J. R. Statist. Soc. B (2004)
66, *Part* 1, *pp.* 63–78

Estimation and testing stationarity for double-autoregressive models

Shiqing Ling

Hong Kong University of Science and Technology, People's Republic of China

[Received June 2002. Revised May 2003]

Summary. The paper considers the double-autoregressive model $y_t = \phi y_{t-1} + \varepsilon_t$ with $\varepsilon_t = \eta_t \surd(\omega + \alpha y_{t-1}^2)$. Consistency and asymptotic normality of the estimated parameters are proved under the condition $E \ln|\phi + \surd\alpha\eta_t| < 0$, which includes the cases with $|\phi| = 1$ or $|\phi| > 1$ as well as $E(\varepsilon_t^2) = \infty$. It is well known that all kinds of estimators of ϕ in these cases are not normal when ε_t are independent and identically distributed. Our result is novel and surprising. Two tests are proposed for testing stationarity of the model and their asymptotic distributions are shown to be a function of bivariate Brownian motions. Critical values of the tests are tabulated and some simulation results are reported. An application to the US 90-day treasury bill rate series is given.

Keywords: Asymptotic normality; Brownian motion; Consistency; Double-autoregressive model; Lagrange multiplier test; Maximum likelihood estimator; Stationarity

1. Introduction

Consider the first-order autoregressive model with conditional heteroscedasticity

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \eta_t \surd(\omega + \alpha y_{t-1}^2), \end{aligned} \tag{1.1}$$

where $\omega, \alpha > 0$, $t \in \{1, 2, \ldots\}$, η_t is a sequence of independent and identically distributed (IID) random variables with mean 0 and variance 1, and y_0 is independent of $\{\eta_t : t \geqslant 1\}$. Let $\mathcal{F}_0$ and $\mathcal{F}_t$ be the σ-fields that are generated by $\{y_0\}$ and $\{\eta_t, \ldots, \eta_1, y_0\}$ respectively. The conditional variance of y_t is $\operatorname{var}(y_t|\mathcal{F}_{t-1}) = \omega + \alpha y_{t-1}^2$, which is changing over time. Model (1.1) is a special case of the autoregressive moving average–autoregressive conditional heteroscedastic (ARCH) models in Weiss (1986) and an example of weak autoregressive moving average models in Francq and Zakoïan (1998, 2000), but it differs from Engle's (1982) ARCH model. We call model (1.1) the double-autoregressive (DAR) model. Guégan and Diebolt (1994) showed that the condition $\phi^2 + \alpha < 1$ is sufficient for $E(y_t^2) < \infty$. Borkovec and Kluppelberg (1998) further proved that this condition is also necessary. By theorem 3.3 in Borkovec and Kluppelberg (1998), $\{y_t\}$ is geometrically ergodic and has a unique stationary distribution if the following assumptions 1 and 2 hold. Furthermore, when y_0 is initialized from the stationary distribution, $\{y_t\}$ is strictly stationary. Thus, there is a unique strictly stationary and ergodic solution $\{y_t\}$ to model (1.1).

Assumption 1. η_t has a symmetric, positive and continuous Lebesgue density in $R \equiv (-\infty, \infty)$ with mean 0 and variance 1.

Address for correspondence: Shiqing Ling, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, People's Republic of China.
E-mail: maling@ust.hk

 1369–7412/04/66063

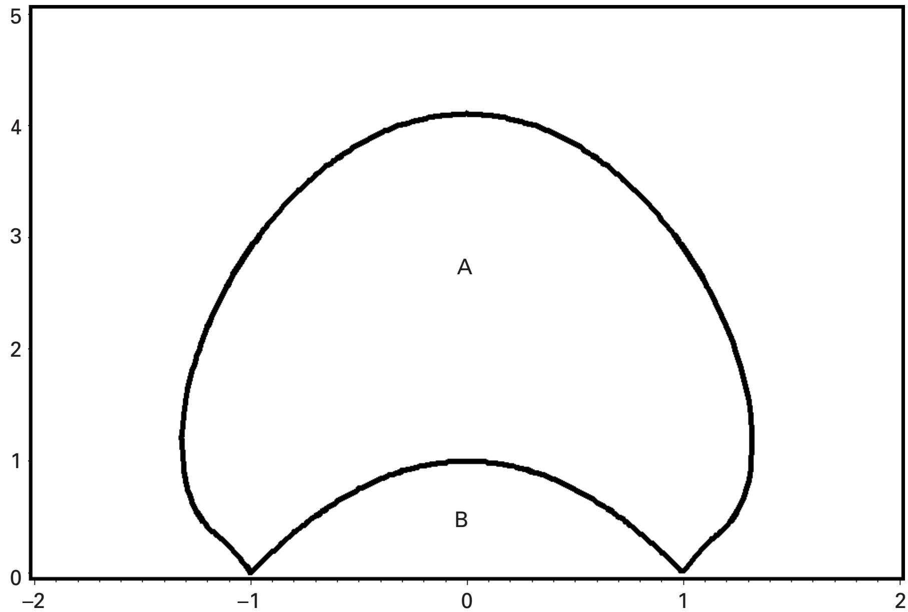

Fig. 1. $E(\ln|\phi+\eta_t\sqrt{\alpha}|)<0$ as $(\phi,\alpha)\in A\cup B$ and $\phi^2+\alpha<1$ as $(\phi,\alpha)\in B$

Assumption 2. The parametric space is $\Theta=\{\lambda\equiv(\phi,\omega,\alpha):E(\ln|\phi+\eta_t\sqrt{\alpha}|)<0$ with $|\phi|\leqslant\tilde{\phi}$, $\underline{\omega}\leqslant\omega\leqslant\tilde{\omega}$, and $\underline{\alpha}\leqslant\alpha\leqslant\tilde{\alpha}\}$, where $\tilde{\phi}$, $\underline{\omega}$, $\tilde{\omega}$, $\underline{\alpha}$ and $\tilde{\alpha}$ are some finite positive constants, and the true parameter value λ_0 is an interior point in Θ.

When η_t is normal, assumption 1 is satisfied. Fig. 1 gives the regions of (ϕ,α) such that $\phi^2+\alpha<1$ and $E(\ln|\phi+\eta_t\sqrt{\alpha}|)<0$. Model (1.1) in this case is equivalent in distribution to the random coefficient autoregressive (RCAR) model

$$y_t=(\phi+\alpha_t)y_{t-1}+\omega_t, \tag{1.2}$$

where (α_t,ω_t) are IID bivariate normal with mean 0 and covariance $\text{diag}(\alpha,\omega)$. The proof of this equivalence for the general case is given in Ling (2003). Model (1.2) is a special RCAR model of Nicholls and Quinn (1982). The second-order stationarity condition of models (1.1) and (1.2) are identical (see Nicholls and Quinn (1982)). Bougerol and Picard (1992) showed that the necessary and sufficient condition for strict stationarity and ergodicity of model (1.2) is $E(\ln|\phi+\alpha_t|)<0$. Tsay (1987) and Tong (1990) found the connection between the RCAR model and the ARCH model. Tsay (1987) also proposed a class of random-coefficient ARCH-type models. The asymptotic normality of the estimated parameters in Nicholls and Quinn (1982) and Weiss (1986) requires $E(y_t^4)<\infty$, whereas those in Tsay (1987) require $E(y_t^6)<\infty$. Weiss's ARCH models, the RCAR models and Tsay's ARCH-type models are useful. Some empirical examples can be found in the literature; see, for example, Wong and Li (1997) and Li *et al.* (2001).

In Section 2, we show that the estimated parameters in model (1.1) are consistent and asymptotically normal under assumptions 1 and 2. The condition $E(\ln|\phi+\eta_t\sqrt{\alpha}|)<0$ in assumption 2 includes the cases with $|\phi|\geqslant 1$ as well as $E(\varepsilon_t^2)=\infty$ (see remark 1). It is well known that all kinds of estimators of ϕ in these cases are not normal when ε_t are IID. Our result is therefore novel and surprising. Model (1.1) is non-stationary at the boundary points $(\phi,\alpha)=(\pm 1,0)$ and nests the unit root models at the two points. In Section 3, two Lagrange multiplier (LM) tests are proposed for testing stationarity of model (1.1). Their asymptotic distributions are shown to be functions of bivariate Brownian motions and critical values are tabulated. Some simulation results are reported and an application to the US 90-day treasury bill rate series are given in Sections 4 and 5 respectively. We present some concluding remarks in Section 6.

