Econometric Theory, **15**, 1999, 1–23. Printed in the United States of America.

UNIT ROOT TESTS BASED ON ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION

DONG WAN SHIN AND BEONG SOO SO
Ewha Womans University

Adaptive maximum likelihood estimators of unit roots in autoregressive processes with possibly non-Gaussian innovations are considered. Unit root tests based on the adaptive estimators are constructed. Limiting distributions of the test statistics are derived, which are linear combinations of two functionals of Brownian motions. A Monte Carlo simulation reveals that the proposed tests have improved powers over the classical Dickey–Fuller tests when the distribution of the innovation is not close to normal. We also compare the proposed tests with those of Lucas (1995, *Econometric Theory* 11, 331–346) based on *M*-estimators.

1. INTRODUCTION

Testing for autoregressive unit roots has attracted a great deal of attention since the influential works of Fuller (1976, 1996) and Dickey and Fuller (1979). Various extensions have been and are currently being made in several directions. One research direction is to widen the applicability of the unit root tests under broader correlation structures for the innovations. Phillips (1987) and Phillips and Perron (1988) developed tests that are applicable under weakly dependent innovations. Said and Dickey (1984, 1985), Pantula and Hall (1991), and others constructed tests that work for the innovations of stationary autoregressive moving averages (ARMA). Another direction is to improve powers of tests as studied by Pantula et al. (1994), Park and Fuller (1995), Shin and So (1997), and others. The power improvements are achieved by a proper handling of the initial observation or by considering symmetric estimation of the unit root parameter.

All of the above-mentioned tests are based on the ordinary least-squares estimator (OLSE), the Gaussian maximum likelihood estimator, or variants of it, which is optimal if we assume normality of the innovation. However, if the distribution of the innovation is not normal, one can improve the power performances of the tests. The OLSE-based tests have the same limiting distributions regardless of the distribution of the innovation provided the distribution has a

We greatly appreciate the editor, the anonymous associate editor, the three anonymous referees, Professor Yoon-Jae Whang, and Professor Mi-Jin Kim for their valuable comments, which led to a substantial improvement of the previous version of the article. This research was supported by the grant for promotion of scientific research in women's universities. Address correspondence to: Dong Wan Shin, Department of Statistics, Ewha Womans University, Seoul, South Korea, 120-750.

© 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

finite variance. However, when the innovation distribution has fat tails, the OLSE-based tests have lower powers compared with tests that take advantage of non-normality. Lower power implies that the OLSE-based tests tend to give false evidence for unit roots even though the series under study may not have a stochastic trend.

There is abundant evidence that certain economic variables have non-normal distributions, as discussed by Judge et al. (1985, Chap. 20), Fama (1965), and others. Also Franses and Haldrup (1994), Shin et al. (1996), and Lucas (1995b) pointed out that many economic time series are subject to sudden changes. They handled the sudden changes by introducing the concepts of outliers. In fact, the occurrences of sudden changes or outliers are the rule rather than the exception in empirical data. In these cases, distributions of the innovations are non-normal, and thus non-Gaussian maximum likelihood estimations are required. Cox and Llatas (1991) studied non-Gaussian *M*-estimation for unit roots in AR(1) processes and derived the distribution of the *M*-estimator. Their result, despite its theoretical elegance, is not of direct practical use for unit root testing for two reasons. One reason is that they assumed that the true score function is known, which is unrealistic because in practice the true distribution is rarely known. The other reason is that their expression of the distribution depends on this unknown score function. Lucas (1995a) constructed unit root tests based on *M*-estimator that also require specification of a score-type function. Herce (1996) studied the asymptotic theory of the estimator obtained from minimizing the sum of absolute deviations.

In this paper, we develop a new class of unit root tests based on adaptive maximum likelihood estimation. We observe that the score function can be estimated by a nonparametric density estimator of the innovation distribution based on ordinary least-squares residuals. In maximum likelihood estimation, the unknown density function is replaced by its estimator. Under the assumptions stated in the present paper, the resulting estimator is asymptotically equivalent to the maximum likelihood estimator computed under the known density function. In the literature, this approach is called an adaptive procedure because it works well under a broad class of unknown innovation distributions. Some results on adaptive estimation in time series models are available by Kreiss (1987) and Gassiat (1993) for the stationary AR(1) model, Koul and Pflug (1990) for the explosive AR(1) model, and Hidalgo (1992) and Steigerwald (1992) for linear regression with ARMA errors. Drost, Klaassen, and Werker (1997) obtained general LAN (locally asymptotic normal) results that can be applied to adaptive estimation in stationary time series models such as ARMA and ARCH models. Also, Hodgson (1996, 1998a, 1998b) applied adaptive maximum likelihood estimation to a class of multivariate time series models: cointegrating regressions, error correction models, and reduced rank error correction models. Jeganathan (1995) developed general asymptotic frameworks that can be applied to adaptive estimation in nonstationary time series models such as AR models with autoregressive unit roots.

The asymptotic distributions of the proposed adaptive tests are linear combinations of two well-known distributions: the distributions of the OLSE-based

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

unit root tests of Dickey and Fuller (1979) and the standard normal distribution. The weights are functions of one parameter that can be consistently estimated from the density estimates. This permits us to construct tables of null percentiles of the test statistics. Hence our tests are applicable under a wide class of innovations as in Assumption 1 in Section 2. Our results do not assume the knowledge of the density or the score function of the innovation distribution.

In a Monte-Carlo study, the tests based on the adaptive estimator are shown to have improved powers over the classical Dickey–Fuller tests, especially when the distribution of the innovation differs significantly from the normal distribution. Also, the Monte-Carlo study reveals that the adaptive tests have better powers than the *M*-estimator-based tests of Lucas (1995a) if the quasi score of the *M*-estimator differs significantly from the true score.

The remainder of the paper is organized as follows. In Section 2, adaptive estimation for the unit root in an AR(1) model with symmetric errors is introduced. In Section 3, tests for a unit root based on the adaptive estimators are developed. Also tables of null percentiles of the test statistics are constructed. In Section 4, an extension to an AR(p) model is made. In Section 5, a Monte-Carlo simulation is conducted. In Section 6, some discussion is offered. In the Appendix, proofs of theoretical results are provided.

2. ADAPTIVE MAXIMUM LIKELIHOOD ESTIMATION

Consider an AR(1) model

$$y_t = \rho y_{t-1} + e_t, \qquad t = 2, \ldots, n,$$

where $\{y_t\}_{t=1}^n$ is a set of observations, e_t is a sequence of innovations, and ρ is an AR parameter of interest. We are interested in testing the unit root hypothesis

$$H_0 : \rho = 1$$

against the stationarity alternative $H_1 : \rho < 1$. Let $f(\cdot)$ be the probability density function of e_t, let $\psi(x) = -f'(x)/f(x)$ be the score function, and let $I(f) = \int_{-\infty}^{\infty} (f')^2/f\, dx$ be the Fisher information number. We assume the following conditions for the innovation.

Assumption 1.

(i) e_t are independent and identically distributed (i.i.d.) random variables with $E(e_t) = 0$, $\sigma^2 = \mathrm{var}(e_t) < \infty$, $E(|e_t|^{2+\delta}) < \infty$ for some $\delta > 0$, and $0 < I(f) < \infty$,
(ii) f is differentiable and $\psi(x)$ is first-order Lipschitz, that is, there is a constant $M < \infty$ such that $|\psi'(x) - \psi'(y)| \leq M|x - y|$ for all $x, y \in \mathbb{R}$,
(iii) f is symmetric, that is, $f(x) = f(-x)$, $f(x) > 0$ for all $x \in (-\infty, \infty)$.

Conditions (i) and (ii) are used to characterize the limiting distribution of the maximum likelihood estimator. Condition (ii) is a Lipschitz condition that is essential in establishing the asymptotic equivalence of the maximum likelihood estimator and the adaptive estimator. We impose the symmetry condition (iii) in

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

our study for two reasons. One reason is that (iii) simplifies the derivation of the asymptotic equivalence of the maximum likelihood estimator and the adaptive estimator. The other, more important, reason is that, if the innovation is symmetrically distributed, sizes of the tests based on symmetric density estimators are better than those based on the density estimators that are not necessarily symmetric. However, in a Monte-Carlo experiment in Section 5, it is shown that the tests based on the symmetric density estimator still work well for a skewed distribution.

Usually the autoregressive coefficient ρ is estimated by the OLSE

$$\hat{\rho}_L = \sum_{t=2}^{n} y_{t-1} y_t \Big/ \sum_{t=2}^{n} y_{t-1}^2,$$

which maximizes the conditional Gaussian log-likelihood $\Sigma_{t=2}^{n} \log f_N(y_t - \rho y_{t-1})$, where $f_N(x) = (2\pi\sigma^2)^{-1/2} \exp(-2^{-1}x^2/\sigma^2)$ is the Gaussian density. As was shown by Dickey and Fuller (1979), the limiting distribution of $n(\hat{\rho}_L - 1)$ under $\rho = 1$ is given by

$$n(\hat{\rho}_L - 1) \Rightarrow \int_0^1 W(r)\, dW(r) \Big/ \int_0^1 W^2(r)\, dr,$$

where $\Rightarrow$ denotes convergence in distribution and $W(r)$ is a standard Brownian motion on $[0,1]$.

