Econometric Theory, **24**, 2008, 631–650. Printed in the United States of America.
doi: 10.1017/S0266466608080262

GAUSSIAN INFERENCE IN AR(1) TIME SERIES WITH OR WITHOUT A UNIT ROOT

PETER C.B. PHILLIPS
Cowles Foundation, Yale University
University of York
and
University of Auckland

CHIROK HAN
University of Auckland

This paper introduces a simple first-difference-based approach to estimation and inference for the AR(1) model. The estimates have virtually no finite-sample bias and are not sensitive to initial conditions, and the approach has the unusual advantage that a Gaussian central limit theory applies and is continuous as the autoregressive coefficient passes through unity with a uniform $\sqrt{n}$ rate of convergence. En route, a useful central limit theorem (CLT) for sample covariances of linear processes is given, following Phillips and Solo (1992, *Annals of Statistics*, 20, 971–1001). The approach also has useful extensions to dynamic panels.

1. MAIN RESULTS

We consider a simple AR(1) model in which $y_t = \alpha + u_t$, $u_t = \rho u_{t-1} + \varepsilon_t$, with $\rho \in (-1,1]$ and $\varepsilon_t \sim iid(0,\sigma^2)$. When $|\rho| < 1$, the process u_t may be initialized in the infinite past. When $\rho = 1$, we may initialize at $t = -1$, and u_{-1} may be any random variable and may even depend on n, as it does in distant past initializations where, e.g., $u_{-1} = \sum_{j=1}^{[n\tau]} \varepsilon_{-j} = O_p(\sqrt{n})$ where $[n\tau]$ is the integer part of $n\tau$ for some $\tau > 0$. In both stationary and nonstationary cases, observations on y_t satisfy

$$y_t = (1-\rho)\alpha + \rho y_{t-1} + \varepsilon_t, \qquad \rho \in (-1,1]. \tag{1}$$

The authors thank two referees, the co-editor Bruce Hansen, and Douglas Steigerwald, Peter Thomson, Jin Seo Cho, John Owens, and John Randal for helpful comments. Phillips acknowledges partial support from a Kelly Fellowship and the NSF under Grant SES 04-142254. Address correspondence to Peter Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA; e-mail: peter.phillips@yale.edu.

 0266-4666/08 $15.00

Model (1) is equivalent to the conventional formulation $y_t = \alpha + \rho y_{t-1} + \varepsilon_t$ for all $|\rho| < 1$. At the boundary value $\rho = 1$, the intercept produces a time trend in y_t for the latter model, as is well known. In contrast, under (1), the data are either stationary about a fixed mean (α) when $|\rho| < 1$ or form a simple unit root process when $\rho = 1$.

The present paper provides an estimator of the autoregressive coefficient ρ in (1) that has a Gaussian limit distribution that is continuous as ρ passes through unity. We start by differencing (1) to $\Delta y_t = \rho \Delta y_{t-1} + \Delta\varepsilon_t$ where Δ is the usual difference operator. Least squares on this differenced equation yields an inconsistent estimate because Δy_{t-1} and $\Delta\varepsilon_t$ are correlated. Higher lags of Δy_{t-1} provide valid instruments if $\rho < 1$ but are uncorrelated with the regressor if $\rho = 1$.

Further transformation of the differenced equation produces the new regression equation

$$2\Delta y_t + \Delta y_{t-1} = \rho \Delta y_{t-1} + \eta_t, \qquad \eta_t = 2\Delta y_t + (1-\rho)\Delta y_{t-1}. \tag{2}$$

This transformation is justified by the fact that $E[(\Delta y_{t-1})^2] = 2\sigma^2/(1+\rho)$ and $E\Delta y_{t-1}\Delta y_t = -(1-\rho)\sigma^2/(1+\rho)$, implying that $E[2\Delta y_{t-1}\Delta y_t + (1-\rho)(\Delta y_{t-1})^2] = E\Delta y_{t-1}\eta_t = 0$. It is basically the same as the transformation used by Paparoditis and Politis (2000)[1] but was derived independently.

Least squares on (2) yields the following estimator of ρ:

$$\hat\rho_n = \frac{\sum_{t=1}^{n} \Delta y_{t-1}(2\Delta y_t + \Delta y_{t-1})}{\sum_{t=1}^{n} (\Delta y_{t-1})^2}, \tag{3}$$

where it is assumed that $\{y_t : t = -1, 0, \dots, n\}$ are observed. The following limit theory applies.[2]

THEOREM 1. *For all* $\rho \in (-1,1]$, $\sqrt{n}(\hat\rho_n - \rho) \Rightarrow N(0, 2(1+\rho))$ *as* $n \to \infty$.[3]

The limit theory changes when $\rho > 1$, and the system becomes explosive. In fact, $\hat\rho_n$ is inconsistent, and the limit distribution is nonnormal, and no invariance principle applies, as in the case of the conventional serial correlation coefficient (cf. White, 1958). More particularly, because $\Delta y_{t-1} = O_p(\rho^t)$ when $\rho > 1$, it is clear from (2) that in this case $\hat\rho_n \to_p 1 + 2\rho$. However, when ρ is in the local vicinity of unity and the system is only mildly explosive, the limiting distribution is still Gaussian, as we now show. So, the result in the theorem does not have a hard boundary at unity.

Let $\rho = \rho_n$ and $a_n = \rho_n - 1$ depend on the sample size n, so that a_n measures local deviations from unity and $a_n \to 0$ as $n \to \infty$. The system is now formally a triangular array, but it is convenient to omit the additional subscript in the notation. Because $u_t = \rho_n u_{t-1} + \varepsilon_t$, we have

$$u_{t-2} = \rho_n^t u_{-2} + \sum_{j=1}^{t} \rho_n^{t-j} \varepsilon_{j-2}, \tag{4}$$

and because $y_t = \alpha + u_t$,

$$\Delta y_{t-1} = a_n u_{t-2} + \varepsilon_{t-1}. \tag{5}$$

Using (5), we write

$$\sum_{t=1}^{n} (\Delta y_{t-1})^2 = a_n^2 \sum_{t=1}^{n} u_{t-2}^2 + 2a_n \sum_{t=1}^{n} u_{t-2}\varepsilon_{t-1} + \sum_{t=1}^{n} \varepsilon_{t-1}^2, \tag{6}$$

and because $\eta_t = 2(\Delta\varepsilon_t + \Delta y_{t-1}) + a_n \Delta y_{t-1} = 2(a_n u_{t-2} + \varepsilon_t) + a_n \Delta y_{t-1}$ as a result of (5), we have

$$\sum_{t=1}^{n} \Delta y_{t-1}\eta_t = 2\sum_{t=1}^{n} (a_n u_{t-2} + \varepsilon_{t-1})(a_n u_{t-2} + \varepsilon_t) + a_n \sum_{t=1}^{n} (\Delta y_{t-1})^2. \tag{7}$$

So from (6) and (7),

$$\hat{\rho}_n = \rho_n + (a_n + \delta_n) + \xi_n, \tag{8}$$

where

$$\delta_n = \frac{2a_n^2 \sum_{t=1}^{n} u_{t-2}^2}{a_n^2 \sum_{t=1}^{n} u_{t-2}^2 + 2a_n \sum_{t=1}^{n} u_{t-2}\varepsilon_{t-1} + \sum_{t=1}^{n} \varepsilon_{t-1}^2}, \tag{9}$$

$$\xi_n = \frac{2\sum_{t=1}^{n} [a_n u_{t-2}(\varepsilon_{t-1} + \varepsilon_t) + \varepsilon_t \varepsilon_{t-1}]}{a_n^2 \sum_{t=1}^{n} u_{t-2}^2 + 2a_n \sum_{t=1}^{n} u_{t-2}\varepsilon_{t-1} + \sum_{t=1}^{n} \varepsilon_{t-1}^2}. \tag{10}$$

Here, $a_n + \delta_n$ explains the transition of the bias from zero to $\rho_n + 1$ as ρ_n increases beyond unity, and the quantity ξ_n is instrumental in determining the asymptotic distribution.