2. Quasi-maximum-likelihood estimator and its asymptotics

Suppose that $y_1, \ldots, y_n$ are generated by model (1.1). The conditional log-likelihood function (ignoring the constant) can be written as

$$
\begin{aligned}
L_n(\lambda) &= \sum_{t=2}^{n} l_t(\lambda), \\
l_t(\lambda) &= -\frac{1}{2}\ln(\omega + \alpha y_{t-1}^2) - \frac{(y_t - \phi y_{t-1})^2}{2(\omega + \alpha y_{t-1}^2)}.
\end{aligned}
\tag{2.1}
$$

Since we do not assume that η_t is normal, $L_n(\lambda)$ is the quasi-log-likelihood function. The maximizer $\hat{\lambda}_n$ of $L_n(\lambda)$ is called the quasi-maximum-likelihood estimator (QMLE) of λ_0. The following theorem gives the asymptotic properties of the QMLE and its proof is given in Appendix A.

Theorem 1. Suppose that $\{y_t : t = 0, 1, \ldots\}$ is the strictly stationary and ergodic solution of model (1.1) and assumptions 1 and 2 hold. Then, as $n \to \infty$,

(a) $\hat{\lambda}_n \to \lambda_0$ in probability and
(b) furthermore, if $E(\eta_t^4) < \infty$, then

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \underset{\mathcal{L}}{\to} N\{0, \operatorname{diag}(\Sigma^{-1}, \kappa\Omega^{-1})\},$$

where $\to_{\mathcal{L}}$ denotes convergence in distribution, $\kappa = E(\eta_t^4) - 1$, $\Sigma = E\{y_t^2/(\omega_0 + \alpha_0 y_t^2)\}$ and

$$\Omega = E\left\{ \begin{pmatrix} 1 & y_t^2 \\ y_t^2 & y_t^4 \end{pmatrix} \Big/ (\omega_0 + \alpha_0 y_t^2)^2 \right\}$$

is positive definite. In particular, when η_t is normal, $\kappa = 2$.

Remark 1. The assumption on the symmetry of η_t is one of the sufficient conditions for the existence of a strictly stationary and ergodic solution to model (1.1). Apart from resulting in the block diagonal form of the information matrix, it is not used in the proof in Appendix A. The compactness of the parameter space Θ may be strong. We suspect that it can be relaxed by some other arguments. If only that there is a sequence of estimators $\hat{\lambda}_n$ to maximize $L_n(\lambda)$ is required, then the asymptotic theory in Tjøstheim (1986) can be used to establish strong consistency. Since $E(\varepsilon_t^2) = \omega + \alpha\, E(y_{t-1}^2)$, $E(\varepsilon_t^2) = \infty$ if and only if $E(y_t^2) = \infty$. Even if $E(\eta_t^4) < \infty$, the region of (ϕ, α) such that $E(\varepsilon_t^2) = \infty$ may be very large (see Fig. 1). The lower bounds $\underline{\omega}$ and $\underline{\alpha}$ in assumption 2 are only for the technical reason in the proof of theorem 1. In practice, we can select them to be very close to 0. By lemmas 3 and 5 in Appendix A, Ω and Σ can be estimated consistently by

$$
\begin{aligned}
\hat{\Sigma}_n &= \frac{1}{n}\sum_{t=1}^{n} \frac{y_t^2}{\hat{\omega}_n + \hat{\alpha}_n y_t^2}, \\
\hat{\Omega}_n &= \frac{1}{n}\sum_{t=1}^{n} \frac{1}{(\hat{\omega}_n + \hat{\alpha}_n y_t^2)^2} \begin{pmatrix} 1 & y_t^2 \\ y_t^2 & y_t^4 \end{pmatrix}.
\end{aligned}
\tag{2.2}
$$

Similarly, it is straightforward to show that κ can be estimated consistently by

$$\frac{1}{n}\sum_{t=1}^{n}\frac{(y_t-\hat{\phi}_n y_{t-1})^4}{(\hat{\omega}_n+\hat{\alpha}_n y_{t-1}^2)^2}-1.$$

The asymptotic theory of the estimated ϕ_0 in model (1.1) with IID ε_t is well known in the literature. When $|\phi_0|<1$ and $E(\varepsilon_t^2)<\infty$, all types of estimators of ϕ_0 are asymptotically normal, such as the least squares estimator (LSE) and the M-estimator, among many others. When $|\phi_0|<1$ and $E(\varepsilon_t^2)=\infty$, all types of estimators of ϕ_0 follow stable laws asymptotically with a convergence rate that is much faster than $\sqrt{n}$ (see Davis *et al.* (1992) and Mikosch *et al.* (1995)). When $|\phi_0|=1$, all types of estimators of ϕ_0 are asymptotically a function of standard Brownian motions with convergence rate n if $E(\varepsilon_t^2)<\infty$, or they are asymptotically a function of a bivariate Lévy process if $E(\varepsilon_t^2)=\infty$ (see Chan and Tran (1989)). When $|\phi_0|>1$, the LSE and the adaptive estimator of ϕ_0 are distributed as a mixed normal with convergence rate ϕ_0^n if $E(\log^+|\varepsilon_t|)<\infty$ (see Jeganathan (1988) and Koul and Pflug (1990)). Theorem 1 shows that the QMLE of ϕ_0 in model (1.1) may still be asymptotically normal in the cases with $|\phi_0|=1$ or $|\phi_0|>1$ as well as $E(\varepsilon_t^2)=\infty$ (see Fig. 1). This is a striking contrast with the results when ε_t are IID. Our results may provide some new research insights in the field of time series.

The main reason for our result is that the conditional variance $\omega+\alpha y_{t-1}^2$ can control the log-likelihood function, score function and information matrix in such a way that they are bounded. This point is quite similar to that of the integrated generalized autoregressive conditional heteroscedastic (GARCH) model, which has an infinite variance, but its QMLE is still asymptotically normal (see Lee and Hansen (1994) and Lumsdaine (1996)). When ε_t is an ARCH-type error and $E(\varepsilon_t^4)=\infty$, the asymptotic properties of the QMLE and the LSE of the parameters in the AR part are quite unclear. When y_t follows a non-trivial AR model, only the consistency of the QMLE was proved by Ling and McAleer (2003). Otherwise, all the results on the QMLE require $E(\varepsilon_t^4)<\infty$, such as those in Nicholls and Quinn (1982), Weiss (1986), Tsay (1987), Ling and Li (1997, 1998) and Ling and McAleer (2003). From Borkovec's (2001) results, the conjecture is that the LSE for the AR part would no longer be asymptotically normal if $E(\varepsilon_t^2)<\infty$ but $E(\varepsilon_t^4)=\infty$. However, theorem 1 shows that the QMLE is still asymptotically normal in this case. Our method here may provide some new clues for further research on ARCH-type time series models.

We should mention that ε_t in model (1.1) is different from IID white noise because it is only uncorrelated and depends on y_{t-1}, and hence the structure of model (1.1) is rather different from the classical AR(1) model. Model (1.1) is also different from the models in Ling and Li (1997, 1998, 2003), where ε_t is specified as the standard GARCH model and does not directly depend on the observation y_{t-1}.

3. Lagrange multiplier tests for stationarity

In economic and financial studies, a very important issue is to test the non-stationarity of a time series. This is related to testing whether or not a financial market is efficient. The standard method for this in the simple case is to test the null hypothesis $H_0:\phi=1$ in the AR(1) model. Under H_0, the time series is interpreted as non-stationary in both the strict and the weak senses. For model (1.1), this null hypothesis tells us only that the series is not weakly stationary (i.e. covariance stationary, which implies $E(y_t^2)<\infty$), but it may still be strictly stationary. All the unit root tests are no longer valid for model (1.1) if $\alpha\neq 0$. For example, the tests in Phillips (1987) require $E(\varepsilon_t^2)<\infty$, whereas this is not satisfied because $\phi^2+\alpha>1$ results in $E(\varepsilon_t^2)=\infty$ in this case. The condition $E(\ln|\phi+\eta_t\sqrt{\alpha}|)<0$ in assumption 2 may not be necessary for the strict stationarity of model (1.1). However, it is clear that it is neither strictly nor weakly stationary at the boundary points $(\phi, \alpha) = (\pm 1, 0)$ and nests the standard unit root model at these points. In this case, $\ln|\phi + \eta_t \sqrt{\alpha}| = 0$, so assumption 2 is not satisfied. Thus, it is interesting to test

$$H_0 : (\phi, \alpha) = (\pm 1, 0) \quad \text{against} \quad H_1 : (\phi, \alpha) \neq (\pm 1, 0). \tag{3.1}$$

Under hypothesis H_0, $y_t = \pm y_{t-1} + \eta_t \sqrt{\omega}$ and hence it is the standard unit root processes. Under hypothesis H_1, y_t is stationary if assumption 2 is satisfied. As the classical unit root tests, we cannot claim that y_t is stationary when H_0 is rejected. In this case, a further study is needed. If $E(\ln|\phi + \eta_t \sqrt{\alpha}|) < 0$, we can construct a Wald test to test whether or not $\phi = 1$ by using theorem 1. If $\phi = 1$ in this case, we can claim that model (1.1) is strictly stationary but not weakly stationary. How to test $E(\ln|\phi + \eta_t \sqrt{\alpha}|) < 0$ remains an interesting issue.

