Ann Inst Stat Math (2019) 71:1121–1142
https://doi.org/10.1007/s10463-018-0671-z

Maximum likelihood estimation of autoregressive models with a near unit root and Cauchy errors

Jungjun Choi[1] · **In Choi**[2]

Received: 20 December 2016 / Revised: 29 April 2018 / Published online: 14 June 2018

Abstract This paper studies maximum likelihood estimation of autoregressive models of order 1 with a near unit root and Cauchy errors. Autoregressive models with an intercept and with an intercept and a linear time trend are also considered. The maximum likelihood estimator (MLE) for the autoregressive coefficient is $n^{3/2}$-consistent with n denoting the sample size and has a mixture-normal distribution in the limit. The MLE for the scale parameter of Cauchy distribution is $n^{1/2}$-consistent, and its limiting distribution is normal. The MLEs of the intercept and the linear time trend are $n^{1/2}$- and $n^{3/2}$-consistent, respectively. It is also shown that the t statistic for the null hypothesis of a unit root based on the MLE has a standard normal distribution in the limit. In addition, finite-sample properties of the MLE are compared with those of the least square estimator (LSE). It is found that the MLE is more efficient than the LSE when the errors have a Cauchy distribution or a distribution which is a mixture of Cauchy and normal distributions. It is also shown that empirical power of the MLE-based t test for a unit root is much higher than that of the Dickey–Fuller t test.

This paper stemmed from the MA thesis of the first author which was supervised by the second author.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-018-0671-z) contains supplementary material, which is available to authorized users.

✉ In Choi
inchoi@gmail.com
http://inchoi.sogang.ac.kr

Jungjun Choi
jchoi@economics.rutgers.edu

[1] Department of Economics, Rutgers University, New Brunswick, NJ 08901-1248, USA

[2] School of Economics, Sogang University, Seoul 04107, Korea

Keywords Autoregressive model · Near unit root · Cauchy distribution · Maximum likelihood estimator · Infinite variance

1 Introduction

Since the original works of Fama (1965) and Mandelbrot (1963, 1967), researchers have long been interested in studying properties of stochastic processes with infinite variance. Empirical papers applying econometric models with infinite variance to financial data include, among others, Akgiray and Booth (1988), Lau et al. (1990), Falk and Wang (2003), and Koedijk and Kool (1992).

Because asymptotic properties of the stochastic processes with infinite variance are significantly different from those with finite variance, it is important to study properties of econometric estimators and tests under the assumption of infinite variance. For the AR(1) model (autoregressive model of order 1) with a unit root and errors having infinite variance, Chan and Tran (1989) and Phillips (1990) study the least squares estimator (LSE). Chan and Tran (1989) derive limiting distributions of the Dickey–Fuller test statistics, while Phillips (1990) studies those of the Phillips–Perron test statistics. Extensions of these works to the AR(1) model with an intercept and with an intercept and a linear time trend are made in Ahn et al. (2001) and Callegari et al. (2003), respectively. See Choi (2015, pp.75–78) for further discussions.

It is well known that LSEs are not robust to outliers and less efficient than M- and least absolute deviation (LAD) estimators for fat-tailed errors. Pollard (1991) derives convergence rate of the LAD estimator for the causal AR(1) model with Cauchy errors. Knight (1989, 1991) studies the M- and LAD estimators for the unit root AR(1) model with i.i.d. errors that lie in the domain of attraction of a stable law with the index of stability $\alpha \in (0, 2)$. Andrews et al. (2009) study asymptotics of the maximum likelihood estimator (MLE) for AR coefficients in stationary processes with $\alpha \in (0, 2)$. Furthermore, Zhang and Chan (2012) study the MLE for a near unit root AR(1) process with i.i.d. α-stable errors. They estimate the AR(1) coefficient and the parameters of the characteristic function of stable errors jointly by MLE. They find that the rate of convergence of the MLE of the AR(1) coefficient depends on the stability parameter α and the mean parameter of errors and that the MLE is more efficient than LSE in finite samples. But the case of $\alpha = 1$ (i.e., Cauchy errors) is not considered in their paper, probably because they estimate the stable parameter α and because there is a discontinuity of the characteristic function with respect to α at $\alpha = 1$.

The main purpose of this paper is to derive limiting distributions of MLEs for nearly non-stationary AR(1) models with Cauchy errors. This is not considered by Zhang and Chan (2012) as mentioned above. Although the assumption of Cauchy errors is more specialized than that of α-stable errors, the Cauchy distribution has a long history in statistics and it seems worthwhile to study the MLE under Cauchy errors. Another difference of this paper and Zhang and Chan (2012) is that this paper also considers the AR(1) models with an intercept and with an intercept and a linear time trend, while Zhang and Chan study only the AR(1) model without nonstochastic regressors.

There are some contributions of this paper. First, it shows that the MLE of the AR(1) coefficient converges faster than the LSE (the convergence rate of MLE is $n^{3/2}$, while that of LSE is n with n denoting the sample size) and that it has a mixture-normal distribution in the limit. The former finding implies that the MLE is more efficient than the LSE for Cauchy errors. Simulation results also show that the MLE based on the assumption of Cauchy errors performs better than the LSE.

Second, it is shown that the MLE-based t statistic for the AR(1) coefficient has a standard normal distribution in the limit. The t statistic can be used to test the null hypothesis of a unit root. By contrast, the LSE-based test statistics for a unit root have a nonstandard distribution in the limit (cf. Phillips 1990). Samarakoon and Knight (2009) propose test statistics for the unit root null hypothesis that are based on M-estimators. These statistics converge in distribution to normal distributions under some regularity conditions. But these limiting normal distributions include nuisance parameters which have to be estimated. The t-ratio of this paper, however, does not involve any nuisance parameters in its limiting distribution.

This paper is planned as follows. Section 2 introduces the MLE for the AR(1) model without nonstochastic terms and derives its limiting distribution. Section 3 extends the results of Sect. 2 to the AR(1) models with nonstochastic regressors. Section 4 proposes MLE-based t test for the AR(1) coefficient and derives its limiting distribution. Section 5 contains a summary of simulation results, the details of which are relegated to a supplementary file. Section 6 provides a summary of the paper and further remarks. Proofs are in appendices.

There are a few words on our notation. $\mathbb{R}$, $\mathbb{R}^+$ and $\mathbb{N}$ denote the sets of real numbers, positive real numbers and natural numbers, respectively. Weak convergence in the Skorokhod topology and convergence in probability are denoted by $\overset{d}{\to}$ and $\overset{p}{\to}$, respectively. $A \overset{d}{=} B$ signifies that A and B have the same distribution. $D[0, 1]$ denotes the space of right-continuous functions with left limit on the unit interval. The (i, i)-th element of the matrix X is written as X_{ii}.

2 Maximum likelihood estimation of the AR(1) model with Cauchy errors

This section studies asymptotic properties of the MLE for the AR(1) model with a near unit root and Cauchy errors. The model we are concerned with is

$$Y_t = \rho_n Y_{t-1} + \varepsilon_t \ (t = 1, 2, \dots, n), \tag{1}$$

where $\rho_n = 1 - \gamma/n$ ($\gamma \in \mathbb{R}$), Y_0 is a fixed constant and $\{\varepsilon_t\}$ is a sequence of i.i.d. Cauchy errors with pdf $f(\varepsilon_1, \sigma) = \frac{\sigma}{\pi(\varepsilon_1^2+\sigma^2)}$ ($\sigma \in \mathbb{R}^+$, $\varepsilon_1 \in \mathbb{R}$).

Given observations $Y_0, Y_1, \dots, Y_n$, the likelihood function $L(\rho_n, \sigma)$ of the observations is written as

$$\begin{aligned} L(\rho_n, \sigma) &= \prod_{t=2}^{n} f(y_t - \rho_n y_{t-1}, \sigma) \\ &= \prod_{t=2}^{n} \frac{\sigma}{\pi((y_t - \rho_n y_{t-1})^2 + \sigma^2)}. \end{aligned}$$

The MLE of (ρ_n, σ) is defined by

$$(\hat{\rho}_n, \hat{\sigma}_n) = \arg\max_{\rho_n, \sigma} L(\rho_n, \sigma).$$

Using the standard method for MLEs (cf. Newey and McFadden 1994), we derive the limiting properties of $\hat{\rho}_n$ and $\hat{\sigma}_n$ as follows.

Theorem 1 *Suppose that the true value of (ρ_n, σ) is (ρ_n^o, σ^o), where $\rho_n^o = 1 - \gamma^o/n$, $\gamma^o \in \mathbb{R}$ and $\sigma^o \in \mathbb{R}^+$. Assume $(\rho_n^o, \sigma^o) \in \Theta$ for every n, where Θ is a compact subset of $\mathbb{R} \times \mathbb{R}^+$. Then, as $n \to \infty$,*

(i) $\hat{\rho}_n - \rho_n^o \xrightarrow{p} 0$ *and* $\hat{\sigma}_n \xrightarrow{p} \sigma^o$;

(ii)

$$\begin{pmatrix} n^{3/2}(\hat{\rho}_n - \rho_n^o) \\ n^{1/2}(\hat{\sigma}_n - \sigma^o) \end{pmatrix} \xrightarrow{d} 2\sigma^{o2} \begin{pmatrix} \int_0^1 S^2(r)\mathrm{d}r & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 S(r)\mathrm{d}B(r) \\ Q(1) \end{pmatrix},$$

where $S(r) = Z(r) - \gamma^o \int_0^r e^{\gamma^o(s-r)} Z(s)\mathrm{d}s$, $Z(s)$ is a Cauchy process with characteristic function $\phi(t) = \exp\{-(s\sigma^o)|t|\}$ and $(B(r), Q(s))'$ is a Brownian motion with mean zero and the covariance matrix $\begin{pmatrix} \frac{r}{2\sigma^{o2}} & 0 \\ 0 & \frac{s}{2\sigma^{o2}} \end{pmatrix}$.