In the unit root case where $a_n = 0$, the bias term $a_n + \delta_n$ is zero, and $\sqrt{n}\xi_n = 2n^{-1/2}\sum_{t=1}^{n} \varepsilon_t\varepsilon_{t-1}/n^{-1}\sum_{t=1}^{n} \varepsilon_{t-1}^2 \Rightarrow N(0,4)$, giving the result of Theorem 1. If $\rho_n = \rho > 1$, i.e., if u_t is explosive, then $\sum_{t=1}^{n} u_{t-2}^2$ dominates the other terms related to ε_t, and so $\delta_n \to_p 2$, and ξ_n converges to zero at an exponential rate, as can be shown using analytical tools similar to those in recent work by Phillips and Magdalinos (2005).

When $\rho_n \downarrow 1$ at a rate such that neither $a_n^2 \sum_{t=1}^n u_{t-2}^2$ nor $\sum_{t=1}^n \varepsilon_{t-1}^2$ dominates the other, the asymptotics will be located somewhere in between those two extreme cases. The exact borderline rate of ρ_n is determined by the condition that

$$c_n = \rho_n^n/\sqrt{n} \to c \in [0,\infty) \quad \text{as } n \to \infty, \qquad \rho_n \geq 1. \tag{11}$$

One example that satisfies (11) with $c > 0$ is $\rho_n = (c\sqrt{n})^{1/n}$, in which case $c_n \equiv c$. This ρ_n converges to unity at a rate slower than n^{-1} and faster than $n^{-\beta}$ for any $\beta < 1$ when n is large.

Now suppose that the u_t series is initialized at $t = -2$ and the effect of the initial status is negligible in the sense that

$$\tilde{u}_{-2} = a_n^{1/2}\sigma^{-1}u_{-2} \to_p 0. \tag{12}$$

Let $c_* = \lim_{n\to\infty} \rho_n^{-n} \in [0,1]$ and $c_{**} = \lim n a_n \rho_n^{-n}$. If $c > 0$, then $c_* = c_{**} = 0$, and if $a_n = o(n^{-1})$, then $c_* = 1$ because $\log \rho_n^{-n} = -n\log\rho_n = -n\log(1 + a_n) = -n[a_n + o(a_n)] \to 0$, and $c_{**} = 0$. Note that c_{**} is not always zero. One example is $\rho_n = 1 + c/n$, in which case $na_n = c$ and $\rho_n^n \to e^c$; therefore $c_{**} = c/e^c$. Using (8)–(10) and Lemma 7 in Section 4, we have the following result.

THEOREM 2. *When $\rho_n \geq 1$, under (11) and (12),*

$$\hat{\rho}_n = \rho_n + (\rho_n - 1) + \delta_n + \xi_n,$$

where

(i) $\delta_n \Rightarrow \frac{1}{2}c^2X^2/(\frac{1}{4}c^2X^2 + 1)$,
(ii) $\sqrt{n}\xi_n \Rightarrow (cXY + 2Z)/(\frac{1}{4}c^2X^2 + 1)$,

with

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N\left(0, \begin{bmatrix} 1 - c_*^2 & c_{**} & 0 \\ c_{**} & 1 - c_*^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right). \tag{13}$$

Note that the covariance c_{**} of X and Y is irrelevant for the limit distribution of $\sqrt{n}(\hat{\rho}_n - \rho_n)$ because if $c = 0$ then the XY term disappears from $\sqrt{n}\xi_n$ and if $c > 0$ then $c_{**} = 0$.

If $c = \infty$, then $\delta_n \to_p 2$ and thus $\hat{\rho}_n = 2\rho_n + 1 + o_p(1)$, implying that

$$\frac{\sum_{t=1}^n \Delta y_{t-1}\Delta\varepsilon_t}{\sum_{t=1}^n (\Delta y_{t-1})^2} = \rho_n + o_p(1).$$

Furthermore, in this case the limit distribution of $\sqrt{n}\xi_n$ is degenerate, and when it is appropriately (i.e., exponentially) scaled, it can be shown that the limit distribution is Cauchy-like. We do not pursue this case in the present paper.

On the other hand, if $c = 0$ and $\rho_n \downarrow 1$ sufficiently fast, we still have a Gaussian limit distribution, as follows.

THEOREM 3. *If* $\rho_n^{2n}/\sqrt{n} \to 0$, *then under (12),* $\sqrt{n}(\hat{\rho}_n - \rho_n - a_n) \Rightarrow N(0,4)$.

An obvious example satisfying the condition for Theorem 3 is the conventional local to unity case, where $\rho_n = 1 + c/n$, $\rho_n^n \to e^c$, and hence $n^{-1/2}\rho_n^n \to 0$. In this case $\sqrt{n}a_n \to 0$, and so the bias does not affect the limit distribution, giving $\sqrt{n}(\hat{\rho}_n - \rho_n) \Rightarrow N(0,4)$, as in Theorem 1 when $\rho = 1$. Thus, Theorem 1 holds with the same $\sqrt{n}$ rate as ρ passes through unity to locally explosive values.

The novelty in this result is that the limit distribution is clearly continuous as ρ passes through unity. So the Gaussian limit theory may be used to construct confidence intervals for ρ that are valid across stationary, nonstationary, and even locally explosive cases. However, such confidence intervals are wide compared with those that are based on the usual serial correlation coefficient, and clearly the $N(0,4)$ limit theory is insensitive to local departures from unity.

Differencing in the regression equation (2) reduces the signaling information carried by the regressor Δy_{t-1} in comparison to the usual levels-based approach. The effects are most obvious when $\rho_n \to 1$, in which case the conventional serial correlation coefficient has a higher rate of convergence (cf. Phillips and Magdalinos, 2005), so that $\hat{\rho}_n$ is infinitely deficient over this band of ρ_n values. On the other hand, the limit theory is not sensitive to initial conditions at all when $\rho = 1$, as is the limit theory for the conventional serial correlation.

Because of the inefficiency, the practical usefulness of the proposed approach may be limited. However, as explained in Section 3, the differencing approach eliminates the intercept term and is therefore particularly useful in dynamic panel data models with fixed effects.[4] Furthermore, because the regressor and the error of (2) are uncorrelated, there would be little bias even with a very small time dimension.

Simulation results are provided in Table 1. The limit theory is apparently quite accurate even for small n. Noticeably, there is virtually no bias in the estimator, unlike conventional serial correlations, and the t-ratio is well approximated by the standard normal.

2. MODELS WITH TREND

Next consider the corresponding model with a linear trend. Define $y_t = \alpha + \gamma t + u_t$, where $u_t = \rho u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0,\sigma^2)$, and $\rho \in (-1,1]$, with the

Table 1. Simulation evidence from 50,000 replications

	$\rho = 0$			$\rho = 0.3$			$\rho = 0.5$		
n	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$
40	0.023	2.009	1.025	0.317	2.548	1.020	0.512	2.896	1.019
80	0.012	2.001	1.007	0.310	2.554	1.001	0.506	2.937	1.006
160	0.007	2.031	1.017	0.304	2.560	0.995	0.503	2.958	0.999
320	0.003	2.012	1.007	0.303	2.569	0.993	0.502	2.973	0.997
	$\rho = 0.9$			$\rho = 0.95$			$\rho = 1$		
n	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$	$E(\hat{\rho})$	$nv(\hat{\rho})$	$v(t)$
40	0.903	3.651	1.026	0.951	3.706	1.018	1.001	3.848	1.028
80	0.902	3.672	0.997	0.951	3.782	0.999	1.001	3.910	1.009
160	0.901	3.759	1.004	0.950	3.833	1.000	0.999	3.964	1.010
320	0.901	3.729	0.989	0.950	3.847	0.995	1.000	3.977	1.003