We use the LM principle for testing hypothesis H_0. This principle has been widely applied in statistics and its explanation can be found in Hamilton (1994). For hypotheses (3.1), the LM test can be defined as

$$\begin{aligned} \mathrm{LM}_n &= \left\{ \sum_{t=1}^{n} \frac{\partial l_t(\lambda)}{\partial m'} \right\} \left\{ \sum_{t=1}^{n} \frac{\partial^2 l_t(\lambda)}{\partial m \partial m'} \right\}^{-1} \sum_{t=1}^{n} \frac{\partial l_t(\lambda)}{\partial m} \Bigg|_{\lambda = (\hat{\omega}_n, \pm 1, 0)} \\ &= \frac{\left(\sum_{t=1}^{n} y_{t-1} \varepsilon_t \right)^2}{\hat{\omega}_n \sum_{t=1}^{n} y_t^2} + \frac{1}{2} \frac{\left\{ \sum_{t=1}^{n} y_{t-1}^2 (\hat{\omega}_n - \varepsilon_t^2) \right\}^2}{\hat{\omega}_n^2 \sum_{t=1}^{n} y_t^4}, \end{aligned} \tag{3.2}$$

where $m = (\phi, \alpha)$, $\hat{\omega}_n = \Sigma_{t=1}^n \varepsilon_t^2 / n$ and $\varepsilon_t = y_t - y_{t-1}$. Many time series have non-zero means. In these cases, we may consider the model

$$x_t = \mu + y_t, \tag{3.3}$$

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \eta_t \sqrt{(\omega + \alpha y_{t-1}^2)}. \end{aligned} \tag{3.4}$$

Let $x_t^* = x_t - \Sigma_{t=1}^n x_t / n$ be the demeaned variable of x_t. Since x_t^* has the mean removed, unit root tests based on x_t^* are invariant to the mean. The LM test, denoted by $\mathrm{LM}_{\mu n}$, based on the demeaned variable x_t^* for the null hypothesis (3.1) is defined as in equation (3.2) with y_t replaced by x_t^*. The following theorem gives the limiting distribution of LM_n and $\mathrm{LM}_{\mu n}$.

Theorem 2. If $E(\eta_t^3) = 0$ and $E(\eta_t^4) < \infty$, then under the null hypothesis H_0,

(a) $$\mathrm{LM}_n \underset{\mathcal{L}}{\rightarrow} \frac{\left\{ \int_0^1 B_1(\tau) \, \mathrm{d}B_1(\tau) \right\}^2}{\int_0^1 B_1^2(\tau) \, \mathrm{d}\tau} + \frac{\kappa}{2} \frac{\left\{ \int_0^1 B_1^2(\tau) \, \mathrm{d}B_2(\tau) \right\}^2}{\int_0^1 B_1^4(\tau) \, \mathrm{d}\tau},$$

(b) $$\mathrm{LM}_{\mu n} \underset{\mathcal{L}}{\rightarrow} \frac{\left\{ \int_0^1 \tilde{B}_1(\tau) \, \mathrm{d}\tilde{B}_1(\tau) \right\}^2}{\int_0^1 \tilde{B}_1^2(\tau) \, \mathrm{d}\tau} + \frac{\kappa}{2} \frac{\left\{ \int_0^1 \tilde{B}_1^2(\tau) \, \mathrm{d}B_2(\tau) \right\}^2}{\int_0^1 \tilde{B}_1^4(\tau) \, \mathrm{d}\tau},$$

as $n \to \infty$, where $B_1(\tau)$ and $B_2(\tau)$ are two independent and standard Brownian motions, $\tilde{B}(\tau) = B_1(\tau) - B_1(1)$ and κ is defined as in theorem 1.

Proof. By the invariance principle (e.g. theorem 4.4 of Hall and Heyde (1980)), it is easy to show that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \{(\pm 1)^t \eta_t, 1-\eta_t^2\} \underset{\mathcal{L}}{\to} \{B_1(\tau), B_2(\tau)\sqrt{\kappa}\} \qquad \text{in } D^2, \tag{3.5}$$

as $n \to \infty$, where $D = D[0,1]$ denotes the space of functions $f(s)$ on [0,1], which is defined and equipped with the Skorokhod topology (Billingsley, 1968). By expression (3.5) and the continuity mapping theorem, we have

$$\frac{1}{\sqrt{n}} (\pm 1)^{[n\tau]} y_{[n\tau]} = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} (\pm 1)^t \varepsilon_t \underset{\mathcal{L}}{\to} B_1(\tau)\sqrt{\omega} \qquad \text{in } D. \tag{3.6}$$

By expressions (3.5) and (3.6) and theorem 2.2 in Kurtz and Protter (1991), we have

$$\frac{1}{n} \sum_{t=1}^{n} \{(\pm 1)^{t-1} y_{t-1}\}(\pm 1)^t \varepsilon_t \underset{\mathcal{L}}{\to} \omega \int_0^1 B_1(\tau)\, \mathrm{d}B_1(\tau), \tag{3.7}$$

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} y_{t-1}^2 (\omega - \varepsilon_t^2) \underset{\mathcal{L}}{\to} \omega^2 \int_0^1 B_1^2(\tau)\, \mathrm{d}B_2(\tau). \tag{3.8}$$

By expressions (3.6)–(3.8) and the continuity mapping theorem, we can show that part (a) holds. Similarly, we can show that part (b) holds. This completes the proof. □

Remark 2. It is easy to show that

$$\left\{ \int_0^1 B_1^4(\tau)\, \mathrm{d}\tau \right\}^{-1/2} \int_0^1 B_1^2(\tau)\, \mathrm{d}B_2(\tau)$$

and

$$\left\{ \int_0^1 \tilde{B}_1^4(\tau)\, \mathrm{d}\tau \right\}^{-1/2} \int_0^1 \tilde{B}_1^2(\tau)\, \mathrm{d}B_2(\tau)$$

are standard normal variables and are independent of $B_1(\tau)$ and $\tilde{B}_1(\tau)$ respectively; see, for example, Phillips (1989) and Kluppelberg *et al.* (2002). Thus, we can simplify the asymptotic distributions of LM_n and $\mathrm{LM}_{\mu n}$ respectively as

$$\frac{\left\{ \int_0^1 B_1(\tau)\, \mathrm{d}B_1(\tau) \right\}^2}{\int_0^1 B_1^2(\tau)\, \mathrm{d}\tau} + \frac{\kappa \xi^2}{2}$$

and

$$\frac{\left\{ \int_0^1 \tilde{B}_1(\tau)\, \mathrm{d}\tilde{B}_1(\tau) \right\}^2}{\int_0^1 \tilde{B}_1^2(\tau)\, \mathrm{d}\tau} + \frac{\kappa \xi^2}{2},$$

where $\xi \sim N(0,1)$ and independent of $B_1(\tau)$ and $\tilde{B}_1(\tau)$. Kluppelberg *et al.* (2002) investigated the likelihood ratio test for the null hypothesis $H_0 : \phi = 1, \omega > 0, \alpha = 0$. Their null hypothesis is different from that in expression (3.1). Thus, the asymptotic distributions of LM_n and their likelihood ratio test are different. Moreover, the LM tests are very simple because they require

Table 1. Critical values for LM tests

Sample size n	*Probability that the LM value is greater than the following values:*					
	0.500	*0.200*	*0.100*	*0.050*	*0.025*	*0.010*
LM_n (without mean adjustment)						
100	1.509	3.369	4.761	6.125	7.572	9.328
250	1.529	3.408	4.770	6.189	7.605	9.522
500	1.534	3.417	4.824	6.195	7.552	9.327
∞	1.545	3.440	4.851	6.238	7.680	9.523
$LM_{\mu n}$ (with mean adjustment)						
100	3.367	6.086	7.886	9.554	11.247	13.321
250	3.390	6.138	7.937	9.690	11.354	13.555
500	3.406	6.173	8.009	9.769	11.495	13.646
∞	3.441	6.233	8.118	9.827	11.604	13.686

only the estimation of ω_0 under the null hypothesis H_0. Hypotheses (3.1) are also different from those in Ling and Li (2003), where they test for $H_0 : \phi_0 = 1$ and assume that ε_t is the standard GARCH(1,1) process. Hypothesis (3.1) is actually a composite hypothesis for the unit root and for heteroscedasticity. Under H_0 in expression (3.1), model (1.1) reduces to the classical AR(1) model and hence the Dickey–Fuller tests are still valid. However, H_1 in expression (3.1) is different from the alternative of the Dickey–Fuller tests. Under H_1 with $\alpha \neq 0$, it is expected that the LM tests are more powerful than the Dickey–Fuller tests. Under some other alternatives for testing the unit root, we refer to Robinson (1994). Under the local alternative $H_{1n} : (\phi, \alpha) = (\pm 1 + u/n, v/\sqrt{n})$ with constants u and v, the limiting distributions of LM_n and $\mathrm{LM}_{\mu n}$ can be obtained. Since it involves more technical details, we do not give them here.