This theorem shows that the MLE for the autoregressive coefficient is $n^{3/2}$-consistent and converges in distribution to a random variable involving a Cauchy process and a Brownian motion and that the MLE for the scale parameter of Cauchy distribution has the conventional $\sqrt{n}$-consistency property and is asymptotically normal. Since the convergence rate of the LSE for the AR(1) coefficient is n (cf. Chan and Tran 1989; Phillips 1990), we find that the MLE is more efficient than the LSE when the errors are from a Cauchy distribution.

Since $S(r)$ and $B(s)$ are independent for $0 \leq r \leq 1$ and $0 \leq s \leq 1$ (cf. Resnick and Greenwood 1979), the result for the MLE of ρ_n^o can be written as

$$\begin{aligned} n^{3/2}(\hat{\rho}_n - \rho_n^o) &\xrightarrow{d} 2\sigma^{o2} \frac{\int_0^1 S(r)\mathrm{d}B(r)}{\int_0^1 S(r)^2 \mathrm{d}r} \\ &\stackrel{d}{=} N\left(0, \frac{2\sigma^{o2}}{\int_0^1 S(r)^2 \mathrm{d}r}\right)_{|\{S(r):0\leq r\leq 1\}} \\ &\stackrel{d}{=} N\left(0, \frac{2}{\int_0^1 S^*(r)^2 \mathrm{d}r}\right)_{|\{S^*(r):0\leq r\leq 1\}}, \end{aligned} \tag{2}$$

where $S^*(r) = S(r)/\sigma^o = Z^*(r) - \gamma^o \int_0^r e^{\gamma^o(s-r)} Z^*(s)\mathrm{d}s$, $Z^*(s)$ is a standard Cauchy process with characteristic function $\phi(t) = \exp\{-s|t|\}$. The first equality relation follows from Arnold (1974, p. 77). Relation (2) shows that $n^{3/2}(\hat{\rho}_n - \rho_n^o)$ has a mixture-normal distribution in the limit and does not depend on the scale parameter σ^o.

In addition, we deduce from part (ii) of Theorem 1

$$n^{1/2}(\hat{\sigma}_n - \sigma^o) \xrightarrow{d} N(0, 2\sigma^{o2}).$$

By applying the delta method to this relation, the asymptotic distribution of $\hat{\sigma}_n^2$ can be derived as

$$n^{1/2}(\hat{\sigma}_n^2 - \sigma^{o2}) \xrightarrow{d} 2\sigma^o \cdot N(0, 2\sigma^{o2}) \stackrel{d}{=} N(0, 8\sigma^{o4}).$$

The limiting distributions of the MLEs of ρ_n^o and σ^o are independent.

Assuming a stable distribution for $\{\varepsilon_t\}$, Zhang and Chan (2012) derive the limiting distributions of the MLEs of ρ_n^o and other parameters associated with the error distribution. But they do not consider the case of Cauchy errors, perhaps because it brings complications in estimating the parameters associated with the error distribution (see DuMouchel 1973, for discussion on the parameter space for the MLE of the parameters related to stable distributions). Thus, Theorem 1 is not a special case of Theorem 1 of Zhang and Chan (2012). Moreover, it seems difficult to find any clue regarding asymptotic distribution of the MLE of ρ_n^o under Cauchy errors ffrom the results of Zhang and Chan (2012).

3 Extensions to the AR(1) models with nonstochastic regressors

This section extends the results of the previous section to the AR(1) models with a near unit root and nonstochastic regressors.

3.1 AR(1) model with an intercept

This subsection considers an unobserved components model

$$Y_t = a_0 + X_t,\ X_t = \rho_n X_{t-1} + \varepsilon_t, (a_0 \in \mathbb{R}).$$

This model can be written as

$$Y_t = \mu + \rho_n Y_{t-1} + \varepsilon_t,\ \ (t = 1, 2, \ldots, n), \tag{3}$$

where $\mu = a_0(1 - \rho_n)$.

The likelihood function for model (3) is given as

$$\begin{aligned} L(\rho_n, \mu, \sigma) &= \prod_{t=2}^{n} f(y_t - \mu - \rho_n y_{t-1}, \sigma) \\ &= \prod_{t=2}^{n} \frac{\sigma}{\pi((y_t - \mu - \rho_n y_{t-1})^2 + \sigma^2)}. \end{aligned}$$

The MLE of (ρ_n, μ, σ) is defined by

$$(\hat{\rho}_n, \hat{\mu}_n, \hat{\sigma}_n) = \arg \max_{\rho_n, \mu, \sigma} L(\rho_n, \mu, \sigma).$$

The limiting properties of $\hat{\rho}_n$, $\hat{\mu}_n$ and $\hat{\sigma}_n$ are reported in the following theorem.

Theorem 2 *Suppose that the true value of* (ρ_n, μ, σ) *is* $(\rho_n^o, \mu^o, \sigma^o)$, *where* $\rho_n^o = 1 - \gamma^o/n$, $\gamma^o \in \mathbb{R}$, $\mu^o \in \mathbb{R}$ *and* $\sigma^o \in \mathbb{R}^+$. *Assume* $(\rho_n^o, \mu^o, \sigma^o) \in \Theta$ *for every n, where* Θ *is a compact subset of* $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$. *Then, as* $n \to \infty$,

(i) $\hat{\rho}_n - \rho_n^o \xrightarrow{p} 0$, $\hat{\mu}_n \xrightarrow{p} \mu^o$ *and* $\hat{\sigma}_n \xrightarrow{p} \sigma^o$;
(ii)

$$\begin{pmatrix} n^{3/2}(\hat{\rho}_n - \rho_n^o) \\ n^{1/2}(\hat{\mu}_n - \mu^o) \\ n^{1/2}(\hat{\sigma}_n - \sigma^o) \end{pmatrix} \xrightarrow{d} 2\sigma^{o2} \begin{pmatrix} \int_0^1 S^2(r)\mathrm{d}r & \int_0^1 S(r)\mathrm{d}r & 0 \\ \int_0^1 S(r)\mathrm{d}r & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 S(r)\mathrm{d}B(r) \\ B(1) \\ Q(1) \end{pmatrix},$$

where $S(r)$ *and* $(B(r), Q(s))$ *are as defined in Theorem 1.*

This theorem shows that the asymptotic distribution of the MLE of ρ_n^o changes when an intercept term is included in the model while that of σ^o does not. But the rates of convergence of $\hat{\rho}_n$ and $\hat{\sigma}_n$ are the same as in Theorem 1. The MLE of the intercept term is $\sqrt{n}$-consistent.

Letting $V(r) = (S(r), 1)'$, the results of Theorem 2 can be rewritten as

$$\begin{pmatrix} n^{3/2}(\hat{\rho}_n - \rho_n^o) \\ n^{1/2}(\hat{\mu}_n - \mu^o) \end{pmatrix} \xrightarrow{d} 2\sigma^{o2} \left(\int_0^1 V(r)V(r)'\mathrm{d}r\right)^{-1} \int_0^1 V(r)\mathrm{d}B(r)$$
$$\stackrel{d}{=} N\left(0, 2\sigma^o \left(\int_0^1 V(r)V(r)'\mathrm{d}r\right)^{-1}\right)_{|\{S(r):0\le r\le 1\}}. \tag{4}$$

This shows that the MLE of (ρ_n^o, μ^o) has a multivariate mixture-normal distribution in the limit. Moreover, we obtain from relation (4)

$$n^{3/2}(\hat{\rho}_n - \rho_n^o) \xrightarrow{d} 2\sigma^{o2} \left(\int_0^1 \bar{S}(r)^2\mathrm{d}r\right)^{-1} \int_0^1 \bar{S}(r)\mathrm{d}B(r)$$
$$\stackrel{d}{=} N\left(0, \frac{2}{\int_0^1 \bar{S}^*(r)^2\mathrm{d}r}\right)_{|\{S^*(r):0\le r\le 1\}},$$

where $\bar{S}(r) = S(r) - \int_0^1 S(r)\mathrm{d}r$ and $\bar{S}^*(r) = S^*(r) - \int_0^1 S^*(r)\mathrm{d}r$. This shows that the limiting distribution of $\hat{\rho}_n$ is free of nuisance parameters as in Theorem 1.

3.2 AR(1) model with an intercept and a linear time trend

This subsection considers an unobserved components model

$$Y_t = a_0 + b_0 t + X_t,\ X_t = \rho_n X_{t-1} + \varepsilon_t,\ (a_0, b_0 \in \mathbb{R}),$$

which can be rewritten as

$$Y_t = \mu + \beta t + \rho_n Y_{t-1} + \varepsilon_t,\ (t = 1, 2, \ldots, n), \tag{5}$$

where $\mu = a_0(1-\rho_n) + \rho_n b_0$, $\beta = b_0(1-\rho_n)$.

The likelihood function $L(\rho_n, \mu, \beta, \sigma)$ for model (5) is given as

$$L(\rho_n, \mu, \beta, \sigma) = \prod_{t=2}^{n} f(y_t - \mu - \beta t - \rho_n y_{t-1}, \sigma) = \prod_{t=2}^{n} \frac{\sigma}{\pi((y_t - \mu - \beta t - \rho_n y_{t-1})^2 + \sigma^2)},$$

and the MLE of $(\rho_n, \mu, \beta, \sigma)$ is defined by

$$(\hat{\rho}_n, \hat{\mu}_n, \hat{\beta}_n, \hat{\sigma}_n) = \arg \max_{\rho_n, \mu, \beta, \sigma} L(\rho_n, \mu, \beta, \sigma).$$

The limiting properties of $\hat{\rho}_n$, $\hat{\mu}_n$,$\hat{\beta}_n$ and $\hat{\sigma}_n$ are reported in the following theorem.

