INFERENCE IN AUTOREGRESSION UNDER HETEROSKEDASTICITY

BY

PETER C. B. PHILLIPS AND KE-LI XU

COWLES FOUNDATION PAPER NO. 1159

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2006

http://cowles.econ.yale.edu/

INFERENCE IN AUTOREGRESSION UNDER HETEROSKEDASTICITY

BY PETER C. B. PHILLIPS AND KE-LI XU

Yale University

First Version received February 2005

Abstract. A scalar pth-order autoregression (AR(p)) is considered with heteroskedasticity of the unknown form delivered by a transition function of time. A limit theory is developed and three heteroskedasticity-robust test statistics are proposed for inference, one of which is based on the nonparametric estimation of the variance function. The performance of the resulting testing procedures in finite samples is compared in simulations and some suggestions for practical application are given.

Keywords. Autoregression; heterogeneity; nonparametric estimation; robust inference.
JEL classification. C22.
AMS 2000 subject classifications. 62M10.

1. INTRODUCTION

Time-series models, where the nonstationarity is due to time-dependent parameters, have been considered by several authors. Dahlhaus (1997) proposed minimum distance estimation for locally stationary time series under possible model misspecification. Basawa and Lund (2001) investigated maximum likelihood and least squares estimation of autoregressive and moving average (ARMA) models with periodically varying coefficients. Their models were extended recently by Francq and Gautier (2004) to consider ARMA models with nonperiodic time-varying coefficients which are subject to irregular regime changes at known time points. In contrast, the present paper focuses on inference for autoregressive models with time-varying innovation variances, while the autoregressive coefficients are constant over time.

The study is motivated by increasing attention in the literature on the effects of violations of error homoskedasticity in unit-root testing. Typical deviations occur when there is a variance break at an unknown point of time within sample, when there is periodic heteroskedasticity, or when the errors follow some specific model of conditional heterogeneity such as generalized autoregressive conditionally heteroskedastic (GARCH). The size and power properties of standard unit-root tests can be affected moderately, or even significantly, in such cases, depending on the pattern of the variance changes and where they occur in the sample. Kim and Schmidt (1993), Hamori and Tokihisa (1997), Seo (1999), Burridge and Taylor (2001), Nelson *et al.* (2001), Cavaliere (2003), Ling *et al.* (2003) and Boswijk

0143-9782/06/02 289–308 JOURNAL OF TIME SERIES ANALYSIS Vol. 27, No. 2

(2005) all studied various forms of departures from homoskedasticity and a general framework of analysis is provided in Cavaliere (2004). Recently, studies have been undertaken to develop modified unit-root tests which are robust to the presence of unknown variance changes (Kim *et al.*, 2002; Cavaliere and Taylor, 2004; Beare, 2004). The tests considered in this study are asymptotically valid under certain types of heteroskedastic errors and are found to perform better in small samples than in standard uncorrected unit-root tests.

In contrast to the studies dealing with unit-root regressions, there has been little work of this type on stable autoregressions. One obvious explanation is that standard regression procedures can be used to make the inference robust in such situations by using heteroskedasticity-consistent (HC) covariance matrix estimates, as explored in early studies by Eicker (1963) and White (1980). However, there may be advantages in considering alternative, more extensive methods. Accordingly, this paper develops some new procedures for making inference robust to general forms of heteroskedasticity in the context of a finite-order stable autoregression.

Our framework is a pth-order autoregression (AR(p)), the errors of which are martingale differences with conditional variances that involve an unknown nonstochastic time-varying function. This model accommodates a wide range of heterogeneous error models, allowing for discrete shifts in the variance function. We focus on unconditional heteroskedasticity and the kind of specification used here is routinely employed in the nonparametric curve-fitting literature and in parametric models with time-dependent coefficients (see, e.g. Robinson, 1989; Dahlhaus, 1997; Cavaliere, 2004). Linear regression with nonstationary variances has also been discussed by Hansen (1995). In his model, the conditional variance of the error is a nonlinear, continuous function of a nearly integrated process, with a suitable normalization that facilitates an asymptotic development using functional limits and continuous mappings. The present paper allows the variance function to be continuous or discontinuous but is assumed to be deterministic.

The layout of the remainder of the paper is as follows. Section 2 introduces the model and assumptions. Sections 3 and 4 develop the limit theory and heteroskedasticity-robust inference for the autoregressive coefficients respectively. Section 5 reports some simulation results and Section 6 concludes. Mathematical proofs are presented in the Appendix.

2. MODEL AND ASSUMPTIONS

We consider the AR(p)

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + \varepsilon_t, \tag{1}$$

where $\theta_p \neq 0$ and the lag order p is finite and known. The innovations are heterogeneously distributed according to

$$\varepsilon_t = g\left(\frac{t}{T}\right)u_t, \tag{2}$$

where $g(\cdot)$ is an unknown non-negative scale function. Suppose the data generating process (1) is doubly infinite, and we observe a sample containing $T + p$ observations, denoted by $\{Y_{-p+1}, Y_{-p+2}, \ldots, Y_0, Y_1, \ldots, Y_T\}$. $\theta = (\theta_0, \theta_1, \ldots, \theta_p)'$ is the parameter vector of interest. In view of eqn (2), $\{\varepsilon_t\}$ and $\{Y_t\}$ constitute triangular arrays and an additional subscript might accordingly be added, but the array notation is unnecessary and can be neglected in what follows. Assumption 1 holds throughout the paper.

Assumption 1.

(*i*) *All roots of the polynomial* $1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_p z^p = 0$ *lie outside the unit circle.*

(ii) $g(\cdot)$ *is non-stochastic, measurable and uniformly bounded on the interval* $(-\infty, 1]$, *with a finite number of points of discontinuity,* $g(\cdot) > 0$ *and satisfies a Lipschitz condition except at points of discontinuity.*

(iii) $\{u_t\}$ *is a strong mixing* (α*-mixing*) *martingale difference process with* $E(u_t|\mathcal{F}_{t-1}) = 0$, $E(u_t^2|\mathcal{F}_{t-1}) = 1$, *a.s., for all* t, *with the natural filtration* $\mathcal{F}_t = \sigma(u_s, s \leq t)$. *There exist* $\delta > 1$ *and* $C > 0$, *such that* $\sup_t Eu_t^{4\delta} < C < \infty$.

Remarks 1. Under Assumption 1(i) the autoregressive coefficients are assumed to satisfy the usual stability conditions which, if g were a constant function and ε_t homoskedastic, would ensure that Y_t is stationary or asymptotically covariance-stationary, depending on initial conditions. In this case, the mean of Y_t exists and is given by

$$\mu = \frac{\theta_0}{1 - \theta_1 - \theta_2 - \cdots - \theta_p}.$$

When $p = 1$ and $\theta_0 = 0$, Assumption 1(i) requires $|\theta_1| < 1$. In the unit-root case where $\theta_1 = 1$, testing issues have been studied widely in the literature (see Phillips and Xiao, 1998, for a review) and recent studies by Cavaliere (2004), Cavaliere and Taylor (2004) and Beare (2004) consider unit-root testing under heterogeneous errors of the form (2).