Note: The t-ratios are computed using $\sqrt{n}(\hat{\rho}_n - \rho)/\sqrt{2(1+\hat{\rho}_n)}$. The simulated variances of the t-ratios are given in the $v(t)$ columns.

initial conditions at $t = -2$ and the same specifications as before. The implied model is

$$y_t = (1-\rho)\alpha + \rho\gamma + (1-\rho)\gamma t + \rho y_{t-1} + \varepsilon_t. \tag{14}$$

For this model, we use double differencing to eliminate the intercept and the trend, leading to

$$\Delta^2 y_t = \rho \Delta^2 y_{t-1} + \Delta^2 \varepsilon_t. \tag{15}$$

By recursion, we have

$$\Delta^2 y_{t-1} = \sum_{j=0}^{\infty} \rho^j \Delta^2 \varepsilon_{t-j-1} = \varepsilon_{t-1} - (2-\rho)\varepsilon_{t-2} + (1-\rho)^2 \sum_{j=2}^{\infty} \rho^{j-2} \varepsilon_{t-j-1}, \tag{16}$$

and then

$$E(\Delta^2 y_{t-1})^2 = \left\{ 1 + (2-\rho)^2 + \frac{(1-\rho)^4}{1-\rho^2} \right\} \sigma^2 = \frac{2(3-\rho)\sigma^2}{1+\rho}. \tag{17}$$

Further, because $\Delta^2 \varepsilon_t = \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$, we have

$$E\Delta^2 y_{t-1} \Delta^2 \varepsilon_t = [-2 - (2-\rho)]\sigma^2 = -(4-\rho)\sigma^2. \tag{18}$$

It follows from (17) and (18) that

$$E\Delta^2 y_{t-1}\tilde{\eta}_t = 0, \quad \text{where } \tilde{\eta}_t = 2\Delta^2\varepsilon_t + \phi\Delta^2 y_{t-1}, \qquad \phi = \frac{(4-\rho)(1+\rho)}{3-\rho}. \tag{19}$$

The orthogonality condition (19) leads to the regression model

$$\tilde{\eta}_t = 2(\Delta^2 y_t - \rho\Delta^2 y_{t-1}) + \phi\Delta^2 y_{t-1} = (2\Delta^2 y_t + \Delta^2 y_{t-1}) - \theta\Delta^2 y_{t-1}, \tag{20}$$

where

$$\theta = 1 + 2\rho - \phi = -\frac{(1-\rho)^2}{3-\rho}. \tag{21}$$

Least squares regression on (20) produces the estimator

$$\hat{\theta}_n = \frac{\sum_{t=1}^{n} \Delta^2 y_{t-1}(2\Delta^2 y_t + \Delta^2 y_{t-1})}{\sum_{t=1}^{n} (\Delta^2 y_{t-1})^2}, \tag{22}$$

where it is assumed that $\{y_t : t = -2, -1, 0, \ldots, n\}$ are observed. The estimate $\hat{\theta}_n$ is consistent for θ, and because $\Delta^2 y_t$ is stationary for all $\rho \in (-1,1]$, it is asymptotically normal, as shown in the following theorem.

THEOREM 4. *For all* $\rho \in (-1,1]$, $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, V_\rho)$ *with*

$$V_\rho = \left(\frac{1+\rho}{3-\rho}\right)^2 \sum_1^\infty b_k^2,$$

$$b_1 = 2(3-\rho) + (1-\rho)^2 - \{(2-\rho) + 2(1-\rho)^2/(1+\rho)\}\phi,$$

$$b_2 = -(2-\rho)[1 + (1-\rho)^2] + (1-\rho)^3\phi/(1+\rho),$$

$$b_k = \rho^{k-3}(1-\rho)^3[(1-\rho) + \rho\phi/(1+\rho)], \qquad k \geq 3,$$

where ϕ *and* θ *are defined in (19) and (21), respectively.*

Some simulations are reported in Table 2, where the data for y_t are generated by (14) with $\alpha = \gamma = 1$ and $\varepsilon_t \sim N(0,1)$. For small sample sizes, $\hat{\theta}_n$ seems to be slightly biased upward. Note that $\hat{\theta}_n$ needs to be transformed back to $\hat{\rho}$ to estimate V_ρ and compute the t-ratio. The recovery of $\hat{\rho}$ is conducted according to $\rho = \frac{1}{2}[2 + \theta - \sqrt{\theta(\theta - 8)}]$ if $\theta < 0$ and $\rho = 1$ if $\theta \geq 0$. This right censoring seems to cause slightly large variations in the t-ratio for $\rho \simeq 1$ with small sample sizes.

Table 2. Simulation evidence relating to Theorem 4 with 50,000 replications

	$\rho = 0$ ($\theta = -0.333$, $V_\rho = 1.210$)			$\rho = 0.3$ ($\theta = -0.181$, $V_\rho = 1.547$)			$\rho = 0.5$ ($\theta = -0.1$, $V_\rho = 1.751$)		
n	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$
40	−0.308	1.267	1.026	−0.159	1.594	1.119	−0.078	1.756	1.166
80	−0.321	1.240	0.993	−0.170	1.569	1.052	−0.089	1.746	1.092
160	−0.326	1.240	0.998	−0.176	1.554	1.015	−0.094	1.780	1.074
320	−0.330	1.224	1.000	−0.179	1.573	1.019	−0.097	1.764	1.034
	$\rho = 0.9$ ($\theta = -0.005$, $V_\rho = 1.990$)			$\rho = 0.95$ ($\theta = -0.001$, $V_\rho = 1.997$)			$\rho = 1$ ($\theta = 0$, $V_\rho = 2$)		
n	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$	$E(\hat{\theta})$	$nv(\hat{\theta})$	$v(t)$
40	0.019	1.973	1.228	0.023	2.013	1.253	0.024	2.002	1.244
80	0.007	2.001	1.175	0.010	2.005	1.177	0.013	1.994	1.170
160	0.001	1.985	1.114	0.005	1.988	1.115	0.005	1.997	1.120
320	−0.002	1.977	1.075	0.003	2.010	1.090	0.004	2.000	1.083

Note that $\sum_{k=3}^{\infty} b_k^2 = (1-\rho)^6[(1-\rho) + \rho\phi/(1+\rho)]^2/(1+\rho)$, which is continuous in ρ. Thus V_ρ is continuous in ρ, and the asymptotic distribution is again continuous as ρ passes through unity. When $\rho = 1$, we have $\theta = 0$, $b_1 = 1$, $b_2 = -1$, and $b_k = 0$ for all $k \geq 3$, and the next result follows directly.

COROLLARY 5. *If $\rho = 1$, then $\sqrt{n}\hat{\theta}_n \Rightarrow N(0,2)$.*

3. EXTENSIONS AND APPLICATIONS

The difference-based approach to estimation that is explored previously in this paper can be particularly useful in dynamic panel data models with fixed effects. For dynamic panels, the transformation (2) effectively eliminates fixed effects, and because information about the autoregressive coefficient accumulates as the number of both individual and time series observations increases, the cost of first differencing disappears rather quickly. Moreover, as the simulations indicate, there is virtually no time series autoregressive bias in this approach, so that the dynamic panel bias is correspondingly small. Furthermore, there is no weak instrument problem as $\rho \to 1$ in this new approach, as there is with conventional generalized method of moments (GMM) approaches. Moreover, the Gaussian limit theory with estimable variances also holds with the time span T fixed and large N, not just for large T. Also, in view of Theorem 4, incidental linear trends can be eliminated, while still retaining standard Gaussian asymptotics with estimable variances. This last fact allows for the construction of valid panel unit root tests in the presence of incidental trends. These issues are being explored by the authors in other work.