LM_n and $\mathrm{LM}_{\mu n}$ are new test statistics and their critical values with $\kappa = 2$ are tabulated in Table 1 with sample sizes $n = 100, 250, 500, 5000\,(\infty)$. Table 1 is constructed by using 60000 replications with a sequence of the bivariate IID normal vectors $N(0, I_2)$ via simulation. From Table 1, we see that its entries do not change much for $n \geqslant 250$. This indicates that the limiting distributions in theorem 2 provide good approximations to those of LM_n and $\mathrm{LM}_{\mu n}$ in finite samples.

4. Simulation studies

This section examines the performance of the asymptotic results in finite samples through Monte Carlo experiments. We first study the biases and standard deviations of the maximum likelihood estimates (MLEs). The true observations were generated through model (1.1) with $\eta_t \sim^{\mathrm{IID}} N(0, 1)$. The true parameters are $\omega = 1$ and $(\phi, \alpha) = (0.0, 3.0)$, (0.5, 0.5), (1.0, 0.5), (1.0, 1.0), (1.0, 1.5), (1.2, 0.7) and (1.2, 1.0). The sample sizes are $n = 200$ and $n = 400$. 1000 replications were used. Table 2 summarizes the empirical biases, empirical standard deviations SD and asymptotic standard deviations AD of the MLEs of (ϕ, ω, α). The ADs are calculated by using the estimated covariances in expression (2.2). Table 2 shows that all the biases of the MLEs are very small. The SDs and ADs are very close, particularly, when the sample has $n = 400$. When α becomes larger, all SDs and ADs become larger. As the sample size n is increased from 200 to 400, all SDs and ADs become smaller. The simulation results indicate that the MLE performs very well in the finite samples.

Table 2. Bias and standard deviations of MLEs for DAR models†

ϕ	ω	α		Results for the following values of n:					
				n = 200			n = 400		
				$\hat{\phi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\omega}$	$\hat{\alpha}$
0.0	1	3.0	Bias	0.003	0.119	−0.072	0.000	0.060	−0.043
			SD	0.128	0.841	0.337	0.090	0.517	0.235
			AD	0.127	0.688	0.318	0.090	0.409	0.225
0.5	1	0.5	Bias	−0.009	0.006	−0.016	−0.007	0.006	−0.013
			SD	0.087	0.169	0.131	0.059	0.118	0.092
			AD	0.083	0.142	0.121	0.059	0.101	0.086
1.0	1	0.5	Bias	−0.011	0.002	−0.001	−0.007	0.005	−0.004
			SD	0.067	0.210	0.080	0.046	0.146	0.059
			AD	0.065	0.192	0.077	0.046	0.135	0.054
1.0	1	1.0	Bias	−0.011	−0.004	−0.006	−0.008	0.005	−0.007
			SD	0.085	0.268	0.133	0.059	0.190	0.097
			AD	0.082	0.249	0.129	0.058	0.174	0.091
1.0	1	1.5	Bias	−0.011	−0.005	0.010	−0.007	−0.001	−0.009
			SD	0.097	0.352	0.179	0.066	0.261	0.127
			AD	0.094	0.468	0.170	0.067	0.257	0.119
1.2	1	0.7	Bias	−0.020	−0.070	0.002	−0.011	−0.027	−0.001
			SD	0.061	0.398	0.084	0.044	0.324	0.059
			AD	0.064	0.418	0.080	0.045	0.293	0.055
1.2	1	1.0	Bias	−0.022	−0.084	−0.001	−0.013	−0.032	−0.005
			SD	0.076	0.333	0.113	0.051	0.278	0.083
			AD	0.077	0.361	0.112	0.053	0.285	0.077

†1000 replications.

We now investigate the size and power of LM_n and $\mathrm{LM}_{\mu n}$. Again, 1000 replications were used. The sample size is $n = 250$. DF and DF_μ are respectively the $\hat{\tau}$- and $\hat{\tau}_\mu$-test statistics in Dickey and Fuller (1979) for the cases without means and with adjusted means. Table 3 reports the sizes and powers of DF, DF_μ, LM_n and $\mathrm{LM}_{\mu n}$ at significance levels 5% and 10%. All the sizes are reasonably close to the nominal values 5% and 10% and are somewhat conservative. When $\phi = 0.98$, 0.95, 0.90, and $\alpha = 0$, the Dickey–Fuller tests have a little more power than the LM tests. This is reasonable since the alternatives in these cases are exactly AR models, and α makes no contribution to the powers of LM_n and $\mathrm{LM}_{\mu n}$. When $\phi = 0.98$, and $\alpha = 0.01$, LM_n and $\mathrm{LM}_{\mu n}$ are respectively more powerful than DF and DF_μ. However, when $\phi = 0.95$, 0.90, and $\alpha = 0.01$, DF and DF_μ are respectively more powerful than LM_n and $\mathrm{LM}_{\mu n}$. When $\phi = 0.98$, 0.95, 0.90, and $\alpha = 0.02$, LM_n and $\mathrm{LM}_{\mu n}$ are respectively more powerful than DF and DF_μ. This means that the powers in these cases are controlled by the alternatives of α. When $\phi = 1$, and $\alpha = 0.01$, 0.02, 0.03, LM_n and $\mathrm{LM}_{\mu n}$ are quite powerful, whereas DF and DF_μ have very low power. These experiments show that the LM tests are useful for testing the stationarity of model (1.1).

5. An empirical example

This section applies model (1.1) to the US monthly interest rate series over the period July 1972–August 2001. The series is the 90-day treasury bill rate and has 350 observations in total. We take x_t to be the logarithms of the 90-day treasury bill rate and $y_t = x_t - x_{t-1}$. Figs 2 and 3 are

Table 3. Sizes and powers of LM tests for the null hypothesis H_0: $(\phi, \alpha) = (1, 0)$†

ϕ	α	*Results at the following significance levels:*							
		5%				*10%*			
		DF	LM_n	DF_μ	$LM_{\mu n}$	DF	LM_n	DF_μ	$LM_{\mu n}$
Sizes									
1.00	0.00	0.041	0.037	0.046	0.049	0.081	0.068	0.103	0.097
Powers									
0.98	0.00	0.324	0.196	0.113	0.115	0.533	0.318	0.218	0.197
0.95	0.00	0.903	0.674	0.451	0.404	0.986	0.837	0.648	0.584
0.90	0.00	0.999	0.993	0.966	0.943	1.000	0.999	0.997	0.986
0.98	0.01	0.389	0.546	0.131	0.584	0.609	0.689	0.244	0.698
0.95	0.01	0.906	0.772	0.477	0.668	0.982	0.875	0.660	0.803
0.90	0.01	0.999	0.984	0.961	0.959	1.000	0.999	0.995	0.992
0.98	0.02	0.471	0.693	0.159	0.895	0.691	0.811	0.287	0.946
0.95	0.02	0.917	0.984	0.503	0.904	0.978	0.996	0.690	0.961
0.90	0.02	0.998	1.000	0.955	0.983	1.000	1.000	0.994	0.997
1.00	0.01	0.075	0.869	0.059	0.682	0.166	0.919	0.108	0.757
1.00	0.02	0.132	0.985	0.072	0.927	0.282	0.995	0.137	0.956
1.00	0.03	0.216	1.000	0.091	0.977	0.470	1.000	0.178	0.989

†1000 replications.