If the true density f is known, by maximizing the true log-likelihood $\Sigma_{t=2}^{n} \log f(y_t - \rho y_{t-1})$, one can construct a better estimator than the OLSE $\hat{\rho}_L$. One-step application of the Newton–Raphson procedure for maximization of the true likelihood with an initial estimator, say, $\tilde{\rho}$, gives the following one-step maximum likelihood estimator:

$$\hat{\rho}_M = \tilde{\rho} + \Delta_n(\tilde{\rho}) \Big/ \left\{ I(f) \sum_{t=2}^{n} y_{t-1}^2 \right\}, \tag{1}$$

where

$$\Delta_n(\rho) = \sum_{t=2}^{n} \psi(e_t(\rho)) y_{t-1}, \qquad e_t(\rho) = y_t - \rho y_{t-1}.$$

Cox and Llatas (1991) and Lucas (1995a), under regularity conditions similar to (i) and (ii) of Assumption 1, derived the limiting distribution of $n(\hat{\rho}_M - 1)$. Their estimators are slightly different from estimator (1) in that they are solutions of the score equations. However, their estimators are asymptotically equivalent to estimator (1) if the true score function is used in the M-estimation. In the following theorem, we derive expressions for the limiting distributions of $n(\hat{\rho}_M - 1)$ and its pivotal statistic that are useful for unit root testing.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

THEOREM 1. *Under the regularity conditions of Assumption* 1 *and* $n(\tilde{\rho} - 1) = O_p(1)$, *we have*

(i) $n(\hat{\rho}_M - 1) \Rightarrow D(\lambda)$,
(ii) $\{I(f)\sum_{t=2}^{n} y_{t-1}^2\}^{1/2}(\hat{\rho}_M - 1) \Rightarrow T(\lambda)$,

where

$$D(\lambda) = \lambda T(\lambda) \Big/ \left[\int_0^1 W^2(r)\, dr \right]^{1/2},$$

$$T(\lambda) = \lambda \int_0^1 W(r)\, dW(r) \Big/ \left\{ \int_0^1 W^2(r)\, dr \right\}^{1/2} \tag{2}$$

$$+ (1 - \lambda^2)^{1/2} \int_0^1 W(r)\, dV(r) \Big/ \left\{ \int_0^1 W^2(r)\, dr \right\}^{1/2},$$

$$\lambda = \{\sigma^2 I(f)\}^{-1/2},$$

and $W(r)$ *and* $V(r)$, $0 \le r \le 1$ *are two independent standard Brownian motions*.

By the Cauchy–Schwartz inequality, we obtain $\sigma^2 I(f) \ge 1$ with equality if and only if f is the normal density. The quantity λ can be interpreted as the correlation coefficient between e_t and $\psi(e_t)$. In case $\lambda = 1$, the distributions $D(\lambda)$ and $T(\lambda)$ reduce to those of Dickey and Fuller (1979). We can interpret λ as an intrinsic measure of normality for the innovation e_t. The closer λ is to zero, the more the distribution of the innovations differs from the normal distribution. For innovations with infinite variances, $\lambda = 0$ because $\sigma^2 = \infty$ and $I(f) < \infty$.

In the real world, the density $f(x)$ is usually unknown, and hence the one-step maximum likelihood estimator $\hat{\rho}_M$ cannot be evaluated. However, if a good estimator $\hat{f}(x)$ of $f(x)$ is available, we can construct an estimator like

$$\hat{\rho}_A = \tilde{\rho} + \hat{\Delta}_n(\tilde{\rho}) \Big/ \left\{ \hat{I}(\hat{f}) \sum_{t=2}^{n} y_{t-1}^2 \right\}, \tag{3}$$

where $\hat{\Delta}_n(\tilde{\rho})$ and $\hat{I}(\hat{f})$ are estimators of $\Delta_n(\tilde{\rho})$ and $I(f)$, respectively, constructed from an appropriate density estimator $\hat{f}(x)$. We call the estimator $\hat{\rho}_A$ an adaptive estimator of ρ. For an initial estimator $\tilde{\rho}$, we use $\tilde{\rho} = 1$ as Said and Dickey (1985) did in their estimation of unit roots in the ARMA models. The main reason for using the null value $\rho = 1$ as the initial estimator is that, as is discussed at the end of Section 5, it provides us with stable sizes of the unit root tests based on the adaptive estimator.

Usually, the density $f(x)$ is estimated by the kernel method which is well summarized in Silverman (1986). The kernel estimator of $f(x)$ is defined by

$$\hat{f}_n^s(x) = 2^{-1}\{\hat{f}_n(x) + \hat{f}_n(-x)\},$$

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

where

$$\hat{f}_n(x) = (nh_n)^{-1} \sum_{t=2}^{n} K(h_n^{-1}\{x - e_t(\tilde{\rho})\}), \tag{4}$$

K is a function called kernel satisfying $\int_{-\infty}^{\infty} K(x)\,dx = 1$, $K(x) > 0$, and h_n is a sequence of positive numbers called the bandwidths. The function $\hat{f}_n(x)$ in (4) is the usual kernel-based density estimator under no symmetry assumption on $f(\cdot)$. The function $\hat{f}_n^s(x)$ is a symmetrized version of $\hat{f}_n(x)$. Because $\hat{f}_n^s(x)$ uses the information that $f(\cdot)$ is symmetric, when the error distribution is symmetric, $\hat{f}_n^s(x)$ is a better estimator than $\hat{f}_n(x)$ in the sense that $\hat{f}_n^s(x)$ gives more stable sizes for the adaptive unit root tests. The assumption of symmetry and the use of a symmetrized estimator are quite important to the proof that our estimator is adaptive. It is apparently much more difficult, if not impossible, to get an adaptivity result without the assumption of symmetric errors. We take the kernel and the bandwidth to be

$$K(x) = e^{-x}/(1 + e^{-x})^2 \tag{5}$$

and

$$h_n = k_2^{-2/5} \left\{ \int_{-\infty}^{\infty} K(x)^2\,dx \right\}^{1/5} \left\{ \int_{-\infty}^{\infty} f''(x)^2\,dx \right\}^{-1/5} n^{-1/5}, \tag{6}$$

respectively, where $k_2 = \int_{-\infty}^{\infty} x^2 K(x)\,dx$. We have taken the logistic kernel (5) for two reasons. First, density estimation is not too sensitive to the type of kernel (Silverman, 1986, p. 43). Second, as noted by Schick (1987), it simplifies asymptotic analysis because of simple bounds of all derivatives of it. We have chosen the bandwidth (6) because it asymptotically minimizes the mean integrated square error $E\int_{-\infty}^{\infty}\{\hat{f}_n(x) - f(x)\}^2\,dx$, (Silverman, 1986, p. 40). Note that the bandwidth h_n involves the integral of $(f'')^2$, which is unknown. Later, in Section 5, we suggest a way to deal with this problem.

The terms $\Delta_n(\rho)$ and $I(f)$ are now estimated by

$$\hat{\Delta}_n^s(\tilde{\rho}) = \sum_{t=2}^{n} \hat{\psi}_{nt}^s(e_t(\tilde{\rho});\tilde{\rho}) y_{t-1} \quad \text{and} \quad \hat{I}_n^s = \sum_{t=2}^{n} \{\hat{\psi}_{nt}^s(e_t(\tilde{\rho});\tilde{\rho})\}^2/n, \tag{7}$$

respectively, where

$$\hat{\psi}_{nt}^s(x;\rho) = -\{\partial \hat{f}_{nt}^s(x;\rho)/\partial x\}/\hat{f}_{nt}^s(x;\rho), \tag{8}$$

$$\hat{f}_{nt}^s(x;\rho) = \{\hat{f}_{nt}(x;\rho) + \hat{f}_{nt}(-x;\rho)\}/2, \tag{9}$$

$$\hat{f}_{nt}(x;\rho) = s_n + (nh_n)^{-1} \sum_{i=2, i\neq t}^{n} K(h_n^{-1}\{x - e_i(\rho)\}), \tag{10}$$

and $s_n > 0$ is a sequence of positive real numbers. The reason for using the delete-one-version $\hat{f}_{nt}^s(x)$ instead of $\hat{f}_n^s(x)$ is for technical simplicity, because it

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

allows independence of $\hat{f}^s_{nt}(x)$ and e_t, as noted by Schick (1987). The term s_n in (10) is included to prevent $\hat{f}_{nt}(x)$ from being close to zero, and it guarantees better numerical performance of the score function estimator $\hat{\psi}^s_{nt}(x) = -\hat{f}^{s\prime}_{nt}(x)/\hat{f}^s_{nt}(x)$. We now, in (3) and (5)–(10), have fully implemented an adaptive estimation procedure. In Theorem 2, we show that the adaptive estimator and the maximum likelihood estimator are asymptotically equivalent.

THEOREM 2. *Assume* $s_n = o(1)$, $h_n = o(1)$, *and* $h_n^{-4} s_n^{-2} n^{-1} = o(1)$. *Then, under the regularity conditions of Assumption* 1 *and* $\rho = 1$, *we have* $n(\hat{\rho}_A - \hat{\rho}_M) = o_p(1)$.

Theorem 2 implies that $n(\hat{\rho}_A - 1)$ and $n(\hat{\rho}_M - 1)$ have the same limiting distribution, which is given in (2). Because $\lambda = \{\sigma^2 I(f)\}^{-1/2}$ can be consistently estimated from the estimated density, the null distribution of $n(\hat{\rho}_A - 1)$ can also be estimated. These facts enable us to construct a class of adaptive unit root tests as discussed in the following section.

