTS estimator. More specifically,

$$\frac{\|\widehat{\boldsymbol{\phi}}_{\text{LTTS}} - \boldsymbol{\phi}_0\|}{\|\widehat{\boldsymbol{\phi}}_{\text{LTTAD}} - \boldsymbol{\phi}_0\|} = \begin{cases} O_P((\ln(n))^{1/2}), & \text{if } \alpha = 2, \\ O_P((n/k_n^Z)^{1/\alpha-1/2}), & \text{if } \alpha \in (0, 2), \end{cases}$$

where k_n^Z is the number of trimmed $|z_t(\boldsymbol{\phi})|$'s as in [10], which satisfies $1 \leq k_n^Z < n - p$ and $1/k_n^Z + k_n^Z/n \rightarrow 0$ as $n \rightarrow \infty$. On the other hand, when $\alpha \in (0, 2)$ the LAD estimator of ϕ_{0i} , $i = 1, \dots, p$, has a $n^{1/\alpha}$ convergence rate so it converges faster than the LTTAD estimator:

$$\frac{|\hat{\phi}_{i,\text{LAD}} - \phi_{0i}|}{|\hat{\phi}_{i,\text{LTTAD}} - \phi_{0i}|} = O_P((k_n)^{1/2-1/\alpha}), i = 1, \dots, p.$$

Remark 2. When the innovation variance is finite (i.e. $\alpha > 2$), the trimming does not affect the efficiency and the asymptotic result coincides with that of the LAD estimator established [7]. When $0 < \alpha \leq 2$, the super- \sqrt{n} convergence is ensured.

4. Numerical studies

4.1. Simulation studies

We conducted simulation studies to evaluate the finite-sample performance of LTTAD estimation. We considered the AR(2) model studied in [10]: $X_t = 0.2 + 0.8X_{t-1} - 0.3X_{t-2} + Z_t$, where $\{Z_t\}$ is a sequence of i.i.d. random variables. Two types of distributions were specified for the innovation $\{Z_t\}$: the two-sided compound Pareto distribution and the stable distribution. The Pareto (type II) distribution has a probability density function (pdf) of the form

$$g(z) = \frac{\alpha}{\sigma} \left(1 + \frac{z - \mu}{\sigma}\right)^{-(\alpha+1)}, z \geq \mu,$$

where $\alpha > 0$ is the tail index, $\mu \in \mathbb{R}$ is the location parameter, and $\sigma > 0$ is the scale parameter. In the studies we set $\mu = 0$ and $\sigma = 1$. The two-sided compound Pareto distribution was defined such that it had a pdf of $0.5g(-z)I(z < 0) + 0.5g(z)I(z \geq 0)$ for $z \in \mathbb{R}$. So, the distribution is symmetric about zero. Three values of α were taken: 0.75 (infinite mean), 1.5 (finite mean but infinite variance), and 2.5 (finite variance). On the other hand, the stable distribution can be described by four parameters: the index parameter $\alpha \in (0, 2]$, the skewness parameter $\tilde{\beta} \in [-1, 1]$, the scale parameter $\tilde{\gamma} > 0$, and the shift parameter $\tilde{\delta}$. We denote the distribution as $S(\alpha, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta})$. We considered two index values, 0.75 (infinite mean) and 1.5 (finite mean but infinite variance); more specifically, we used skewed distributions $S(0.75, 0.5, 1, -0.2565)$ and $S(1.5, 0.5, 1, -0.1339)$, where for each distribution the value of shift parameter ensures a median of zero (i.e. its magnitude equals the 50th percentile of the corresponding distribution with $\tilde{\delta} = 0$), as well as symmetric distributions $S(0.75, 0, 1, 0)$ and $S(1.5, 0, 1, 0)$.

Three sample sizes: 100, 400, and 800 were used. When sample size is small, the objective function $l_n(\boldsymbol{\phi})$ in (2.2) might be sensitive to extreme values of $z_t(\boldsymbol{\phi})$. As a remedial measure, additional trimming on the residuals may be incorporated into $l_n(\boldsymbol{\phi})$ as in [10]. In this regard, we define likewise $z_t^\alpha(\boldsymbol{\phi})$, $t =$

$p + 1, \dots, n$, and $z_{(1)}^a(\boldsymbol{\phi}), \dots, z_{(n-p)}^a(\boldsymbol{\phi})$. Then, we can formulate the following objective function

$$l_n^*(\boldsymbol{\phi}) = \sum_{t=p+1}^n |z_t(\boldsymbol{\phi})| I_{nt} I_{nt}^z(\boldsymbol{\phi}),$$

where $I_{nt}^z(\boldsymbol{\phi}) = I(|z_t(\boldsymbol{\phi})| < z_{(k_n^z)}^a(\boldsymbol{\phi}))$. The minimizer of $l_n^*(\boldsymbol{\phi})$ is denoted as $\widehat{\boldsymbol{\phi}}_n^*$ and is referred to as the LTTADR estimator of $\boldsymbol{\phi}_0$. However, our simulation studies indicated that it would not necessarily have a better finite-sample performance than the LTTAD estimator.

We simulated 5000 replicates of time series from the AR(2) model for each case of setting. To compare the performance of the LTTAD estimator with the LAD, Hill's LTTS, Ling's SLAD, and LTTADR estimators we computed the empirical bias and standard deviation (SD) for each type of estimates. The trimming parameter k_n was set to $\max\{1, [0.2n/(\ln(n))^2]\}$ in the tail-trimmed procedures (i.e. LTTS, LTTAD, and LTTADR); that is, $k_n = 1, 2$, and 4 for $n = 100, 400$, and 800 , respectively. On the other hand, the specification of $k_n^Z = \max\{1, [0.05n/\ln(n)]\}$ in [10] was used in the computation of LTTS and LTTADR estimates, which equals to $1, 3$, and 6 for the three sample sizes, respectively.