Extension of the present approach to the AR(p) model is challenging, and the rest of the section will describe one possibility by considering the AR(2) process. Let the data generating process be $y_t = \alpha + u_t$ where $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$ with $\varepsilon_t \sim iid(0, \sigma^2)$. We assume that there is at most one unit root to ensure that the differenced series is stationary.

We first analyze the estimator (3) fitted under the false assumption that the series is AR(1) to examine how sensitive the proposed estimator is to the misspecification of the lag order. For that purpose, let $\hat{\theta}_1$ denote the estimator of (3), where we use a notation different from $\hat{\rho}$ to clarify that a misspecified model is fitted. Then we get

$$\hat{\theta}_1 \to_p \theta_1 := \rho_1 - \rho_2, \tag{23}$$

which is proved in Section 4.

Next, consider estimation of ρ_1 and ρ_2 for this AR(2) model based on differenced data. Let θ_2 be the probability limit of the estimator $\hat{\theta}_2$ obtained by regressing Δy_t on Δy_{t-2}, i.e., $\theta_2 = E(\Delta y_t \Delta y_{t-2})/E[(\Delta y_{t-2})^2]$. Then $\theta_2 = -\theta_1 + (1 + \theta_1)\rho_1/2$, and therefore by combining it with (23), we have

$$\rho_1 = \frac{2(\theta_1 + \theta_2)}{1 + \theta_1} \quad \text{and} \quad \rho_2 = \rho_1 - \theta_1. \tag{24}$$

(Proofs are found in Sect. 4.) The parameters ρ_1 and ρ_2 are now estimated by replacing θ_1 and θ_2 by $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. In principle, it is possible to extend this method to more general general AR(p) models, but the algebra rapidly becomes involved as the order p increases.

4. PROOFS

Let $X_t = C(L)\varepsilon_t = \sum_0^\infty c_j \varepsilon_{t-j}$ and $Y_t = D(L)\varepsilon_t = \sum_0^\infty d_j \varepsilon_{t-j}$ where $\varepsilon_t \sim iid(0, \sigma^2)$. Let $\sum_0^\infty c_j d_j = 0$, so that X_t and Y_t are uncorrelated. Define $\psi_k = \sum_0^\infty (c_j d_{k+j} + c_{k+j} d_j)$. We first establish a useful central limit theorem (CLT) for $n^{-1/2} \sum_{t=1}^n X_t Y_t$, in a manner similar to Phillips and Solo (1992).

THEOREM 6. *$\sum_1^\infty \psi_k^2 < \infty$ and $n^{-1/2} \sum_{t=1}^n X_t Y_t \Rightarrow N(0, \sigma^4 \sum_1^\infty \psi_k^2)$ if*

$$\sum_1^\infty s(c_s^2 + d_s^2) < \infty. \tag{25}$$

Proof. Let $a_j = |c_j| + |d_j|$. Then $|c_j d_k + c_k d_j| \le |c_j d_k| + |c_k d_j| \le a_j a_k$ for all j and k, and (25) implies that

$$\sum_0^\infty a_s^2 < \infty, \qquad \sum_1^\infty s a_s^2 < \infty \tag{26}$$

because $a_j^2 \le 2(c_j^2 + d_j^2)$. Thus

$$\sum_1^\infty \psi_k^2 = \sum_{k=1}^\infty \left[\sum_{j=0}^\infty (c_j d_{k+j} + c_{k+j} d_j) \right]^2 \le \sum_{k=1}^\infty \left(\sum_{j=0}^\infty a_j a_{k+j} \right)^2$$

$$\le \sum_{k=1}^\infty \left(\sum_{j=0}^\infty a_j^2 \right) \left(\sum_{j=0}^\infty a_{k+j}^2 \right) = \left(\sum_0^\infty a_j^2 \right) \left(\sum_1^\infty s a_s^2 \right) < \infty$$

by (26), and so $\sum_1^\infty \psi_k^2 < \infty$ is proved.

For the CLT, write $X_t Y_t = \sum_0^\infty c_j d_j \varepsilon_{t-j}^2 + \sum_{j=0}^\infty \sum_{k=j+1}^\infty (c_j d_k + c_k d_j) \varepsilon_{t-j} \varepsilon_{t-k} = Z_{at} + Z_{bt}$. We will show that

$$n^{-1/2} \sum_{t=1}^n Z_{at} \to_p 0, \tag{27}$$

$$n^{-1/2} \sum_{t=1}^n Z_{bt} \Rightarrow N\left(0, \sigma^4 \sum_1^\infty \psi_k^2 \right). \tag{28}$$

For (27), let $f_j = c_j d_j$ and $F(L) = f_j L^j$. Then $Z_{at} = F(L)\varepsilon_t^2$. Apply the Phillips–Solo device to $F(L)$ to get $F(L) = F(1) + \tilde{F}(L)(L-1) = \tilde{F}(L)(L-1)$, where $\tilde{F}(L) = \sum_0^\infty \tilde{f}_j L^j$, which simplifies because $F(1) = \sum_0^\infty c_j d_j = 0$ by supposition. Thus, we have

$$n^{-1/2} \sum_{t=1}^n Z_{at} = n^{-1/2}(\tilde{Z}_{a0} - \tilde{Z}_{an}), \qquad \tilde{Z}_{at} = \sum_0^\infty \tilde{f}_j \varepsilon_{t-j}^2. \tag{29}$$

Now (27) follows if $\sup_t E|\tilde{Z}_{at}| < \infty$. But

$$E|\tilde{Z}_{at}| \le E \sum_0^\infty |\tilde{f}_j| \varepsilon_{t-j}^2 = \sigma^2 \sum_0^\infty |\tilde{f}_j| \le \sigma^2 \sum_{j=0}^\infty \sum_{k=j+1}^\infty |f_j| = \sigma^2 \sum_1^\infty s|f_s|,$$

and furthermore

$$\sum_1^\infty s|f_s| = \sum_1^\infty s|c_s d_s| \le \left(\sum_0^\infty s c_s^2 \right)^{1/2} \left(\sum_0^\infty s d_s^2 \right)^{1/2} < \infty$$

by (26). So (27) is proved. For (28), let $g_{k,j} = c_j d_{k+j} + c_{k+j} d_j$. Then

$$Z_{bt} = \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} (c_j d_k + c_k d_j) \varepsilon_{t-j} \varepsilon_{t-k} = \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} (c_j d_{r+j} + c_{r+j} d_j) \varepsilon_{t-j} \varepsilon_{t-r-j}$$

$$= \sum_{r=1}^{\infty} \sum_{0}^{\infty} g_{r,j} \varepsilon_{t-j} \varepsilon_{t-r-j} = \sum_{r=1}^{\infty} G_r(L) \varepsilon_t \varepsilon_{t-r}, \qquad G_r(L) = \sum_{0}^{\infty} g_{r,j} L^j.$$

Apply the Phillips–Solo device again to each $G_r(L)$ to get $G_r(L) = G_r(1) + \tilde{G}_r(L)(L-1)$. Thus

$$n^{-1/2} \sum_{t=1}^{n} Z_{bt} = n^{-1/2} \sum_{t=1}^{n} \sum_{r=1}^{\infty} G_r(1) \varepsilon_t \varepsilon_{t-r} + n^{-1/2} \sum_{r=1}^{\infty} (\tilde{v}_{r0} - \tilde{v}_{rn}), \tag{30}$$

where $\tilde{v}_{rt} = \tilde{G}_r(L) \varepsilon_t \varepsilon_{t-r} = \sum_0^\infty \tilde{g}_{r,j} \varepsilon_{t-j} \varepsilon_{t-r-j}$ with $\tilde{g}_{r,j} = \sum_{s=j+1}^{\infty} g_{r,s}$. But

$$\sum_{r=1}^{\infty} \tilde{v}_{rt} = \sum_{r=1}^{\infty} \sum_{j=0}^{\infty} \tilde{g}_{r,j} \varepsilon_{t-j} \varepsilon_{t-r-j} = \sum_{j=0}^{\infty} \varepsilon_{t-j} \left(\sum_{r=1}^{\infty} \tilde{g}_{r,j} \varepsilon_{t-r-j} \right).$$