3. TESTING FOR A UNIT ROOT

To test the unit root hypothesis, we consider three estimators corresponding to the regression models

$$y_t = \hat{\rho}_A y_{t-1} + \hat{e}_t, \tag{11}$$

$$y_t - \bar{y} = \bar{\rho}_A (y_{t-1} - \bar{y}) + \bar{e}_t, \tag{12}$$

and

$$y_t - \bar{\mu} - \bar{\gamma}(t/n - 0.5) = \rho^*_A [y_{t-1} - \bar{\mu} - \bar{\gamma}\{(t-1)/n - 0.5\}] + e^*_t, \tag{13}$$

where $\hat{e}_t, \bar{e}_t, e^*_t$ are the corresponding residuals, $\bar{y} = \sum_{t=1}^n y_t/n$, and $(\bar{\mu}, \bar{\gamma})$ is the OLSE obtained by regressing y_t on $(1, t - n/2)$, $t = 2, \ldots, n$. The coefficients $\bar{\rho}_A$ and ρ^*_A are estimated by the adaptive procedure (3) and (5)–(10) using the mean-adjusted observations $y_{t,\mu} = y_t - \bar{y}$ and the trend-adjusted observations $y_{t,\tau} = y_t - \bar{\mu} - \bar{\gamma}(t/n - 0.5)$, respectively, instead of y_t, $t = 2, \ldots, n$. For testing the unit root hypothesis, the statistics $n(\hat{\rho}_A - 1), n(\bar{\rho}_A - 1), n(\rho^*_A - 1)$ can be used. Also, the pivotal statistics

$$\hat{\tau}_A = (\hat{\rho}_A - 1)/\mathrm{se}(\hat{\rho}_A), \qquad \bar{\tau}_A = (\bar{\rho}_A - 1)/\mathrm{se}(\bar{\rho}_A), \qquad \tau^*_A = (\rho^*_A - 1)/\mathrm{se}(\rho^*_A)$$

can be used, where

$$\mathrm{se}(\hat{\rho}_A) = (\hat{I}^s_n \Sigma y^2_{t-1})^{-1/2}, \qquad \mathrm{se}(\bar{\rho}_A) = (\hat{I}^s_n \Sigma y^2_{t-1,\mu})^{-1/2},$$

$$\mathrm{se}(\rho^*_A) = (\hat{I}^s_n \Sigma y^2_{t-1,\tau})^{-1/2}$$

are the standard errors of the estimators. The limiting distributions of the test statistics are given in Theorem 3.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

THEOREM 3. *Under the conditions of Theorem* 2,

(i) $n(\hat{\rho}_A - 1) \Rightarrow D(\lambda)$, $\quad n(\bar{\rho}_A - 1) \Rightarrow D_\mu(\lambda)$, $\quad n(\rho_A^* - 1) \Rightarrow D_\tau(\lambda)$,

where $D_\mu(\lambda)$ *and* $D_\tau(\lambda)$ *are the same as* $D(\lambda)$ *in* (2) *except, instead of* $W(r)$ *in* $D(\lambda)$,

$$W_\mu(r) = W(r) - \int_0^1 W(s)\, ds$$

and

$$W_\tau(r) = W(r) - 4\left\{ \int_0^1 W(s)\, ds - 1.5 \int_0^1 sW(s)\, ds \right\} + 6r\left\{ \int_0^1 W(s)\, ds - 2\int_0^1 sW(s)\, ds \right\}$$

are used in $D_\mu(\lambda)$ *and* $D_\tau(\lambda)$, *respectively*,

(ii) $\hat{\tau}_A \Rightarrow T(\lambda)$, $\quad \bar{\tau}_A \Rightarrow T_\mu(\lambda)$, $\quad \tau_A^* \Rightarrow T_\tau(\lambda)$,

where $T_\mu(\lambda)$ *and* $T_\tau(\lambda)$ *are the same as* $T(\lambda)$ *in Theorem* 1 *except, instead of* $W(r)$ *in* $T(\lambda)$, $W_\mu(r)$ *and* $W_\tau(r)$ *are used in* $T_\mu(\lambda)$ *and* $T_\tau(\lambda)$, *respectively*.

The expressions of the limiting distributions in Theorem 3 are similar to those of the *M*-estimator-based tests of Lucas (1995a, Theorem 1) and the LAD-estimator-based tests of Herce (1996, Theorem 1). In their results, the parameter λ is the correlation coefficient between the innovation and the quasi score (the *M*-type score for the *M*-estimator of Lucas; the sign for the LAD estimator of Herce). However, because of the use of the quasi scores, the limiting distributions of their $n(\rho - 1)$-type tests depend on the other distributional parameters. In our approach, the limiting distributions depend only on λ because a consistent estimator of the true score is used. The *M*-estimators and the LAD estimators of Lucas, Herce, and Cox and Llatas are robust in that their performances are not sensitive to outliers, because they discount wild observations in the case of heavy-tailed error distributions. However, when the error distributions are not close to those of the specified score functions, the robust methods may lose efficiency. Now, our adaptive method attains both robustness and efficiency by selecting a consistent estimator of the true score automatically. However, the adaptive method is not necessarily easy to implement.

Note that, in contrast to the limiting distributions of the regression biases $n(\hat{\rho}_A - 1)$, $n(\bar{\rho}_A - 1)$, and $n(\rho_A^* - 1)$, those of the pivotal statistics $\hat{\tau}_A$, $\bar{\tau}_A$, and τ_A^* do not degenerate to zero and converge to the standard normal distribution as λ approaches zero. The distributions $T(\lambda)$, $T_\mu(\lambda)$, and $T_\tau(\lambda)$ of the pivotal statistics are mixtures of the Dickey–Fuller distributions (DF) and a standard normal distribution (Z) independent of DF because they are of the form

$$\lambda \mathrm{DF} + (1 - \lambda^2)^{1/2} Z.$$

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

When σ^2 is infinite and $I(f) > 0$, $\lambda = 0$ and the distributions $T(\lambda)$, $T_\mu(\lambda)$, and $T_\tau(\lambda)$ become standard normal as Knight (1989) observed for the M-estimator-based tests. On the other hand, the limiting distributions of the OLSE-based pivotal statistics of Chan and Tran (1989) and Phillips (1990) are not normal. The fact that $D(\lambda) \to 0$ as $\lambda \to 0$ (or $\sigma^2 \to \infty$) means that, in the infinite variance case, the OLSE gets infinitely less efficient than the M-estimator or the adaptive estimator. This agrees with the result of Knight (1989), who proved that, in the infinite variance model, the rate of convergence of the OLSE is n^{-1} whereas that of the M-estimator is $n^{-(1/\alpha+1/2)}$ when the error distribution belongs to the domain of attraction of a stable distribution with index $0 < \alpha < 2$.

To test the unit root hypothesis, we need percentiles of the null distributions. For finite sample sizes, the distributions of the test statistics depend on the unknown distribution $f(\cdot)$ of the innovation. However, by Theorems 1 and 3, the limiting distributions of the test statistics depend on the unknown distribution only through λ. Hence finite-sample distributions of the statistics $n(\hat{\rho}_A - 1)$ and $\hat{\tau}_A$ should be close to those of

$$n\lambda C_n(\lambda) \Big/ \sum_{t=2}^{n} y_{t-1}^2 \quad \text{and} \quad C_n(\lambda) \Big/ \left(\hat{\sigma} \sum_{t=2}^{n} y_{t-1}^2 \right)^{1/2}, \tag{14}$$

respectively, where

$$C_n(\lambda) = \left\{ \lambda \sum_{t=2}^{n} y_{t-1} e_t + (1-\lambda^2)^{1/2} \sum_{t=2}^{n} y_{t-1} \delta_t \right\},$$

$$\hat{\sigma}^2 = (n-2)^{-1} \sum_{t=2}^{n} (y_t - \hat{\rho}_L y_{t-1})^2,$$

$y_t = y_{t-1} + e_t$, $\{e_t\}$ is an independent $N(0,1)$ sequence, and $\{\delta_t\}$ is an independent $N(0,1)$ sequence independent of $\{e_t\}$. Therefore, we can construct empirical percentiles for various values of λ between (0,1) by a Monte-Carlo simulation of (14). Using 100,000 replications of (14) with $y_0 = 0$, empirical 1%, 5%, 10% left percentiles are computed. Independent (e_t, δ_t) were used for different combinations of (n, λ). Also empirical percentiles of the other statistics $n(\bar{\rho}_A - 1), \bar{\tau}_A$, $n(\rho_A^* - 1), \tau_A^*$ are simulated by (14) using $y_{t,\mu}$ and $y_{t,\tau}$, respectively, instead of y_t, and proper adjustment for degrees of freedom for $\hat{\sigma}^2$. Because the statistics $n(\hat{\rho}_A - 1)$ and $\hat{\tau}_A$ are rarely used in practice, we only report the percentiles of the mean-adjusted statistics and the trend-adjusted statistics in Table 1.