Theorem 3 *Suppose that the true value of $(\rho_n, \mu, \beta, \sigma)$ is $(\rho_n^o, \mu^o, \beta^o, \sigma^o)$, where $\rho_n^o = 1 - \gamma^o/n$, $\gamma^o \in \mathbb{R}$, $\mu^o \in \mathbb{R}$, $\beta^o \in \mathbb{R}$ and $\sigma^o \in \mathbb{R}^+$. Assume $(\rho_n^o, \mu^o, \beta^o, \sigma^o) \in \Theta$ for every n, where Θ is a compact subset of $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$. Then, as $n \to \infty$,*

(i) $\hat{\rho}_n - \rho_n^o \xrightarrow{p} 0$, $\hat{\mu}_n \xrightarrow{p} \mu^o$, $\hat{\beta}_n \xrightarrow{p} \beta^o$ *and* $\hat{\sigma}_n \xrightarrow{p} \sigma^o$;
(ii)

$$\begin{pmatrix} n^{3/2}\left(\hat{\rho}_n - \rho_n^o\right) \\ n^{1/2}\left(\hat{\mu}_n - \mu^o\right) \\ n^{3/2}(\hat{\beta}_n - \beta^o) \\ n^{1/2}\left(\hat{\sigma}_n - \sigma^o\right) \end{pmatrix} \xrightarrow{d} 2\sigma^{o2} \begin{pmatrix} \int_0^1 U^2(r)\mathrm{d}r & \int_0^1 U(r)\mathrm{d}r & \int_0^1 rU(r)\mathrm{d}r & 0 \\ \int_0^1 U(r)\mathrm{d}r & 1 & \frac{1}{2} & 0 \\ \int_0^1 rU(r)\mathrm{d}r & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \times \begin{pmatrix} \int_0^1 U(r)\mathrm{d}B(r) \\ B(1) \\ \int_0^1 r\mathrm{d}B(r) \\ Q(1) \end{pmatrix}$$

where $U(r) = S(r) + b_0 r$, $S(r)$ and $(B(r), Q(s))$ are as defined in Theorem 1.

Because $n^{1/2}(\hat{\sigma}_n - \sigma^o) \xrightarrow{d} 2\sigma^{o2}Q(1)$, $\hat{\sigma}_n$ has the same limiting distribution as in Theorems 1 and 2. But the rest have limiting distributions different from those in Theorems 1 and 2.

Let $R(r) = (U(r), 1, r)$. Then, the results of Theorem 2 can be rewritten as

$$\begin{pmatrix} n^{3/2}(\hat{\rho}_n - \rho_n^o) \\ n^{1/2}(\hat{\mu}_n - \mu^o) \\ n^{3/2}(\hat{\beta}_n - \beta^o) \end{pmatrix} \xrightarrow{d} 2\sigma^{o2} \left(\int_0^1 R(r)R(r)'\mathrm{d}r\right)^{-1} \int_0^1 R(r)\mathrm{d}B(r) \stackrel{d}{=} N\left(0, 2\sigma^{o2}\left(\int_0^1 R(r)R(r)'\mathrm{d}r\right)^{-1}\right)_{|\{S(r):0\le r\le 1\}}, \quad (6)$$

which shows that the MLEs of ρ_n^o, μ^o and β^o have a multivariate mixture-normal distribution in the limit. The marginal limiting distribution of ρ_n^o is

$$\begin{aligned} n^{3/2}(\hat{\rho}_n - \rho_n^o) &\xrightarrow{d} 2\sigma^{o2}\left(\int_0^1 \tilde{S}(r)^2 \mathrm{d}r\right)^{-1} \int_0^1 \tilde{S}(r)\mathrm{d}B(r) \\ &\stackrel{d}{=} N\left(0, \frac{2}{\int_0^1 \tilde{S}^*(r)^2\mathrm{d}r}\right)_{|\{S^*(r):0\le r\le 1\}}, \end{aligned} \tag{7}$$

where $\tilde{S}(r) = S(r) - 4\left(\int_0^1 S(r)\mathrm{d}r - \frac{3}{2}\int_0^1 rS(r)\mathrm{d}r\right) + 6r\left(\int_0^1 S(r)\mathrm{d}r - 2\int_0^1 rS(r)\mathrm{d}r\right)$ and $\tilde{S}^*(r)$ is similarly defined. This shows that no nuisance parameters are involved in the limiting distribution of $\hat{\rho}_n$ as in Theorems 1 and 2. Note that $\tilde{S}(r)$ is the residual from the continuous time regression

$$S(r) = \hat{\alpha}_0 + \hat{\alpha}_1 r + \tilde{S}(r)$$

where $(\hat{\alpha}_0, \hat{\alpha}_1)$ minimizes the least squares criterion $\int_0^1 (S(r) - \alpha_0 - \alpha_1 r)^2 \mathrm{d}r$ (cf. Park and Phillips 1988) and that the Frisch-Waugh theorem is used to derive relation (7) from relation (6).

4 Tests for autoregressive coefficients

This section studies t test for the autoregressive coefficient ρ_n using MLEs. The t statistics for the null hypothesis $H_o : \rho_n = \rho_n^o$ are defined as

$$t(\hat{\rho}_n) = \frac{\hat{\rho}_n - \rho_n^o}{\hat{\sigma}_{\hat{\rho}_n}},$$

where

$$\hat{\sigma}^2_{\hat{\rho}_n} = \begin{cases} \left(-\sum_{t=1}^n \frac{\partial^2 \ln f}{\partial \rho_n^2}{}_{|\theta_n = \hat{\theta}_n}\right)^{-1} & \text{for model (1)} \\ \left[-\sum_{t=1}^n \begin{pmatrix} \frac{\partial^2 \ln f}{\partial \rho_n^2} & \frac{\partial^2 \ln f}{\partial \rho_n \partial \mu} \\ \frac{\partial^2 \ln f}{\partial \mu \partial \rho_n} & \frac{\partial^2 \ln f}{\partial \mu^2} \end{pmatrix}_{|(\rho_n,\mu)=(\hat{\rho}_n,\hat{\mu}_n)}\right]^{-1}_{11} & \text{for model (3)} \\ \left[-\sum_{t=1}^n \begin{pmatrix} \frac{\partial^2 \ln f}{\partial \rho_n^2} & \frac{\partial^2 \ln f}{\partial \rho_n \partial \mu} & \frac{\partial^2 \ln f}{\partial \rho_n \partial \beta} \\ \frac{\partial^2 \ln f}{\partial \mu \partial \rho_n} & \frac{\partial^2 \ln f}{\partial \mu^2} & \frac{\partial^2 \ln f}{\partial \mu \partial \beta} \\ \frac{\partial^2 \ln f}{\partial \beta \partial \rho_n} & \frac{\partial^2 \ln f}{\partial \beta \partial \mu} & \frac{\partial^2 \ln f}{\partial \beta^2} \end{pmatrix}_{|(\rho_n,\mu,\beta)=(\hat{\rho}_n,\hat{\mu}_n,\hat{\beta}_n)}\right]^{-1}_{11} & \text{for model (5)} \end{cases}.$$

The limiting distributions of the t statistics are reported in the following theorem.

Theorem 4 *For models* (1), (3) *and* (5), *we have under* $H_0 : \rho_n = \rho_n^o$, *as* $n \to \infty$,

$$t(\hat{\rho}_n) \xrightarrow{d} N(0, 1).$$

It is shown that the t statistics for the autoregressive coefficient ρ_n using MLEs have a standard normal distribution in the limit. This follows from the mixture-normal limiting distributions of the MLEs and the block-diagonal structure of the information matrices in the limit.

When $\rho_n^o = 1$ (i.e., $\gamma^o = 0$), this theorem can be used to test the null hypothesis of a unit root. In addition, we can construct confidence intervals for γ by using this theorem as described below.

There are a couple of remarkable aspects of the t test of this section under the null hypothesis of a unit root. First, the limiting distribution of the t-ratio is a standard normal. This is in contrast to the LSE-based t statistics which have nonnormal distributions in the limit whether or not the errors have finite variances (cf. Dickey and Fuller 1979; Phillips 1990). Second, the limiting distribution of the t test of this section does not change with the inclusion of nonstochastic regressors. Finite-sample power of the unit root test of this section is not expected to decrease much with the inclusion of nonstochastic regressors because of that aspect. This is confirmed by the simulation results in the supplementary file to this paper. By contrast, the limiting distributions of the Dickey–Fuller and Phillips-Perron tests shift leftward as higher-order time polynomials are included as regressors, making the tests' power decrease with the inclusion of time polynomials as regressors.

What implications does presence of a unit root have for the observed data $\{Y_t\}$? First, when the null hypothesis is true, the errors have permanent effects on the trajectory of $\{Y_t\}$ (see Choi 2015, p.4). Second, $\{Y_t\}$ has stochastic trends in the presence of a unit root, while it does not when AR(1) coefficient is less than 1. This is illustrated in Figure 1 which plots typical trajectories of $\{Y_t\}$ with standard Cauchy errors and $\rho_n = 0.5, 1$ $(n = 500)$. We observe many outliers in the simulated data, but stochastic trends are found only for the data with a unit root.