Under Assumption 1(i), Y_t has Wold representation

$$Y_t = \mu + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i}, \tag{3}$$

where the coefficients $\{\alpha_i\}$ satisfy the recursion

$$\alpha_i - \theta_1 \alpha_{i-1} - \cdots - \theta_p \alpha_{i-p} = 0 \quad \text{for } i > 0,$$

$$\alpha_0 = 1, \quad \alpha_i = 0 \text{ for } i < 0,$$

and

$$\sum_{i=0}^{\infty} |\alpha_i| < \infty. \tag{4}$$

Define Ω to be the $p \times p$ matrix with (i, j)-th element $\gamma_{|i-j|}$, where

$$\gamma_k = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} < \infty, \tag{5}$$

for $k = 0, 1, \ldots, p-1$, which is implied by eqn (4) as

$$\left| \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} \right| \leq \left(\sum_{i=0}^{\infty} |\alpha_i|^2 \right).$$

In the case of homoskedastic errors [when $g(\cdot) = \sigma$ is a constant over t], $\sigma^2 \gamma_k (k = 0, 1, \ldots)$ is the autocovariance sequence of the covariance-stationary process $\{Y_t\}$.

Both γ_k and Ω can be expressed in terms of $\theta_1, \ldots, \theta_p$, more explicitly $\gamma_0, \gamma_1, \ldots, \gamma_{p-1}$ are the first p elements in the first column of the $(p^2 \times p^2)$ matrix $[I_{p^2} - F \otimes F]^{-1}$, where '$\otimes$' denotes the Kronecker product and

$$F = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \\ & & & 0 \\ & I_{p-1} & & \vdots \\ & & & 0 \end{pmatrix}.$$

For the zero-mean AR(1) process with homoskedastic errors and $g(\cdot) = \sigma$, we have $\Omega = \gamma_0 = \left(1 - \theta_1^2\right)^{-1}$.

Assumption 1(ii) is the same as Assumption $\mathcal{V}$ of Cavaliere (2004). We require the definition of the function $g(r)$ for $r < 0$ since initial conditions are in the infinite past and the MA(∞) representation of Y_t is used in eqn (3). Note that under Assumption 1(ii) the function g is integrable on the interval [0,1] up to any finite order. For brevity, we write $\int_0^1 g^m(r)\mathrm{d}r$ as $\int g^m$ for any finite-positive integer m.

Under Assumption 1(iii) $Eu_s u_t = 0$ for $s \neq t$. We impose no restriction on the mixing decay rate of $\{u_t\}$, and by Assumption 1(iii) and Lyapunov's inequality, $E(|u_t|^\eta)$ exists for all $\eta \leq 4\delta$, as do all expectations involving up to any four combinations of u_t. We set $E(u_t^2 | \mathcal{F}_{t-1}) = 1$ for normalization, so that the function g is identified. Then,

$$E(\varepsilon_t^2) = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = g^2\left(\frac{t}{T}\right),$$

and model heteroskedasticity is characterized in eqn (2) as being systematically dependent on the relative position of the observation in terms of the scale function g. This includes cases where the conditional error variance evolves over time, slowly transitions, or abruptly or periodically changes across the sample. Cavaliere (2004) provides further discussion of the variance patterns that are allowed. Since g is taken to be unknown, the formulation is nonparametric.

3. LIMIT THEORY

Write (1) in regression form as

$$Y_t = X'_{t-1}\theta + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $X_{t-1} = (1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$, and applying ordinary least squares (OLS) estimation, the scaled error is

$$\sqrt{T}(\hat{\theta} - \theta) = \left(\frac{1}{T}\sum_{t=1}^{T} X'_{t-1}X_{t-1}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}\varepsilon_t\right). \tag{6}$$

Lemma 1 contains some preliminary results, which lead to the limit theory for $\hat{\theta}$.

Lemma 1. *Let ℓ_p be the column p-vector $(1, 1, \ldots, 1)'$. Under the stated assumptions as $T \to \infty$,*

(i) $\frac{1}{T}\sum_{t=1}^{T} X_{t-1}\varepsilon_t \xrightarrow{p} 0$;

(ii) $\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X'_{t-1} \xrightarrow{p} \Omega_1$;

(iii) $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 X_{t-1}X'_{t-1} \xrightarrow{p} \Omega_2$;

(iv) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}\varepsilon_t \xrightarrow{d} \mathcal{N}(0, \Omega_2)$,

where the $(p+1) \times (p+1)$ matrix Ω_1 and Ω_2 are defined as

$$\Omega_1 = \begin{pmatrix} 1 & \mu\ell'_p \\ \mu\ell_p & \mu^2 + (\int g^2)\Omega \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} \int g^2 & \mu(\int g^2)\ell'_p \\ \mu(\int g^2)\ell_p & \mu^2(\int g^2) + (\int g^4)\Omega \end{pmatrix}.$$

The proof of Lemma 1 is given in the Appendix. In this proof, the conditional homoskedasticity assumption, $E(u_t^2|\mathcal{F}_{t-1}) = 1$, a.s., is used to obtain the limit of $T^{-1}\sum_{t=1}^{T} E(\varepsilon_t^2 X_{t-1}X'_{t-1})$, which involves the quarticity function $\int g^4$ of g. By contrast, the probability limit of $T^{-1}\sum_{t=1}^{T} X_{t-1}X'_{t-1}$ involves only the integrated volatility function $\int g^2$. The following result follows directly from Lemma 1.

Theorem 1. *Let Ω_1 and Ω_2 as defined in Lemma 1. Under the stated assumptions, $\hat{\theta}$ is consistent and asymptotically normally distributed, satisfying*

$$\sqrt{T}(\hat{\theta} - \theta) \to_d \mathcal{N}(0, \Lambda), \tag{7}$$

as $T \to \infty$, *where* $\Lambda = \Omega_1^{-1}\Omega_2\Omega_1^{-1}$.

REMARKS. (1) Theorem 1 shows that the asymptotic covariance matrix of $\hat{\theta}$ has the traditional sandwich form involving the integration of the second- and fourth-order powers of the variance function, which is very simple and convenient to use. We observe that the variance matrix depends on both integrated volatility and quarticity through Ω_1 and Ω_2. Integrated volatility and quarticity measures have recently been popularized in recent studies on the realized volatility of financial asset returns (see, e.g. Barndorff-Nielsen and Shephard, 2002, 2004; Andersen *et al.*, 2003).

(2) Two special cases are easily derived from Theorem 1. When $g(r) = \sigma$ is constant, eqn (7) reduces to the standard homoskedastic limit theory,

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \begin{pmatrix} 1 & \mu\ell_p' \\ \mu\ell_p & \mu^2 + \sigma^2\Omega \end{pmatrix}^{-1}\right).$$

When $\theta_0 = 0$ and $\{Y_t\}$ is a zero-mean, AR(p) process with heteroskedastic errors in eqns (2) and (7) leads to

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{\int g^4}{(\int g^2)^2}\Omega^{-1}\right). \tag{8}$$

(3) Hansen (1995) studied autoregressive models with nonstationary volatility. He assumed the conditional variances to be a function of a nearly integrated time series of

$$E(\varepsilon_t^2|\mathcal{F}_{t-1}) = g(T^{-1/2}S_t^0), \tag{9}$$

where $S_t^0 = \rho S_{t-1}^0 + z_t$ with $\rho = 1 - c/T$, c a constant, and z_t a conditionally homoskedastic martingale difference sequence. Under mild assumptions, the standardized and centred OLS estimator then converges weakly to a variance mixture-normal limit distribution. If S_t were nearly integrated with drift and of the form $S_t = \alpha t + S_t^0$, say, then extending this approach would lead to the following conditional variance function

$$E(\varepsilon_t^2|\mathcal{F}_{t-1}) = g(T^{-1}(\alpha t + S_t^0)), \tag{10}$$

with a natural normalization factor of T^{-1} in place of $T^{-1/2}$. For such extensions, Hansen's (1995) model is asymptotically equivalent to eqn (8).