We now state our test procedure. Compute the test statistics $n(\hat{\rho}_A - 1)$, $n(\bar{\rho}_A - 1)$, $n(\rho_A^* - 1), \hat{\tau}_A, \bar{\tau}_A, \tau_A^*$. Also compute the corresponding estimates of $\hat{\lambda} = \text{Min}[\{\hat{\sigma}^2 \hat{I}(\hat{f})\}^{-1/2}, 1]$ of λ in the estimated models (11), (12), and (13). From Table 1, using interpolation, find the percentiles corresponding to $\hat{\lambda}$. Compare the test statistics with the corresponding percentiles. Reject the unit root hypothesis if the statistics are smaller than the tabulated percentiles.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

TABLE 1. Left percentage points of null distributions of the adaptive test statistics for the hypothesis of unit root $H_0: \rho = 1$ in the model $y_t = \rho y_{t-1} + e_t$

	Mean adjusted						Mean and trend adjusted					
	$n(\bar{\rho}_A - 1)$			$\bar{\tau}_A$			$n(\rho_A^* - 1)$			τ_A^*		
	Probability of smaller value											
λ	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
						$n = 25$						
1.0	−17.1	−12.5	−10.2	−3.7	−3.0	−2.6	−22.6	−18.0	−15.6	−4.3	−3.6	−3.2
0.8	−12.7	−8.9	−7.2	−3.6	−2.8	−2.4	−16.9	−12.9	−10.9	−4.1	−3.3	−3.0
0.6	−8.5	−5.9	−4.6	−3.4	−2.6	−2.2	−11.3	−8.4	−7.1	−3.9	−3.0	−2.6
0.4	−5.0	−3.4	−2.6	−3.2	−2.3	−1.9	−6.8	−4.8	−3.9	−3.6	−2.7	−2.2
0.2	−2.2	−1.4	−1.1	−2.9	−2.1	−1.7	−2.9	−2.0	−1.6	−3.1	−2.3	−1.8
0.0	−0.0	−0.0	−0.0	−2.6	−1.7	−1.3	−0.0	−0.0	−0.0	−2.7	−1.8	−1.4
						$n = 50$						
1.0	−18.8	−13.2	−10.7	−3.5	−2.9	−2.6	−25.4	−19.7	−16.8	−4.1	−3.5	−3.2
0.8	−13.5	−9.3	−7.4	−3.4	−2.7	−2.4	−18.4	−13.8	−11.5	−3.9	−3.2	−2.9
0.6	−9.0	−6.0	−4.7	−3.2	−2.5	−2.2	−12.0	−8.9	−7.4	−3.7	−2.9	−2.5
0.4	−5.1	−3.4	−2.6	−3.0	−2.3	−1.9	−7.0	−5.0	−4.0	−3.3	−2.6	−2.2
0.2	−2.2	−1.4	−1.0	−2.7	−2.0	−1.6	−2.9	−2.0	−1.6	−2.9	−2.2	−1.8
0.0	−0.0	−0.0	−0.0	−2.4	−1.7	−1.3	−0.0	−0.0	−0.0	−2.5	−1.7	−1.3

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

						$n = 100$						
1.0	−19.4	−13.5	−10.9	−3.5	−2.9	−2.6	−27.3	−20.5	−17.4	−4.0	−3.5	−3.1
0.8	−14.1	−9.6	−7.6	−3.4	−2.7	−2.4	−19.3	−14.3	−12.0	−3.8	−3.2	−2.9
0.6	−9.1	−6.1	−4.8	−3.2	−2.5	−2.1	−12.5	−9.1	−7.6	−3.5	−2.9	−2.5
0.4	−5.2	−3.4	−2.6	−3.0	−2.3	−1.9	−7.1	−5.0	−4.1	−3.2	−2.5	−2.2
0.2	−2.2	−1.4	−1.0	−2.7	−2.0	−1.6	−2.9	−2.0	−1.6	−2.8	−2.1	−1.7
0.0	−0.0	−0.0	−0.0	−2.4	−1.7	−1.3	−0.0	−0.0	−0.0	−2.4	−1.7	−1.3
						$n = 250$						
1.0	−20.2	−13.9	−11.1	−3.5	−2.9	−2.6	−28.5	−21.3	−17.9	−4.0	−3.4	−3.1
0.8	−14.3	−9.7	−7.7	−3.3	−2.7	−2.4	−20.1	−14.6	−12.2	−3.8	−3.2	−2.8
0.6	−9.2	−6.2	−4.9	−3.1	−2.5	−2.1	−12.9	−9.3	−7.7	−3.5	−2.8	−2.5
0.4	−5.2	−3.4	−2.6	−2.9	−2.2	−1.9	−7.2	−5.1	−4.1	−3.2	−2.5	−2.1
0.2	−2.2	−1.4	−1.0	−2.6	−2.0	−1.6	−2.9	−2.0	−1.6	−2.8	−2.1	−1.7
0.0	−0.0	−0.0	−0.0	−2.3	−1.7	−1.3	−0.0	−0.0	−0.0	−2.4	−1.7	−1.3
						$n = 5{,}000$						
1.0	−20.5	−14.1	−11.2	−3.4	−2.9	−2.6	−29.1	−21.6	−18.3	−4.0	−3.4	−3.1
0.8	−14.4	−9.8	−7.7	−3.3	−2.7	−2.4	−20.4	−14.9	−12.4	−3.7	−3.1	−2.8
0.6	−9.3	−6.2	−4.9	−3.1	−2.5	−2.1	−13.0	−9.4	−7.7	−3.5	−2.8	−2.5
0.4	−5.3	−3.4	−2.6	−2.9	−2.2	−1.9	−7.3	−5.1	−4.1	−3.1	−2.5	−2.1
0.2	−2.2	−1.4	−1.0	−2.6	−1.9	−1.6	−2.9	−2.0	−1.6	−2.7	−2.1	−1.7
0.0	−0.0	−0.0	−0.0	−2.3	−1.7	−1.3	−0.0	−0.0	−0.0	−2.3	−1.7	−1.3

Note: This table is constructed from a Monte-Carlo simulation with 100,000 replications; $\lambda = \{\sigma^2 I(f)\}^{-1/2}$ where $\sigma^2 = \mathrm{var}(e_t)$, f is the probability density function of e_t, and $I(f)$ is the Fisher information number of f.

An alternative estimator of λ can be obtained by estimating σ^2 from the density estimator

$$\tilde{\sigma}^2 = \int_{-\infty}^{\infty} x^2 \hat{f}(x)\, dx. \tag{15}$$

Then $\tilde{\lambda} = (\tilde{\sigma}^2 \hat{I}(\hat{f}))^{-1/2}$ is an estimator of λ that is always less than or equal to one.

4. AN EXTENSION

An extension of our approach to a pth-order autoregression is obvious. Consider an AR(p) model

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + e_t.$$

To estimate the possible unit root, we reparametrize the model into

$$y_t = \rho y_{t-1} + \beta_1 z_{t-1} + \cdots + \beta_{p-1} z_{t-p+1} + e_t,$$

where $z_t = y_t - y_{t-1}$ and β_k, $k = 1, \ldots, p-1$ are linear combinations of α_k, $k = 1, \ldots, p$. The adaptive estimator of $\boldsymbol{\theta} = (\rho, \beta_1, \ldots, \beta_{p-1})'$ is given by

$$\hat{\boldsymbol{\theta}}_A = \tilde{\boldsymbol{\theta}}_n + \left(\sum_{t=p+1}^{n} \mathbf{X}_t \mathbf{X}_t' \right)^{-1} \hat{\Delta}_n^s(\tilde{\boldsymbol{\theta}}_n)/\hat{I}_n^s, \tag{16}$$

where $\mathbf{X}_t = (y_{t-1}, z_{t-1}, \ldots, z_{t-p+1})'$, $\tilde{\boldsymbol{\theta}}_n = (1, \tilde{\beta}_1, \ldots, \tilde{\beta}_{p-1})'$, and $(\tilde{\beta}_1, \ldots, \tilde{\beta}_{p-1})'$ is the OLSE in the regression model

$$z_t = \beta_1 z_{t-1} + \cdots + \beta_{p-1} z_{t-p+1} + e_t.$$

Also,

$$\hat{\Delta}_n^s(\tilde{\boldsymbol{\theta}}_n) = \sum_{t=p+1}^{n} \hat{\psi}_{nt}^s(e_t(\tilde{\boldsymbol{\theta}}_n); \tilde{\boldsymbol{\theta}}_n) \mathbf{X}_t$$

and $\hat{I}_n^s$ are estimators of $\Delta_n(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} \psi(e_t(\boldsymbol{\theta})) \mathbf{X}_t$ and $I(f)$, respectively, based on the residuals

$$e_t(\tilde{\boldsymbol{\theta}}_n) = z_t - \tilde{\beta}_1 z_{t-1} - \cdots - \tilde{\beta}_{p-1} z_{t-p+1}, \qquad t = p+1, \ldots, n.$$

Let $\hat{\rho}_A$ be the first component of $\hat{\boldsymbol{\theta}}_A$. The pivotal statistic of $\hat{\rho}_A$ is

$$\hat{\tau}_A = (\hat{\rho}_A - 1)/\mathrm{se}(\hat{\rho}_A), \qquad \mathrm{se}(\hat{\rho}_A) = (\hat{I}_n^s / c_{11})^{-1/2},$$

where c_{11} is the (1,1) element of $(\Sigma \mathbf{X}_t \mathbf{X}_t')^{-1}$. Define the mean-adjusted statistics $\bar{\rho}_A$ and $\bar{\tau}_A$ and the trend-adjusted statistics ρ_A^* and τ_A^* similarly. The limiting distributions of the test statistics are given in Theorem 4.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

THEOREM 4. *Assume that the characteristic roots of* $m^{p-1} - \beta_1 m^{p-2} - \cdots - \beta_{p-1} = 0$ *lie inside the unit circle. Then, under the conditions of Theorem* 2, *we have*

(i) $n(\hat{\rho}_A - 1) \Rightarrow cD(\lambda)$, $n(\bar{\rho}_A - 1) \Rightarrow cD_\mu(\lambda)$, $n(\rho_A^* - 1) \Rightarrow cD_\tau(\lambda)$,
(ii) $\hat{\tau}_A \Rightarrow T(\lambda)$, $\bar{\tau}_A \Rightarrow T_\mu(\lambda)$, $\tau_A^* \Rightarrow T_\tau(\lambda)$,

where $c = (1 - \beta_1 - \cdots - \beta_{p-1})$.