Table 1 reported the results when $\alpha = 0.75$ and Z_t follows a compound Pareto distribution; a similar pattern was displayed when Z_t follows a Stable distribution. The table exposed us to several distinct features. First, while all five procedures performed very well in estimation of ϕ_{01} and ϕ_{02} , the estimates of the intercept ϕ_{00} were quite off except for the LTTAD, SLAD, and LAD (when $n = 400$ and 800) procedures. The irregular bias was resulted from a large number of extremely outlying estimates. It seemed that the undefined mean of time series somehow caused difficulty in numerical minimization to obtain $\widehat{\phi}_{00}$. Note that, this abnormality vanished when there existed finite mean; see Tables 2 and 3. Second, LTTAD estimation significantly outperformed LTTS estimation. Third, additional trimming on the residuals did not necessarily improve the performance of estimation. This may be partially due to the non-convexity of the LTTADR criterion function. Note that, LAD estimates are expected to have a better performance than LTTAD estimates; see Remark 1. The motivation of studying LTTAD estimation is to trade a small loss of precision for asymptotic normality so that statistical inferences can be conducted in the classical fashion. The magnitudes of the reported standard deviations in Table 1 supported the relative convergence rates of the LAD, LTTS, and LTTAD estimators. Table 2 contains the results when the distribution of Z_t has a tail index $\alpha = 1.5$ (finite mean but infinite variance). The LTTAD estimator again outperformed the LTTS estimator. We also examined the impact of skewness of the innovation distribution on estimation. It is seen from the results that skewness made little difference on the performance of LTTAD estimation but resulted in a biased estimate of the intercept ϕ_{00} for the LTTS method. Table 3 reported the results when the time series has a finite variance. On the one hand, the LTTAD estimator outperformed the LTTS estimator. On the other hand, there was not much

TABLE 1
 Coefficient estimates for the AR(2) model $X_t = \phi_{00} + \phi_{01}X_{t-1} + \phi_{02}X_{t-2} + Z_t$, where $(\phi_{00}, \phi_{01}, \phi_{02}) = (0.2, 0.8, -0.3)$ and Z_t is compound Pareto with tail index $\alpha = 0.75$.

	LAD	SLAD	LTTTS	LTTAD	LTTADR
	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)
<i>n</i> = 100					
$\hat{\phi}_0$	30.41 (19.49)	0.360 (0.488)	113.4 (87.68)	0.502 (0.745)	3.7×10^4 (2.6×10^4)
$\hat{\phi}_1$	-0.081 (0.013)	-0.061 (0.015)	-1.085 (0.082)	0.028 (0.029)	0.111 (0.055)
$\hat{\phi}_2$	0.004 (0.010)	-0.042 (0.013)	-0.128 (0.063)	-0.048 (0.017)	0.005 (0.031)
<i>n</i> = 400					
$\hat{\phi}_0$	-0.239 (0.242)	0.030 (0.079)	-148.1 (258.1)	-0.404 (0.192)	-110.4 (1646)
$\hat{\phi}_1$	0.001 (0.004)	-0.005 (0.003)	-0.121 (0.026)	-0.002 (0.003)	-0.139 (0.045)
$\hat{\phi}_2$	-0.004 (0.002)	-0.003 (0.003)	-0.020 (0.021)	0.000 (0.002)	0.084 (0.027)
<i>n</i> = 800					
$\hat{\phi}_0$	-0.190 (0.176)	0.068 (0.054)	3796 (4344)	0.372 (0.391)	3245 (4327)
$\hat{\phi}_1$	-0.002 (0.002)	-0.001 (0.002)	0.027 (0.026)	-0.014 (0.006)	0.012 (0.013)
$\hat{\phi}_2$	-0.000 (0.001)	-0.000 (0.002)	-0.071 (0.021)	0.010 (0.005)	-0.027 (0.014)

difference in performance between the LTTAD and LTTADR estimators. In all the cases the LTTAD and SLAD methods had comparable performances. This is not surprising because the SLAD estimation implemented in Ling’s simulation studies is essentially a variant of the LTTAD method. To see it, note that the constant trimming threshold in Ling’s equation (2.3) was replaced by the 95% quantile of X_1^a, \dots, X_n^a . As a consequence, Ling weighted down substantially the relevant residual terms in the objective function while we simply assigned a zero weight to them.

To get a feeling of normality of the LTTAD and LTTTS estimators, in Figure 1 we constructed normal probability plots of LTTAD and LTTTS estimates of ϕ_{00} , ϕ_{01} , and ϕ_{02} , respectively, where Z_t is Stable with tail index $\alpha = 1.5$ and $n = 400$. The Gaussian approximation was fairly reasonable for both estimators if the few outliers due to the computer glitch were ignored.

To explore the performance of the Wald test, we considered testing AR(1) model against AR(2) model, where $H_0 : (0, 0, 1)(\phi_0, \phi_1, \phi_2)^T = 0$. Data were simulated from an AR(1)/AR(2) model with $\phi_0 = 0.2$, $\phi_1 = 0.8$, and ϕ_2 ranging from -0.10 to 0.10 with an increment of 0.05 , while other settings remained the same as before. We computed the empirical power/size of the Wald test based on both LTTAD and LTTTS methods, respectively, at nominal significance levels of 0.01 , 0.05 , and 0.1 . The results when Z_t is a two-sided compound Pareto distribution with tail index $\alpha = 1.5$ were displayed in Figure 2, where the dotted horizontal line represents the nominal size. In estimation of $f(0)$ a Gaussian kernel was used with the default normal scale bandwidth selector. For both methods, the empirical size was close to the nominal size. But the Wald test

TABLE 2
Coefficient estimates for the AR(2) model $X_t = \phi_{00} + \phi_{01}X_{t-1} + \phi_{02}X_{t-2} + Z_t$, where $(\phi_{00}, \phi_{01}, \phi_{02}) = (0.2, 0.8, -0.3)$ and Z_t is Stable with tail index $\alpha = 1.5$.