(4) Goncalves and Kilian (2004) investigated the AR(p) model assuming unconditional homoskedasticity of the errors, allowing for some degree of conditional heteroskedasticity. These authors derived the limit distribution of the OLS estimator, which involves fourth-order cumulants of the errors (see also Kuersteiner, 2001).

4. INFERENCE

In the asymptotic variance formula (7), μ and Ω can be consistently estimated by $\hat{\mu}$ and $\hat{\Omega}$, formed by replacing θ with $\hat{\theta}$ in the expression for μ and Ω [as in Remark (1) of Assumption 1]. If g were known, then heteroskedasticity-robust t-ratios and Wald tests could be constructed using eqn (7) directly. Since g is typically unknown, the asymptotic variance in eqn (7) must be estimated and this can be done in several ways. First, the Eicker–White correction for heteroskedasticity can be used. We also modify the Eicker–White estimator by first estimating the integrated volatility factor $\int g^2$ and using this estimate in the expression for Ω_1 to get a consistent estimator of Ω_1. As a third method, we can estimate the function g^2 nonparametrically by weighted sum of squared OLS residuals using kernel smoothing, i.e. for $r \in [0, 1]$,

$$\hat{g}^2(r) = \sum_{t=1}^{T} w_{rt}\hat{\varepsilon}_t^2, \tag{11}$$

where $\hat{\varepsilon}_t = Y_t - \hat{\theta}' X_{t-1}$ is the OLS residual and the weights w_{rt}, $t = 1, \ldots, T$, are defined as

$$w_{rt} = \left(\sum_{t=1}^{T} K\left(\frac{[Tr] - t}{Tb} \right) \right)^{-1} K\left(\frac{[Tr] - t}{Tb} \right).$$

The kernel function $K(\cdot) : \mathbb{R}^1 \to [0, \infty)$ is assumed to satisfy $0 \leq K(z) < C_1 < \infty$ uniformly in z and

$$\int_{-\infty}^{\infty} K(z)\mathrm{d}z < C_2 < \infty$$

for some constants C_1 and C_2; and b is a bandwidth parameter, depending on T, such that $b + (1/Tb) \to 0$, as $T \to \infty$. The estimator (11) was originally proposed by Nadaraya (1964) and Watson (1964) for estimation of regression functions, and it was used by Robinson (1989) for estimation of time-varying regression models. Lemma 2 shows that eqn (11) estimates $g^2(r)$ consistently when $r \in [0, 1]$ is a point of continuity of the function g.

LEMMA 2. *Let Ω_2 as defined in Lemma 1. Under the stated assumptions, as $T \to \infty$*

(i) $\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 \xrightarrow{p} \int g^2$;

(ii) $\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^4 \xrightarrow{p} \omega_4 \int g^4$, *where* $\omega_4 = E(u_t^4)$;

(iii) $\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 X_{t-1}X_{t-1}' \xrightarrow{p} \Omega_2$;

(iv) $\hat{g}(r) \xrightarrow{p} g(r)$, *for all* $r \in [0, 1]$ *for which the function* g *is continuous.*

REMARKS. (1) Lemma 2(i) shows that $\int g^2$ is consistently estimated by $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$, and $\int g^4$ by $\frac{1}{\omega_4}T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^4$, which is $(3T)^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^4$ in the Gaussian case. In practice, of course, the distribution of the error, and therefore its fourth-moment ω_4, is unknown, so this latter estimate of $\int g^4$ is generally infeasible. Nonparametric estimation of the variance function g then offers a more general solution. After this study was completed, we found that estimation of the variance function has been investigated independently by Dahl and Levine (2005) in a similar model but with serially dependent errors and $\theta = 0$.

(2) In practice, when estimating g, the bandwidth parameter b can be chosen using cross-validation on the average squared error (see Wong, 1983). Define

$$\hat{g}_{-s}^2(r) = \sum_{t=1,t\neq s}^{T} w_{rt}\hat{\varepsilon}_t^2.$$

The cross-validation choice of b is the value b^* which minimizes

$$\hat{\mathrm{CV}}(b) = \frac{1}{T}\sum_{s=1}^{T}\left(\hat{\varepsilon}_s^2 - \hat{g}_{-s}^2\left(\frac{s}{T}\right)\right)^2.$$

(3) The limits in Lemma 2 suggest the following three consistent estimates of the asymptotic variance matrix Λ when g is unknown:

$$\hat{\Lambda}_1 = T\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t^2 X_{t-1}X_{t-1}'\right)\left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}, \tag{12}$$

which is the usual Eicker–White heteroskedasticity robust covariance matrix estimator,

$$\hat{\Lambda}_2 = \hat{\Omega}_1^{-1}\left(T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 X_{t-1}X_{t-1}'\right)\hat{\Omega}_1^{-1}, \tag{13}$$

where the $(p+1)\times(p+1)$ matrix $\hat{\Omega}_1$ is defined as

$$\hat{\Omega}_1 = \begin{pmatrix} 1 & \hat{\mu}\ell_p' \\ \hat{\mu}\ell_p & \hat{\mu}^2 + \left(T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)\hat{\Omega} \end{pmatrix},$$

and

$$\hat{\Lambda}_3 = \hat{\Omega}_1^{-1}\widetilde{\Omega}_2\hat{\Omega}_1^{-1}, \tag{14}$$

where

$$\widetilde{\Omega}_2 = \begin{pmatrix} \int \hat{g}^2 & \hat{\mu}(\int \hat{g}^2)\ell_p' \\ \hat{\mu}(\int \hat{g}^2)\ell_p & \hat{\mu}^2(\int \hat{g}^2) + (\int \hat{g}^4)\hat{\Omega} \end{pmatrix}.$$

The following result is an immediate consequence of Theorem 1 and Lemma 2.

THEOREM 2. *For* Λ *and* $\hat{\Lambda}_j (j = 1, 2, 3)$ *defined in eqns* (12), (13) *and* (14), *and under the stated assumptions as* $T \to \infty$*:*

(i) $\hat{\Lambda}_j \xrightarrow{p} \Lambda$.

(ii) *Under* $H_0 : \theta_k = \bar{\theta}_k (k = 0, 1, \ldots, p)$,

$$t_j = \frac{\sqrt{T}(\hat{\theta}_k - \bar{\theta}_k)}{[(\hat{\Lambda}_j)_{kk}]^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $(\hat{\Lambda}_j)_{kk}$ *is the* (k, k)*-th element of* $\hat{\Lambda}_j$.

(iii) *Under* $H_0 : a(\theta) = 0$, *where* $a(\theta)$ *is an s-vector of continuously differentiable restrictions for which the* $s \times (p + 1)$ *matrix* $A(\theta)$ *of first derivatives is of full row rank*,

$$W_j = T \cdot a(\hat{\theta})'[A(\hat{\theta})\hat{\Lambda}_j A(\hat{\theta})']^{-1} a(\hat{\theta}) \xrightarrow{d} \chi^2_s,$$

where χ^2_s *is chi-square distribution with s degrees of freedom.*

In the zero-mean AR(1) case, we have $\theta = \theta_1$ and $\hat{\Omega} = \hat{\gamma}_0 = 1/(1 - \hat{\theta}^2)$. So, the three statistics in Theorem 2(ii) for testing $H_0 : \hat{\theta} = \bar{\theta}$ against $H_1 : \hat{\theta} \neq \bar{\theta}$ can be written in simpler forms and their LM (Lagrange multiplier) alternatives (obtained by substituting $\hat{\theta}$ with $\bar{\theta}$ in $\hat{\Omega}$) are:

$$t_1 = \frac{(\hat{\theta} - \bar{\theta})\left(\sum_{t=1}^T Y_{t-1}^2\right)}{\left(\sum_{t=1}^T Y_{t-1}^2 \hat{\varepsilon}_t^2\right)^{1/2}}, \quad t_2 = \frac{(\hat{\theta} - \bar{\theta})\hat{\varepsilon}'\hat{\varepsilon}}{(1 - \bar{\theta}^2)\left(\sum_{t=1}^T Y_{t-1}^2 \hat{\varepsilon}_t^2\right)^{1/2}}, \quad t_3 = \frac{(\hat{\theta} - \bar{\theta})\hat{\varepsilon}'\hat{\varepsilon}}{\left(T(1 - \bar{\theta}^2) \int \hat{g}^4\right)^{1/2}}, \tag{15}$$

which are asymptotically distributed as standard normal.