It is possible to construct confidence intervals for γ using Theorem 4. Because the limiting distribution of the t-ratio is a standard normal, there are no complications in constructing the confidence intervals. The $(1 - \alpha)$-level confidence interval for γ is

$$[n(1 - \hat{\rho}_n - c_{\alpha/2}\hat{\sigma}_{\hat{\rho}_n}),\ n(1 - \hat{\rho}_n + c_{\alpha/2}\hat{\sigma}_{\hat{\rho}_n})],$$

where $c_{\alpha/2}$ denotes the percentile from a standard normal distribution with the tail probability $\alpha/2$. By contrast, constructing confidence intervals of γ for the AR(1) model with errors having a finite variance is quite complicated as reported in Stock (1991) and Phillips (2014). See Choi (2015, pp. 140–142) for further discussions.

5 Simulation

This section reports a summary of simulation results that examine finite-sample properties of the estimators and test statistics of the previous sections. The tables that

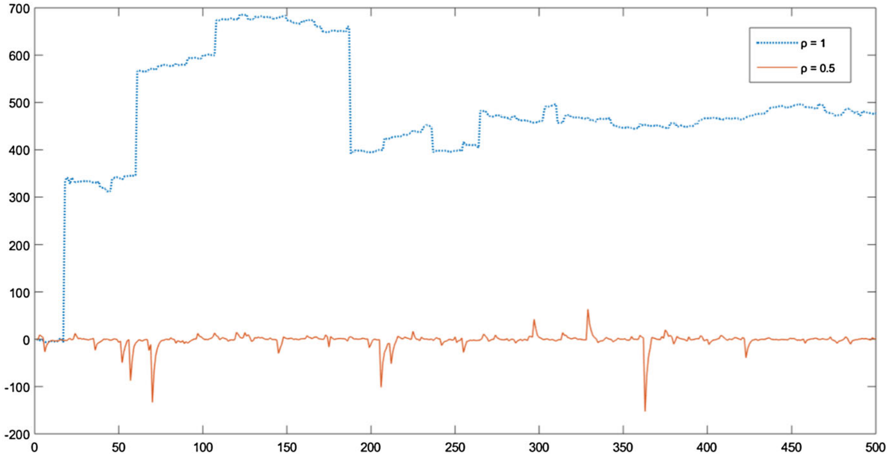

Fig. 1 Simulated AR(1) processes with $\rho_n = 0.5$ and 1 ($\sigma^o = 1, n = 500$)

contain the results can be found in the supplementary file to this paper. The simulation results we have obtained indicate that the MLE of the AR coefficient for a nearly nonstationary AR(1) model performs better than the LSE when the errors have a Cauchy distribution or a distribution which is a mixture of standard normal and Cauchy distributions at sample sizes 50, 100 and 250. Furthermore, the finite-sample distribution of the MLE-based t statistic becomes closer to a standard normal distribution as sample size increases. Last, the MLE-based t test works reasonably well in finite samples and is more powerful than the Dickey–Fuller t test.

6 Summary and further remarks

We have studied the MLE of the AR(1) models with a near unit root and Cauchy errors. The MLE of the AR coefficient is $n^{3/2}$-consistent and has a mixture-normal distribution in the limit. The MLE of the scale parameter is $n^{1/2}$-consistent and asymptotically normal. The MLEs of the intercept and the time trend are $n^{1/2}$ and $n^{3/2}$-consistent, respectively. An MLE-based t test for the null hypothesis of a unit root is also proposed. It is shown that the t statistic has a standard normal distribution asymptotically. Simulation results show that the MLE of the AR coefficient is more efficient in finite samples than the LSE and that the MLE-based t test for a unit root is more powerful than the Dickey–Fuller test.

7 Appendix A: Proofs

The following lemma will be used to prove Lemmas 2, 3, 5 which are used for the proofs of all the theorems of this paper.

Lemma 1 *Define* $Z_n(r) = n^{-1}\sum_{t=1}^{[nr]} \varepsilon_t$, $B_n(r) = n^{-1/2}\sum_{t=1}^{[nr]} \frac{\partial \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t}$ *and* $Q_n(r) = n^{-1/2}\sum_{t=1}^{[nr]} \frac{\partial \ln f(\varepsilon_t, \sigma)}{\partial \sigma}_{|\sigma=\sigma^o}$ *for* $0 \le r \le 1$. *Then,* $(Z_n(r_1), B_n(r_2),$

$Q_n(r_3)) \xrightarrow{d} (Z(r_1), B(r_2), Q(r_3))$ *in* $D[0, 1]^3$, *where* $Z(r_1)$, $B(r_2)$ *and* $Q(r_3)$ *are defined in Theorem 1.*

Proof This lemma is similar to Lemma 1 of Zhang and Chan (2012) except that the limit results for $B_n(r_2)$ and $Q_n(r_3)$ under the assumption of Cauchy errors are not dealt with there. Thus, we only need to show marginal weak convergence results for $B_n(r_2)$ and $Q_n(r_3)$ under the assumption of Cauchy errors. Because $E\big(\frac{\partial \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t}\big) = 0$ by standard theory and $E\big(\frac{\partial \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t}\big)^2 = \frac{1}{2\sigma^{o2}} < \infty$ by Lemma 7, we can apply the classical functional central limit theorem for the sequence of i.i.d. random variables $\big\{\frac{\partial \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t}\big\}$, obtaining

$$B_n(r) \xrightarrow{d} B(r) \text{ in } D[0, 1].$$

Likewise, we have

$$Q_n(r) \xrightarrow{d} Q(r) \text{ in } D[0, 1].$$

Since $E\big(\frac{\partial \ln f(\varepsilon_i, \sigma^o)}{\partial \varepsilon_i}\big)\left(\frac{\partial \ln f(\varepsilon_j, \sigma)}{\partial \sigma}_{|\sigma=\sigma^o}\right) = 0$ for every i and j, $B(r)$ and $Q(s)$ are independent for $0 \le r \le 1$ and $0 \le s \le 1$. □

The following lemma will be used to prove Theorem 1.

Lemma 2 (i) $n^{-3/2} \sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t} Y_{t-1} \xrightarrow{d} \int_0^1 S(r) \mathrm{d}B(r)$;
(ii) $n^{-2} \sum_{t=1}^n Y_{t-1} \xrightarrow{d} \int_0^1 S(r) \mathrm{d}r$;
(iii) $n^{-3} \sum_{t=1}^n Y_{t-1}^2 \xrightarrow{d} \int_0^1 S^2(r) \mathrm{d}r$;
(iv) $n^{-3} \sum_{t=1}^n Y_{t-1}^2 \left[\frac{\partial^2 \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t^2} - E\big(\frac{\partial^2 \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t^2}\big)\right] \xrightarrow{p} 0$;
(v) $n^{-2} \sum_{t=1}^n Y_{t-1} \left[\frac{\partial^2 \ln f(\varepsilon_t, \sigma)}{\partial \varepsilon_t \partial \sigma}_{|\sigma=\sigma^o} - E\left(\frac{\partial^2 \ln f(\varepsilon_t, \sigma)}{\partial \varepsilon_t \partial \sigma}_{|\sigma=\sigma^o}\right)\right] \xrightarrow{p} 0$;
(vi) $n^{-1} \sum_{t=1}^n \left[\frac{\partial^2 \ln f(\varepsilon_t, \sigma)}{\partial \sigma^2}_{|\sigma=\sigma^o} - E\left(\frac{\partial^2 \ln f(\varepsilon_t, \sigma)}{\partial \sigma^2}_{|\sigma=\sigma^o}\right)\right] \xrightarrow{p} 0$.

Proof Because $S_n(r) \xrightarrow{d} S(r)$ in $D[0, 1]$, where $S_n(r) = n^{-1} Y_{[nr]}$ and $S(r)$ is defined in Theorem 1, as shown in Chan et al. (2006), parts (i), (ii) and (iii) follow as in Lemma 4 of Zhang and Chan (2012). Because the variances of $\big\{\frac{\partial^2 \ln f(\varepsilon_t, \sigma^o)}{\partial \varepsilon_t^2}\big\}$ and $\left\{\frac{\partial^2 \ln f(\varepsilon_t, \sigma)}{\partial \varepsilon_t \partial \sigma}_{|\sigma=\sigma^o}\right\}$ are finite due to Lemma 10, we obtain (iv) and (v) by using the same method as for equation (16) of Zhang and Chan (2012). Part (vi) is trivial. □

Proof of Theorem 1 (i) We need to check the four conditions of Newey and McFadden (1994) Theorem 2.5. Conditions (i), (ii) and (iii) are trivially satisfied. To check condition (iv), write

$$\begin{aligned} f_t(\rho_n, \sigma) &= \frac{\sigma}{\pi\{\sigma^2 + (Y_t - \rho_n Y_{t-1})^2\}} \\ &= \frac{\sigma}{\pi\{\sigma^2 + \varepsilon_t^2 + 2(\rho_n^o - \rho_n)\varepsilon_t Y_{t-1} + (\rho_n^o - \rho_n)^2 Y_{t-1}^2\}}. \end{aligned}$$

Suppose that Y_{t-1} is given. Then, as in Newey and McFadden (1994; p.2125),

$$\begin{aligned}|\ln f_t(\rho_n,\sigma)| &\leq \ln\sigma - \ln\pi - \ln\{\sigma^2 + \varepsilon_t^2 + 2(\rho_n^o - \rho_n)\varepsilon_t Y_{t-1} + (\rho_n^o - \rho_n)^2 Y_{t-1}^2\}\\ &\leq C_1 + \ln(C_2 + C_3\varepsilon_t^2)\end{aligned}$$

for some positive constants $C_1,\ C_2$ and C_3. Since $E\left[\ln(C_2 + C_3\varepsilon_t^2)\right] < \infty$ for every $t,$ we obtain

$$E\left(\sup_{\rho_n,\sigma}|\ln f_t(\rho_n,\sigma)|\right) = EE\left(\sup_{\rho_n,\sigma}|\ln f_t(\rho_n,\sigma)| \mid Y_{t-1}\right) < \infty,$$

as desired.