5. A MONTE-CARLO STUDY

In this section, we compare the adaptive tests with the OLSE-based tests of Dickey and Fuller (1979) and the *M*-estimator-based tests of Lucas (1995a) for AR(1) models. We select four distributions with the relevant parameters being

D1: $N(0,1)$,	$I(f) = 1$,	$\lambda = 1$,
D2: $t(3)$,	$I(f) = \frac{2}{3}$,	$\lambda = .796$,
D3: $0.5N(-3,1) + 0.5N(3,1)$,	$I(f) = .963$,	$\lambda = .322$,
D4: $\chi_1^2 - 1$,	$I(f) = \infty$,	$\lambda = 0$,

where $N(a,b)$ denotes the normal distribution with mean a and variance b, $t(k)$ denotes the t-distribution with k degrees of freedom, and χ_1^2 is the chi-square distribution with one degree of freedom. The three distributions D1, D2, and D3 are symmetric, and the last distribution D4 is skewed. The distribution D2 shows a moderate departure from the normal distribution in that λ is around 0.7; D3 shows a substantial departure from the normal distribution while retaining symmetry; D4 is not symmetric and reveals a big departure from normality.

We describe our choice of the bandwidth (6) in more detail because adaptive estimators are sensitive to the bandwidth. According to Silverman (1986, p. 50), the key ingredient in (6), $\int_{-\infty}^{\infty} f''(x)^2\, dx$ can be consistently approximated by

$$n^{-2}h_0^{-1} \sum_{i=2}^{n} \sum_{j=2}^{n} K^{(2)}(h_0^{-1}\{e_i(\tilde{\rho}) - e_j(\tilde{\rho})\}), \tag{17}$$

where h_0 is an initial bandwidth and $K^{(2)}(x) = \exp(-x^2/4)$. As discussed in Silverman (1986), expressions (6) and (17) can be iterated, starting with an initial value h_0. We observed, through a Monte-Carlo simulation, that two iterations are sufficient. We used $h_0 = 0.4$. We tried other initial bandwidths $h_0 = 0.1, 0.7$. The results are almost the same as those for $h_0 = 0.4$ as a result of the fact that iteration of (6) and (17) gives a stabilized new bandwidth for a wide range of the initial bandwidths h_0. The term s_n in (10) is chosen to be $0.001h_n^{1/3}$. However, the value 0.001 is not critical. In a Monte-Carlo experiment not reported here, it was shown that the estimator $\hat{\rho}_A$ is not sensitive to the value of s_n. Values ranging from $s_n = h_n^{1/3}$ to $s_n = 10^{-5}h_n^{1/3}$ give almost the same performance in the adaptive tests. For an estimator of λ, we used $\hat{\lambda} = \text{Min}[\hat{\sigma}^2\hat{I}(\hat{f})\}^{-1/2},1]$ because $\hat{\lambda}$ is simpler to compute than $\tilde{\lambda}$ of (15).

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

The data are generated according to $y_t = \rho y_{t-1} + e_t$, $t = 1, 2, \ldots, n$ with $n = 50$, 100; $\rho = 1, 0.95, 0.9, 0.8$; $y_0 = 0$. We consider two cases of adjustments: the mean-adjusted case and the trend-adjusted case. The mean-adjusted Dickey–Fuller tests $n(\bar{\rho}_L - 1)$ and $\bar{\tau}_L$ and the trend-adjusted Dickey–Fuller tests $n(\rho_L^* - 1)$ and τ_L^* are computed. For a quasi-score function of the *M*-estimators of Lucas (1995a), we selected the $t(5)$ distribution. The mean-adjusted and trend-adjusted *M*-estimators $\bar{\rho}_m$ and ρ_m^* of Lucas (1995a) are computed by solving the *M*-equations

$$\Sigma\psi(s^{-1}(y_{t,\mu} - \bar{\rho}_m y_{t-1,\mu}))y_{t-1,\mu} = 0$$

and

$$\Sigma\psi(s^{-1}(y_{t,\tau} - \rho_m^* y_{t-1,\tau}))y_{t-1,\tau} = 0,$$

respectively, where $\psi(x) = 6x/(5 + x^2)$ is the score function of the $t(5)$ distribution and s is the median absolute deviation of the OLS residuals divided by 0.6745. The $n(\hat{\rho} - 1)$-type test statistics based on the *M*-estimator are adjusted for nuisance parameters as in (8) of Lucas (1995a) to have the same limiting distributions as the Dickey–Fuller tests and are given by $n(\hat{\rho} - 1)m_\psi s_e/s_\psi$, where $\hat{\rho} = \bar{\rho}_m$ or ρ_m^*, $m_\psi = n^{-1}\Sigma\psi'(s^{-1}\hat{e}_t)$, $s_e^2 = n^{-1}\Sigma\hat{e}_t^2$, and $s_\psi^2 = n^{-1}\Sigma\{\psi(s^{-1}\hat{e}_t)\}^2 s^2$. The mean-adjusted and the trend-adjusted pivotal statistics of the *M*-estimators of Lucas (1995a) are $\hat{\tau}_m = (\hat{\rho}_m - 1)/\mathrm{se}(\bar{\rho}_m)$ and $\tau_m^* = (\rho_m^* - 1)/\mathrm{se}(\rho_m^*)$, respectively, where $\mathrm{se}(\bar{\rho}_m)$ is the standard error of $\bar{\rho}_m$ given by

$$\{\mathrm{se}(\bar{\rho}_m)\}^2 = \frac{s^2\Sigma\{\psi(s^{-1}(y_{t,\mu} - \bar{\rho}_m y_{t-1,\mu}))\}^2 y_{t-1,\mu}^2}{[\Sigma\psi'(s^{-1}(y_{t,\mu} - \bar{\rho}_m y_{t-1,\mu}))y_{t-1,\mu}^2]^2}$$

and $\mathrm{se}(\rho_m^*)$ is defined analogously. The nominal level is set to 5%. Corresponding critical values are from Fuller (1996, App. A) and Table 1.

In Table 2, rejection frequencies in 1,000 independent Monte-Carlo samples are reported. We see that sizes of the Dickey–Fuller tests and the adaptive tests do not deviate much from the nominal 5% level for all cases, whereas those of the *M*-estimator-based tests have size distortions in some cases, especially for the $\hat{\tau}$-statistics for D3.

One interesting point is that, for distributions D2–D4, the adaptive tests have significantly better powers than the Dickey–Fuller tests. For example, under D3, when $n = 100$, $\rho = 0.9$, we note that the trend-adjusted Dickey–Fuller test τ_L^* has power 20.4%, whereas the corresponding adaptive test has power 98.1%. For all innovations D1–D4, the Dickey–Fuller tests have about the same finite-sample powers against a given alternative. On the other hand, powers of the adaptive tests get higher as λ decreases.

For D3 and D4, the true scores are substantially different from the quasi score of the $t(5)$-based *M*-estimator. For these distributions, except in a few cases for D4, our adaptive tests have higher powers than the *M*-estimator-based tests of Lucas (1995a). However, for D1 and D2, which are relatively closer to $t(5)$ than D3 and D4, the adaptive tests do not outperform the tests of Lucas (1995a). Es-

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

pecially for D2 which is $t(3)$ and is close to $t(5)$, the tests of Lucas (1995a) have substantially higher powers than the adaptive tests. Even though our adaptive tests are constructed using the symmetric kernel estimator $f_n^s(x)$, they still work well for the skewed distribution D4 having sizes close to the nominal level 5% and higher powers than the other two tests.

Because the initial estimator $\tilde{\rho} = 1$ in (3) is consistent only under the null hypothesis, the adaptive tests based on the null value would have lower powers compared to adaptive tests based on consistent initial estimators. In a Monte-Carlo simulation similar to that used for constructing Table 2, we computed sizes and powers of the adaptive tests based on consistent initial estimators, the OLSE. We found undesirable size distortions for the OLSE-based adaptive tests for the non-normal innovations. For example, the empirical sizes (%) of the mean-adjusted test $\bar{\tau}_A$ for (D1, D2, D3, D4) are (4.3, 17.1, 23.7, 12.4) for $n = 50$ and are (5.1, 18.4, 13.5, 9.5) for $n = 100$. Also, for all $\rho = 0.95, 0.9, 0.8$, we found no significant power improvements for the OLSE-based adaptive tests over the null-based adaptive tests. However, when ρ is not close to one, the null-based adaptive tests seemed to lose some power. For example, when $\rho = 0.7$, $n = 50$, and the innovation is D1, we found the same size 4.3% for both the OLSE-based adaptive test $\bar{\tau}_A$ and the null-based adaptive test $\bar{\tau}_A$ and found smaller powers for the null-based test $\bar{\tau}_A$ (47.5%) compared to the OLSE-based test $\bar{\tau}_A$ (52.1%). From this analysis we may conclude that the null value $\rho = 1$ as the initial estimator provides us with more stable sizes than the OLSE, without big power losses when ρ is not much different from one.

6. DISCUSSION

In our research, we found that the adaptive estimators provide us with higher powers than the Dickey–Fuller tests based on the OLSE when the error distribution is different from a normal distribution, whereas our tests have comparable performances when the error distribution is near normal. Also, if the score function is suitably specified, *M*-estimation would give us tests with higher powers than adaptive estimation. However, if the score function is not well specified, the performance of the *M*-estimator-based tests would not be as good as that of the adaptive tests.