5. SIMULATIONS

This section reports a brief simulation experiment comparing the finite-sample performance of the three asymptotically correct tests (15), along with the usual OLS t-test without heteroskedasticity correction, in the zero-mean AR(1) case:

$$Y_t = \theta Y_{t-1} + g\left(\frac{t}{T}\right) u_t,$$

where $u_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. For the semiparametric test t_3 the kernel function K is the Gaussian density

$$K(z) = (2\pi)^{-1/2} \exp(-z^2/2) \quad \text{for } -\infty < z < \infty,$$

and the bandwidth parameter b is selected by the cross-validation criterion described in Section 4.

The designs used are the same as in Cavaliere (2004) and Cavaliere and Taylor (2004). Two models for the variance function $g(\cdot)$ are used:

$$1.\ g(r)^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbf{1}_{\{r \geq \tau\}}, \quad r \in [0, 1]$$
$$2.\ g(r)^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)r^m, \quad r \in [0, 1]$$

Model 1 corresponds to the case of a single abrupt change of error variance from σ_0^2 to σ_1^2 at time $[\tau T]$, and the steepness of the break is measured by the ratio of post-break and pre-break standard deviations $\delta := \sigma_1/\sigma_0$. The break date τ takes values within the set $\{0.1, 0.5, 0.9\}$ and δ within $\{0.2, 5\}$, so that both early and late breaks, and positive ($\delta > 1$) and negative ($\delta < 1$) shifts are considered. We also check the performance of four tests when the innovations are homoskedastic, i.e. when $\delta = 1$. In model 2, the variance of the errors changes continuously from σ_0^2 to σ_1^2. We let $m \in \{1, 2\}$ and δ takes values on $\{0.2, 5\}$, so that positive ($\delta > 1$) and negative ($\delta < 1$) trending variances are allowed.

Without loss of generality we set $\sigma_0^2 = 1$ in all cases. We consider the actual size of the standard OLS t-test without heteroskedasticity correction (labelled 'OLS'), and the three tests in eqn (15) when the true value of θ is taken from the set $\{0.1, 0.5, 0.9\}$ and the nominal size is 5%. The sample sizes are set to $T = 60$, 200 and the number of replications is 10,000. Results are reported in Tables I and II.

TABLE I

ACTUAL SIZE OF THREE TESTS IN EQN (15) UNDER MODEL 1 FOR $\theta \in \{0.1, 0.5, 0.9\}$, $\tau \in \{0.1, 0.5, 0.9\}$, $\delta \in \{0.2, 5\}$ AND THE SAMPLE SIZE $T = \{60, 200\}$, BASED ON 10,000 REPLICATIONS

			$T = 60$				$T = 200$			
θ	τ	δ	OLS	t_1	t_2	t_3	OLS	t_1	t_2	t_3
0.1	0.1	0.2	0.340	0.145	0.061	0.103	0.344	0.094	0.058	0.054
		1	0.049	0.082	0.062	0.043	0.051	0.067	0.061	0.053
		5	0.059	0.076	0.059	0.047	0.066	0.061	0.063	0.052
	0.5	0.2	0.148	0.090	0.047	0.048	0.157	0.057	0.044	0.045
		5	0.148	0.094	0.071	0.072	0.130	0.057	0.052	0.051
	0.9	0.2	0.070	0.084	0.057	0.039	0.059	0.060	0.055	0.048
		5	0.334	0.149	0.158	0.141	0.395	0.096	0.090	0.082
0.5	0.1	0.2	0.321	0.124	0.116	0.095	0.375	0.095	0.086	0.063
		1	0.053	0.072	0.084	0.058	0.048	0.052	0.052	0.047
		5	0.054	0.070	0.088	0.057	0.059	0.055	0.059	0.049
	0.5	0.2	0.138	0.085	0.083	0.048	0.163	0.065	0.062	0.064
		5	0.144	0.080	0.092	0.065	0.136	0.058	0.059	0.045
	0.9	0.2	0.069	0.083	0.092	0.060	0.051	0.047	0.054	0.044
		5	0.324	0.147	0.174	0.135	0.408	0.088	0.109	0.091
0.9	0.1	0.2	0.196	0.082	0.258	0.135	0.309	0.064	0.185	0.076
		1	0.052	0.063	0.194	0.116	0.040	0.045	0.102	0.056
		5	0.063	0.070	0.219	0.124	0.045	0.045	0.117	0.071
	0.5	0.2	0.126	0.082	0.222	0.123	0.156	0.069	0.158	0.104
		5	0.147	0.090	0.278	0.182	0.135	0.058	0.140	0.089
	0.9	0.2	0.066	0.072	0.191	0.117	0.064	0.063	0.110	0.079
		5	0.269	0.109	0.402	0.253	0.332	0.060	0.225	0.101

TABLE II

Actual Size of Three Tests in eqn (15) under Model 2 for $\theta \in \{0.1, 0.5, 0.9\}$, $m \in \{1, 2\}$, $\delta \in \{0.2, 5\}$ and the Sample Size $T = \{60, 200\}$, Based on 10,000 Replications

			$T = 60$				$T = 200$			
θ	m	δ	OLS	t_1	t_2	t_3	OLS	t_1	t_2	t_3
0.1	1	0.2	0.099	0.090	0.057	0.047	0.080	0.059	0.051	0.049
		5	0.085	0.090	0.077	0.067	0.080	0.066	0.068	0.058
	2	0.2	0.073	0.080	0.056	0.044	0.080	0.060	0.054	0.054
		5	0.110	0.075	0.067	0.069	0.122	0.059	0.055	0.057
0.5	1	0.2	0.078	0.079	0.073	0.049	0.070	0.055	0.059	0.054
		5	0.070	0.063	0.085	0.066	0.081	0.058	0.067	0.055
	2	0.2	0.081	0.078	0.083	0.057	0.063	0.052	0.056	0.046
		5	0.096	0.079	0.097	0.077	0.111	0.063	0.065	0.062
0.9	1	0.2	0.071	0.075	0.191	0.107	0.094	0.063	0.137	0.086
		5	0.075	0.070	0.204	0.131	0.077	0.051	0.118	0.083
	2	0.2	0.066	0.073	0.187	0.105	0.078	0.062	0.113	0.078
		5	0.105	0.070	0.220	0.139	0.113	0.050	0.150	0.099

Some interesting phenomena emerge. All four tests tend to overreject the null hypothesis when the null is true, but the degree of overrejection depends on the pattern of the variance dynamics. For model 1, all tests have larger distorted size under the early negative shift or late positive shift, and smaller size distortion under the early positive shift or late negative shift. For instance, take the case where the true value of the autoregressive parameter θ is 0.1, and there is an abrupt shift near the beginning of the sample, i.e. $\tau = 0.1$. Here, when the shift is negative ($\delta = 0.2$), and the sample size is 60, OLS overrejects overwhelmingly, which is not surprising, with actual size 34% compared with a nominal size of 5%. The Eicker–White test t_1 has size 14.5%. The tests t_2 and t_3 have less size distortion, with actual sizes 6.1% and 10.3% respectively. When the sample size increases from 60 to 200, the size distortions of the three asymptotically valid tests t_1, t_2, and t_3 decrease, while this is not the case for OLS. When the early shift is positive ($\delta = 5$), all four tests considered work reasonably well in both cases $T = 60$ and 200. When the error variance dynamics follow a polynomial shape as in model 2, the size distortions of the four tests are less general than those in model 1.