Thus, we have

$$E\left(\sum_{r=1}^{\infty} \tilde{v}_{rt} \right)^2 = \sigma^4 \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \tilde{g}_{r,j}^2 = \sigma^4 \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \left(\sum_{k=j+1}^{\infty} g_{r,k} \right)^2, \tag{31}$$

and again because $|g_{k,j}| \le a_j a_{k+j}$,

$$\sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \left(\sum_{k=j+1}^{\infty} g_{r,k} \right)^2 \le \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \left(\sum_{k=j+1}^{\infty} a_k a_{r+k} \right)^2$$

$$\le \sum_{j=0}^{\infty} \sum_{r=1}^{\infty} \left(\sum_{k=j+1}^{\infty} a_k^2 \right) \left(\sum_{k=j+1}^{\infty} a_{r+k}^2 \right) = \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^{\infty} a_k^2 \right) \left(\sum_{r=1}^{\infty} \sum_{k=j+1}^{\infty} a_{r+k}^2 \right)$$

$$= \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^{\infty} a_k^2 \right) \left(\sum_{s=j+1}^{\infty} (s-j) a_{s+1}^2 \right) \le \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^{\infty} a_k^2 \right) \left(\sum_{s=1}^{\infty} s a_s^2 \right)$$

$$= \left(\sum_{1}^{\infty} s a_s^2 \right)^2 < \infty, \quad \text{by (26)}.$$

Thus (31) is finite uniformly in t, and so the second term of (30) converges in probability to zero.

The first term of (30) is $n^{-1/2} \sum_t \varepsilon_t \varepsilon_{t-1}^g$ where $\varepsilon_{t-1}^g = \sum_{r=1}^\infty G_r(1) \varepsilon_{t-r}$ with $G_r(1) = \psi_r = \sum_0^\infty (c_j d_{r+j} + c_{r+j} d_j)$. This term will be shown to follow the martingale CLT, which holds if (i) a version of the Lindeberg condition holds and (ii) $n^{-1} \sum_{t=1}^n (\varepsilon_t \varepsilon_{t-1}^g)^2 \to_p \sigma^4 \sum_1^\infty \psi_k^2$. The Lindeberg condition follows directly from stationarity and integrability. The convergence in probability (ii) holds if $n^{-1} \sum_1^n (\varepsilon_{t-1}^g)^2 \to_p \sigma^2 \sum_1^\infty \psi_k^2$, which is satisfied by Lemma 5.11 of Phillips and Solo. (See Sect. 5.10 of Phillips and Solo for details.) Now the stated CLT follows by (27) and (28) because $X_t Y_t = Z_{at} + Z_{bt}$. ■

It is of some independent interest from a technical point of view that only a finite second moment is assumed for ε_t in the preceding derivation.

Theorem 1 is now proved using Phillips and Solo (1992, Thm. 3.7) for the denominator and our Theorem 6 for the numerator. In what follows, the regularity conditions are automatically satisfied because of the exponentially decaying coefficients in the lag polynomials.

Proof of Theorem 1. First note that

$$\sqrt{n}(\hat{\rho}_n - \rho) = \frac{n^{-1/2} \sum_{t=1}^{n} \Delta y_{t-1} \eta_t}{n^{-1} \sum_{t=1}^{n} (\Delta y_{t-1})^2}. \tag{32}$$

When $\rho = 1$, we have $\Delta y_t = \varepsilon_t$ and $\eta_t = 2\Delta\varepsilon_t + 2\Delta y_{t-1} = 2\varepsilon_t$, and so

$$\sqrt{n}(\hat{\rho}_t - \rho) = \frac{2n^{-1/2} \sum_t \varepsilon_{t-1} \varepsilon_t}{n^{-1} \sum_t \varepsilon_{t-1}^2} \Rightarrow \frac{2N(0,\sigma^4)}{\sigma^2} =_d N(0,4),$$

as stated. For $|\rho| < 1$, because

$$\Delta y_{t-1} = \rho \Delta y_{t-2} + \Delta\varepsilon_{t-1} = \sum_0^\infty \rho^j \Delta\varepsilon_{t-j-1} = \varepsilon_{t-1} - (1-\rho) \sum_1^\infty \rho^{j-1} \varepsilon_{t-j-1}, \tag{33}$$

we find $E(\Delta y_{t-1})^2 = \sigma^2[1 + (1-\rho)^2 \sum_1^\infty \rho^{2(j-1)}] = 2\sigma^2/(1+\rho)$, and hence

$$\frac{1}{n} \sum_{t=1}^{n} (\Delta y_{t-1})^2 \to_{a.s.} \frac{2\sigma^2}{1+\rho} \tag{34}$$

by Theorem 3.7 of Phillips and Solo (1992). We also have

$$\eta_t = 2\Delta\varepsilon_t + (1+\rho)\Delta y_{t-1} = 2\varepsilon_t - (1-\rho)\varepsilon_{t-1} - (1-\rho^2) \sum_1^\infty \rho^{j-1} \varepsilon_{t-j-1}. \tag{35}$$

Let $\Delta y_{t-1} = \sum_0^\infty c_j \varepsilon_{t-j}$ and $\eta_t = \sum_0^\infty d_j \varepsilon_{t-j}$, where

$$c_0 = 0, \qquad d_0 = 2,$$

$$c_1 = 1, \qquad d_1 = -(1-\rho),$$

$$c_k = -\rho^{k-2}(1-\rho), \qquad d_k = -\rho^{k-2}(1-\rho^2), \quad k \geq 2,$$

because of (33) and (35). Clearly $E\Delta y_{t-1}\eta_t = -(1-\rho) + (1-\rho)(1-\rho^2)(1+\rho^2+\cdots) = 0$, and by Theorem 6, we have

$$n^{-1/2}\sum_{t=1}^{n} \Delta y_{t-1}\eta_t \Rightarrow N\left(0, \sigma^4 \sum_1^\infty \psi_k^2\right), \quad \psi_k = \sum_{j=0}^{\infty}(c_j d_{k+j} + c_{k+j} d_j), \tag{36}$$

and so it only remains to calculate the ψ_k's. After some algebra, we get

$$\sum_{j=0}^{\infty} c_j d_{k+j} = -\rho^{k-1}(1-\rho), \qquad k \geq 1,$$

and

$$\sum_{j=0}^{\infty} c_{1+j} d_j = 3-\rho, \qquad \sum_{j=0}^{\infty} c_{k+j} d_j = -\rho^{k-2}(1-\rho)(2-\rho), \qquad k \geq 2,$$

implying that

$$\psi_1 = 2, \qquad \psi_k = -2\rho^{k-2}(1-\rho), \qquad r \geq 2.$$

So the variance in (36) is $\sigma^4[4 + 4(1-\rho)^2 \sum_2^\infty \rho^{2(k-2)}] = 8\sigma^4/(1+\rho)$. Finally, from (32), (34), and (36) with the calculated variance, we have

$$\sqrt{n}(\hat{\rho}_n - \rho) \Rightarrow \left(\frac{1+\rho}{2\sigma^2}\right) N\left(0, \frac{8\sigma^4}{1+\rho}\right) =_d N(0, 2(1+\rho)),$$

as stated. ■

Theorem 4 will be proved next because it involves similar algebra. Recall that $\phi = (4-\rho)(1+\rho)/(3-\rho)$ and $\theta = -(1-\rho)^2/(3-\rho)$.