(ii) Let $\theta_n = (\rho_n,\sigma)$, $\hat{\theta}_n = (\hat{\rho}_n,\hat{\sigma}_n)$, $\theta_n^o = (\rho_n^o,\sigma^o)$ and θ_n^* be on the line joining θ_n^o and $\hat{\theta}_n$. Because $\sum_{t=1}^n \frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\widehat{\theta}_n} = 0$, we obtain by the mean value theorem

$$0 = \sum_{t=1}^n \frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o} + \sum_{t=1}^n \frac{\partial^2 \ln f}{\partial\theta_n\partial\theta_n'}_{|\theta_n=\theta_n^*}\begin{pmatrix}\hat{\rho}_n - \rho_n^o\\ \hat{\sigma}_n - \sigma^o\end{pmatrix},$$

which gives

$$\begin{aligned}&\begin{pmatrix}n^{3/2}\left(\hat{\rho}_n - \rho_n^o\right)\\ n^{1/2}\left(\hat{\sigma}_n - \sigma^o\right)\end{pmatrix}\\ &= -\left[J_n^{-1}\sum_{t=1}^n \frac{\partial^2 \ln f}{\partial\theta_n\partial\theta_n'}_{|\theta_n=\theta_n^*} J_n^{-1}\right]^{-1} J_n^{-1}\sum_{t=1}^n \frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o}\\ &= -\left(A_{n|_{\theta_n=\theta_n^*}}\right)^{-1} J_n^{-1}\sum_{t=1}^n \frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o}, \text{ say,}\end{aligned}$$

where $J_n = diag(n^{3/2}, n^{1/2})$. Because $\theta_n^* - \theta_n^o \xrightarrow{p} 0$ as $n \to \infty$, it follows that

$$A_{n|_{\theta_n=\theta_n^*}} - A_{n|_{\theta_n=\theta_n^o}} \xrightarrow{p} 0. \tag{8}$$

Since

$$A_{n|_{\theta_n=\theta_n^o}} = \begin{bmatrix} n^{-3}\sum_{t=1}^n Y_{t-1}^2 \frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2} & -n^{-2}\sum_{t=1}^n Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}\\ -n^{-2}\sum_{t=1}^n Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma} & n^{-1}\sum_{t=1}^n \frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\sigma^2}\end{bmatrix}_{|\theta_n=\theta_n^o},$$

parts (iv), (v) and (vi) of Lemma 2 show that $A_{n|_{\theta_n=\theta_n^o}}$ has the same limiting distribution as

$$\begin{bmatrix} n^{-3}\sum_{t=1}^n Y_{t-1}^2 E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \varepsilon_t^2}\big) & -n^{-2}\sum_{t=1}^n Y_{t-1} E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \varepsilon_t \partial\sigma}\big) \\ -n^{-2}\sum_{t=1}^n Y_{t-1} E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \varepsilon_t \partial\sigma}\big) & n^{-1}\sum_{t=1}^n E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \sigma^2}\big) \end{bmatrix}_{|\theta_n=\theta_n^o},$$

which is equal to, due to Lemmas 8 and 9,

$$-\frac{1}{2\sigma^{o2}}\begin{bmatrix} n^{-3}\sum_{t=1}^n Y_{t-1}^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, using relation (8) and part (iii) of Lemma 2, we obtain

$$A_{n|_{\theta_n=\theta_n^*}} \xrightarrow{d} -\frac{1}{2\sigma^{o2}}\begin{bmatrix} \int_0^1 S^2(r)\mathrm{d}r & 0 \\ 0 & 1 \end{bmatrix}. \tag{9}$$

In addition, Lemmas 1 and 2 yield

$$\begin{aligned} J_n^{-1}\sum_{t=1}^n \frac{\partial \ln f}{\partial \theta_n}_{|_{\theta_n=\theta_n^o}} &= \begin{pmatrix} -n^{-3/2}\sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t} Y_{t-1} \\ n^{-1/2}\sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t,\sigma)}{\partial \sigma}_{|\sigma=\sigma^o} \end{pmatrix} \\ &\xrightarrow{d} \begin{pmatrix} -\int_0^1 S(r)\mathrm{d}B(r) \\ Q(1) \end{pmatrix}. \end{aligned} \tag{10}$$

The stated result follows, once the continuous mapping theorem is applied to relations (9) and (10). □

It is straightforward to show that $S_n(r) = n^{-1}Y_{[nr]} \xrightarrow{d} S(r)$ for model (3). Thus, we continue to use Lemma 2 to prove Theorem 2. In addition to Lemma 2, we need the following lemma to prove Theorem 2.

Lemma 3 (i) $n^{-2}\sum_{t=1}^n Y_{t-1}\Big[\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t^2} - E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t^2}\big)\Big] \xrightarrow{p} 0$;
(ii) $n^{-1}\sum_{t=1}^n \Big[\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \varepsilon_t \partial \sigma}_{|\sigma=\sigma^o} - E\Big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial \varepsilon_t \partial \sigma}_{|\sigma=\sigma^o}\Big)\Big] \xrightarrow{p} 0$;
(iii) $n^{-1}\sum_{t=1}^n \Big[\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t^2} - E\big(\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t^2}\big)\Big] \xrightarrow{p} 0$.

Proof Using the same method as for Eq. (16) of Zhang and Chan (2012), we obtain (i). Parts (ii) and (iii) are trivial. □

Proof of Theorem 2 (i) Use the same method as for the proof of Theorem 1 (i).
(ii) Let $\theta_n = (\rho_n, \mu, \sigma)$, $\hat{\theta}_n = (\hat{\rho}_n, \hat{\mu}_n, \hat{\sigma}_n)$, $\theta_n^o = (\rho_n^o, \mu^o, \sigma^o)$ and θ_n^* be on the line joining θ_n^o and $\hat{\theta}_n$. Because $\sum_{t=1}^n \frac{\partial \ln f}{\partial \theta_n}_{|_{\theta_n=\hat{\theta}_n}} = 0$, the mean value theorem yields

$$0 = \sum_{t=1}^n \frac{\partial \ln f}{\partial \theta_n}_{|_{\theta_n=\theta_n^o}} + \sum_{t=1}^n \frac{\partial^2 \ln f}{\partial \theta_n \partial \theta_n'}_{|_{\theta_n=\theta_n^*}} \begin{pmatrix} \hat{\rho}_n - \rho_n^o \\ \hat{\mu}_n - \mu^o \\ \hat{\sigma}_n - \sigma^o \end{pmatrix},$$

which gives

$$\begin{pmatrix} n^{3/2}\left(\hat{\rho}_n - \rho_n^o\right) \\ n^{1/2}\left(\hat{\mu}_n - \mu^o\right) \\ n^{1/2}\left(\hat{\sigma}_n - \sigma^o\right) \end{pmatrix} = -\left[K_n^{-1}\sum_{t=1}^{n}\frac{\partial^2 \ln f}{\partial\theta_n\partial\theta_n'}_{|\theta_n=\theta_n^*} K_n^{-1}\right]^{-1} K_n^{-1}\sum_{t=1}^{n}\frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o} = -\left(E_{n|\theta_n=\theta_n^*}\right)^{-1} K_n^{-1}\sum_{t=1}^{n}\frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o}, \text{ say,}$$

where $K_n = diag(n^{3/2}, n^{1/2}, n^{1/2})$. Because $\theta_n^* - \theta_n^o \xrightarrow{p} 0$ as $n \to \infty$, it follows that

$$E_{n|\theta_n=\theta_n^*} - E_{n|\theta_n=\theta_n^o} \xrightarrow{p} 0. \tag{11}$$

Since

$$E_{n|\theta_n=\theta_n^o} = \begin{bmatrix} n^{-3}\sum_{t=1}^{n} Y_{t-1}^2 \frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2} & n^{-2}\sum_{t=1}^{n} Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2} & -n^{-2}\sum_{t=1}^{n} Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma} \\ n^{-2}\sum_{t=1}^{n} Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2} & n^{-1}\sum_{t=1}^{n}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2} & -n^{-1}\sum_{t=1}^{n}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma} \\ -n^{-2}\sum_{t=1}^{n} Y_{t-1}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma} & -n^{-1}\sum_{t=1}^{n}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma} & n^{-1}\sum_{t=1}^{n}\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\sigma^2} \end{bmatrix}_{|\theta_n=\theta_n^o},$$

parts (iv), (v) and (vi) of Lemma 2 and parts (i), (ii) and (iii) of Lemma 3 show that $E_{n|\theta_n=\theta_n^o}$ has the same limiting distribution as

$$\begin{bmatrix} n^{-3}\sum_{t=1}^{n} Y_{t-1}^2 E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2}\right) & n^{-2}\sum_{t=1}^{n} Y_{t-1}E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2}\right) & -n^{-2}\sum_{t=1}^{n} Y_{t-1}E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}\right) \\ n^{-2}\sum_{t=1}^{n} Y_{t-1}E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2}\right) & n^{-1}\sum_{t=1}^{n} E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t^2}\right) & -n^{-1}\sum_{t=1}^{n} E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}\right) \\ -n^{-2}\sum_{t=1}^{n} Y_{t-1}E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}\right) & -n^{-1}\sum_{t=1}^{n} E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}\right) & n^{-1}\sum_{t=1}^{n} E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\sigma^2}\right) \end{bmatrix}_{|\theta_n=\theta_n^o},$$

which is equal to, due to Lemmas 8 and 9,

$$-\frac{1}{2\sigma^{o2}}\begin{bmatrix} n^{-3}\sum_{t=1}^{n} Y_{t-1}^2 & n^{-2}\sum_{t=1}^{n} Y_{t-1} & 0 \\ n^{-2}\sum_{t=1}^{n} Y_{t-1} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, using relation (11) and parts (ii) and (iii) of Lemma 2, we obtain

$$E_{n|\theta_n=\theta_n^*} \xrightarrow{d} -\frac{1}{2\sigma^{o2}}\begin{bmatrix} \int_0^1 S^2(r)\mathrm{d}r & \int_0^1 S(r)\mathrm{d}r & 0 \\ \int_0^1 S(r)\mathrm{d}r & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