When the true value of θ is small ($\theta = 0.1$, 0.5), the semiparametric test t_3 performs relatively well for all patterns of the variance dynamics, and it has the smallest size distortion overall. The cross-validation criterion also seems to work well in choosing the bandwidth parameter. Cases worth special attention occur when the late variance shift is positive in model 1, e.g. $\tau = 0.9$ and $\theta = 5$, all tests considered have severe size distortions, with over 14% actual size when the nominal size is 5% and the sample size $T = 60$. When the true value of θ is close to unity ($\theta = 0.9$), the tests t_2 and t_3 work poorly, and the traditional Eicker–White robust test t_1 seems to be the better choice.

In summary, the simulation results generally corroborate the limit theory. The OLS-based test is unreliable when the innovations are heteroskedastic and the traditional heteroskedasticity-robust test may still be the preferred general procedure for testing, especially when there is a near unit root in the system. The approach works well for a large range of true values of the autoregressive parameters. The explicit form approach produces a competitive alternative method for some subsets of the parameter space, but has the disadvantage of having to specify the full parameter vector when LM versions are used. In some cases, it appears that none of the tests is sufficiently reliable for practical work.

6. CONCLUSION

This paper develops a limit theory for least squares estimates of stable autoregressions in the presence of unconditional heteroskedasticity of unknown form and considers several methods of robustifying inference. Some extensions are possible. First, there is some scope for improving finite-sample properties by wild bootstrap techniques that allow for heteroskedastic errors. Second, efficient estimates may be obtained by feasible weighted least squares methods that employ nonparametric estimates of the residual variances. The second extension is considered in Xu and Phillips (2005) and others will be explored in later studies.

APPENDIX

PROOFS

In what follows, C is a generic positive constant. We use $\|\cdot\|$ to denote the Euclidean norm $\|X\| = (X_1^2 + \cdots + X_n^2)^{1/2}$, for $X = (X_1, \ldots, X_n)'$, and $\|\cdot\|_K$ to denote the L^K-norm, so that $\|\xi\|_K = (E\|\xi\|^K)^{1/K}$, for a random vector ξ.

LEMMA A. *Under the stated assumptions,*

(i) $$\sup_{1 \le t \le T} E(Y_t - \mu)^{4\delta} < \infty, \quad \textit{for } \delta > 1$$

specified in Assumption 1(iii);

(ii) *Let* $1 \le h \le p$, $0 \le k \le p - h$. *Then*

$$\lim_{T\to\infty} E(Y_{[Tr]-h} - \mu)(Y_{[Tr]-h-k} - \mu) = g^2(r)\gamma_k,$$

for values $r \in (0, 1]$ *at which the function* g *is continuous, where* $[\cdot]$ *refers to the integer part and* γ_k *is defined in eqn* (5).

PROOF.

(i) Note that

$$E(Y_t-\mu)^{4\delta} \overset{\text{def}}{=} \|Y_t-\mu\|_{4\delta}^{4\delta} \overset{\text{Minkowski}}{\leq} \left(\sum_{i=0}^{\infty}|\alpha_i|\|\varepsilon_{t-i}\|_{4\delta}\right)^{4\delta} \leq C\left(\sum_{i=0}^{\infty}|\alpha_i|\right)^{4\delta} < \infty$$

uniformly in $1 \leq t \leq T$ by eqn (4). So (i) holds.

(ii) By eqn (3),

$$E(Y_{t-h}-\mu)(Y_{t-h-k}-\mu) = \sum_{i=0}^{\infty}\alpha_i\alpha_{i+k}g^2\left(\frac{t-h-k-i}{T}\right).$$

Let $t=[Tr]$, where $r \in (0, 1]$, so that

$$\begin{aligned} &E(Y_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu)\\ &\quad=\sum_{i=0}^{\infty}\alpha_i\alpha_{i+k}g^2\left(\frac{[Tr]-h-k-i}{T}\right)\\ &\quad=\sum_{i=0}^{L}\alpha_i\alpha_{i+k}g^2\left(\frac{[Tr]-h-k-i}{T}\right)+\sum_{i=L+1}^{\infty}\alpha_i\alpha_{i+k}g^2\left(\frac{[Tr]-h-k-i}{T}\right), \end{aligned} \tag{16}$$

where $L > 0$ is chosen so that

$$\frac{L}{T}+\frac{1}{L}\to 0.$$

As $T\to\infty$, the first term in eqn (16) satisfies

$$\sum_{i=0}^{L}\alpha_i\alpha_{i+k}g^2\left(\frac{[Tr]-h-k-i}{T}\right)\to\sum_{i=0}^{\infty}\alpha_i\alpha_{i+k}g^2(r)=g^2(r)\gamma_k,$$

since

$$\frac{[Tr]-h-k+1-i}{T}\to r$$

uniformly in i, for $i \leq L$ and fixed h and k. The modulus of the second term in eqn (16) is bounded by

$$C\sum_{i=L+1}^{\infty}|\alpha_i\alpha_{i+k}| = C\left(\sum_{i=0}^{\infty}|\alpha_i\alpha_{i+k}|-\sum_{i=0}^{L}|\alpha_i\alpha_{i+k}|\right)\to 0,$$

as $T\to\infty$. It follows that

$$\lim_{T\to\infty}E(Y_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu) = g^2(r)\gamma_k,$$

as required. □

PROOF OF LEMMA 1.

(i) It suffices to show that $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \xrightarrow{p} 0$ *and* $\frac{1}{n}\sum_{t=1}^{T}Y_{t-h}\varepsilon_t \xrightarrow{p} 0$, *for* $1 \leq h \leq p$. Note that

$$E(\varepsilon_t|\mathcal{F}_{t-1})=0 \quad\text{and}\quad E(Y_{t-h}\varepsilon_t|\mathcal{F}_{t-1})=Y_{t-h}E(\varepsilon_t|\mathcal{F}_{t-1})=0.$$

By the stated assumptions and Lemma A(i), we have

$$E\varepsilon_t^2 < \infty \quad \text{and} \quad EY_{t-h}^2\varepsilon_t^2 \le \sqrt{EY_{t-h}^4 E\varepsilon_t^4} \le \sqrt{\sup_t EY_{t-h}^4 \cdot \sup_t E\varepsilon_t^4} < \infty.$$

Then (i) follows directly from the law of large numbers for martingale differences.