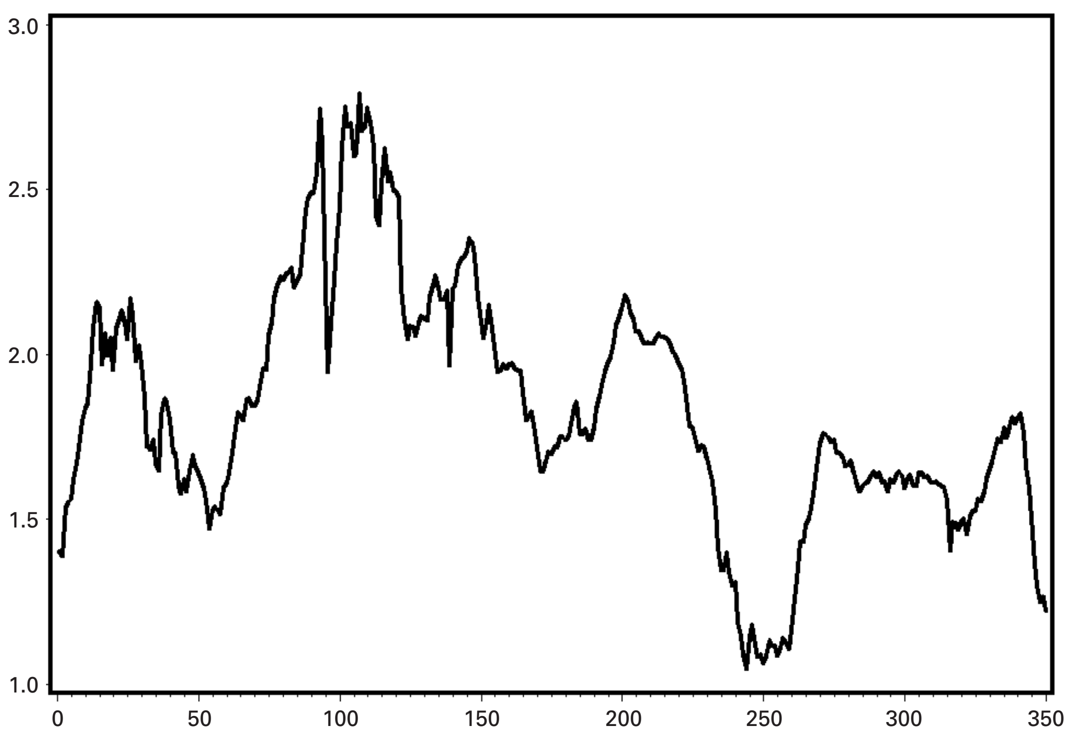

Fig. 2. Time plot of x_t (logarithms of the 90-day treasury bill rate)

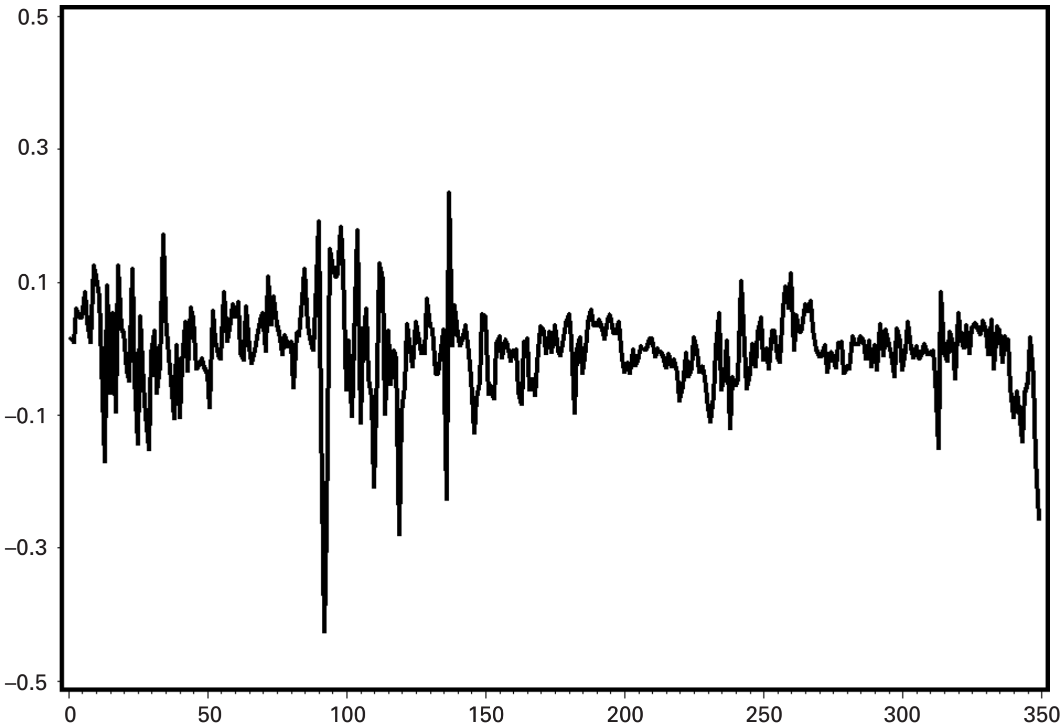

Fig. 3. Time plot of y_t (first difference of x_t)

time plots of x_t and y_t respectively. We first fit an AR(1) model with normal errors to the data y_t. The estimated ϕ is 0.3877 and its estimated standard deviation is 0.052. The corresponding value of the log-likelihood is 405.96. The portmanteau test statistic $Q(M)$ in Li (1992) is used for checking the adequacy of the model. When $M = 3, 6, 12$, their values are $Q(3) = 0.7$, $Q(6) = 2.90$ and $Q(12) = 10.31$. These suggest that the AR(1) model is adequate for the data y_t at significance level 5%.

We then use model (1.1) to fit the data. In estimation, we take $\underline{\omega} = \underline{\alpha} = 0.00001$. The result is

$$y_t = 0.3810(0.0777)y_{t-1} + \varepsilon_t, \tag{5.1}$$

$$\varepsilon_t = \eta_t\sqrt{\{0.0023(0.0002) + 0.6209(0.0908)y_{t-1}^2\}}, \tag{5.2}$$

where the values in parentheses are the corresponding standard deviations calculated by using the estimated covariances in expression (2.2). The value of the log-likelihood is 483.87. $Q(M)$ and the portmanteau test statistic $Q^2(M)$ in Li and Mak (1994) with $M = 3, 6, 12$ are used for checking the adequacy of the model. Their values are $Q(3) = 1.75$, $Q(6) = 3.77$, $Q(12) = 19.65$, $Q^2(3) = 2.61$, $Q^2(6) = 3.50$ and $Q^2(12) = 15.86$. All these test statistics suggest that model (5.1)–(5.2) is adequate for the data at significance level 5%. Since $\lambda \notin \Theta$ if $\alpha = 0$, the Wald test cannot be used to test $H_0 : \alpha_0 = 0$. However, the null hypothesis $H_0 : \alpha_0 = 0.1$ is strongly rejected by this test at significance level 5%. We may claim that $\alpha_0 \neq 0$. From the values of the log-likelihood, model (5.1)–(5.2) is much better than the AR(1) model. Fig. 4 shows time plots of data x_t with 95% confidence intervals for the one-step-ahead forecast errors based on the AR(1) model and model (5.1)–(5.2), starting from February 1989. Fig. 4 shows that the confidence intervals based on model (5.1)–(5.2) are narrower than those based on the AR(1)

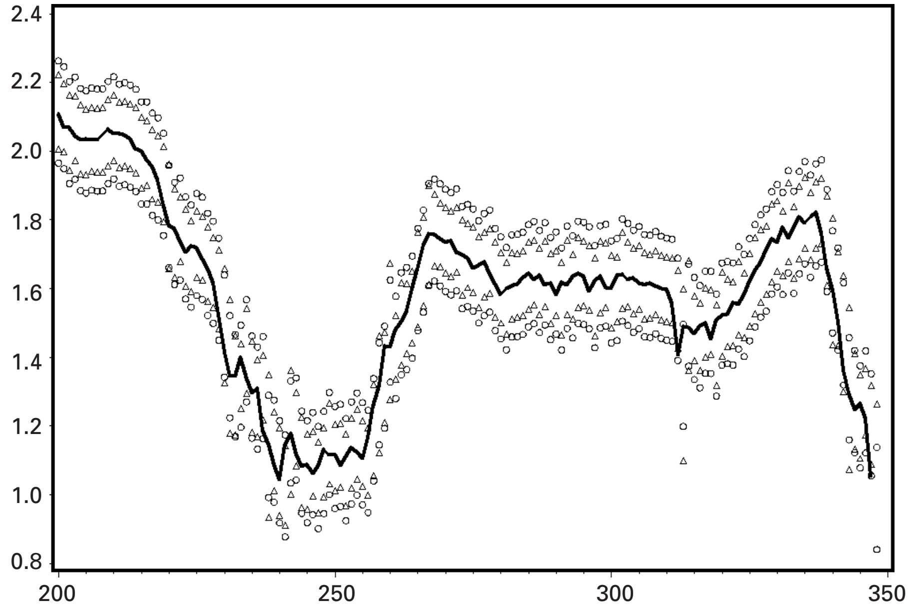

Fig. 4. Plot of the data x_t (———), 95% forecasting confidence intervals based on an AR model (○) and 95% forecasting confidence intervals based on a DAR model (△)

model at nearly all the data points. From this, model (5.1)–(5.2) is also clearly superior to the AR(1) model.