We have seen that when the innovation has heavy tails in the sense that $\lambda = \{\sigma^2 I(f)\}^{-1/2}$ is close to zero, the OLSE-based tests have significantly lower powers than the adaptive tests. By contrast, for heavy-tailed error distributions, *M*-estimators take full advantage of good leverage points, and this gives them an edge in terms of better efficiency than the OLSE. The adaptive estimator is even asymptotically better than the *M*-estimator because it guesses the true score.

We should mention that Beelders (1996) also independently obtained similar adaptive tests utilizing asymptotic theories from Jeganathan (1995). He obtained the results in a more general setting of ARMA processes and discussed some optimality properties of the adaptive estimator. In comparison to his tests, our

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

TABLE 2. Empirical sizes (%) and powers (%) of tests for $H_0: \rho = 1$ against $H_1: \rho < 1$

		Mean adjusted						Mean and trend adjusted					
		$n(\hat{\rho}-1)$			$\hat{\tau}$			$n(\hat{\rho}-1)$			$\hat{\tau}$		
ρ		DF	AD	M5	DF	AD	M5	DF	AD	M5	DF	AD	M5
						$n = 50$							
1.0	D1	4.4	6.0	3.0	5.3	4.3	9.4	4.8	6.6	2.0	5.5	2.5	14.8
	D2	3.9	5.2	2.6	4.5	4.5	5.2	3.9	3.3	2.0	4.6	2.6	6.2
	D3	5.6	5.0	4.0	5.7	4.0	17.8	5.5	2.4	3.2	5.0	2.4	29.2
	D4	4.9	4.5	2.4	5.3	4.8	5.8	3.6	2.8	1.2	4.5	2.6	6.4
0.95	D1	9.9	11.9	7.5	7.3	4.9	13.4	6.0	9.2	2.6	6.5	2.9	16.1
	D2	8.1	13.1	6.6	5.7	11.3	14.2	5.1	6.2	2.4	5.3	7.4	12.1
	D3	11.4	45.1	8.6	7.9	46.4	18.5	7.1	22.0	3.6	7.3	28.3	25.9
	D4	9.3	20.0	5.3	6.3	22.2	16.2	5.1	9.8	2.4	5.7	12.5	14.5
0.9	D1	19.2	22.1	13.8	12.2	9.1	19.8	9.0	12.6	4.9	9.0	4.9	20.4
	D2	15.5	23.9	18.4	9.6	19.7	29.1	7.6	12.1	5.8	7.8	12.0	23.1
	D3	19.2	85.8	13.8	13.3	79.0	21.7	11.9	59.0	6.3	11.1	62.8	25.8
	D4	16.3	37.5	15.3	10.1	36.0	33.2	8.1	20.8	4.5	7.7	25.1	27.7
0.8	D1	47.9	48.2	37.9	34.6	24.2	44.5	23.9	27.0	13.9	20.7	12.6	37.8
	D2	47.6	54.9	55.4	29.7	39.5	63.5	18.2	28.8	23.8	16.2	25.6	49.3
	D3	48.7	95.8	31.7	34.0	93.5	31.7	24.9	88.5	13.8	22.0	86.0	33.3
	D4	47.3	63.5	56.6	31.4	55.3	68.5	22.3	40.4	21.4	18.0	37.5	55.6
						Critical values							
		−13.3		−14.1	−2.90		−2.90	−19.9		−21.6	−3.49		−3.40

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

						$n = 100$							
1.0	D1	4.9	7.7	4.3	5.0	4.9	6.9	4.3	6.0	3.6	4.4	3.5	7.8
	D2	5.8	7.3	3.9	6.1	6.7	3.7	4.6	4.6	3.0	4.3	3.9	3.6
	D3	5.5	4.9	5.1	5.2	4.9	13.0	4.9	3.3	3.5	4.6	2.7	19.2
	D4	4.7	5.1	2.1	5.4	4.7	3.6	4.0	4.0	1.0	5.2	3.8	4.8
0.95	D1	16.9	20.0	13.9	10.8	9.5	13.7	9.4	9.7	6.4	8.3	5.4	12.5
	D2	17.6	31.4	21.2	11.4	26.9	25.5	8.5	16.6	7.7	8.5	19.2	18.5
	D3	19.3	92.9	14.6	11.9	88.1	15.1	9.7	76.0	7.4	9.4	76.7	18.1
	D4	15.9	52.0	16.6	10.3	55.6	27.4	9.2	31.6	4.7	7.6	41.0	18.2
0.9	D1	46.0	48.1	41.8	31.4	26.3	34.6	22.2	23.9	16.8	19.0	12.9	25.1
	D2	44.0	65.8	62.7	31.0	53.5	64.9	21.2	41.2	30.4	18.2	38.6	49.5
	D3	48.2	100.0	36.5	34.2	99.1	28.1	24.2	98.8	16.7	20.4	98.1	25.5
	D4	45.3	81.2	61.3	30.0	79.6	66.6	20.6	61.4	25.8	16.9	64.7	50.6
0.8	D1	94.1	93.4	91.9	86.0	74.8	81.0	72.5	72.9	63.5	64.4	51.2	64.7
	D2	96.0	95.2	98.8	87.0	84.5	94.9	72.0	82.6	86.8	64.7	72.4	87.9
	D3	96.2	100.0	83.2	87.8	100.0	63.1	72.5	99.9	53.0	66.3	99.7	52.2
	D4	94.5	96.4	99.1	86.1	91.4	95.7	70.7	90.6	88.4	62.9	84.8	88.2
						Critical values							
		−13.8		−14.1	−2.86		−2.90	−20.9		−21.6	−3.45		−3.40

Note: Nominal level = 5%. Number of replications = 1,000. D1: $N(0,1)$; D2: $t(3)$; D3: 0.5 $N(-3,1)$ + 0.5 $N(3,1)$; D4: $\chi_1^2 - 1$; DF: Dickey–Fuller tests; AD: tests based on adaptive estimator; M5: tests based on *M*-estimation of Lucas (1995a) using $t(5)$ as a quasi score. Critical values of DF are from Fuller (1996, App. A); those of AD depend on $\hat{\lambda}$, are from Table 1, and are not given in this table; those of M5 are the critical values of DF for large n (5,000) as suggested in Lucas (1995a). The $n(\hat{\rho} - 1)$ for M5 is adjusted for nuisance parameters as in (8) of Lucas (1995a) and is $n(\hat{\rho} - 1)m_\psi s_e / s_\psi$, where $m_\psi = n^{-1}\Sigma\psi'(s^{-1}\hat{e}_t)$, $s_e^2 = n^{-1}\Sigma\hat{e}_t^2$, and $s_\psi^2 = n^{-1}\Sigma\{\psi(s^{-1}\hat{e}_t)\}^2 s^2$.

tests have more desirable size properties than those of Beelders because, in our adaptive estimation, the unity is used as an initial estimator of the unit root and an iterated estimate of the nearly optimal bandwidth (6) is used.

An interesting future research topic would be to extend our results to the case of serially correlated errors as Lucas (1995a) and Herce (1996) did for their *M*-estimation and LAD estimation. Even though a rigorous justification seems nontrivial, we expect that the limiting distributions of the adaptive tests for the dependent error model would be similar to those in Theorem 3, except for nuisance parameters of scale factors and location shifts due to the serial correlation. Such an extension would permit us to construct test statistics adjusted for the nuisance parameters as equations (8) and (9) of Lucas (1995a) and equations (21)–(24) of Herce (1996). An extension to an ARMA model is relatively straightforward. The arguments of Theorem 4 in Section 4 for an AR(p) model easily extend to an ARMA model $y_t = \rho y_{t-1} + \beta_1 z_{t-1} + \cdots + \beta_{p-1} z_{t-p+1} + e_t + \gamma_1 e_{t-1} + \cdots + \gamma_q e_{t-q}$ with $\mathbf{X}$ and $\tilde{\boldsymbol{\theta}}_n$ in (16) being replaced by $\mathbf{X}_t = (y_{t-1}, z_{t-1}, \ldots, z_{t-p+1}, \tilde{e}_{t-1}, \ldots, \tilde{e}_{t-q})'$ and $\tilde{\boldsymbol{\theta}}_n = (1, \tilde{\beta}_1, \ldots, \tilde{\beta}_{p-1}, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_q)'$ where $\tilde{e}_t$ and $(\tilde{\beta}_1, \ldots, \tilde{\beta}_{p-1}, \tilde{\gamma}_1, \ldots, \tilde{\gamma}_q)'$ are estimated residuals and a consistent estimator, respectively, in an ARMA$(p-1, q)$ model fitting to z_t.

REFERENCES

Beelders, O. (1996) Adaptive Estimation and Unit Root Tests. Manuscript, Department of Economics, Rochester University.

Bickel, P. (1982) On adaptive estimation. *Annals of Statistics* 10, 647–671.

Bickel, P., C.A. Klaassen, Y. Ritov., & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. London: Johns Hopkins University Press.

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

Cox, D.D. & I. Llatas (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series. *Annals of Statistics* 19, 1109–1128.

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

Drost, F.C., C.A. Klaassen, & J.H. Werker (1997) Adaptive estimation in time series models. *Annals of Statistics* 25, 786–817.

Fama, E. (1965) The behavior of stock price markets. *Journal of Business* 38, 34–105.

Franses, P.H. & N. Haldrup (1994) The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics* 12, 471–478.

Fuller, W.A. (1976) *Introduction to Statistical Time Series.* New York: Wiley.