	LAD	SLAD	LTTS	LTTAD	LTTADR
	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)
$Z_t \sim S(1.5, 0, 1, 0)$; symmetric case					
$n = 100$					
$\hat{\phi}_0$	0.505 (0.181)	0.632 (0.186)	0.557 (0.250)	0.601 (0.185)	0.670 (0.184)
$\hat{\phi}_1$	-0.671 (0.069)	-0.760 (0.085)	-1.188 (0.095)	-0.659 (0.085)	-0.705 (0.085)
$\hat{\phi}_2$	-0.217 (0.068)	-0.457 (0.082)	-0.433 (0.086)	-0.355 (0.077)	-0.291 (0.077)
$n = 400$					
$\hat{\phi}_0$	0.031 (0.088)	0.036 (0.091)	-0.086 (0.112)	0.053 (0.089)	0.057 (0.088)
$\hat{\phi}_1$	-0.099 (0.027)	-0.202 (0.043)	-0.197 (0.041)	-0.135 (0.034)	-0.148 (0.034)
$\hat{\phi}_2$	-0.080 (0.027)	-0.127 (0.041)	-0.156 (0.038)	-0.098 (0.032)	-0.078 (0.032)
$n = 800$					
$\hat{\phi}_0$	-0.083 (0.062)	-0.072 (0.064)	-0.131 (0.081)	-0.089 (0.062)	-0.110 (0.062)
$\hat{\phi}_1$	-0.023 (0.017)	-0.061 (0.030)	-0.115 (0.028)	-0.086 (0.023)	-0.092 (0.023)
$\hat{\phi}_2$	-0.020 (0.017)	-0.080 (0.029)	-0.068 (0.027)	-0.013 (0.022)	-0.001 (0.023)
$Z_t \sim S(1.5, 0.5, 1, -0.1339)$; skewed case with median zero					
$n = 100$					
$\hat{\phi}_0$	1.675 (0.192)	2.109 (0.195)	22.20 (0.264)	1.962 (0.194)	0.927 (0.194)
$\hat{\phi}_1$	-0.601 (0.070)	-0.713 (0.088)	-1.135 (0.096)	-0.652 (0.087)	-0.671 (0.087)
$\hat{\phi}_2$	-0.152 (0.069)	-0.480 (0.084)	-0.496 (0.085)	-0.333 (0.078)	-0.273 (0.078)
$n = 400$					
$\hat{\phi}_0$	0.274 (0.093)	0.468 (0.097)	21.18 (0.120)	0.355 (0.094)	-0.372 (0.095)
$\hat{\phi}_1$	-0.072 (0.028)	-0.183 (0.043)	-0.168 (0.041)	-0.145 (0.035)	-0.130 (0.035)
$\hat{\phi}_2$	-0.064 (0.027)	-0.150 (0.042)	-0.129 (0.039)	-0.079 (0.033)	-0.077 (0.033)
$n = 800$					
$\hat{\phi}_0$	0.018 (0.065)	0.103 (0.068)	20.87 (0.090)	0.053 (0.067)	-0.645 (0.067)
$\hat{\phi}_1$	0.010 (0.017)	-0.101 (0.030)	-0.040 (0.029)	-0.052 (0.024)	-0.045 (0.024)
$\hat{\phi}_2$	-0.011 (0.017)	-0.050 (0.030)	-0.066 (0.027)	-0.022 (0.023)	-0.011 (0.023)

based on the LTTAD method had much larger empirical power, especially when the true ϕ_2 value is close to the hypothesized value of zero.

4.2. Empirical example

As an empirical example, we considered the log-returns of Hang Seng Index (HSI) between June 3, 1996 and May 31, 1998. The same data set has been

TABLE 3
Coefficient estimates for the AR(2) model $X_t = \phi_{00} + \phi_{01}X_{t-1} + \phi_{02}X_{t-2} + Z_t$, where $(\phi_{00}, \phi_{01}, \phi_{02}) = (0.2, 0.8, -0.3)$ and Z_t is compound Pareto with tail index $\alpha = 2.5$.

	LAD	SLAD	LTTS	LTTAD	LTTADR
	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)	Bias $\times 10^2$ (SD)
<i>n</i> = 100					
$\hat{\phi}_0$	0.271 (0.052)	0.322 (0.053)	0.843 (0.115)	0.291 (0.053)	0.296 (0.053)
$\hat{\phi}_1$	-0.517 (0.049)	-0.615 (0.056)	-1.276 (0.098)	-0.500 (0.057)	-0.511 (0.057)
$\hat{\phi}_2$	-0.193 (0.047)	-0.261 (0.055)	-0.673 (0.089)	-0.268 (0.052)	-0.239 (0.052)
<i>n</i> = 400					
$\hat{\phi}_0$	0.031 (0.023)	0.062 (0.024)	0.242 (0.053)	0.041 (0.023)	0.041 (0.023)
$\hat{\phi}_1$	-0.056 (0.019)	-0.071 (0.024)	-0.289 (0.045)	-0.099 (0.021)	-0.102 (0.022)
$\hat{\phi}_2$	-0.043 (0.018)	-0.102 (0.023)	-0.124 (0.043)	-0.028 (0.020)	-0.021 (0.020)
<i>n</i> = 800					
$\hat{\phi}_0$	0.036 (0.016)	0.043 (0.016)	0.114 (0.037)	0.044 (0.016)	0.036 (0.016)
$\hat{\phi}_1$	-0.038 (0.012)	-0.023 (0.016)	-0.089 (0.031)	-0.025 (0.014)	-0.022 (0.014)
$\hat{\phi}_2$	0.009 (0.012)	-0.016 (0.015)	-0.014 (0.030)	-0.005 (0.014)	0.002 (0.014)

investigated in [10, 11]. Figure 3(a) displayed the time series. The normal probability plot in Figure 3(b) showed strong evidence for heavy-tailedness, which was further confirmed by the Jarque–Bera and Shapiro–Wilk tests. Figure 3(c) depicted the kernel density estimate for the log-returns, superimposed with normal and stable density functions. For the normal density we took the corresponding empirical mean (-4.356×10^{-4}) and standard deviation (2.250×10^{-2}) as the distributional mean and standard deviation. For the stable density we estimated the distribution parameters by fitting a stable distribution to the log-returns using the maximum likelihood method via the R package “fBasics”. The fitted stable distribution was $S(1.404, -0.204, 0.009, 0.001)$. It is seen that the estimated density was well approximated by the fitted stable density. Thus, it is plausible to model the log-returns as a non-Gaussian stable time series.