In addition, Lemmas 1 and 2 yield

$$K_n^{-1}\sum_{t=1}^{n}\frac{\partial \ln f}{\partial\theta_n}_{|\theta_n=\theta_n^o} = \begin{pmatrix} -n^{-3/2}\sum_{t=1}^{n}\frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t}Y_{t-1} \\ -n^{-1/2}\sum_{t=1}^{n}\frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t} \\ n^{-1/2}\sum_{t=1}^{n}\frac{\partial \ln f(\varepsilon_t,\sigma)}{\partial\sigma}_{|\sigma=\sigma^o} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} -\int_0^1 S(r)\mathrm{d}B(r) \\ -B(1) \\ Q(1) \end{pmatrix}. \tag{13}$$

The stated result is obtained by applying the continuous mapping theorem to relations (12) and (13). □

The following two lemmas will be used to prove Theorem 3.

Lemma 4 *For model (5), we have* $S_n(r) = n^{-1}Y_{[nr]} \xrightarrow{d} U(r) = S(r) + b_0 r$ *in* $D[0, 1]$.

Proof Because $X_t = \rho_n X_{t-1} + \varepsilon_t$, $n^{-1}X_{[nr]} \xrightarrow{d} S(r)$ in $D[0, 1]$, we obtain

$$S_n(r) = n^{-1}Y_{[nr]} = n^{-1}\left(a_0 + b_0[nr] + X_{[nr]}\right) \xrightarrow{d} U(r) = b_0 r + S(r) \text{ in } D[0, 1],$$

as stated. □

Lemma 5 (i) $n^{-3/2}\sum_{t=1}^{n}\frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t}Y_{t-1} \xrightarrow{d} \int_0^1 U(r)\mathrm{d}B(r)$;

(ii) $n^{-2}\sum_{t=1}^{n} Y_{t-1} \xrightarrow{d} \int_0^1 U(r)\mathrm{d}r$;

(iii) $n^{-3}\sum_{t=1}^{n} Y_{t-1}^2 \xrightarrow{d} \int_0^1 U^2(r)\mathrm{d}r$;

(iv) $n^{-3}\sum_{t=1}^{n} tY_{t-1} \xrightarrow{d} \int_0^1 rU(r)\mathrm{d}r$;

(v) $n^{-3/2}\sum_{t=1}^{n} t\frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t} \xrightarrow{d} \int_0^1 r\mathrm{d}B(r)$;

(vi) $n^{-3}\sum_{t=1}^{n} tY_{t-1}\left[\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2} - E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2}\right)\right] \xrightarrow{p} 0$;

(vii) $n^{-2}\sum_{t=1}^{n} t\left[\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}_{|\sigma=\sigma^o} - E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma)}{\partial\varepsilon_t\partial\sigma}_{|\sigma=\sigma^o}\right)\right] \xrightarrow{p} 0$;

(viii) $n^{-2}\sum_{t=1}^{n} t\left[\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2} - E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2}\right)\right] \xrightarrow{p} 0$;

(ix) $n^{-3}\sum_{t=1}^{n} t^2\left[\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2} - E\left(\frac{\partial^2 \ln f(\varepsilon_t,\sigma^o)}{\partial\varepsilon_t^2}\right)\right] \xrightarrow{p} 0$.

Proof Because $S_n(r) \xrightarrow{d} U(r)$ in $D[0, 1]$ by Lemma 4, parts (i), (ii), (iii) and (iv) follow as in Lemma 4 of Zhang and Chan (2012). Part (v) follows from the central limit theorem. Using the same method as for Eq. (16) of Zhang and Chan (2012), we obtain (vi), (vii), (viii) and (ix). □

Proof of Theorem 3 (i) Using the same method as for the proof of Theorem 1 (i), it is straightforward to prove this.

(ii) Let $\theta_n = (\rho_n, \mu, \beta, \sigma)$, $\hat{\theta}_n = (\hat{\rho}_n, \hat{\mu}_n, \hat{\beta}_n, \hat{\sigma}_n)$, $\theta_n^o = (\rho_n^o, \mu^o, \beta^o, \sigma^o)$ and θ_n^* be on the line joining θ_n^o and $\hat{\theta}_n$. Because $\sum_{t=1}^n \frac{\partial \ln f}{\partial \theta_n}_{|\theta_n=\widehat{\theta}_n} = 0$, we obtain by the mean value theorem

$$0 = \sum_{t=1}^{n} \frac{\partial \ln f}{\partial \theta_n}_{|\theta_n=\theta_n^o} + \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \theta_n \partial \theta_n'}_{|\theta_n=\theta_n^*} \begin{pmatrix} \hat{\rho}_n - \rho_n^o \\ \hat{\mu}_n - \mu^o \\ \hat{\beta}_n - \beta^o \\ \hat{\sigma}_n - \sigma^o \end{pmatrix}.$$

Thus, we have

$$\begin{aligned}
&\begin{pmatrix} n^{3/2}\left(\hat{\rho}_n - \rho_n^o\right) \\ n^{1/2}\left(\hat{\mu}_n - \mu^o\right) \\ n^{3/2}\left(\hat{\beta}_n - \beta^o\right) \\ n^{1/2}\left(\hat{\sigma}_n - \sigma^o\right) \end{pmatrix} \\
&= -\left[L_n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \theta_n \partial \theta_n'}_{|\theta_n=\theta_n^*} L_n^{-1}\right]^{-1} L_n^{-1} \sum_{t=1}^{n} \frac{\partial \ln f}{\partial \theta_n}_{|\theta_n=\theta_n^o} \\
&= -\left(C_{n|_{\theta_n=\theta_n^*}}\right)^{-1} L_n^{-1} \sum_{t=1}^{n} \frac{\partial \ln f}{\partial \theta_n}_{|\theta_n=\theta_n^o}, \text{ say,}
\end{aligned}$$

where $L_n = diag(n^{3/2}, n^{1/2}, n^{3/2}, n^{1/2})$. Because $\theta_n^* - \theta_n^o \xrightarrow{p} 0$ as $n \to \infty$, it follows that

$$C_{n|_{\theta_n=\theta_n^*}} - C_{n|_{\theta_n=\theta_n^o}} \xrightarrow{p} 0. \tag{14}$$

Since

$$C_{n|_{\theta_n=\theta_n^o}} = \begin{bmatrix}
n^{-3} \sum_{t=1}^{n} Y_{t-1}^2 \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-2} \sum_{t=1}^{n} Y_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-3} \sum_{t=1}^{n} tY_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & -n^{-2} \sum_{t=1}^{n} Y_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} \\
n^{-2} \sum_{t=1}^{n} Y_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-2} \sum_{t=1}^{n} t \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & -n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} \\
n^{-3} \sum_{t=1}^{n} tY_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-2} \sum_{t=1}^{n} t \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & n^{-3} \sum_{t=1}^{n} t^2 \frac{\partial^2 \ln f}{\partial \varepsilon_t^2} & -n^{-2} \sum_{t=1}^{n} t \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} \\
-n^{-2} \sum_{t=1}^{n} Y_{t-1} \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} & -n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} & -n^{-2} \sum_{t=1}^{n} t \frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma} & n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \ln f}{\partial \sigma^2}
\end{bmatrix}_{|\theta_n=\theta_n^o},$$

Lemmas 2, 3 and 5 show that $C_{n|_{\theta_n=\theta_n^o}}$ has the same limiting distribution as

$$\begin{bmatrix} n^{-3}\sum_{t=1}^{n} Y_{t-1}^2 G & n^{-2}\sum_{t=1}^{n} Y_{t-1} G & n^{-3}\sum_{t=1}^{n} tY_{t-1} G & -n^{-2}\sum_{t=1}^{n} Y_{t-1} H \\ n^{-2}\sum_{t=1}^{n} Y_{t-1} G & n^{-1}\sum_{t=1}^{n} G & n^{-2}\sum_{t=1}^{n} tG & -n^{-1}\sum_{t=1}^{n} H \\ n^{-3}\sum_{t=1}^{n} tY_{t-1} G & n^{-2}\sum_{t=1}^{n} tG & n^{-3}\sum_{t=1}^{n} t^2 G & -n^{-2}\sum_{t=1}^{n} tH \\ -n^{-2}\sum_{t=1}^{n} Y_{t-1} H & -n^{-1}\sum_{t=1}^{n} H & -n^{-2}\sum_{t=1}^{n} tH & n^{-1}\sum_{t=1}^{n} J \end{bmatrix}_{|\theta_n=\theta_n^o},$$

where $G = E[\frac{\partial^2 \ln f}{\partial \varepsilon_t^2}]$, $H = E[\frac{\partial^2 \ln f}{\partial \varepsilon_t \partial \sigma}]$ and $J = E[\frac{\partial^2 \ln f}{\partial \sigma^2}]$. Due to Lemmas 8 and 9, the above matrix is equal to

$$-\frac{1}{2\sigma^{o2}}\begin{bmatrix} n^{-3}\sum_{t=1}^n Y_{t-1}^2 & n^{-2}\sum_{t=1}^n Y_{t-1} & n^{-3}\sum_{t=1}^n tY_{t-1} & 0 \\ n^{-2}\sum_{t=1}^n Y_{t-1} & 1 & n^{-2}\sum_{t=1}^n t & 0 \\ n^{-3}\sum_{t=1}^n tY_{t-1} & n^{-2}\sum_{t=1}^n t & n^{-3}\sum_{t=1}^n t^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus, using relation (14) and parts (ii), (iii) and (iv) of Lemma 5, we obtain

$$C_{n|_{\theta_n=\theta_n^*}} \xrightarrow{d} -\frac{1}{2\sigma^{o2}}\begin{bmatrix} \int_0^1 U^2(r)\mathrm{d}r & \int_0^1 U(r)\mathrm{d}r & \int_0^1 rU(r)\mathrm{d}r & 0 \\ \int_0^1 U(r)\mathrm{d}r & 1 & \frac{1}{2} & 0 \\ \int_0^1 rU(r)\mathrm{d}r & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