Proof of Theorem 4. Write

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{n^{-1/2}\sum_t \Delta^2 y_{t-1}\tilde{\eta}_t}{n^{-1/2}\sum_t (\Delta^2 y_{t-1})^2}. \tag{37}$$

The denominator of (37) converges almost surely to $2(3-\rho)\sigma^2/(1+\rho)$ by (17) and Theorem 3.7 of Phillips and Solo (1992). As for the numerator of (37), because of the exponential decay in the coefficients of the lag polynomials we may invoke Theorem 6. Let $\Delta^2 y_{t-1} = \sum_0^\infty c_j \varepsilon_{t-j}$ with

$$c_0 = 0, \qquad c_1 = 1, \qquad c_2 = -(2-\rho), \qquad c_k = \rho^{k-3}(1-\rho)^2, \qquad k \geq 3, \tag{38}$$

because of (16). Because $\tilde{\eta}_t = 2\Delta^2\varepsilon_t + \phi\Delta^2 y_{t-1} = 2\varepsilon_t - 4\varepsilon_{t-1} + 2\varepsilon_{t-2} + \phi\sum_0^\infty c_j\varepsilon_{t-j}$, we also have $\tilde{\eta}_t = \sum_0^\infty d_j\varepsilon_{t-j}$ with

$$d_0 = 2 + \phi c_0, \qquad d_1 = -4 + \phi c_1, \qquad d_2 = 2 + \phi c_2, \qquad d_k = \phi c_k, \qquad k \geq 3. \tag{39}$$

Note that $\sum_0^\infty c_j d_j = 0$. We can show that $\sum_{j=2}^\infty c_j c_{k+j} = \rho^{k-1}\mu$ where $\mu = -2(1-\rho)^2/(1+\rho)$ for $k \geq 1$. (First show the result for $k = 1$ and then use the recursion $c_{j+1} = \rho c_j$, $j \geq 3$.) Using this fact and (38) and (39), we can show that

$$\sum_{j=0}^\infty c_j d_{k+j} = 2c_1\{k=1\} + (c_{k+1} + \rho^{k-1}\mu)\phi,$$

$$\sum_{j=0}^\infty c_{k+j} d_j = 2(c_k - 2c_{k+1} + c_{k+2}) + (c_{k+1} + \rho^{k-1}\mu)\phi,$$

for $k \geq 1$. Adding term by term, we get

$$\psi_k = 2b_k = 2[c_1\{k=1\} + (c_k - 2c_{k+1} + c_{k+2}) + (c_{k+1} + \rho^{k-1}\mu)\phi],$$

and by Theorem 6, $n^{-1/2}\sum_{t=1}^n \Delta^2 y_{t-1}\tilde{\eta}_t \Rightarrow N(0, 4\sigma^4\sum_1^\infty b_k^2)$. The result then follows by combining the limits for the numerator and denominator. ∎

Next we prove the theorem for the mildly explosive case (11). Define

$$X_{n,k} = \left(\frac{2a_n}{\sigma^2}\right)^{1/2}\sum_{t=1}^n \rho_n^{-t}\varepsilon_{t-k}, \qquad Y_{n,k} = \left(\frac{2a_n}{\sigma^2}\right)^{1/2}\sum_{t=1}^n \rho_n^{-(n-t)}\varepsilon_{t-k}, \tag{40}$$

$W_n = n^{-1}\sum_{t=1}^n \varepsilon_{t-1}^2/\sigma^2$, and $Z_n = n^{-1/2}\sum_{t=1}^n \varepsilon_t\varepsilon_{t-1}/\sigma^2$. We will deal with the individual terms of (6) and (7) one by one. Note that $\varepsilon_t \sim iid(0,\sigma^2)$.

LEMMA 7. *The following result is true:*

$$\frac{a_n^2}{n\sigma^2}\sum_{t=1}^n u_{t-2}^2 = \frac{c_n^2 X_{n,2}^2}{2(\rho_n+1)} + \frac{\rho_n^2(c_n^2 - n^{-1})\tilde{u}_{-2}^2}{\rho_n+1} + \frac{\sqrt{2}c_n\tilde{u}_{-2}}{\rho_n+1}(\rho_n^2 c_n X_{n,2} - n^{-1/2}Y_{n,2}) - \nu_{1n};$$

$$\frac{2a_n}{n^{1/2}\sigma^2}\sum_{t=1}^n u_{t-2}\varepsilon_{t-k} = c_n X_{n,2}Y_{n,k} + \sqrt{2}c_n Y_{n,k}\tilde{u}_{-2} - \nu_{2n,k}, \qquad k = 0,1,$$

where

$$\nu_{1n} = \frac{a_n}{n(\rho_n+1)}\sum_{t=1}^{n}\varepsilon_{t-2}^2 + \frac{2a_n}{n(\rho_n+1)}\sum_{j=1}^{n}\sum_{k=j+1}^{n}\rho_n^{k-j}\varepsilon_{j-2}\varepsilon_{k-2}$$
$$= O_p(a_n) + O_p(n^{-1/2}c_n),$$
$$\nu_{2n,k} = \frac{a_n}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j=t+1}^{n}\rho_n^{t-j}\varepsilon_{j-2}\varepsilon_{t-k} = O_p(a_n^{1/2}), \qquad k = 0,1.$$

Proof of Lemma 7. The expressions are worked out using the identities $\sum_{t=1}^{n}\sum_{j=1}^{t}x_{tj} = \sum_{j=1}^{n}\sum_{t=j}^{n}x_{tj}$, $\sum_{t=2}^{n}\sum_{j=1}^{t-1}x_{tj} = \sum_{j=1}^{n-1}\sum_{t=j+1}^{n}x_{tj}$, and $\sum_{t=j+1}^{n}\sum_{k=j+1}^{n}x_{tk} = \sum_{k=j+1}^{n}\sum_{t=k}^{n}x_{tk}$. ■

The next lemma provides a useful simplification.

LEMMA 8. $X_{n,2} = X_{n,0} + o_p(1)$ *and* $Y_{n,1} = Y_{n,0} + o_p(1)$.

Proof. Let $k_{1n} = (2a_n/\sigma^2)^{1/2}$ (only for notational simplicity). Then

$$X_{n,2} = k_{1n}\sum_{t=1}^{n}\rho_n^{-t}\varepsilon_{t-2} = k_{1n}\sum_{t=-1}^{n-2}\rho_n^{-(t+2)}\varepsilon_t$$
$$= k_{1n}\left(\rho_n^{-2}\sum_{t=1}^{n}\rho_n^{-t}\varepsilon_t + \rho_n^{-1}\varepsilon_{-1} + \rho_n^{-2}\varepsilon_0 - \rho_n^{-(n-1)}\varepsilon_{n-1} - \rho_n^{-n}\varepsilon_n\right)$$
$$= \rho_n^{-2}X_{n,0} + k_{1n}(\rho_n^{-1}\varepsilon_{-1} + \rho_n^{-2}\varepsilon_0 - \rho_n^{-(n-1)}\varepsilon_{n-1} - \rho_n^{-n}\varepsilon_n)$$
$$= \rho_n^{-2}X_{n,0} + O_p(k_n) = X_{n,0} + O_p(a_n^{1/2}),$$

under (11), because $X_{n,k} = O_p(1)$ and $\rho_n^{-2} = 1 - a_n(\rho_n+1)/\rho_n^2 = 1 - O(a_n)$. Similarly,

$$Y_{n,1} = k_{2n}\sum_{t=1}^{n}\rho_n^{t-n}\varepsilon_{t-1} = k_{2n}\sum_{t=0}^{n-1}\rho_n^{t+1-n}\varepsilon_t, \qquad k_{2n} = (2a_n/\sigma^2)^{1/2}$$
$$= k_{2n}\left(\rho_n\sum_{t=1}^{n}\rho_n^{t-n}\varepsilon_t + \rho_n^{1-n}\varepsilon_0 - \rho_n\varepsilon_n\right)$$
$$= \rho_n Y_{n,0} + k_{2n}(\rho_n^{1-n}\varepsilon_0 - \rho_n\varepsilon_n) = Y_{n,0} + O_p(a_n^{1/2})$$

under (11). ■

Now let $X_n = X_{n,0} = (2a_n/\sigma^2)^{1/2}\sum_{t=1}^{n}\rho_n^{-t}\varepsilon_t$ and $Y_n = Y_{n,0} = (2a_n/\sigma^2)^{1/2}\sum_{t=1}^{n}\rho_n^{t-n}\varepsilon_t$ for notational simplicity. We obtain the limit distribution of (X_n, Y_n, Z_n). The following lemma will be useful in proving the CLT in Lemma 10.