Large-scale simulations showed [1] that sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) are still excellent tools for tentative identification of the orders of a stable ARMA model. The sample PACF plot of the log-returns was given in Figure 3(d). The 95% bounds in the figure are modified bounds for the stable ARMA models computed based on Table 3 of [1]. To be specific, with a fitted index value of 1.404, the bounds were approximately $\pm 2.814(\log n/n)^{1/1.4} = \pm 0.124$, where $n = 490$ is the length of the time series. For comparison, the standard bounds under the classical Gaussian framework are $\pm 1.96/\sqrt{n} = \pm 0.089$. From the plot the sample partial autocorrelation is significantly outside the bounds at lag 3, slightly outside the bounds at lag 4, and insignificant at other lags. So it is reasonable to consider fitting the following three AR models:

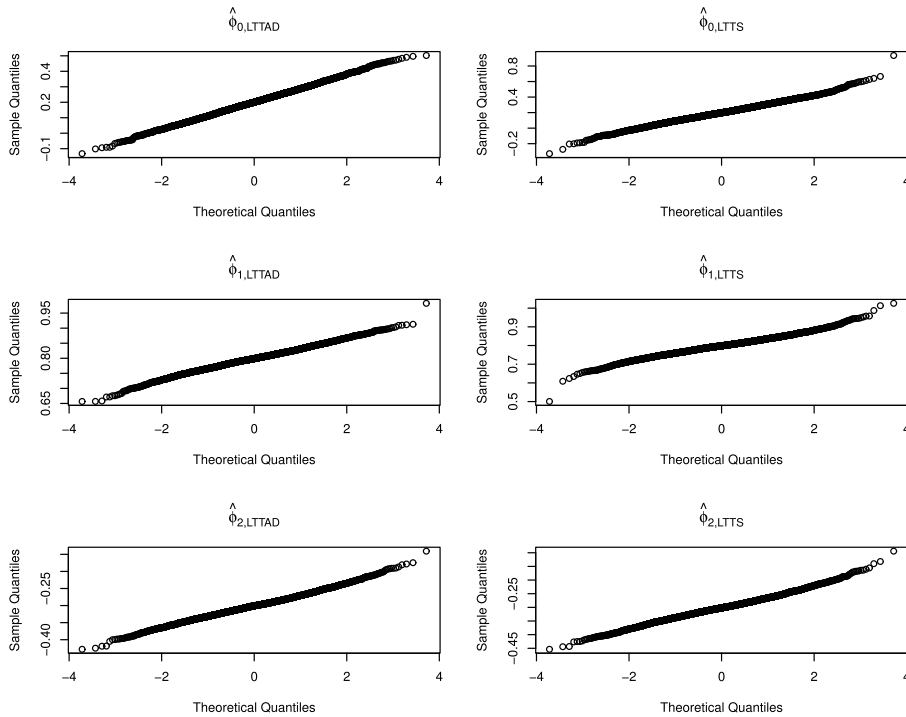


FIG 1. The normal probability plots of LTTAD and LTTTS estimates, where Z_t is Stable with tail index $\alpha = 1.5$ and $n = 400$

(M1) AR(4) model, $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + Z_t$,

(M2) AR(3) model, $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Z_t$,

(M3) subset AR(3) model, $X_t = \phi_3 X_{t-3} + Z_t$.

The intercept term was dropped (i.e. zero-mean AR models for the log-returns) due to the insignificance of the empirical mean. This is in agreement with the conclusion of [10, 11] that ϕ_0 is insignificant. Also, the model (M3) is the final model chosen by [11].

We first fitted the model (M1) using both the LTTAD and LTTTS procedures with the trimming parameters as specified in the simulation studies. The resulting LTTAD model was

$$X_t = \underset{(0.035)}{0.060} X_{t-1} - \underset{(0.035)}{0.121} X_{t-2} + \underset{(0.036)}{0.147} X_{t-3} + \underset{(0.037)}{0.026} X_{t-4} + Z_t$$

where $Z_t \sim S(1.372, -0.171, 0.009, 0.001)$. The standard errors beneath the coefficient estimates were computed based on Theorem 2.1. The fitted LTTTS model was

$$X_t = \underset{(0.045)}{0.080} X_{t-1} - \underset{(0.045)}{0.093} X_{t-2} + \underset{(0.046)}{0.122} X_{t-3} - \underset{(0.048)}{0.037} X_{t-4} + Z_t$$

where $Z_t \sim S(1.398, -0.190, 0.009, 0.001)$. The standard errors were computed