In addition, Lemmas 1 and 5 yield

$$L_n^{-1}\sum_{t=1}^{n}\frac{\partial \ln f}{\partial \theta_n}_{|\theta_n=\theta_n^o} = \begin{pmatrix} -n^{-3/2}\sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t} Y_{t-1} \\ -n^{-1/2}\sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t} \\ -n^{-3/2}\sum_{t=1}^n t\frac{\partial \ln f(\varepsilon_t,\sigma^o)}{\partial \varepsilon_t} \\ n^{-1/2}\sum_{t=1}^n \frac{\partial \ln f(\varepsilon_t,\sigma)}{\partial \sigma}_{|\sigma=\sigma^o} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} -\int_0^1 U(r)\mathrm{d}B(r) \\ -B(1) \\ -\int_0^1 r\mathrm{d}B(r) \\ Q(1) \end{pmatrix}. \tag{16}$$

The stated result follows from relations (15) and (16). □

Proof of Theorem 4 This follows straightforwardly from the mixture normality results (2), (4) and (6), and the block-diagonal structure of the information matrices in the limit. □

8 Appendix B: Auxiliary lemmas

Lemma 6

$$D_1 = \int_{\mathbb{R}} \frac{1}{\sigma^{o2} + \varepsilon_1^2} \mathrm{d}\varepsilon_1 = \frac{\pi}{\sigma^o},\ D_2 = \int_{\mathbb{R}} \frac{1}{\left(\sigma^{o2} + \varepsilon_1^2\right)^2} \mathrm{d}\varepsilon_1 = \frac{\pi}{2\sigma^{o3}},$$
$$D_3 = \int_{\mathbb{R}} \frac{1}{\left(\sigma^{o2} + \varepsilon_1^2\right)^3} \mathrm{d}\varepsilon_1 = \frac{3\pi}{8\sigma^{o5}},\ D_4 = \int_{\mathbb{R}} \frac{1}{\left(\sigma^{o2} + \varepsilon_1^2\right)^4} \mathrm{d}\varepsilon_1 = \frac{5\pi}{16\sigma^{o7}},$$
$$D_5 = \int_{\mathbb{R}} \frac{1}{\left(\sigma^{o2} + \varepsilon_1^2\right)^5} \mathrm{d}\varepsilon_1 = \frac{35\pi}{128\sigma^{o9}}.$$

Proof The first result follows because $\int_{\mathbb{R}} \frac{1}{x+\varepsilon_1^2} \mathrm{d}\varepsilon_1 = \pi x^{-1/2}$. The rest are obtained by successively differentiating both sides of this equation with respect to x and setting $x = \sigma^{o2}$. Differentiating within the integral sign is allowed, because for $k \in \mathbb{N}$ and $x \in \mathbb{R}^+$, $\left|\frac{\partial\left(x+\varepsilon_1^2\right)^{-k}}{\partial x}\right| = \left|\frac{k}{\left(x+\varepsilon_1^2\right)^{k+1}}\right| \leq \left|\frac{k}{\varepsilon_1^{2(k+1)}}\right|$ and $\left|\frac{k}{\varepsilon_1^{2(k+1)}}\right|$ is integrable. □

Lemma 7 *(i)* $E\left[\left(\frac{\partial \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1}\right)^2\right] = \frac{1}{2\sigma^{o2}}$;
(ii) $E\left[\left(\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial \sigma}_{|\sigma=\sigma^o}\right)^2\right] = \frac{1}{2\sigma^{o2}}$.

Proof Using Lemma 6, we obtain

$$\begin{aligned} E\left[\left(\frac{\partial \ln f(\varepsilon_1, \sigma^o)}{\partial \varepsilon_1}\right)^2\right] &= \int_{\mathbb{R}} \left(\frac{2\varepsilon_1}{\varepsilon_1^2 + \sigma^{o2}}\right)^2 \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2 + \sigma^{o2})} \mathrm{d}\varepsilon_1 \\ &= \frac{4\sigma^o}{\pi} \int_{\mathbb{R}} \frac{\varepsilon_1^2}{(\varepsilon_1^2 + \sigma^{o2})^3} \mathrm{d}\varepsilon_1 \\ &= \frac{4\sigma^o}{\pi} \left[\int_{\mathbb{R}} \frac{\varepsilon_1^2 + \sigma^{o2}}{(\varepsilon_1^2 + \sigma^{o2})^3} \mathrm{d}\varepsilon_1 - \int_{\mathbb{R}} \frac{\sigma^{o2}}{(\varepsilon_1^2 + \sigma^{o2})^3} \mathrm{d}\varepsilon_1\right] \\ &= \frac{4\sigma^o}{\pi} \left[D_2 - \sigma^{o2} D_3\right] = \frac{1}{2\sigma^{o2}}, \end{aligned}$$

where D_2 and D_3 are defined in Lemma 6. In the same manner, we have

$$\begin{aligned} E\left[\left(\frac{\partial \ln f(\varepsilon_1, \sigma)}{\partial \sigma}_{|\sigma=\sigma^o}\right)^2\right] &= \int_{\mathbb{R}} \frac{(\varepsilon_1^2 - \sigma^{o2})^2}{\sigma^{o2}(\varepsilon_1^2 + \sigma^{o2})^2} \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2 + \sigma^{o2})} \mathrm{d}\varepsilon_1 \\ &= \frac{1}{\sigma^o \pi} \int_{\mathbb{R}} \frac{(\varepsilon_1^2 + \sigma^{o2})^2 - 4\varepsilon_1^2\sigma^{o2}}{(\varepsilon_1^2 + \sigma^{o2})^3} \mathrm{d}\varepsilon_1 \end{aligned}$$

$$= \frac{1}{\sigma^o \pi}\left[\int_{\mathbb{R}} \frac{1}{\varepsilon_1^2 + \sigma^{o2}} \mathrm{d}\varepsilon_1 - 4\sigma^{o2} \int_{\mathbb{R}} \frac{\varepsilon_1^2}{(\varepsilon_1^2 + \sigma^{o2})^3} \mathrm{d}\varepsilon_1\right]$$
$$= \frac{1}{\sigma^o \pi}\left[D_1 - 4\sigma^{o2}(D_2 - \sigma^{o2} D_3)\right] = \frac{1}{2\sigma^{o2}}.$$

□

Lemma 8 $E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\varepsilon_1 \partial\sigma}_{|\sigma=\sigma^o}\right] = 0.$

Proof Because

$$\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\varepsilon_1 \partial\sigma}_{|\sigma=\sigma^o} = \frac{4\varepsilon_1 \sigma^o}{\left(\varepsilon_1^2 + \sigma^{o2}\right)^2},$$

we have

$$E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\varepsilon_1 \partial\sigma}_{|\sigma=\sigma^o}\right] = \frac{4\sigma^{o2}}{\pi} \int_{\mathbb{R}} \frac{\varepsilon_1}{\left(\varepsilon_1^2 + \sigma^{o2}\right)^3} \mathrm{d}\varepsilon_1.$$

Because the integrand is an odd function, the stated result follows. □

Lemma 9 $E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\sigma^2}_{|\sigma=\sigma^o}\right] = -\frac{1}{2\sigma^{o2}}$ *and* $E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial\varepsilon_1^2}\right] = -\frac{1}{2\sigma^{o2}}.$

Proof We begin with the identity

$$1 = \int_{\mathbb{R}} f(\varepsilon_1, \sigma) \mathrm{d}\varepsilon_1.$$

By Leibniz's rule, taking the derivative of the both sides of the above equation with respect to σ results in

$$0 = \int_{\mathbb{R}} \frac{\partial f(\varepsilon_1,\sigma)}{\partial\sigma} \mathrm{d}\varepsilon_1$$
$$= \int_{\mathbb{R}} \frac{\partial f(\varepsilon_1,\sigma)/\partial\sigma}{f(\varepsilon_1,\sigma)} f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1$$
$$= \int_{\mathbb{R}} \frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial\sigma} f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1$$
$$= E\left[\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial\sigma}\right].$$

Differentiating this again, we obtain by Leibniz's rule

$$0 = \int_{\mathbb{R}} \frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\sigma^2} f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1 + \int_{\mathbb{R}} \left(\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial\sigma}\right)^2 f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1,$$
$$\int_{\mathbb{R}} \frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\sigma^2} f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1 = -\int_{\mathbb{R}} \left(\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial\sigma}\right)^2 f(\varepsilon_1,\sigma) \mathrm{d}\varepsilon_1,$$
$$E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial\sigma^2}\right] = -E\left[\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial\sigma}\right]^2.$$