It is interesting to see the result when we use the AR(1)–GARCH(1,1) model to fit the data y_t. It is

$$y_t = 0.3394(0.0603)y_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \eta_t\sqrt{h_t}, \tag{5.3}$$

$$h_t = 0.00001(0.00004) + 0.3946(0.0422)\varepsilon_{t-1}^2 + 0.6913(0.0252)h_{t-1}, \tag{5.4}$$

and the value of the log-likelihood is 503.90. Since the summation of the coefficients of ε_{t-1}^2 and h_{t-1} is greater than 1, y_t does not have a finite variance. As discussed in Section 2, no asymptotic normality result is available for the AR(1)–GARCH(1,1) model if $\phi \neq 0$ and y_t does not have a finite fourth moment and hence it is difficult to use model (5.3)–(5.4) for statistical inferences. It also does not makes sense in this case to compare the values of the log-likelihood of models (5.1)–(5.2) and (5.3)–(5.4). In this example, the return series y_t are correlated. The high (or low) return associated with the high (or low) risk is directly reflected in model (5.1)–(5.2) through its conditional variance. Thus, model (5.1)–(5.2) seems to be more reasonable and natural than model (5.3)–(5.4) for the data y_t.

6. Concluding remarks

This paper has established an asymptotic theory for the first-order DAR model under some quite weak conditions. LM tests have been proposed for testing the stationarity of the DAR model and their critical values have been tabulated. From the simulation studies and the empirical example, it seems that our results are useful. Although our results are only for the first-order DAR model, they can be easily extended to some higher order cases. When the order is higher than 1, it is difficult to obtain a strict stationarity and ergodicity condition as weak as assumption 2. A stronger sufficient condition can be found by using Markov chain theory such as that in Tjøstheim (1990) and the similar method in Ling (1999). However, it is a challenging open problem to find the necessary and sufficient condition for stationarity of the higher order DAR models.

Acknowledgements

The author thanks Professor Ross Maller of the University of Western Australia for providing a preprint of Borkovec and Kluppelberg (1998) and Dr C. Y. Sin, two referees, the Associate Editor and the Joint Editor for their very helpful comments and suggestions. This research is supported by Competitive Earmarked Research Grants HKUST6113/02P and HKUST6273/03H.

Appendix A: Lemmas and proof of theorem 1

The proof of theorem 1 is to verify the regular conditions in Amemiya (1985). Before doing this, we first establish several basic lemmas.

Lemma 1. Under assumptions 1 and 2, $E\{\sup_{\lambda\in\Theta}|l_t(\lambda)|\}<\infty$ and $\sup_{\lambda\in\Theta}|L_n(\lambda)/n-E\{l_t(\lambda)\}|=o_p(1)$, where $o_p(1)$ converges to 0 in probability as $n\to\infty$.

Proof. Let $q(u)=E|\phi_0+\eta_t\sqrt{\alpha_0}|^u$. Since $E(\eta_t^2)=1$ (the assumption of model (1.1)), it is easy to show that $q(u)$ is differentiable on $[0,2)$ and

$$q'(u)=E(|\phi_0+\eta_t\sqrt{\alpha_0}|^u\ \ln|\phi_0+\eta_t\sqrt{\alpha_0}|).$$

Note that $|\ln(x^s)|\leqslant\max(x^s,x^{-s})-1$ for all $x>0$ and $s\in[0,u/2]$. We can show that

$$E\Big\{\sup_{u\in[0,1]}(|\phi_0+\eta_t\sqrt{\alpha_0}|^u|\ln|\phi_0+\eta_t\sqrt{\alpha_0}||)\Big\}<\infty.$$

By the dominated convergence theorem and assumption 2, $\lim_{u\to 0}\{q'(u)\}=E(\ln|\phi_0+\eta_t\sqrt{\alpha_0}|)<0$. Thus, there is a constant $\tilde\delta\in(0,2)$ such that $q(u)$ is strictly decreasing on $[0,\tilde\delta]$. Hence, it follows that $E|\phi_0+\eta_t\sqrt{\alpha_0}|^{\tilde\delta}<q(0)=1$.

Choose $\delta\in(0,\tilde\delta)$ and let $g(x)=1+|x|^\delta$, where $x\in R$. Borkovec and Kluppelberg (1998) showed that there is an M that is sufficiently large that

$$\sup_{|x|\leqslant M}[E\{g(y_t)|y_{t-1}=x\}]<\infty, \tag{A.1}$$

$$E\{g(y_t)|y_{t-1}=x\}\leqslant(1-\delta)\,g(x), \qquad \text{if } |x|>M. \tag{A.2}$$

Since y_t is a Markov chain, by inequalities (A.1) and (A.2) and by theorem 2 in Feigin and Tweedie (1985), we know that $E|y_t|^\delta$ is a constant less than ∞ for all t. This implies that the δth moment of y_t is finite if assumption 2 holds. Let $\tilde\omega^*=\max(1,\tilde\omega)$. By Jensen's inequality, we have

$$\begin{aligned}E\{\ln(\tilde\omega^*+\tilde\alpha y_{t-1}^2)\}&=\frac{2}{\delta}E\{\ln(\tilde\omega^*+\tilde\alpha y_{t-1}^2)^{\delta/2}\}\\&\leqslant\frac{2}{\delta}\ln(\tilde\omega^{*\delta/2}+\tilde\alpha^{\delta/2}\,E|y_{t-1}|^\delta)<\infty,\end{aligned}\tag{A.3}$$

where the following elementary relationship is used: $(a+b)^s\leqslant a^s+b^s$ for all $a,b>0$ and $s\in[0,1]$. By inequality (A.3), we have

$$\begin{aligned}E\Big\{\sup_{\lambda\in\Theta}|\ln(\omega+\alpha y_{t-1}^2)|\Big\}&\leqslant E\Big[\sup_{\lambda\in\Theta}\{I(\omega+\alpha y_{t-1}^2\geqslant 1)\ \ln(\omega+\alpha y_{t-1}^2)\}\Big]\\&\quad+E\sup_{\lambda\in\Theta}\{-I(\omega+\alpha y_{t-1}^2\leqslant 1)\ \ln(\omega+\alpha y_{t-1}^2)\}\\&\leqslant E\{\ln(\tilde\omega^*+\tilde\alpha y_{t-1}^2)\}-I(\underline{\omega}<1)\ \ln(\underline{\omega})<\infty.\end{aligned}$$

Furthermore, since $y_t - \phi y_{t-1} = \varepsilon_t - (\phi - \phi_0) y_{t-1}$, it can be shown that

$$E\left[\sup_{\lambda\in\Theta}\left\{\frac{(y_t-\phi y_{t-1})^2}{\omega+\alpha y_{t-1}^2}\right\}\right] \leqslant 2\,E\left[\sup_{\lambda\in\Theta}\left\{\frac{(\phi-\phi_0)^2 y_{t-1}^2}{\omega+\alpha y_{t-1}^2}\right\}\right] + 2\,E\left\{\sup_{\lambda\in\Theta}\left(\frac{\omega_0+\alpha_0 y_{t-1}^2}{\omega+\alpha y_{t-1}^2}\right)\right\} < \infty. \tag{A.4}$$

Thus, $E\{\sup_{\lambda\in\Theta} |l_t(\lambda)|\} < \infty$. Because y_t is a sequence of strictly stationary and ergodic time series, we have $\sup_{\lambda\in\Theta} |L_n(\lambda)/n - E\{l_t(\lambda)\}| = o_p(1)$ by theorem 3.1 in Ling and McAleer (2003). □

Lemma 2. Under assumptions 1 and 2, $E\{l_t(\lambda)\}$ has a unique maximum at λ_0.