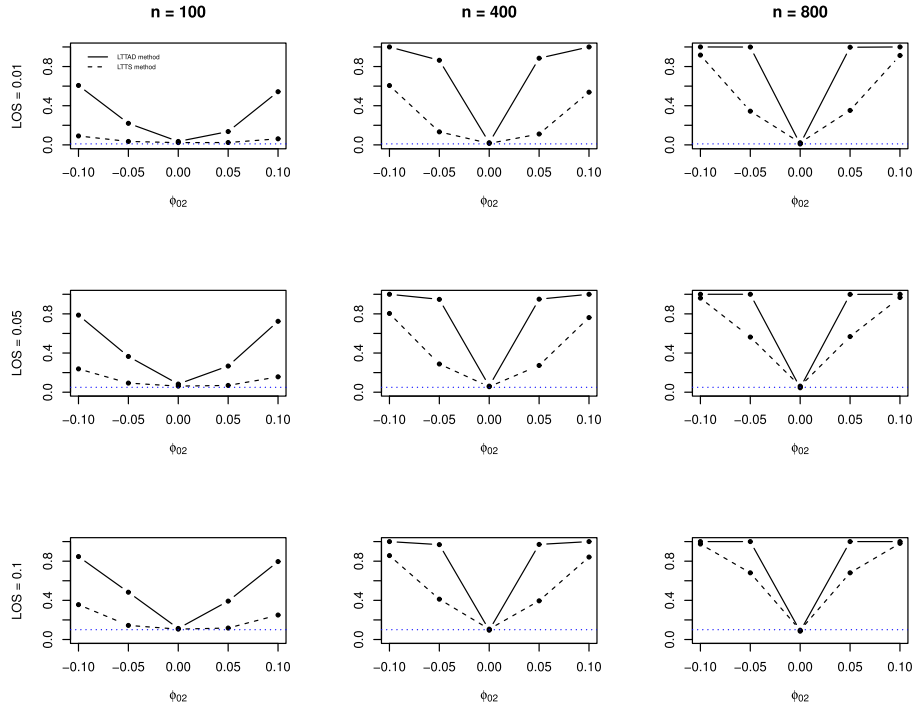


FIG 2. The empirical power/size of the Wald test, where Z_t is a two-sided compound Pareto distribution with tail index $\alpha = 1.5$

based on the asymptotic results of [10]. The two fitted models were comparable other than the opposite signs of $\hat{\phi}_4$, which, however, do not truly imply a contradiction. Indeed, ϕ_4 was insignificantly different from zero, as was not surprising according to the sample PACF plot of the log-returns. Thus we removed the term X_{t-4} and re-fitted the model (M2) to the data, leading to the LTAD model of

$$X_t = \underset{(0.034)}{0.063} X_{t-1} - \underset{(0.034)}{0.126} X_{t-2} + \underset{(0.035)}{0.152} X_{t-3} + Z_t,$$

$$Z_t \sim S(1.379, -0.179, 0.009, 0.001)$$

with $AIC = -7.623$, and the LTTS model of

$$X_t = \underset{(0.044)}{0.063} X_{t-1} - \underset{(0.045)}{0.096} X_{t-2} + \underset{(0.045)}{0.126} X_{t-3} + Z_t,$$

$$Z_t \sim S(1.378, -0.173, 0.009, 0.001)$$

with $AIC = -7.619$. So, both fitted AR(3) models were adequate but the LTAD model-fit was slightly better, which is also seen from the sample ACF plot of the residuals in Figure 4. The bounds in the sample ACF plot are again the modified bounds.

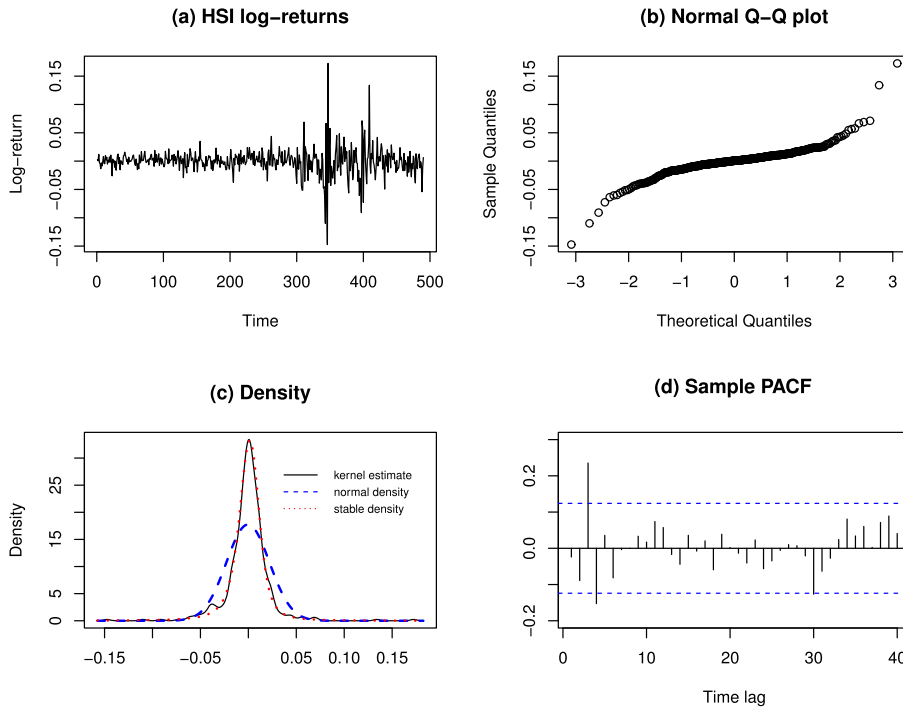


FIG 3. (a) The log-returns of Hang Seng Index, (b) normal q-q plot, (c) kernel density estimate, and (d) sample partial autocorrelation function.

Finally we considered fitting the log-returns using the model (M3). By the LTTAD method we obtained the model

$$X_t = \underset{(0.034)}{0.163} X_{t-3} + Z_t, Z_t \sim S(1.399, -0.185, 0.009, 0.001)$$

with $AIC = -7.634$, and the fitted model by the LTTS method was

$$X_t = \underset{(0.045)}{0.104} X_{t-3} + Z_t, Z_t \sim S(1.392, -0.185, 0.009, 0.001)$$

with $AIC = -7.621$. The diagnostic plots were presented in Figure 5. Again, the LTTAD model-fit was better than the LTTS model-fit although both were adequate. Also, based on the AIC values, the subset AR(3) model was a bit more plausible than the usual AR(3) model. So, we concluded, in agreement with [11], with a subset AR(3) model for fitting the log-returns of Hang Seng Index.