Fuller, W.A. (1996) *Introduction to Statistical Time Series*, 2nd ed. New York: Wiley.

Gassiat, E. (1993) Adaptive estimation in noncausal stationary AR processes. *Annals of Statistics* 21, 2022–2042.

Hall, W.J. & R.M. Loynes (1977) On the concept of contiguity. *Annals of Probability* 5, 278–282.

Herce, M.A. (1996) Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory* 12, 129–153.

Hidalgo, F.J. (1992) Adaptive semiparametric estimation in the presence of autocorrelation of unknown form. *Journal of Time Series Analysis* 13, 47–78.

Hodgson, D.J. (1996) Adaptive Estimation of Cointegrated Models: Simulation Evidence and an Application to the Forward Exchange Market. Manuscript, Department of Economics, University of Rochester.

Hodgson, D.J. (1998a) Adaptive estimation of cointegrating regressions with ARMA errors. *Journal of Econometrics* 85, 231–267.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

Hodgson, D.J. (1998b) Adaptive estimation of error correction models. *Econometric Theory* 14, 44–69.
Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.
Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, & T. Lee (1985) *The Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.
Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261–278.
Koul, H. & G.C. Pflug (1990) Weakly adaptive estimators in explosive autoregression. *Annals of Statistics* 18, 939–960.
Kreiss, J.P. (1987) On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15, 112–133.
Lucas, A. (1995a) Unit root tests based on *M* estimators. *Econometric Theory* 11, 331–346.
Lucas, A. (1995b) An outlier robust unit root test with an application to the extended Nelson–Plosser data. *Journal of Econometrics* 66, 153–173.
Pantula, S.G., G. Gonzalez-Farias, & W.A. Fuller (1994) A comparison of unit-root test criteria. *Journal of Business and Economic Statistics* 12, 449–459.
Pantula, S.G. & A. Hall (1991) Testing for unit roots in autoregressive moving average models. An instrument variable approach. *Journal of Econometrics* 48, 325–353.
Park, H.J. & W.A. Fuller (1995) Alternative estimators for the parameters of the autoregressive process. *Journal of Time Series Analysis* 16, 415–430.
Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
Phillips, P.C.B. (1990) Time series regression with a unit root and infinite-variance errors. *Econometric Theory* 6, 44–62.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Said, D.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models with unknown order. *Biometrika* 71, 599–607.
Said, D.E. & D.A. Dickey (1985) Hypothesis testing in ARIMA$(p,1,q)$ models. *Journal of the American Statistical Association* 80, 369–374.
Schick, A. (1987) A note on the construction of asymptotically linear estimates. *Journal of Statistical Planning and Inferences* 16, 89–105.
Shin, D.W., S. Sarkar, & L.J. Lee (1996) Unit root tests for time series with outliers. *Statistics and Probability Letters* 30, 189–197.
Shin, D.W. & B.S. So (1997) Semiparametric tests for unit root tests based on symmetric estimators. *Statistics and Probability Letters* 33, 177–184.
Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall.
Steigerwald, D.G. (1992) Adaptive estimation in time series regression models. *Journal of Econometrics* 54, 251–275.

APPENDIX

Proof of Theorem 1. This is due to Lucas (1995a). ■

To prove other theorems, we need some results on adaptive estimation. So, before giving the proofs, we give some preliminary steps. We first need the concept of contiguity. Let $\{P_n\}$ and $\{Q_n\}$ be two sequences of probability measures defined on the sample space $\mathbb{R}^n$.

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

DEFINITION 1 (Contiguity; Bickel, Klaassen, Ritov, and Wellner, 1993, p. 498). *We say P_n is contiguous to Q_n if, for every sequence of Borel sets B_n in $\mathbb{R}^n$, $Q_n(B_n) \to 0$ implies $P_n(B_n) \to 0$ as $n \to \infty$.*

Let $\boldsymbol{\theta}$ be a vector of parameters in the regression model (13) except ρ or subvector of parameters in reduced models of (13) and let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$. Because we use the initial estimator $\tilde{\rho} = 1$, the parameter ρ is not included in $\boldsymbol{\theta}$. Throughout the proof, ρ is set to one. Let $\{\boldsymbol{\theta}_n\}$ be a deterministic sequence satisfying

$$\boldsymbol{\delta}_n := n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = O(1). \tag{A.1}$$

In the sequel, P_n and Q_n denote the probability measures of $\{y_2, \ldots, y_n\}$ conditional on y_1 that are defined by the following AR(p) model with a linear trend

$$y_t = \mu + \gamma(t/n - 0.5) + \rho y_{t-1} + \beta_1 z_{t-1} + \cdots + \beta_{p-1} z_{t-p+1} + e_t,$$

under $\boldsymbol{\theta} = (\mu, \gamma, \beta_1, \ldots, \beta_{p-1})' = \boldsymbol{\theta}_0 = (0, 0, \beta_{1,0}, \ldots, \beta_{p-1,0})'$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\delta}_n$, respectively, where $\boldsymbol{\delta}_n \in \mathbb{R}^{p+1}$. In Lemma 2, we show that under the regularity conditions of Theorem 4, P_n is contiguous to Q_n for any deterministic sequence $\{\boldsymbol{\theta}_n\}$ satisfying (A.1). We first need a preliminary lemma.

LEMMA 1 (Hall and Loynes, 1977, Proposition 4). *If the likelihood ratio, $\log dQ_n/dP_n$, is bounded in probability under P_n, then P_n is contiguous to Q_n.*

LEMMA 2. *Assume the conditions of Theorem 4. Then P_n is contiguous to Q_n.*

Proof. By Lemma 1, we have only to show that $\ell_n(\boldsymbol{\delta}_n) = \log L_n(\boldsymbol{\delta}_n)$ is bounded in probability under P_n, where $L_n = dQ_n/dP_n$ and $\boldsymbol{\delta}_n = O(1)$. First we note that

$$\begin{aligned}\ell_n(\boldsymbol{\delta}_n) &= \Sigma\{\log f(e_t - n^{-1/2}\boldsymbol{\delta}_n'\mathbf{Z}_{\tau,t}) - \log f(e_t)\} \\ &= n^{-1/2}\Sigma\psi(e_t)\boldsymbol{\delta}_n'\mathbf{Z}_{\tau,t} + 2^{-1}n^{-1}\Sigma\psi'(e_t)(\boldsymbol{\delta}_n'\mathbf{Z}_{\tau,t})^2 + R_n,\end{aligned}$$

where $\mathbf{Z}_{\tau,t} = (1, (t/n - 0.5), z_{t-1}, \ldots, z_{t-p+1})'$ and R_n is the remainder term. Now, under conditions (i) and (ii) of Assumption 1, the strong law of large numbers and the central limit theorem for martingale sequence imply that, under P_n,

$$n^{-1/2}\Sigma\psi(e_t)\mathbf{Z}_{\tau,t} \Rightarrow N(0, I(f)\boldsymbol{\Omega}),$$

$$n^{-1}\Sigma\psi'(e_t)\mathbf{Z}_{\tau,t}\mathbf{Z}_{\tau,t}' = I(f)\boldsymbol{\Omega} + o_p(1),$$

and

$$|R_n| \le Kn^{-1}\Sigma(\boldsymbol{\delta}_n'\mathbf{Z}_{\tau,t})^2 n^{-1/2} \max_{1\le t\le n} |\boldsymbol{\delta}_n'\mathbf{Z}_{\tau,t}| = O_p(1)o_p(1) = o_p(1),$$

where $\boldsymbol{\Omega} = \lim_{n\to\infty} n^{-1}\Sigma E(\mathbf{Z}_{\tau,t}\mathbf{Z}_{\tau,t}')$. This completes the proof. ■

Let $S_n(\boldsymbol{\theta}) = n^{-1}\{\hat{\Delta}_n^s(\boldsymbol{\theta}) - \Delta_n(\boldsymbol{\theta}_0)\}$ and let $\hat{I}_n^s(\boldsymbol{\theta})$ be the value of $\hat{I}_n^s$ evaluated at $\boldsymbol{\theta}$.

LEMMA 3. *Under the conditions of Theorem 4, $S_n(\boldsymbol{\theta}_n) \to 0$ in Q_n probability implies $S_n(\boldsymbol{\theta}_n) \to 0$ in P_n probability.*

Proof. This follows from the contiguity of P_n to Q_n (Lemma 2) and Borel measurability of the set $\{|S_n(\boldsymbol{\theta}_n)| > \epsilon\}$ for all $\epsilon > 0$. ■

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

DEFINITION 2. *Let $\tilde{\boldsymbol{\theta}}$ be an $n^{1/2}$-consistent estimator of $\boldsymbol{\theta}_0$. A discrete $n^{1/2}$-consistent estimator $\bar{\boldsymbol{\theta}}_n$ of $\tilde{\boldsymbol{\theta}}_n$ is defined as the point in $n^{-1/2}Z^d$ closest to $\tilde{\boldsymbol{\theta}}_n$, where Z denotes the set of integers and d denotes the dimension of $\boldsymbol{\theta}$.*

LEMMA 4. *Assume the conditions of Theorem* 4. *Let $\bar{\boldsymbol{\theta}}_n$ be a discrete $n^{1/2}$-consistent estimator of $\tilde{\boldsymbol{\theta}}$. If $S_n(\boldsymbol{\theta}_n) \to 0$ in Q_n probability for all $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\boldsymbol{\delta}_n$ with $\boldsymbol{\delta}_n = O(1)$, then $S_n(\bar{\boldsymbol{\theta}}_n) \to 0$ in P_n probability.*