Hence, by Lemma 7,

$$E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial \sigma^2}_{|\sigma=\sigma^o}\right] = -E\left[\frac{\partial \ln f(\varepsilon_1,\sigma)}{\partial \sigma}_{|\sigma=\sigma^o}\right]^2$$
$$= -\frac{1}{2\sigma^{o2}}.$$

To prove the second result, consider the relation

$$\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1^2} = \frac{2\varepsilon_1^2 - 2\sigma^{o2}}{\left(\sigma^{o2}+\varepsilon_1^2\right)^2},$$

which gives

$$\begin{aligned}
E\left[\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1^2}\right] &= \int_{\mathbb{R}} \frac{2\varepsilon_1^2 - 2\sigma^{o2}}{(\varepsilon_1^2+\sigma^{o2})^2} \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2+\sigma^{o2})} \mathrm{d}\varepsilon_1 \\
&= \frac{\sigma^o}{\pi} \int_{\mathbb{R}} \frac{2\varepsilon_1^2 - 2\sigma^{o2}}{(\varepsilon_1^2+\sigma^{o2})^3} \mathrm{d}\varepsilon_1 \\
&= \frac{\sigma^o}{\pi}\left(2\int_{\mathbb{R}} \frac{\varepsilon_1^2}{(\varepsilon_1^2+\sigma^{o2})^3}\mathrm{d}\varepsilon_1 - 2\sigma^{o2}\int_{\mathbb{R}} \frac{1}{(\varepsilon_1^2+\sigma^{o2})^3}\mathrm{d}\varepsilon_1\right) \\
&= \frac{\sigma^o}{\pi}(2(D_2 - \sigma^{o2}D_3) - 2\sigma^{o2}D_3) \\
&= -\frac{1}{2\sigma^{o2}},
\end{aligned}$$

as stated. □

Lemma 10 $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1^2}\right)^2\right]$, $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial \sigma^2}\right)^2_{|\sigma=\sigma^o}\right]$ *and* $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial \varepsilon_1 \partial \sigma}\right)^2_{|\sigma=\sigma^o}\right]$ *are finite.*

Proof First, we show that $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1 \partial \sigma}\right)^2_{|\sigma=\sigma^o}\right] = \frac{5}{8}\sigma^{o-4}$. By Lemma 6,

$$\begin{aligned}
E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1 \partial \sigma}\right)^2_{|\sigma=\sigma^o}\right] &= \int_{\mathbb{R}} \left(\frac{4\varepsilon_1\sigma^o}{\left(\varepsilon_1^2+\sigma^{o2}\right)^2}\right)^2 \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2+\sigma^{o2})}\mathrm{d}\varepsilon_1 \\
&= \frac{16\sigma^{o3}}{\pi}\left(D_4 - \sigma^{o2}D_5\right) \\
&= \frac{5}{8}\sigma^{o-4}.
\end{aligned}$$

Second, $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1^2}\right)^2\right] = \frac{7}{8}\sigma^{o-4}$. By Lemma 6,

$$
\begin{aligned}
E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma^o)}{\partial \varepsilon_1^2}\right)^2\right] &= \int_{\mathbb{R}} \left(\frac{2\varepsilon_1^2 - 2\sigma^{o2}}{(\varepsilon_1^2+\sigma^{o2})^2}\right)^2 \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2+\sigma^{o2})} d\varepsilon_1 \\
&= \frac{4\sigma^o}{\pi}\left(\int_{\mathbb{R}} \frac{\left(\varepsilon_1^2+\sigma^{o2}\right)^2}{(\varepsilon_1^2+\sigma^{o2})^5} d\varepsilon_1 - 4\sigma^{o2}\int_{\mathbb{R}} \frac{\varepsilon_1^2}{(\varepsilon_1^2+\sigma^{o2})^5} d\varepsilon_1\right) \\
&= \frac{4\sigma^o}{\pi}\left(D_3 - 4\sigma^{o2}\left(D_4 - \sigma^{o2} D_5\right)\right) \\
&= \frac{7}{8}\sigma^{o-4}.
\end{aligned}
$$

To prove $E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial \sigma^2}\right)^2_{|\sigma=\sigma^o}\right] = \frac{7}{8}\sigma^{o-4}$, we should note the following equations

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{\varepsilon_1^2}{\left(\varepsilon_1^2+\sigma^{o2}\right)^5} d\varepsilon_1 &= D_4 - \sigma^{o2} D_5, \\
\int_{\mathbb{R}} \frac{\varepsilon_1^4}{\left(\varepsilon_1^2+\sigma^{o2}\right)^5} d\varepsilon_1 &= D_3 - 2\sigma^{o2} D_4 + \sigma^{o4} D_5, \\
\int_{\mathbb{R}} \frac{\varepsilon_1^6}{\left(\varepsilon_1^2+\sigma^{o2}\right)^5} d\varepsilon_1 &= D_2 - 3\sigma^{o2} D_3 + 3\sigma^{o4} D_4 - \sigma^{o6} D_5, \\
\int_{\mathbb{R}} \frac{\varepsilon_1^8}{\left(\varepsilon_1^2+\sigma^{o2}\right)^5} d\varepsilon_1 &= D_1 - 4\sigma^{o2} D_2 + 6\sigma^{o4} D_3 - 4\sigma^{o6} D_4 + \sigma^{o8} D_5.
\end{aligned}
$$

By using these equations, we have

$$
\begin{aligned}
&E\left[\left(\frac{\partial^2 \ln f(\varepsilon_1,\sigma)}{\partial \sigma^2}\right)^2_{|\sigma=\sigma^o}\right] = \int_{\mathbb{R}} \left(\frac{\sigma^{o4} - 4\varepsilon_1^2\sigma^{o2} - \varepsilon_1^4}{\sigma^{o2}(\varepsilon_1^2+\sigma^{o2})^2}\right)^2 \cdot \frac{\sigma^o}{\pi(\varepsilon_1^2+\sigma^{o2})} d\varepsilon_1 \\
&= \frac{1}{\pi\sigma^{o3}} \int_{\mathbb{R}} \frac{\varepsilon_1^8 + 8\sigma^{o2}\varepsilon_1^6 + 14\sigma^{o4}\varepsilon_1^4 - 8\sigma^{o6}\varepsilon_1^2 + \sigma^{o8}}{\left(\varepsilon_1^2+\sigma^{o2}\right)^5} d\varepsilon_1 \\
&= \frac{1}{\pi\sigma^{o3}}\left(D_1 + 4\sigma^{o2} D_2 - 4\sigma^{o4} D_3 - 16\sigma^{o6} D_4 + 16\sigma^{o8} D_5\right) \\
&= \frac{7}{8}\sigma^{o-4}.
\end{aligned}
$$

□

References

Ahn, S. K., Fotopoulos, S. B., He, L. (2001). Unit root tests with infinite variance errors. *Econometric Reviews*, *20*(4), 461–483.

Akgiray, V., Booth, G. G. (1988). The stable-law model of stock returns. *Journal of Business & Economic Statistics*, *6*(1), 51–57.

Andrews, B., Calder, M., Davis, R. A. (2009). Maximum likelihood estimation for α-stable autoregressive processes. *The Annals of Statistics*, *37*(4), 1946–1982.

Arnold, L. (1974). *Stochastic differential equations*: *Theory and applications*. New York: Wiley.

Callegari, F., Cappuccio, N., Lubian, D. (2003). Asymptotic inference in time series regressions with a unit root and infinite variance errors. *Journal of Statistical Planning and Inference*, *116*(1), 277–303.

Chan, N. H., Tran, L. T. (1989). On the first-order autoregressive process with infinite variance. *Econometric Theory*, *5*(03), 354–362.

Chan, N. H., Peng, L., Qi, Y. (2006). Quantile inference for near-integrated autoregressive time series with infinite variance. *Statistica Sinica*, *16*(1), 15.

Choi, I. (2015). *Almost all about unit roots*: *Foundations, developments, and applications*. Cambridge: Cambridge University Press.

Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, *1*(5), 948–957.

Falk, B., Wang, C. H. (2003). Testing long-run PPP with infinite-variance returns. *Journal of Applied Econometrics*, *18*(4), 471–484.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34–105.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics*, *17*(3), 261–278.

Knight, K. (1991). Limit theory for M-estimates in an integrated infinite variance. *Econometric Theory*, *7*(02), 200–212.

Koedijk, K. G., Kool, C. J. (1992). Tail estimates of East European exchange rates. *Journal of Business & Economic Statistics*, *10*(1), 83–96.

Lau, A. H. L., Lau, H. S., Wingender, J. R. (1990). The distribution of stock returns: New evidence against the stable model. *Journal of Business & Economic Statistics*, *8*(2), 217–223.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. (1967). The variation of some other speculative prices. *The Journal of Business*, *40*(4), 393–413.

Newey, W. K., McFadden, D. (1994). *Large sample estimation and hypothesis testing, Handbook of econometrics* (Vol. 4, pp. 2111–2245). Amsterdam: Elsevier.

Park, J. Y., Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*(03), 468–497.

Phillips, P. C. (1990). Time series regression with a unit root and infinite-variance errors. *Econometric Theory*, *6*(01), 44–62.

Phillips, P. C. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, *82*(3), 1177–1195.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*(02), 186–199.

Resnick, S., Greenwood, P. (1979). A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, *9*(2), 206–221.

Samarakoon, D. M., Knight, K. (2009). A note on unit root tests with infinite variance noise. *Econometric Reviews*, *28*(4), 314–334.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics*, *28*(3), 435–459.

Zhang, R. M., Chan, N. H. (2012). Maximum likelihood estimation for nearly non-stationary stable autoregressive processes. *Journal of Time Series Analysis*, *33*(4), 542–553.