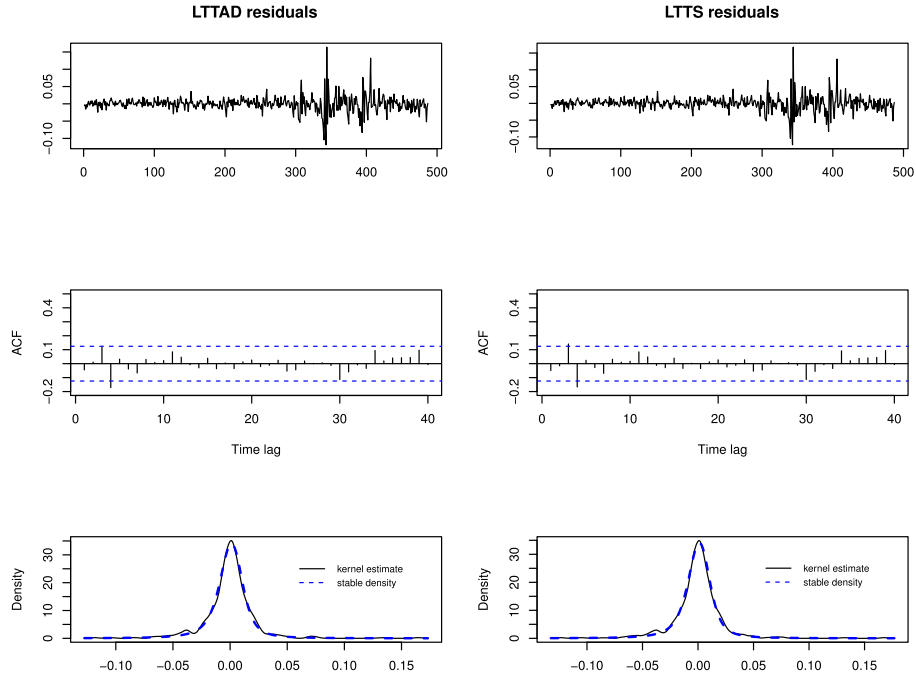


FIG 4. Diagnostic plots of residuals from the fitted AR(3) models (M2)

Appendix A: Technical proofs

We first establish a lemma, from which Theorem 2.1 follows. To be specific, we reparameterize by setting $\nu = n^{1/2}V_n^{1/2}(\phi - \phi_0)$ and hence can rewrite

$$z_t(\phi) = Z_t - n^{-1/2}\nu^T V_n^{-1/2}Y_t.$$

Then, minimizing $l_n(\phi)$ with respect to ϕ is equivalent to minimizing

$$S_n(\nu) := \sum_{t=p+1}^n \left(\left| Z_t - n^{-1/2}\nu^T V_n^{-1/2}Y_t \right| - |Z_t| \right) I_{nt} \tag{A.1}$$

with respect to ν . Stated below is the functional limit theorem for $S_n(\nu)$.

Lemma A.1. Under the Assumptions (A1)–(A3),

$$S_n(\nu) \xrightarrow{d} S(\nu) := -\nu^T N(\mathbf{0}, \mathbf{I}_{p+1}) + f(0)\nu^T \nu$$

on $C(\mathbb{R}^{p+1})$, the space of real-valued continuous functions on \mathbb{R}^{p+1} where convergence means uniform convergence on every compact set in \mathbb{R}^{p+1} .

Proof. We adopt the strategy used in [7]. Using the identity

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

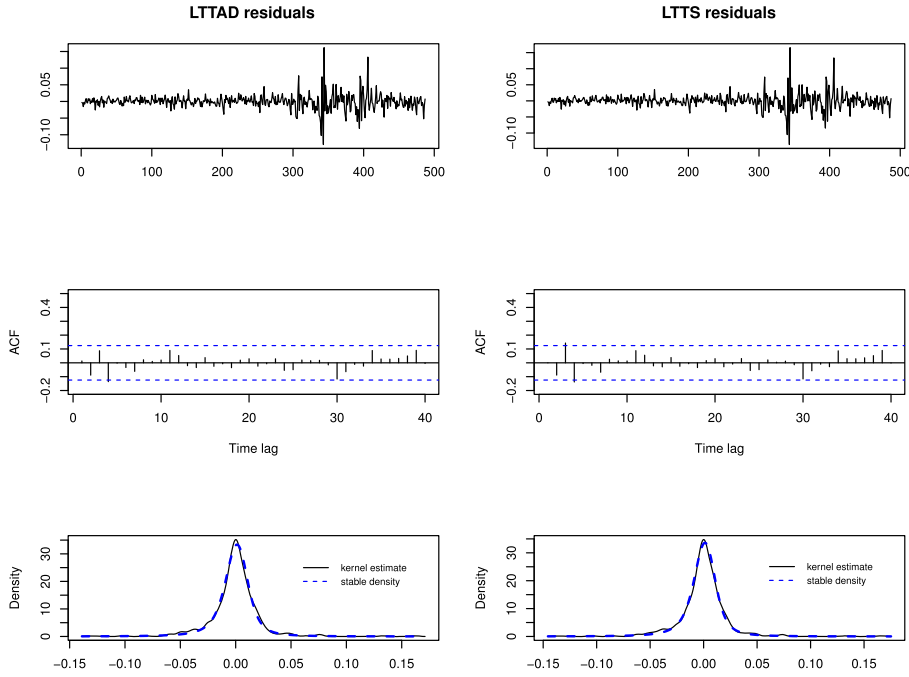


FIG 5. Diagnostic plots of residuals from the fitted subset AR(3) models (M3)

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

$$S_n(\boldsymbol{\nu}) = -\boldsymbol{\nu}^T \mathbf{A}_n + 2 \sum_{t=p+1}^n (\xi_{1nt}(\boldsymbol{\nu}) + \xi_{2nt}(\boldsymbol{\nu})), \tag{A.2}$$

where

$$\begin{aligned} \mathbf{A}_n &= n^{-1/2} \mathbf{V}_n^{-1/2} \sum_{t=p+1}^n \mathbf{Y}_t \text{sgn}(Z_t) I_{nt}, \\ \xi_{1nt}(\boldsymbol{\nu}) &= \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - Z_t \right) I(0 < Z_t < n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t) I_{nt}, \\ \xi_{2nt}(\boldsymbol{\nu}) &= - \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - Z_t \right) I(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t < Z_t < 0) I_{nt}. \end{aligned}$$