(ii) It is sufficient to prove (a)

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-h} \xrightarrow{p} \mu$$

and (b)

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-h}Y_{t-h-k} \xrightarrow{p} \mu^2 + \left(\int g^2\right)\gamma_k, \quad \text{for } 1 \le h \le p, 0 \le k \le p-h.$$

To prove (a), for $m \ge p$ such that $\alpha_{m-p+1} \ne 0$, we write

$$Y_{t-p} - \mu = \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-p-i} = \sum_{i=0}^{m-p} \alpha_i \varepsilon_{t-p-i} + \alpha_{m-p+1}(Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}),$$

where $\varphi_2, \ldots, \varphi_p$ are finite constants. Let $\mathcal{F}_{t-m}^{t+m}$ be the sigma-field generated by $\{u_{t-m}, \ldots, u_{t+m}\}$, then

$$E(Y_{t-p} - \mu|\mathcal{F}_{t-m}^{t+m}) = \sum_{i=0}^{m-p} \alpha_i \varepsilon_{t-p-i} + \alpha_{m-p+1} E(Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}|\mathcal{F}_{t-m}^{t+m}).$$

Hence

$$\begin{aligned}
&\|Y_{t-p} - \mu - E(Y_{t-p} - \mu|\mathcal{F}_{t-m}^{t+m})\|_2 \\
&\quad = |\alpha_{m-p+1}| \cdot \|Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p} \\
&\qquad - E(Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}|\mathcal{F}_{t-m}^{t+m})\|_2 \\
&\quad \le |\alpha_{m-p+1}| \cdot (\|Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}\|_2 \\
&\qquad + \|E(Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}|\mathcal{F}_{t-m}^{t+m})\|_2) \\
&\quad \le 2|\alpha_{m-p+1}| \cdot \|Y_{t-m-1} + \varphi_2 Y_{t-m-2} + \cdots + \varphi_p Y_{t-m-p}\|_2 \\
&\quad \le C|\alpha_{m-p+1}| \cdot \sup_t \|Y_t\|_2,
\end{aligned}$$

where the first inequality is by Minkowski, the second by the conditional Jensen inequality and the law of iterated expectation. Since $|\alpha_{m-p+1}| \to 0$ as $m \to \infty$ by eqn (4) and $\sup_t\|Y_t\|_2 < \infty$ by Lemma A(i), $\{Y_{t-p} - \mu\}$ is mean-zero near-epoch dependent in L^2-norm (L^2-NED) on the α-mixing sequence $\{u_t\}$. By Theorem 17.10 in Davidson (1994), for $1 \le h \le p$, $\{Y_{t-h} - \mu\}$ is also mean-zero L^2-NED on $\{u_t\}$, and necessarily an L^1-mixingale. It is uniformly integrable by Lemma A(i), and then (a) follows from the law of large numbers (e.g. Andrews, 1988, Theorem 1).

To prove (b), it suffices to show

$$\frac{1}{T}\sum_{t=1}^{T}(Y_{t-h} - \mu)(Y_{t-h-k} - \mu) \xrightarrow{p} \left(\int g^2\right)\gamma_k, \quad \text{for } 1 \le h \le p,\ \ 0 \le k \le p-h,$$

in view of (a). Since both $\{Y_{t-h}-\mu\}$ and $\{Y_{t-h-k}-\mu\}$ are L^2-NED on $\{u_t\}$, so by Theorem 17.9 in Davidson (1994), we have $(Y_{t-h}-\mu)(Y_{t-h-k}-\mu) - E(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)$ is mean-zero L^1-NED on $\{u_t\}$ and necessarily an L^1-mixingale. Its uniform integrability is implied by Lemma A(i). Again, by the law of large numbers for L^1-mixingales we have

$$\frac{1}{T}\sum_{t=1}^{T}((Y_{t-h}-\mu)(Y_{t-h-k}-\mu) - E(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)) \xrightarrow{p} 0, \tag{17}$$

so that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\\
&\quad = \frac{1}{T}\sum_{t=1}^{T}E(Y_{t-h}-\mu)(Y_{t-h-k}-\mu) + o_p(1)\\
&\quad = \sum_{t=1}^{T}\int_{\frac{t}{T}}^{\frac{t+1}{T}}(EY_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu)\mathrm{d}r + o_p(1)\\
&\quad = \int_{\frac{1}{T}}^{\frac{T+1}{T}}(EY_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu)\mathrm{d}r + o_p(1) \to \left(\int g^2\right)\gamma_k.
\end{aligned}$$

So (b) is proved and (ii) follows from (a) and (b).

(iii) It suffices to show the three convergence results:

(a) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \int g^2$,

(b) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 Y_{t-h} \xrightarrow{p} \mu(\int g^2)$ for $1 \le h \le p$, and

(c) $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 Y_{t-h}Y_{t-h-k} \xrightarrow{p} \mu^2(\int g^2) + (\int g^4)\gamma_k$, for $1 \le h \le p$, $0 \le k \le p-h$.

For (a), note that $\{\varepsilon_t^2 - g^2(\frac{t}{T}), \mathcal{F}_t\}$ is α-mixing by Theorem 14.1 in Davidson (1994), and $E\left(\varepsilon_t^2 - g^2(\frac{t}{T})\right)^2 < \infty$ by Assumption 1(ii) and 1(iii). By the law of large numbers for L^1-mixingales,

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 = \frac{1}{T}\sum_{t=1}^{T}E(\varepsilon_t^2) + o_p(1) = \frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right) + o_p(1) \to_p \int g^2.$$

For (b), note that $\{\varepsilon_t^2\}$ is mixing and therefore L^2-NED on $\{u_t\}$, and $\{Y_{t-h}-\mu\}$ is L^2-NED on $\{u_t\}$ as shown in (ii). So by Theorem 17.9 in Davidson (1994), $\{\varepsilon_t^2(Y_{t-h}-\mu)\}$ is L^1-NED on $\{u_t\}$. Moreover, we have

$$E|\varepsilon_t^2(Y_{t-h}-\mu)|^{\delta} \le E\varepsilon_t^{4\delta}E(Y_{t-h}-\mu)^{2\delta} < \infty$$

by Assumption 1(iii) and Lemma A(i). So, by the law of large numbers for L^1-mixingales

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(Y_{t-h}-\mu) \xrightarrow{p} 0.$$

By (a),

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 Y_{t-h} = \mu\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(Y_{t-h}-\mu) \xrightarrow{p} \mu\left(\int g^2\right).$$

To prove (c), it suffices to show

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\xrightarrow{p}\left(\int g^4\right)\gamma_k, \quad \text{for } 1\le h\le p,\ \ 0\le k\le p-h,$$

in view of (a) and (b). Note that

$$\{\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)-g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu),\mathcal{F}_t\}$$

is martingale difference sequence as

$$\begin{aligned}&E(\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)-g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)|\mathcal{F}_{t-1})\\&\quad=(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\cdot E(\varepsilon_t^2-g^2\left(\frac{t}{T}\right)|\mathcal{F}_{t-1})=0.\end{aligned}$$

We have

$$\begin{aligned}&\|\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)-g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\|_\delta\\
&\overset{\text{Minkowski}}{\le}\|\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\|_\delta+\|g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\|_\delta\\
&=\left(E\varepsilon_t^{2\delta}(Y_{t-h}-\mu)^\delta(Y_{t-h-k}-\mu)^\delta\right)^{\frac{1}{\delta}}+g^2\left(\frac{t}{T}\right)\left(E(Y_{t-h}-\mu)^\delta(Y_{t-h-k}-\mu)^\delta\right)^{\frac{1}{\delta}}\\
&\overset{\text{CS}}{\le}\left(E\varepsilon_t^{4\delta}E(Y_{t-h}-\mu)^{2\delta}(Y_{t-h-k}-\mu)^{2\delta}\right)^{\frac{1}{2\delta}}+g^2\left(\frac{t}{T}\right)\left(E(Y_{t-h}-\mu)^{2\delta}E(Y_{t-h-k}-\mu)^{2\delta}\right)^{\frac{1}{2\delta}}\\
&\overset{\text{CS}}{\le}\left(E\varepsilon_t^{4\delta}\left(E(Y_{t-h}-\mu)^{4\delta}E(Y_{t-h-k}-\mu)^{4\delta}\right)^{1/2}\right)^{\frac{1}{2\delta}}\\
&\quad+g^2\left(\frac{t}{T}\right)\left(E(Y_{t-h}-\mu)^{2\delta}E(Y_{t-h-k}-\mu)^{2\delta}\right)^{\frac{1}{2\delta}}<\infty,\end{aligned}$$

where the inequalities follow by Minkowski, Cauchy–Schwarz (CS), and Lemma A(i), respectively. By the law of large numbers, we then have