Proof. This follows from the contiguity of P_n to Q_n, Lemma 3, discreteness of $\bar{\boldsymbol{\theta}}_n$, and $n^{1/2}$-consistency of $\bar{\boldsymbol{\theta}}_n$. ■

Lemma 4 simplifies our proof very much. The major part of the proofs of Theorems 2–4 is to show, with $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n$, under P_n,

$$n^{-1}\{\hat{\Delta}_n^s(\boldsymbol{\theta}) - \Delta_n(\boldsymbol{\theta}_0)\} = o_p(1) \quad \text{and} \quad \hat{I}_n^s(\boldsymbol{\theta}) - I(f) = o_p(1), \tag{A.2}$$

which is not simple because $\tilde{\boldsymbol{\theta}}_n$ is random and $e_t(\tilde{\boldsymbol{\theta}}_n)$ is a dependent sequence under P_n. Note that, in the real world, we are using the discrete estimators as a result of rounding off in computing estimators. Thus a practical version of (A.2) with $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_n$ is that based on the discrete estimator $\bar{\boldsymbol{\theta}}_n$ of $\tilde{\boldsymbol{\theta}}_n$. Hence proving (A.2) with $\bar{\boldsymbol{\theta}}_n$ instead of $\tilde{\boldsymbol{\theta}}_n$ is a practical equivalent of proving (A.2). Thanks to Lemma 4, the problem reduces to showing (A.2) with $\boldsymbol{\theta} = \boldsymbol{\theta}_n$ under Q_n, which is simpler because $\boldsymbol{\theta}_n$ is deterministic and, under Q_n, the residuals $\{e_t(\boldsymbol{\theta}_n)\}$ become i.i.d. random errors. Throughout the remainder of the proof, if not specified otherwise, all the probability statements are under the probability Q_n.

We are now ready to state the key substitution lemma that is the most crucial step toward our theorems.

LEMMA 5 (Substitution lemma). *Assume the conditions of Theorem* 4. *Let $\bar{\boldsymbol{\theta}}_n$ be a discrete $n^{1/2}$-consistent estimator of the initial estimator $\tilde{\boldsymbol{\theta}}_n$. Then we have, under P_n,*

$$n^{-1}\Sigma\{\hat{\psi}_{nt}^s(e_t(\bar{\boldsymbol{\theta}}_n);\bar{\boldsymbol{\theta}}_n) - \psi(e_t(\bar{\boldsymbol{\theta}}_n))\}\mathbf{X}_{\tau,t} = o_p(1), \tag{A.3}$$

$$\hat{I}_n^s(\bar{\boldsymbol{\theta}}_n) = I(f) + o_p(1), \tag{A.4}$$

where $\mathbf{X}_{\tau,t} = (1,(t/n - 0.5), y_{t-1}, z_{t-1}, \ldots, z_{t-p+1})'$.

Proof. By Lemma 4, we have only to show that (A.3) and (A.4) hold under the deterministic sequence $\{\boldsymbol{\theta}_n\}$ satisfying $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\delta_n$ instead of under $\bar{\boldsymbol{\theta}}_n$, where $\boldsymbol{\delta}_n = O(1)$. Because our proof follows closely that of Bickel (1982) and Schick (1987) in the semiparametric adaptive estimation for the independent case, we summarize the key results of Schick (1987):

$$n^{-1/2}\Sigma\{\hat{\psi}_{nt}^s(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) - \psi(e_t(\boldsymbol{\theta}_n))\} = o_p(1), \tag{A.5}$$

$$E\int[\hat{\psi}_{nt}^s(x;\boldsymbol{\theta}_n) - \psi(x)]^2 f(x)\,dx = o(1), \tag{A.6}$$

$$\hat{I}_n^s(\boldsymbol{\theta}_n) = I(f) + o_p(1), \tag{A.7}$$

$$\sup_x|\hat{\psi}_{nij}^s(x;\boldsymbol{\theta}_n) - \hat{\psi}_{ni}^s(x;\boldsymbol{\theta}_n)| \le (h_n^2 s_n n)^{-1} = o(n^{-1/2}), \tag{A.8}$$

where $\hat{\psi}_{nij}^s(x;\boldsymbol{\theta})$ is the estimator of the score function $\psi(x)$ constructed from deleting two errors $e_i(\boldsymbol{\theta})$ and $e_j(\boldsymbol{\theta})$. Note that (A.4) directly follows from (A.7). We give proof of (A.3)

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

only for the third elements of $\mathbf{X}_{\tau,t}$, that is, y_{t-1}, because proofs for other elements are similar or easier. We need the decomposition of AR(p) processes

$$y_t = y_0 - \eta_0 + \eta_t + c^{-1}w_t, \tag{A.9}$$

where $w_t = \Sigma_{i=1}^{t} e_t$ and $\{\eta_t\}$ is a stationary process that is negligible compared with w_t (Fuller, 1996, p. 556). The representation (A.9) together with (A.5)–(A.8) takes care of most of the terms in (A.3) except for the last two terms in (A.9), which are investigated in more detail subsequently. Now, by the Cauchy–Schwartz inequality and (A.6),

$$|n^{-1}\Sigma\eta_t\{\hat{\psi}_{nt}^{s}(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) - \psi(e_t(\boldsymbol{\theta}_n))\}|$$
$$\le (n^{-1}\Sigma\eta_t^2)^{1/2}[n^{-1}\Sigma\{\hat{\psi}_{nt}^{s}(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) - \psi(e_t(\boldsymbol{\theta}_n))\}^2]^{1/2} = O_p(1)o_p(1) = o_p(1).$$

As for the last term of (A.9), we observe

$$E[n^{-1}\Sigma\{\hat{\psi}_{nt}^{s}(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n) - \psi(e_t(\boldsymbol{\theta}_n))\}w_{t-1}]^2$$
$$= n^{-2}\Sigma E\left[\int\{\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\, w_{t-1}^2\right] \tag{A.10}$$

because $\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)$ is odd and hence the cross-product terms are all zero. Now (A.10) becomes

$$n^{-2}\sum_{t=p+1}^{n}\sum_{j=p+1}^{t-1} E\left[e_j^2\int\{\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right]$$
$$+ n^{-2}\sum_{t=p+1}^{n}\sum_{j=p+1}^{t-1}\sum_{i=p+1,\,i\neq j}^{t-1} E\left[e_i e_j\int\{\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right].$$

We note that by (A.6) and (A.8),

$$\sup_{j<t} E\left[e_j^2\int\{\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right]$$
$$\le \sup_{j<t}\left\{2E\left[e_j^2\int\{\hat{\psi}_{ntj}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right]\right.$$
$$\left. + 2E\left[e_j^2\int\{\hat{\psi}_{ntj}^{s}(x;\boldsymbol{\theta}_n) - \hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n)\}^2 f(x)\,dx\right]\right\}$$
$$\le 2\sigma^2\sup_{j<t} E\left[\int\{\hat{\psi}_{ntj}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right] + 2\sigma^2(h_n^2 s_n n)^{-2} = o(1). \tag{A.11}$$

Because $\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n)$ is an odd function with respect to each of $\{e_j\}_{j\neq t}$, we get the identity

$$E\left[e_i e_j\int\{\hat{\psi}_{nt}^{s}(x;\boldsymbol{\theta}_n) - \psi(x)\}^2 f(x)\,dx\right] = 0, \qquad i, \qquad j \neq t, \qquad i \neq j.$$

This together with (A.11) finishes the proof for (A.3). ■

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press

Proof of Theorem 2. We get the result from Theorem 1 and observing that, by Lemma 5,

$$\begin{aligned} n(\hat{\rho}_A - 1) &= n^{-1}\Sigma\hat{\psi}^s_{nt}(e_t(\tilde{\rho});\tilde{\rho})y_{t-1}/\{n^{-2}\Sigma y^2_{t-1}\hat{I}^s_n\} \\ &= [n^{-1}\Sigma\psi(e_t)y_{t-1} + o_p(1)]/[n^{-2}\Sigma y^2_{t-1}\{I(f) + o_p(1)\}] = n(\hat{\rho}_M - 1) + o_p(1). \end{aligned}$$

■

Proof of Theorem 3. By Lemma 4, we have only to prove the result under the deterministic sequence of $\boldsymbol{\theta}_n = n^{-1/2}[(\alpha,\gamma) - o(1)]'$ with $(\alpha,\gamma) = (0,0)$. Now the result for $n(\bar{\rho}_A - 1)$ follows from

$$\begin{aligned} n(\bar{\rho}_A - 1) &= n^{-1}\Sigma\hat{\psi}^s_{nt}(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n)(y_{t-1} - \bar{y})/\{n^{-2}\Sigma(y_{t-1} - \bar{y})^2\hat{I}^s_n(\boldsymbol{\theta}_n)\} \\ &= [n^{-1}\Sigma\psi(e_t(\boldsymbol{\theta}_n);\boldsymbol{\theta}_n)(y_{t-1} - \bar{y}) + o_p(1)]/[n^{-2}\Sigma(y_{t-1} - \bar{y})^2\{I(f) + o_p(1)\}] \\ &\Rightarrow D_\mu(\lambda), \end{aligned}$$

which is a consequence of Lemma 5 and the bivariate functional central limit theorem. The result for $n(\rho^*_A - 1)$ follows from similar arguments. ■

Proof of Theorem 4. The results follow from combined arguments for proof of Theorem 3, Lemma 5, and the representation (A.9). ■

https://doi.org/10.1017/S0266466699151016 Published online by Cambridge University Press