Define $\mathbf{A}_n^c = n^{-1/2} \sum_{t=p+1}^n \boldsymbol{\eta}_{nt}^c(\boldsymbol{\nu})$, where $\boldsymbol{\eta}_{nt}^c(\boldsymbol{\nu}) = \mathbf{V}_n^{-1/2} \mathbf{Y}_t \text{sgn}(Z_t) I_{nt}^c$. Applying a martingale central limit theorem to $\{\boldsymbol{\eta}_{nt}^c(\boldsymbol{\nu})\}$, we obtain $\mathbf{A}_n^c \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p+1})$. On the other hand, following the lines of [10] for proving his Lemma A1(a) it is easy to show that $\mathbf{A}_n - \mathbf{A}_n^c \xrightarrow{P} \mathbf{0}$. Therefore,

$$\mathbf{A}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p+1}). \tag{A.3}$$

Next, for given $\boldsymbol{\nu}$, define

$$\begin{aligned} \xi_{1nt}^c(\boldsymbol{\nu}) &= \left(n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t - Z_t\right) I(0 < Z_t < n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t)I_{nt}^c, \\ \xi_{2nt}^c(\boldsymbol{\nu}) &= -\left(n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t - Z_t\right) I(n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t < Z_t < 0)I_{nt}^c. \end{aligned}$$

Let \mathcal{F}_t be the sigma-field generated by $\{Z_s, s \leq t\}$. Then, for $t = p + 1, \dots, n$,

$$\begin{aligned} E(\xi_{1nt}^c(\boldsymbol{\nu}) \mid \mathcal{F}_{t-1}) &= \int_0^{n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t} \left(n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t - z\right) I_{nt}^c I(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t > 0)F(dz) \\ &\sim \int_0^{n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t} \left(n^{-1/2}\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t - z\right) f(0)I_{nt}^c I(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t > 0)dz \\ &= \frac{f(0)}{2n} \left(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t\right)^2 I_{nt}^c I(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t > 0), \end{aligned}$$

where the approximation holds on the set $\left\{0 \leq \boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}\right\}$, where $\varepsilon > 0$ stands for a small number. Note that,

$$\begin{aligned} &P\left(\max_{t=p+1, \dots, n} \left|\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t I_{nt}^c\right| \geq \varepsilon n^{1/2}\right) \\ &\leq nP\left(\left|\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_{p+1} I_{n,p+1}^c\right| \geq \varepsilon n^{1/2}\right) \\ &\leq \varepsilon^{-2}E\left[\left(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_{p+1}\right)^2 I_{n,p+1}^c I\left(\left|\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_{p+1} I_{n,p+1}^c\right| \geq \varepsilon n^{1/2}\right)\right] \\ &\rightarrow 0, \end{aligned}$$

where the convergence in the last step follows from the fact that

$$\begin{aligned} &E\left[\left(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_{p+1}\right)^2 I_{n,p+1}^c\right] \\ &= \boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}E\left(\mathbf{Y}_{p+1}\mathbf{Y}_{p+1}^T I_{n,p+1}^c\right)\mathbf{V}_n^{-1/2}\boldsymbol{\nu} = \boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{V}_n\mathbf{V}_n^{-1/2}\boldsymbol{\nu} = \boldsymbol{\nu}^T\boldsymbol{\nu} < \infty \end{aligned}$$

for given $\boldsymbol{\nu}$. Therefore,

$$\sum_{t=p+1}^n E(\xi_{1nt}^c(\boldsymbol{\nu}) \mid \mathcal{F}_{t-1}) \sim \frac{f(0)}{2n} \sum_{t=p+1}^n \left(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t\right)^2 I_{nt}^c I(\boldsymbol{\nu}^T\mathbf{V}_n^{-1/2}\mathbf{Y}_t > 0). \tag{A.4}$$

Moreover,

$$\begin{aligned}
& E \left[(\xi_{1nt}^c(\boldsymbol{\nu}))^2 \mid \mathcal{F}_{t-1} \right] \\
&= \int_0^{n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t} \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - z \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t > 0) F(dz) \\
&= \int_0^{n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t} \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - z \right)^2 \\
&\quad \cdot I_{nt}^c I(0 < \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}) F(dz) \\
&\quad + \int_0^{n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t} \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - z \right)^2 \\
&\quad \cdot I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c \geq \varepsilon n^{1/2}) F(dz) \\
&=: T_1 + T_2.
\end{aligned}$$

The term T_1 is upper-bounded by

$$\begin{aligned}
& B \int_0^{n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t} \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t - z \right)^2 \\
&\quad \cdot I_{nt}^c I(0 < \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}) dz \\
&= \frac{B}{3} n^{-3/2} \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^3 I_{nt}^c I(0 < \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}) \\
&\leq \frac{\varepsilon B}{3n} \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(0 < \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
T_2 &\leq \int_0^{n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t} \left(n^{-1/2} \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c \geq \varepsilon n^{1/2}) F(dz) \\
&\leq n^{-1} \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c \geq \varepsilon n^{1/2}).
\end{aligned}$$

So,

$$\begin{aligned}
nE \left[(\xi_{1nt}^c(\boldsymbol{\nu}))^2 \right] &= nE(T_1) + nE(T_2) \\
&\leq \frac{\varepsilon B}{3} E \left[\left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(0 < \boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c < \varepsilon n^{1/2}) \right] \\
&\quad + E \left[\left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t I_{nt}^c \geq \varepsilon n^{1/2}) \right]
\end{aligned}$$