Proof. As in inequality (A.4), we can show that

$$\begin{aligned} E\{l_t(\lambda)\} &= -\frac{1}{2} E\left\{\ln(\omega+\alpha y_{t-1}^2) + \frac{(y_t-\phi y_{t-1})^2}{\omega+\alpha y_{t-1}^2}\right\} \\ &= -\frac{1}{2}\left\{E\,\ln(\omega+\alpha y_{t-1}^2) + E\left(\frac{\omega_0+\alpha_0 y_{t-1}^2}{\omega+\alpha y_{t-1}^2}\right)\right\} - \frac{(\phi-\phi_0)^2}{2} E\left(\frac{y_{t-1}^2}{\omega+\alpha y_{t-1}^2}\right). \end{aligned} \tag{A.5}$$

The second term in equation (A.5) reaches its maximum at 0, and this occurs if and only if $\phi = \phi_0$. The first term in equation (A.5) is equal to

$$-\tfrac{1}{2}[-E\{\ln(M_t)\} + E(M_t)] - \tfrac{1}{2} E\{\ln(\omega_0 + \alpha_0 y_{t-1}^2)\}, \tag{A.6}$$

where $M_t = (\omega_0 + \alpha_0 y_{t-1}^2)/(\omega + \alpha y_{t-1}^2)$. For any $x > 0$, $-f(x) \equiv -\ln(x) + x \geqslant 1$, and hence $-E\{\ln(M_t)\} + E(M_t) \geqslant 1$. When $M_t = 1$, we have $f(M_t) = f(1) = -1$. If $M_t \neq 1$, then $f(M_t) < f(1)$, so $E\{f(M_t)\} \leqslant E\{f(1)\}$ with equality only if $M_t = 1$ with probability 1. Thus, expression (A.6) reaches its maximum $-\frac{1}{2} - E\{\ln(\omega_0 + \alpha_0 y_{t-1}^2)/2\}$, and this occurs if and only if $\omega_0 + \alpha_0 y_{t-1}^2 = \omega + \alpha y_{t-1}^2$, which holds if and only if $\omega = \omega_0$ and $\alpha = \alpha_0$. Thus, $E\{l_t(\lambda)\}$ is uniquely maximized at λ_0. □

Lemma 3. Suppose that assumptions 1 and 2 hold and $E(\eta_t^4) < \infty$. Then

(a) $E\left[\sup_{\lambda\in\Theta}\left\{\frac{\partial l_t(\lambda)}{\partial\lambda'}\frac{\partial l_t(\lambda)}{\partial\lambda}\right\}\right] < \infty$ and

(b) $\sup_{\lambda\in\Theta}\left|\frac{1}{n}\sum_{t=2}^{n}\left[\frac{\partial l_t(\lambda)}{\partial\lambda'}\frac{\partial l_t(\lambda)}{\partial\lambda} - E\left\{\frac{\partial l_t(\lambda)}{\partial\lambda'}\frac{\partial l_t(\lambda)}{\partial\lambda}\right\}\right]\right| = o_p(1).$

Proof. By direct differentiation, we have

$$\begin{aligned} \frac{\partial l_t(\lambda)}{\partial\phi} &= \frac{y_{t-1}(y_t-\phi y_{t-1})}{\omega+\alpha y_{t-1}^2}, \\ \frac{\partial l_t(\lambda)}{\partial\omega} &= -\frac{1}{2(\omega+\alpha y_{t-1}^2)}\left\{1-\frac{(y_t-\phi y_{t-1})^2}{\omega+\alpha y_{t-1}^2}\right\}, \\ \frac{\partial l_t(\lambda)}{\partial\alpha} &= -\frac{y_{t-1}^2}{2(\omega+\alpha y_{t-1}^2)}\left\{1-\frac{(y_t-\phi y_{t-1})^2}{\omega+\alpha y_{t-1}^2}\right\}. \end{aligned}$$

As in inequality (A.4), it is easy to show that

$$\begin{aligned} E\left[\sup_{\lambda\in\Theta}\left\{\frac{\partial l_t(\lambda)}{\partial\phi}\right\}^2\right] &= E\left(\sup_{\lambda\in\Theta}\left[\frac{y_{t-1}^2\{\varepsilon_t-(\phi-\phi_0)y_{t-1}\}^2}{(\omega+\alpha y_{t-1}^2)^2}\right]\right) \\ &\leqslant E\left\{\frac{8\tilde{\phi}^2 y_{t-1}^4}{(\underline{\omega}+\underline{\alpha} y_{t-1}^2)^2}\right\} + 2\,E\left\{\frac{y_{t-1}^2\varepsilon_t^2}{(\underline{\omega}+\underline{\alpha} y_{t-1}^2)^2}\right\} \\ &\leqslant C_1 + 2\,E\left\{\frac{y_{t-1}^2(\omega_0+\alpha_0 y_{t-1}^2)}{(\underline{\omega}+\underline{\alpha} y_{t-1}^2)^2}\right\} \leqslant C < \infty, \end{aligned}$$

where C_1 and C are some constants. Similarly, we can show that $E[\sup_{\lambda\in\Theta}\{\partial l_t(\lambda)/\partial\omega\}^2]$, $E[\sup_{\lambda\in\Theta}\{\partial l_t(\lambda)/\partial\alpha\}^2]$ and other cross-product terms are finite in part (a). Thus, part (a) holds. By part (a) of lemma 3 and theorem 3.1 in Ling and McAleer (2003), we can claim that part (b) holds. □

Lemma 4. Suppose that assumptions 1 and 2 hold and $E(\eta_t^4) < \infty$. Then

(a) $E\left\{\frac{\partial l_t(\lambda_0)}{\partial \lambda'}\frac{\partial l_t(\lambda_0)}{\partial \lambda}\right\} = \tilde{\Omega} > 0$ and

(b) $\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\frac{\partial l_t(\lambda_0)}{\partial \lambda} \rightarrow_{\mathcal{L}} N(0, \tilde{\Omega})$,

where $\tilde{\Omega} = \text{diag}[E\{y_t^2/(\omega_0 + \alpha_0 y_t^2)\}, \kappa\Omega/4]$.

Proof. Since the density of η_t is symmetric, it is easy to show that

$$E\left\{\frac{\partial l_t(\lambda_0)}{\partial \lambda'}\frac{\partial l_t(\lambda_0)}{\partial \lambda}\right\} = \text{diag}\left\{E\left(\frac{y_{t-1}^2}{\omega_0 + \alpha_0 y_{t-1}^2}\right), \frac{\kappa\Omega}{4}\right\},$$

$$\Omega = E\left\{\frac{1}{(\omega_0 + \alpha_0 y_{t-1}^2)^2}\begin{pmatrix} 1 & y_{t-1}^2 \\ y_{t-1}^2 & y_{t-1}^4 \end{pmatrix}\right\}.$$

If Ω is not positive definite, then there is a constant $c = (c_1, c_2)' \neq 0$ such that $c'\Omega c = 0$ and hence we shall have $c_1 + c_2 y_{t-1}^2 = 0$ almost surely. This is impossible. Thus, part (a) holds. Using the martingale central limit theorem and the Cramér–Wold device, it is direct to show that part (b) holds. □

Lemma 5. Under assumptions 1 and 2,

(a) $E\left[\sup_{\lambda\in\Theta}\left\{\frac{\partial^2 l_t(\lambda)}{\partial \lambda' \partial \lambda}\right\}\right] < \infty$ and

(b) $\sup_{\lambda\in\Theta}\left|\frac{1}{n}\sum_{t=2}^{n}\left[\frac{\partial^2 l_t(\lambda)}{\partial \lambda' \partial \lambda} - E\left\{\frac{\partial^2 l_t(\lambda)}{\partial \lambda' \partial \lambda}\right\}\right]\right| = o_p(1)$.