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)-\frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\xrightarrow{p}0.$$

To evaluate the probability limit of $\frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)$, by arguments similar to those in eqn (17)

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right)E(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)+o_p(1)\\
&\quad=\sum_{t=1}^{T}\int_{\frac{t}{T}}^{\frac{t+1}{T}}g^2\left(\frac{[Tr]}{T}\right)E(Y_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu)\mathrm{d}r+o_p(1)\\
&\quad=\int_{\frac{1}{T}}^{\frac{T+1}{T}}g^2\left(\frac{[Tr]}{T}\right)E(Y_{[Tr]-h}-\mu)(Y_{[Tr]-h-k}-\mu)\mathrm{d}r+o_p(1)\xrightarrow{p}\left(\int g^4\right)\gamma_k.\end{aligned}$$

Thus

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)=\frac{1}{T}\sum_{t=1}^{T}g^2\left(\frac{t}{T}\right)(Y_{t-h}-\mu)(Y_{t-h-k}-\mu)+o_p(1)\xrightarrow{p}\left(\int g^4\right)\gamma_k.$$

So (c) is proved and (iii) follows from (a), (b) and (c).

(iv) By the Cramér–Wold device, it suffices to show

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\lambda' X_{t-1}\varepsilon_t \xrightarrow{d} \mathcal{N}(0,\lambda'\Omega_2\lambda) \tag{18}$$

for every fixed $(p+1)$-vector $\lambda=(\lambda_0,\lambda_1,\ldots,\lambda_p)'\neq 0$. Note that $\{\lambda' X_{t-1}\varepsilon_t,\mathcal{F}_t\}$ is a martingale difference sequence and by (iii),

$$\frac{1}{T}\sum_{t=1}^{T}\lambda'\varepsilon_t^2 X_{t-1}X_{t-1}'\lambda \xrightarrow{p} \lambda'\Omega_2\lambda.$$

Moreover, we have

$$\begin{aligned}E(\lambda' X_{t-1}\varepsilon_t)^{2\delta}&=\|\lambda_0\varepsilon_t+\lambda_1 Y_{t-1}\varepsilon_t+\cdots+\lambda_p Y_{t-p}\varepsilon_t\|_{2\delta}^{2\delta}\\&\overset{\text{Minkowski}}{\leq}(\|\lambda_0\varepsilon_t\|_{2\delta}+\|\lambda_1 Y_{t-1}\varepsilon_t\|_{2\delta}+\cdots+\|\lambda_p Y_{t-p}\varepsilon_t\|_{2\delta})^{2\delta}<\infty,\end{aligned}$$

since for $1\leq h\leq p$,

$$\|\lambda_h Y_{t-h}\varepsilon_t\|_{2\delta}=|\lambda_h|\left(E(Y_{t-h}\varepsilon_t)^{2\delta}\right)^{1/2\delta}\overset{\text{CS}}{\leq}|\lambda_h|\left(EY_{t-h}^{4\delta}E\varepsilon_t^{4\delta}\right)^{1/4\delta}<\infty$$

by Lemma A(i). Thus by the central limit theorem for the martingale differences (e.g. Corollary 5.25 in White, 1984), eqn (18) is true and then (iv) holds. □

Proof of Lemma 2.

(i) We have $\hat{\varepsilon}_t=Y_t-X_{t-1}'\hat{\theta}=\varepsilon_t-X_{t-1}'(\hat{\theta}-\theta)=\varepsilon_t-\zeta_t$, where

$$\zeta_t\overset{\text{def}}{=}X_{t-1}'(\hat{\theta}-\theta)=O_p(T^{-1/2}), \tag{19}$$

since $\hat{\theta}-\theta=O_p(T^{-1/2})$ and $X_{t-1}=O_p(1)$ [the former from Theorem 1 and the latter from Lemma A(i) and Markov's inequality]. Then

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2&=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2-\frac{2}{T}\sum_{t=1}^{T}\zeta_t\varepsilon_t+\frac{1}{T}\sum_{t=1}^{T}\zeta_t^2\\&=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2-\frac{2}{T}O_p(\sqrt{T})+\frac{1}{T}O_p(1)=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2+o_p(1)\xrightarrow{p}\int_0^1 g^2,\end{aligned}$$

where the last convergence result follows from Lemma 1(iii).

(ii) Since $\{\varepsilon_t^4-\omega_4 g^4\left(\frac{t}{T}\right)\}$ is α-mixing and $\sup_t \varepsilon_t^{4\delta}<\infty$ for some $\delta>1$, it is an L^1-mixingale. By the law of large numbers,

$$\frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_t^4-\omega_4 g^4\left(\frac{t}{T}\right)\right)\xrightarrow{p}0.$$

By arguments similar to (i)

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^4 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^4 + o_p(1) = \omega_4 \cdot \frac{1}{T}\sum_{t=1}^{T} g^4\left(\frac{t}{T}\right) + o_p(1) \xrightarrow{p} \omega_4 \int_0^1 g^4.$$

(iii) Again, by arguments similar to (i)

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 X_{t-1}X_{t-1}' \\ &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 X_{t-1}X_{t-1}' + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\zeta_t X_{t-1}X_{t-1}' + \frac{1}{T}\sum_{t=1}^{T}\zeta_t^2 X_{t-1}X_{t-1}' \\ &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 X_{t-1}X_{t-1}' + o_p(1) \xrightarrow{p} \Omega_2,\end{aligned}$$

where the last convergence result is by Lemma 1(iii) and the second equality holds since

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\zeta_t X_{t-1}X_{t-1}' \right\| \le \frac{1}{T}\sum_{t=1}^{T}|\varepsilon_t|\cdot|\zeta_t|\cdot\|X_{t-1}X_{t-1}'\| = \frac{1}{T}\sum_{t=1}^{T} O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1),$$

and

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\zeta_t^2 X_{t-1}X_{t-1}' \right\| \le \frac{1}{T}\sum_{t=1}^{T}|\zeta_t|^2\cdot\|X_{t-1}X_{t-1}'\| = \frac{1}{T}\sum_{t=1}^{T} O_p\left(\frac{1}{T}\right) = o_p(1).$$

(iv) Let $K_{rt} = K(\frac{[Tr]-t}{Tb})$. First we establish

$$\frac{1}{Tb}\sum_{t=1}^{T} K_{rt} \to \int_{-\infty}^{\infty} K(z)\mathrm{d}z < C < \infty, \tag{20}$$

for $r \in [0, 1]$. Indeed, for $1 \le t \le T$, let $[Tr] - t = [Tx]$, where x is a real number such that $-1 \le x < 1$. Then

$$\begin{aligned}\frac{1}{Tb}\sum_{t=1}^{T} K_{rt} &= \sum_{t=1}^{T}\int_{([Tr]-t)/T}^{([Tr]-t+1)/T} K\left(\frac{[Tx]}{Tb}\right)\mathrm{d}\left(\frac{x}{b}\right) \\ &\overset{z=x/b}{=} \sum_{t=1}^{T}\int_{([Tr]-t)/Tb}^{([Tr]-t+1)/Tb} K\left(\frac{[Tbz]}{Tb}\right)\mathrm{d}z \\ &= \int_{([Tr]-T)/Tb}^{[Tr]/Tb} K\left(\frac{[Tbz]}{Tb}\right)\mathrm{d}z \to \int_{-\infty}^{\infty} K(z)\mathrm{d}z < C < \infty.\end{aligned}$$