Since $E[(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t)^2 I_{nt}^c]$ is finite and $\varepsilon > 0$ is arbitrary, we have

$$\limsup_{n \rightarrow \infty} nE \left[(\xi_{1nt}^c(\boldsymbol{\nu}))^2 \right] = 0. \tag{A.5}$$

It follows that, as $n \rightarrow \infty$,

$$\begin{aligned} & \text{Var} \left\{ \sum_{t=p+1}^n [\xi_{1nt}^c(\boldsymbol{\nu}) - E(\xi_{1nt}^c(\boldsymbol{\nu}) \mid \mathcal{F}_{t-1})] \right\} \\ &= \sum_{t=p+1}^n \text{Var} [\xi_{1nt}^c(\boldsymbol{\nu}) - E(\xi_{1nt}^c(\boldsymbol{\nu}) \mid \mathcal{F}_{t-1})] \leq \sum_{t=p+1}^n E [(\xi_{1nt}^c(\boldsymbol{\nu}))^2] \rightarrow 0. \end{aligned}$$

Therefore, (A.4) implies

$$\sum_{t=p+1}^n \xi_{1nt}^c(\boldsymbol{\nu}) \sim \frac{f(0)}{2n} \sum_{t=p+1}^n \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t > 0). \tag{A.6}$$

Handling the term $\xi_{2nt}^c(\boldsymbol{\nu})$ in the same fashion yields

$$\sum_{t=p+1}^n \xi_{2nt}^c(\boldsymbol{\nu}) \sim \frac{f(0)}{2n} \sum_{t=p+1}^n \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c I(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t < 0). \tag{A.7}$$

Combining (A.6) and (A.7) leads to

$$2 \sum_{t=p+1}^n (\xi_{1nt}^c(\boldsymbol{\nu}) + \xi_{2nt}^c(\boldsymbol{\nu})) \sim \frac{f(0)}{n} \sum_{t=p+1}^n \left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c,$$

which converges in probability to $f(0)E \left[\left(\boldsymbol{\nu}^T \mathbf{V}_n^{-1/2} \mathbf{Y}_t \right)^2 I_{nt}^c \right] = f(0)\boldsymbol{\nu}^T \boldsymbol{\nu}$. Moreover, it can be shown that, for any fixed $\boldsymbol{\nu}$, $\sum_{t=p+1}^n [\xi_{int}(\boldsymbol{\nu}) - \xi_{int}^c(\boldsymbol{\nu})] \xrightarrow{P} 0$, $i = 1, 2$. Therefore,

$$2 \sum_{t=p+1}^n (\xi_{1nt}(\boldsymbol{\nu}) + \xi_{2nt}(\boldsymbol{\nu})) \xrightarrow{P} f(0)\boldsymbol{\nu}^T \boldsymbol{\nu}. \tag{A.8}$$

Putting together (A.2), (A.3), and (A.8), we conclude that $S_n(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ as $n \rightarrow \infty$ for each $\boldsymbol{\nu}$.

Now, by Theorem 29.4 of [2], the finite dimensional distributions of S_n converge to those of S . On the other hand, since S_n has convex sample paths, the tightness of S_n can be established using the result that pointwise convergence of convex functions implies uniform convergence on compact sets [12]. Therefore, by Theorem 7.1 of [3], $S_n(\boldsymbol{\nu}) \xrightarrow{d} S(\boldsymbol{\nu})$ on $C(K)$, where K is any compact set in \mathbb{R}^{p+1} . As this is equivalent to convergence in distribution on $C(\mathbb{R}^{p+1})$, the proof is complete. \square

Proof of Theorem 2.1. Note that, the limiting process $S(\boldsymbol{\nu})$ is a quadratic function of $\boldsymbol{\nu}$ and hence has a unique minimizer $N(\mathbf{0}, \mathbf{I}_{p+1})/(2f(0))$. Therefore, the asymptotic normality of the LTTAD estimator follows immediately from the above lemma and Lemma 2.2 of [8]. \square

Acknowledgements

We thank the two referees and Associate Editor for the valuable comments. The research of Rongning Wu was supported in part by a PSC-CUNY award, jointly funded by The Professional Staff Congress and The City University of New York.

References

- [1] ADLER, R. J., FELDMAN, R. E. AND GALLAGHER, C. (1998). Analysing stable time series. In *A practical guide to heavy tails* (eds R. J. Adler, R. E. Feldman and M. S. Taqqu), 133–158. Birkhäuser, Boston. [MR1652286](#)
- [2] BILLINGSLEY, P. (1995). *Probability and measure*. 3rd ed. John Wiley and Sons, New York. [MR1324786](#)
- [3] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*. 2nd ed. John Wiley and Sons, New York. [MR1700749](#)
- [4] CLINE, D. (1983). Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data. Ph.D. Dissertation, Department of Statistics, Colorado State University, Fort Collins, CO. [MR2633007](#)
- [5] CUI, Y., FISHER, T. J. AND WU, R. (2014). Diagnostic tests for non-causal time series with infinite variance. *Journal of Statistical Planning and Inference* 142, 117–131. [MR3151850](#)
- [6] DAVIS, R. A. (2010). Heavy tails in financial time series. In *Encyclopedia of Quantitative Finance* (ed R. Cont). Wiley, New York.
- [7] DAVIS, R. A. AND DUNSMUIR, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497. [MR1455154](#)
- [8] DAVIS, R. A., KNIGHT, K. AND LIU, J. (1992). M -estimation for autoregressions with infinite variance. *Stochastic Processes and their Applications* 40, 145–180. [MR1145464](#)
- [9] DAVIS, R. A. AND RESNICK, S. I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* 14, 533–558. [MR0840513](#)
- [10] HILL, J. B. (2013). Least tail-trimmed squares for infinite variance autoregressions. *Journal of Time Series Analysis* 34, 168–186. [MR3028364](#)
- [11] LING, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society, Series B* 67, 381–393. [MR2155344](#)
- [12] ROCKAFELLAR, R. T. (1970). *Convex analysis*. Princeton University Press, Princeton, New Jersey. [MR0274683](#)