LEMMA 9. *Under (11),*

(i) $\max_{1\le t\le n} a_n^{1/2}\rho_n^{-t}|\varepsilon_t| \to_p 0$;
(ii) $\max_{1\le t\le n} a_n^{1/2}\rho_n^{t-n}|\varepsilon_t| \to_p 0$;
(iii) $\max_{1\le t\le n} n^{-1/2}|\varepsilon_t\varepsilon_{t-1}| \to_p 0$.

Proof. We prove (i), and (ii) and (iii) follow similarly. Let $b_{nt} = a_n\rho_n^{-2t}$. Then (i) states that $\max_{1\le t\le n} b_{nt}^{1/2}|\varepsilon_t| \to_p 0$ or equivalently that $\max_{1\le t\le n} b_{nt}\varepsilon_t^2 \to_p 0$. So we shall show that

$$P\left\{\max_{1\le t\le n} b_{nt}\varepsilon_t^2 > \delta\right\} \to 0 \quad \text{for all } \delta > 0.$$

Fix $\delta > 0$. Because

$$P\left\{\max_{1\le t\le n} b_{nt}\varepsilon_t^2 > \delta\right\} = 1 - \prod_{t=1}^{n}(1 - P\{b_{nt}\varepsilon_t^2 > \delta\}),$$

this probability converges to zero if and only if $\sum_{t=1}^n P\{b_{nt}\varepsilon_t^2 > \delta\} \to 0$. But

$$\begin{aligned}\sum_{t=1}^n P\{b_{nt}\varepsilon_t^2 > \delta\} &= \sum_{t=1}^n (b_{nt}/\delta)\cdot(\delta/b_{nt})P\{\varepsilon_t^2 > \delta/b_{nt}\}\\ &\le \sum_{t=1}^n (b_{nt}/\delta)E\varepsilon_t^2 1\{\varepsilon_t^2 > \delta/b_{nt}\}\\ &\le \sum_{t=1}^n (b_{nt}/\delta)\max_{1\le j\le n} E\varepsilon_j^2 1\{\varepsilon_j^2 > \delta/b_{nt}\}\\ &\le \delta^{-1}\left(\sum_{t=1}^n b_{nt}\right)E\varepsilon_1^2 1\{\varepsilon_1^2 > \delta/a_n\} \to 0,\end{aligned}$$

because $b_{nt} \le a_n$, $a_n \to 0$, $E\varepsilon_1^2 < \infty$, and $\sum_{t=1}^n b_{nt} \le a_n/(\rho_n^2 - 1) = 1/(\rho_n + 1) = O(1)$. This proves (i). ■

LEMMA 10. *Under (11),* $(X_n, Y_n, Z_n) \Rightarrow (X, Y, Z)$ *with limit distribution (13).*

Proof. The limit variance matrix is straightforwardly obtained by calculation. For the joint Gauss limit, we use the Cramér–Wold device and show that for any constants λ_1^*, λ_2^*, and λ_3^*,

$$U_n = \lambda_1^* X_n + \lambda_2^* Y_n + \lambda_3^* Z_n$$

converges to a normal distribution. Let $\lambda_1 = \sqrt{2}\sigma^{-1}\lambda_1^*$, $\lambda_2 = \sqrt{2}\sigma^{-1}\lambda_2^*$, and $\lambda_3 = \sigma^{-2}\lambda_3^*$. Then by the definition of X_n, Y_n, and Z_n, we have

$$U_n = \sum_{t=1}^n \zeta_{nt}, \qquad \zeta_{nt} = a_n^{1/2}(\lambda_1\rho_n^{-t} + \lambda_2\rho_n^{t-n})\varepsilon_t + \lambda_3 n^{-1/2}\varepsilon_t\varepsilon_{t-1}.$$

Let $\mathcal{F}_{nt}$ be the σ-field generated by $\varepsilon_j, j \le t$. Then ζ_{nt} is a martingale difference array with respect to $\mathcal{F}_{nt}$. We invoke the martingale difference CLT (e.g., Phillips and Solo, 1992, Thm. CLT), which requires that

(i) $\sum_{t=1}^n \zeta_{nt}^2 \to_p \frac{1}{2}[(\lambda_1^2 + \lambda_2^2)(1 - c_*^2) + 2\lambda_1\lambda_2 c_{**}]\sigma^2 + \lambda_3^2\sigma^4$;
(ii) $\max_{1\le t\le n}|\zeta_{nt}| \to_p 0$.

But (ii) is already proved by Lemma 9 because

$$\max_{1\le t\le n}|\zeta_{nt}| \le \lambda_1 \max_{1\le t\le n} a_n^{1/2}\rho_n^{-t}|\varepsilon_t| + \lambda_2 \max_{1\le t\le n} a_n^{1/2}\rho_n^{t-n}|\varepsilon_t|$$
$$+ \lambda_3 \max_{1\le t\le n} n^{-1/2}|\varepsilon_t\varepsilon_{t-1}|,$$

and so it remains to prove (i). Write

$$\sum_{t=1}^n \zeta_{nt}^2 = \sum_{t=1}^n a_n(\lambda_1\rho_n^{-t} + \lambda_2\rho_n^{t-n})^2\varepsilon_t^2 + \lambda_3^2 \sum_{t=1}^n n^{-1}\varepsilon_t^2\varepsilon_{t-1}^2$$
$$+ 2\lambda_3 \sum_{t=1}^n a_n^{1/2} n^{-1/2}(\lambda_1\rho_n^{-t} + \lambda_2\rho_n^{t-n})\varepsilon_t^2\varepsilon_{t-1}$$
$$= Q_{1n} + \lambda_3^2 Q_{2n} + 2\lambda_3 Q_{3n}, \quad \text{say.}$$

We show that $Q_{1n} \to_p \frac{1}{2}[(\lambda_1^2 + \lambda_2^2)(1 - c_*^2) + 2\lambda_1\lambda_2 c_{**}]\sigma^2$, $Q_{2n} \to_p \sigma^4$, and $Q_{3n} \to_p 0$ by invoking Theorem 11 at the end of this section. (It is recommended that readers refer to that theorem before proceeding.)

For Q_{1n}, let $b_{nt} = a_n(\lambda_1\rho_n^{-t} + \lambda_2\rho_n^{t-n})^2$ and $v_{nt} = b_{nt}(\varepsilon_t^2 - \sigma^2)$. Clearly, this v_{nt} is a row-wise martingale difference with respect to the natural σ-field, and condition (i) of Theorem 11 is obviously satisfied because ε_t are independent and identically distributed (i.i.d.). It is just a matter of calculation that $\sum_{t=1}^n b_{nt} \to \frac{1}{2}[(\lambda_1^2 + \lambda_2^2)(1 - c_*^2) + 2\lambda_1\lambda_2 c_{**}]$ and $\sum_{t=1}^n b_{nt}^2 \to 0$, and so by the theorem, $\sum_{t=1}^n v_{nt} \to_p 0$. Now because $Q_{1n} = \sum_{t=1}^n v_{nt} + \sigma^2 \sum_{t=1}^n b_{nt}$, we have the desired convergence for Q_{1n}.