Note that $\hat{\varepsilon}_t = \varepsilon_t - \zeta_t$ by eqn (19), so

$$\begin{aligned}\left| \left(\frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\right)\left(\hat{g}^2(r) - g^2(r)\right) \right| &= \left| \frac{1}{Tb}\sum_{t=1}^{T} K_{rt}((\varepsilon_t - \zeta_t)^2 - g^2(r)) \right| \\ &\le \left| \frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\left(\varepsilon_t^2 - g^2\left(\frac{t}{T}\right)\right) \right| + \left| \frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\left(g^2\left(\frac{t}{T}\right) - g^2(r)\right) \right| \\ &\quad + 2\left(\frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\varepsilon_t^2\right)^{1/2}\left(\frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\zeta_t^2\right)^{1/2} + \left| \frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\zeta_t^2 \right|.\end{aligned} \tag{21}$$

The first term of eqn (21) vanishes in L^2-norm as $T \to \infty$. Actually, if we let $a_t = \varepsilon_t^2 - g^2(t/T)$, then $\{a_t\}$ is an m.d. sequence and

$$E\left(\frac{1}{Tb}\sum_{t=1}^{T} K_{rt} a_t\right)^2 = \frac{1}{(Tb)^2}\sum_{t=1}^{T} K_{rt}^2 E a_t^2 \le \frac{1}{Tb}\left(\sup_t K_{rt}\right)\left(\sup_t E a_t^2\right)\left(\frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\right) = O\left(\frac{1}{Tb}\right) \to 0,$$

in view of eqn (20). The second term of eqn (21) is bounded by

$$\left|\frac{1}{Tb}\sum_{|t-Tr|\le TMb} K_{rt}\left(g^2\left(\frac{t}{T}\right) - g^2(r)\right)\right| + \left|\frac{1}{Tb}\sum_{|t-Tr|> TMb} K_{rt}\left(g^2\left(\frac{t}{T}\right) - g^2(r)\right)\right|,$$

of which the first term is bounded by

$$\overset{\text{Lipschitz}}{\le} \left|\frac{C}{Tb}\sum_{|t-Tr|\le TMb} K_{rt}\left(\frac{t}{T} - r\right)\right| \le \left|\frac{CM}{T}\sum_{|t-Tr|\le TMb} K_{rt}\right| \le \frac{CM}{T}\cdot 2TMb,$$

and the second term is bounded by

$$\left|\frac{C}{Tb}\sum_{|t-Tr|> TMb} K_{rt}\right| \to C\int_{|z|\ge M} K(z)\mathrm{d}z,$$

by similar arguments to those establishing eqn (20). Thus, the second term of eqn (21) converges to zero by letting $T \to \infty$ and $M \to \infty$. Moreover,

$$\frac{1}{Tb}\sum_{t=1}^{T} K_{rt}\zeta_t^2 = O_p\left(\frac{1}{T}\right)$$

by eqns (19) and (20). So the third and fourth terms of eqn (21) also vanish in probability when $T \to \infty$. Therefore, by eqn (20) again, in view of eqn (21), $\hat{g}^2(r) \xrightarrow{p} g^2(r)$ for values of r at which the function g is continuous. □

ACKNOWLEDGEMENTS

Phillips gratefully acknowledges support from the Kelly Foundation at the Business School, University of Auckland, and the NSF under NSF Grant No. SES 04-142254. The authors thank two referees and the associate editor for helpful comments and suggestions.

REFERENCES

ANDERSEN, T. G., BOLLERSLEV, T. and DIEBOLD, F. X. (2003) Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics* (eds Y. AÏT-SAHALIA and L. P. HANSEN). Amsterdam: North-Holland Press.

ANDREWS, D. W. K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458–67.

Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* B 64, 253–80.

Barndorff-Nielsen, O. E. and Shephard, N. (2004) Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2, 1–48.

Basawa, I. V. and Lund, R. B. (2001) Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis* 22, 651–63.

Beare, B. (2004) Robustifying unit root test to permanent changes in innovation variance. mimeograph, Yale University.

Boswijk, H. P. (2005) Adaptive testing for a unit root with nonstationary volatility. Working Paper, Department of Quantitative Economics, Universiteit van Amsterdam.

Burridge, P. and Taylor, A. M. R. (2001) On regression-based tests for seasonal unit roots in the presence of period heteroscedasticity. *Journal of Econometrics* 104, 91–117.

Cavaliere, G. (2003) Asymptotics for unit root tests under Markov-regime switching. *Econometrics Journal* 6, 93–116.

Cavaliere, G. (2004) Unit root tests under time-varying variance shifts. *Econometric Reviews* 23, 259–92.

Cavaliere, G. and Taylor, A. M. R. (2004) Testing for unit roots in time series models with non-stationary volatility. Working Paper, University of Birmingham.

Dahl, C. M. and Levine, M. (2005) Nonparametric estimation of volatility models with serially dependent innovations. Working Paper, Purdue University.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

Eicker, B. (1963) Limit theorems for regression with unequal and dependant errors. *Annals of Mathematics and Statistics* 34, 447–56.

Francq, C. and Gautier, A. (2004) Large sample properties of parameter least squares estimates for time-varying ARMA models. *Journal of Time Series Analysis* 25, 765–83.

Goncalves, S. and Kilian, L. (2004) Bootstrapping autoregression with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.

Hamori, S. and Tokihisa, A. (1997) Testing for a unit root in the presence of a variance shift. *Economics Letters* 57, 245–53.

Hansen, B. E. (1995) Regression with nonstationary volatility. *Econometrica* 63, 1113–32.

Kim, K. and Schmidt, P. (1993) Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 59, 287–300.

Kim, T. H., Leybourne, S. and Newbold, P. (2002) Unit root tests with a break in innovation variance. *Journal of Econometrics* 109, 365–87.

Kuersteiner, G. M. (2001) Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics* 104, 359–405.

Ling, S., Li, W. K. and McAleer, M (2003) Estimation and testing for unit root processes with GARCH(1,1) errors: theory and Monte Carlo evidence. *Econometric Reviews* 22, 179–202.

Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability and Its Applications* 9, 141–2.

Nelson, C. R., Piger, J. and Zivot, E. (2001) Markov regime-switching and unit root tests. *Journal of Business and Economic Statistics* 19, 404–15.

Phillips, P. C. B. and Xiao, Z. (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–69.

Robinson, P. M. (1989) Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (ed. P. Hackl). Amsterdam: North-Holland, pp. 253–64.

Seo, B. (1999) Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–44.

Watson, G. S. (1964) Smooth regression analysis. *Sankhyā Series* A 26, 359–72.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–38.

White, H. (1984) *Asymptotic Theory for Econometricians*. Orlando, Florida: Academic Press.

Wong, W. H. (1983) On the consistency of cross validation in kernel nonparametric regression. *Annals of Statistics* 11, 1136–41.

Xu, K.-L. and Phillips, P. C. B. (2005) Adaptive estimation of autoregressive models with time-varying variances. Working Paper, Yale University.