Journal of the Korean Statistical Society (2002), **31**: *3, pp 301–314*

Unit Root Tests for Autoregressive Moving Average Processes Based on M-estimators

Dong-Wan Shin[1] and Oesook Lee[1]

ABSTRACT

For autoregressive moving average (ARMA) models, robust unit root tests are developed using M-estimators. The tests are parametric in the sense ARMA parameters are estimated jointly with unit roots. A Monte-Carlo experiment reveals superiority of the parametric tests over the semi-parametric tests of Lucas (1995a) in terms of both empirical sizes and powers.

Keywords. ARMA process, M-estimation, parametric tests, robust test.
AMS 2000 subject classifications. Primary 62F03; Secondary 62F35.

1. Introduction

Since Dickey and Fuller (1979), various tests for the unit root hypothesis have been developed. In constructing unit root tests, the ordinary least squares estimator (OLSE) and the Gaussian maximum likelihood estimator (MLE) are frequently used, which presume Gaussian errors for good performance. However, as documented by Fama (1965), Mittnik and Rachev (1993), Lucas (1995a, b), Shin and So (1999a), and others, there are many situations in which error distributions have heavier tails than Gaussian distributions. In this situation, we need robust unit root tests which are more resistant to presence of outliers than those based on the OLSE or the Gaussian MLE.

Recently, some authors applied M-estimation in constructing robust unit root tests. Knight (1989) considered unit root regressions with infinite variance errors and derived limiting distributions of M-estimators under proper normalizations. Cox and Llatas (1991) and Lucas (1995a, b) proposed unit root tests based on M-estimators of unit roots. Herce (1996) developed unit root tests based on

Received November 2000; accepted February 2002.
[1]Department of Statistics, Ewha Womans University, Seoul 120-750, Korea

least absolute deviation (LAD) estimators. Shin and So (1999a) applied adaptive M-estimation for constructing unit root tests. Shin and So (1999b) developed normal tests for models with possibly infinite variance errors. Some authors such as Lucas (1995a, b) and Herce (1996) adopted the semiparametric adjustment of Phillips (1987) to account for autocorrelations of error processes. The other authors focused their attentions on situations of independent errors.

We point out that, in cases of negatively autocorrelated errors, the semiparametric robust tests of Lucas (1995a, b) and Herce (1996) suffer from severe size distortions as reported by Schwert (1989) and Pantula and Hall (1991) for models with Gaussian errors. Instead of the semiparametric approach of Lucas, we adopt the parametric approach of ARMA modeling to account for autocorrelations of error processes. We develop robust unit root tests by estimating the ARMA parameters of error processes jointly with unit roots. A Monte-Carlo simulation shows that the parametric tests have much more stable empirical sizes than the semiparametric tests of Lucas (1995a). Moreover, it also reveals that the parametric tests have substantial power advantages over the semiparametric tests.

Our method follows the spirit of the parametric unit root tests developed for Gaussian ARMA errors : the tests of Said and Dickey (1985), Said (1991), and Yap and Reinsel (1995) based on the one-step Gauss-Newton approximations to the MLE; the tests of Pantula and Hall (1991) based on instrumental variable estimators of unit roots; the Lagrangian multiplier test of Ahn (1993); the tests of Shin and Fuller (1998) based on the unconditional MLE.

In the remainder of this paper, a model and tests are described in Section 2, a Monte-Carlo experiment compares our parametric tests favorably with the semiparametric tests of Lucas (1995a) in Section 3, and proofs of the theoretical results are provided in Appendix.

2. Model and Tests

Consider the following unit root regression

$$y_t = X_t'\beta + z_t,\ t = 1, \cdots, n, \tag{2.1}$$

where $\{y_t\}_{t=1}^n$ is a set of observations, $X_t = (y_{t-1}, 1, \cdots, t^\kappa)'$, $\beta = (\rho,\ \beta_0,\ \beta_1, \cdots, \beta_\kappa)'$, and $\kappa \in \{-1,\ 0,\ 1,\ \cdots\}$ is a given integer. Here and in the sequel, it is understood that $\kappa = -1$ corresponds to no mean model $y_t = \rho y_{t-1} + z_t$ with $(X_t = y_{t-1}, \beta = \rho)$. Note that if $\kappa = 0$, (2.1) is a mean model $y_t = \beta_0 + \rho y_{t-1} + z_t$.