For Q_{2n}, let $v_{nt} = n^{-1}(\varepsilon_t^2 - \sigma^2)(\varepsilon_{t-1}^2 - \sigma^2)$ and $b_{nt} = n^{-1}$. Then v_{nt} constitutes a martingale difference array, and by Theorem 11, $\sum_{t=1}^n v_{nt} \to_p 0$. Now

$$\frac{1}{n}\sum_{t=1}^n \varepsilon_t^2\varepsilon_{t-1}^2 = \sum_{t=1}^n v_{nt} + \frac{\sigma^2}{n}\sum_{t=1}^n(\varepsilon_t^2 - \sigma^2) + \frac{\sigma^2}{n}\sum_{t=1}^n(\varepsilon_{t-1}^2 - \sigma^2) + \sigma^4 \to_p \sigma^4.$$

Next, for Q_{3n}, let $b_{nt} = (a_n/n)^{1/2}(\lambda_1\rho_n^{-t} + \lambda_2\rho_n^{t-n})$ and $v_{nt} = b_{nt}(\varepsilon_t^2 - \sigma^2)\varepsilon_{t-1}$. Then condition (i) of Theorem 11 is obvious, and condition (iii) is also straightforwardly verified. Condition (ii) also holds: If $\lim na_n > 0$, then $\sum_{t=1}^n b_{nt} \le (na_n)^{-1/2}(\lambda_1 + \rho_n\lambda_2)$, which is finite in the limit, and if $na_n \to 0$, then $\sum_{t=1}^n b_{nt} \le (a_n/n)^{1/2} n(\lambda_1 + \lambda_2) = (na_n)^{1/2}(\lambda_1 + \lambda_2) \to 0$. And as a result $\sum_{t=1}^n v_{nt} \to_p 0$. Now

$$Q_{3n} = \sum_{t=1}^n v_{nt} + \sigma^2 \sum_{t=1}^n b_{nt}\, \varepsilon_{t-1} \to_p 0,$$

because the second term converges to zero in L_2. Thus, the conditions for the martingale CLT are all satisfied, and we have the stated result. ■

Proof of Theorem 2. Combine (8)–(10), Lemma 7, Lemma 8, and Lemma 10. ■

Proof of Theorem 3. If $\rho_n^2/\sqrt{n} = \sqrt{n} c_n^2 \to 0$, then $(a_n^2/\sqrt{n}) \sum_{t=1}^n u_{t-2}^2 \to_p 0$ by Lemma 7, and therefore $\sqrt{n}\delta_n \to_p 0$. Also, because $c = 0$ in this case, $\sqrt{n}\xi_n \Rightarrow 2Z \sim N(0,4)$. So

$$\sqrt{n}(\hat{\rho}_n - \rho_n - a_n) = \sqrt{n}\delta_n + \sqrt{n}\xi_n \Rightarrow N(0,4),$$

as stated. ■

The following result is adapted from Theorem 19.7 of Davidson (1994) and is used in the proof of Lemma 10. For a more general and detailed treatment, see Davidson (1994).

THEOREM 11. *Let $\{v_{nt}\}$ be a row-wise martingale difference array and $\{b_{nt}\}$ an array of positive constants. If*

(i) $\{v_{nt}/b_{nt}\}$ *is uniformly integrable,*
(ii) $\limsup_{n\to\infty} \sum_{t=1}^n b_{nt} < \infty$,
(iii) $\lim_{n\to\infty} \sum_{t=1}^n b_{nt}^2 = 0$,

then $\sum_{t=1}^n v_{nt} \to_{L_1} 0$, *and thus* $\sum_{t=1}^n v_{nt} \to_p 0$.

Now we provide technical results for the AR(2) model.

Let $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$ where $\varepsilon_t \sim iid(0,\sigma^2)$. Let $1 - \rho_1 L - \rho_2 L^2 = (1 - \phi_1 L)(1 - \phi_2 L)$ where $\phi_1 \in (-1,1]$ and $|\phi_2| < 1$ so that Δu_t is stationary. Define $\omega_k = E\Delta u_t \Delta u_{t-k}$. Note that $\Delta y_t = \Delta u_t$. The following result is true.

LEMMA 12. *Let* $c = \sigma^2/[(1-\phi_1\phi_2)(1+\phi_1)(1+\phi_2)]$. *Let* $\theta_1 := \rho_1 - \rho_2$. *Then* $\omega_0 = 2c$, $\omega_1 = -c(1-\theta_1)$, *and* $\omega_2 = -c[2\theta_1 - (1+\theta_1)\rho_1]$.

Proof. When $|\phi_j| < 1$, according to Karanasos (1998), the autocovariance functions are

$$Eu_t u_{t-j} = \frac{\sigma^2}{(1-\phi_1\phi_2)(\phi_1-\phi_2)}\left(\frac{\phi_1^{j+1}}{1-\phi_1^2} - \frac{\phi_2^{j+1}}{1-\phi_2^2}\right),$$

implying that

$$\omega_0 = \frac{2\sigma^2}{(1-\phi_1\phi_2)(\phi_1-\phi_2)}\left(\frac{\phi_1}{1+\phi_1} - \frac{\phi_2}{1+\phi_2}\right),$$

$$\omega_1 = \frac{-\sigma^2}{(1-\phi_1\phi_2)(\phi_1-\phi_2)}\left[\frac{\phi_1(1-\phi_1)}{1+\phi_1} - \frac{\phi_2(1-\phi_2)}{1+\phi_2}\right],$$

$$\omega_2 = \frac{-\sigma^2}{(1-\phi_1\phi_2)(\phi_1-\phi_2)}\left[\frac{\phi_1^2(1-\phi_1)}{1+\phi_1} - \frac{\phi_2^2(1-\phi_2)}{1+\phi_2}\right].$$

(If $\phi_1 = \phi_2$, then apply L'Hôpital's rule.) Note that the preceding identities hold when $\phi_1 = 1$ also, where $\Delta y_t = \phi_2 \Delta y_{t-1} + \varepsilon_t$. The results then follow after some straightforward algebra. ■

Proof of (23). The law of large numbers applies to the averages by Phillips and Solo (1992). Thus, the limit of $\hat{\theta}_1$ can be expressed as

$$\hat{\theta}_1 \to_p \frac{2\omega_1 + \omega_0}{\omega_0} = \theta_1$$

by Lemma 12. ■

Proof of (24). Noting that $\theta_2 = \omega_2/\omega_0$, we have $\theta_2 = -\theta_1 + (1+\theta_1)\rho_1/2$ as a result of Lemma 12. The rest is obvious. ■

NOTES

1. We thank an anonymous referee for the reference to this paper.
2. After a change of variable, this result corresponds to Theorem 2.1 in Paparoditis and Politis (2000).
3. Though the parameter space in the theorem is stated with a hard boundary at unity, there is no boundary parameter problem at unity. In fact the limit theory extends in a local sense beyond unity, as explained in the rest of Section 1.
4. As a referee rightly pointed out, the Cauchy estimator based on the moment condition that $E\,\mathrm{sgn}(y_{t-1})\varepsilon_t = 0$, which is $\rho + \sum_t \mathrm{sgn}(y_{t-1})\varepsilon_t / \sum_t |y_{t-1}|$ if no constant term is present, also has an asymptotically normal t-ratio. Although our method attains asymptotic normality by converting a possibly integrated series into a stationary one by first differencing, the Cauchy estimator attains the same goal by using the nonlinear transformation of y_{t-1} to $\mathrm{sgn}(y_{t-1})$. Yet it is unclear how the constant term may be treated in this approach when it is applied to short dynamic panels.

REFERENCES

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

Karanasos, M. (1998) A new method for obtaining the autocovariance of an ARMA model: An exact form solution. *Econometric Theory* 14, 622–640.

Paparoditis, E. & D.N. Politis (2000) Large-sample inference in the general AR(1) model. *Test* 9, 487–509.

Phillips, P.C.B. & T. Magdalinos (2005) Limit Theory for Moderate Deviations from a Unit Root. Cowles Foundation Discussion paper 1471, Yale University.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29, 1188–1197.