Proof. It is easy to show that

$$\frac{\partial^2 l_t(\lambda)}{\partial \phi^2} = -\frac{y_{t-1}^2}{\omega + \alpha y_{t-1}^2},$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \omega^2} = \frac{1}{2(\omega + \alpha y_{t-1}^2)^2}\left\{1 - \frac{2(y_t - \phi y_{t-1})^2}{\omega + \alpha y_{t-1}^2}\right\},$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \alpha^2} = \frac{y_{t-1}^4}{2(\omega + \alpha y_{t-1}^2)^2}\left\{1 - \frac{2(y_t - \phi y_{t-1})^2}{\omega + \alpha y_{t-1}^2}\right\},$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \phi \partial \omega} = -\frac{y_{t-1}(y_t - \phi y_{t-1})}{(\omega + \alpha y_{t-1}^2)^2},$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \phi \partial \alpha} = -\frac{y_{t-1}^3(y_t - \phi y_{t-1})}{(\omega + \alpha y_{t-1}^2)^2},$$

$$\frac{\partial^2 l_t(\lambda)}{\partial \omega \partial \alpha} = \frac{y_{t-1}^2}{(\omega + \alpha y_{t-1}^2)^2}\left\{1 - \frac{2(y_t - \phi y_{t-1})^2}{\omega + \alpha y_{t-1}^2}\right\}.$$

Now, using a similar method to that for lemma 3, we can show that part (a) holds. Furthermore, by theorem 3.1 in Ling and McAleer (2003), we know that part (b) holds. □

A.1. Proof of theorem 1

(a) First, the space Θ is compact and λ_0 is an interior point in Θ. Second, $L_n(\lambda)$ is continuous in $\lambda \in \Theta$ and is a measurable function of $y_t, t = 1, \ldots, n$, for all $\lambda \in \Theta$. Third, by lemma 1, $n^{-1} L_n(\lambda) \rightarrow_p E\{l_t(\lambda)\}$ uniformly in Θ. Fourth, lemma 2 shows that $E\{l_t(\lambda)\}$ has a unique maximum at λ_0. Thus, we have established all the conditions for consistency in theorem 4.1.1 in Amemiya (1985) and hence part (a) of theorem 1 holds. □

(b) First, by part (a) of theorem 1, the MLE $\hat{\lambda}_n$ of λ_0 is consistent in probability. Second,

$$n^{-1}\sum_{t=2}^{n}\frac{\partial^2 l_t(\lambda)}{\partial\lambda'\partial\lambda}$$

exists and is continuous in Θ. Third, by lemma 5, we can obtain that $n^{-1}\Sigma_{t=2}^{n}\partial^2 l_t(\lambda_n)/\partial\lambda'\partial\lambda$ converges to $\tilde{\Sigma}$ for any sequence λ_n, such that $\lambda_n\to\lambda_0$ in probability, where

$$\tilde{\Sigma}=E\left\{\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda'\partial\lambda}\right\}=-\mathrm{diag}\left\{E\left(\frac{y_{t-1}^2}{\omega_0+\alpha_0 y_{t-1}^2}\right),\frac{\Omega}{2}\right\}.$$

Fourth, by lemma 4, $n^{-1/2}\Sigma_{t=2}^{n}\partial l_t(\lambda_0)/\partial\lambda\to_{\mathcal{L}} N(0,\tilde{\Omega})$. Thus, we have established all the conditions in theorem 4.1.3 in Amemiya (1985) and hence $\sqrt{n}(\hat{\lambda}_n-\lambda_0)\to_{\mathcal{L}} N(0,\tilde{\Sigma}^{-1}\tilde{\Omega}\tilde{\Sigma}^{-1})$, where $\tilde{\Sigma}^{-1}\tilde{\Omega}\Sigma^{-1}=\mathrm{diag}[E^{-1}\{y_t^2/(\omega_0+\alpha_0 y_t^2)\},\kappa\Omega^{-1}]$.

References

Amemiya, T. (1985) *Advanced Econometrics*. Cambridge: Harvard University Press.
Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.
Borkovec, M. (2001) Asymptotic behavior of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH(1). *Bernoulli*, **6**, 847–872.
Borkovec, M. and Kluppelberg, C. (1998) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.*, **11**, 1220–1241.
Bougerol, P. and Picard, N. (1992) Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, **20**, 1714–1730.
Chan, N. H. and Tran, L. T. (1989) On the first-order autoregressive process with infinite variance. *Econometr. Theory*, **5**, 354–362.
Davis, R. A., Knight, K. and Liu, J. (1992) M-estimation for autoregressions with infinite variance. *Stoch. Processes Appl.*, **40**, 145–180.
Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Statist. Ass.*, **74**, 427–431.
Engle, R. F. (1982) Autoregressive conditional heteroskedasticity with estimates of variance of U. K. inflation. *Econometrica*, **50**, 987–1008.
Feigin, P. D. and Tweedie, R. L. (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.*, **6**, 1–14.
Francq, C. and Zakoïan, J. M. (1998) Estimating linear representations of nonlinear processes. *J. Statist. Planng Inf.*, **68**, 145–165.
Francq, C. and Zakoïan, J. M. (2000) Covariance matrix estimation for estimators of mixing weak ARMA models. *J. Statist. Planng Inf.*, **83**, 369–394.
Guégan, D. and Diebolt, J. (1994) Probabilistic properties of the β-ARCH-model. *Statist. Sin.*, **4**, 71–87.
Hall, P. and Heyde, C. C. (1980) *Martingale Limit Theory and Its Applications*. San Diego: Academic Press.
Hamilton, J. D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.
Jeganathan, P. (1988) On the strong approximation of the distributions of estimators in linear stochastic models: I, II, Stationary and explosive AR models. *Ann. Statist.*, **16**, 1283–1314.
Kluppelberg, C., Maller, R. A., Van De Vyver, M. and Wee, D. (2002) Testing for reduction to random walk in autoregressive conditional heteroscedasticity models. *Econometr. J.*, **5**, 387–416.
Koul, H. L. and Pflug, G. (1990) Weakly adaptive estimators in explosive autoregression. *Ann. Statist.*, **18**, 939–960.
Kurtz, T. G. and Protter, P. (1991) Weak limit theorems to stochastic integrals and stochastic differential equations. *Ann. Probab.*, **19**, 1035–1070.
Lee, S.-W. and Hansen, B. E. (1994) Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Econometr. Theory*, **10**, 29–52.
Li, W. K. (1992) On the asymptotic standard errors of residual autocorrelations in nonlinear time series modeling. *Biometrika*, **79**, 435–437.
Li, W. K., Ling, S. and Wong, H. (2001) Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity. *Biometrika*, **88**, 1135–1152.
Li, W. K. and Mak, T. K. (1994) On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *J. Time Ser. Anal.*, **15**, 627–636.
Ling, S. (1999) On probability properties of a double threshold ARMA conditional heteroskedasticity model. *J. Appl. Probab.*, **36**, 688–705.

Ling, S. (2003) A double AR(p) model: structure and estimation. *Working Paper*. Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong.
Ling, S. and Li, W. K. (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroskedasticity. *J. Am. Statist. Ass.*, **92**, 1184–1194.
Ling, S. and Li, W. K. (1998) Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Ann. Statist.*, **26**, 84–125.
Ling, S. and Li, W. K. (2003) Asymptotic inference for unit root processes with GARCH (1, 1) errors. *Econometr. Theory*, **19**, 541–564.
Ling, S. and McAleer, M. (2003) Asymptotic theory for a new vector ARMA-GARCH model. *Econometr. Theory*, **19**, 280–310.
Lumsdaine, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica*, **3**, 575–596.
Mikosch, T., Gadrich, T., Kluppelberg, C. and Adler, R. J. (1995) Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.*, **23**, 305–326.
Nicholls, D. F. and Quinn, B. G. (1982) Random coefficient autoregressive models: an introduction. *Lect. Notes Statist.*, **11**.
Phillips, P. C. B. (1987) Time series regression with a unit root. *Econometrica*, **55**, 277–301.
Phillips, P. C. B. (1989) Partially identified econometric models. *Econometr. Theory*, **5**, 181–240.
Robinson, P. M. (1994) Efficient tests of nonstationary hypotheses. *J. Am. Statist. Ass.*, **89**, 1420–1437.
Tjøstheim, D. (1986) Estimation in nonlinear time series models. *Stoch. Processes Appl.*, **21**, 251–273.
Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Adv. Appl. Probab.*, **22**, 587–611.
Tong, H. (1990) *Nonlinear Time Series: a Dynamical System Approach*. Oxford: Clarendon.
Tsay, R. S. (1987) Conditional heteroscedastic time series models. *J. Am. Statist. Ass.*, **82**, 590–604.
Weiss, A. A. (1986) Asymptotic theory for ARCH models: estimation and testing. *Econometr. Theory*, **2**, 107–131.
Wong, H. and Li, W. K. (1997) On a multivariate conditional heteroscedastic model. *Biometrika*, **84**, 111–123.