If $\kappa = 1$, (2.1) is a trend model. In general, (2.1) is a model of κ-th order polynomial trend. The error process z_t is a sequence of unobservable stationary ARMA errors

$$\phi(L)z_t = \theta(L)e_t,$$

where

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \quad \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q,$$

L is the lag operator such that $Lz_t = z_{t-1}$ and e_t is a sequence of independent errors with mean zero and variance σ^2. We are interested in testing the unit root hypothesis

$$H_0 : \rho = 1.$$

Let $\xi = (\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q)'$ and $\varphi = (\beta', \xi')'$. Let ξ_0 and φ_0 be the true values of the parameters. The M-estimator $\hat{\varphi}$ is a solution to the M-equation

$$\sum_{t=1}^{n} \psi(e_t(\varphi)/\tilde{\sigma})e_{\varphi t} = 0, \tag{2.2}$$

where $e_t(\varphi)$ are recursively computed from

$$e_t(\varphi) = y_t - X_t'\beta - \phi_1 \Delta y_{t-1} - \cdots - \phi_p \Delta y_{t-p} - \theta_1 e_{t-1}(\varphi) - \cdots - \theta_q e_{t-q}(\varphi),$$

$e_{\varphi t}(\varphi) = (e'_{\beta t}(\varphi), e'_{\xi t}(\varphi))' = \partial e_t(\varphi)/\,\partial\varphi$ are derivatives of $e_t(\varphi)$, $\tilde{\sigma}$ is an estimator of σ, and $\Delta y_t = y_t - y_{t-1}$. The function ψ is a real-valued function called "quasi-score" and satisfying assumption A1 below. When $\psi(x) = x$, the M-estimator is the Gaussian MLE. One of the well-known function for ψ is the Huber function given by $\psi_h = \min(c, \max(-c, x))$, where $c > 0$ is a given real number. The score function ψ_h discounts x to c if $x > c$ (discount x to $-c$ if $x < -c$). Other examples of quasi-scores are found in Martin and Yohai (1985) and in Section 3. By suppressing large values of error terms $e_t(\varphi)$, the M-equation produces a robust M-estimator which is resistant to wild errors. When ψ is the derivative of a convex function δ, then $\hat{\varphi}$ is the unique minimizer of $\sum \delta(e_t(\varphi))$. The M-estimators $\hat{\rho}$, $\hat{\beta}$ and $\hat{\xi}$ are the corresponding elements of $\hat{\varphi}$.

When the absolute value of an error $e_t(\varphi)$ is large, the quasi-score function ψ discounts the error to a smaller value. In order to determine whether $e_t(\varphi)$ is large or not, an estimate of a scale parameter of e_t is necessary. One of the most commonly used scale estimator is

$$\tilde{\sigma} = \operatorname*{median}_{1<t<n} |e_t(\tilde{\varphi})| /\ 0.6745,$$

where $\tilde{\varphi}$ is a consistent initial estimator of φ such as the Gaussian MLE or the OLSE. The median is divided by 0.6745 in order for $\tilde{\sigma}$ to be a consistent estimator of σ in case of Gaussian error.

We now state conditions required for our study.

A1. ψ has the derivative $\dot{\psi}$ satisfying the first order Lipschitz condition $|\dot{\psi}(x) - \dot{\psi}(y)| \leq M|x - y|$ for all $x, y \in \mathbf{R}$ and for some $M < \infty$ and the following conditions $E\{\psi(e_t)\} = 0$; $0 < \mu_\psi = E\{\dot{\psi}(e_t)\} < \infty$; $0 < \sigma_\psi^2 = E\{\psi(e_t)\}^2 < \infty$.

The conditions in A1 are standard in dealing with the limiting behavior of M-estimators and were imposed in all the studies of the robust unit root tests discussed in Section 1. For many standard functions ψ, all conditions of A1 except the differentiability are satisfied. The assumption that ψ is everywhere differentiable is not satisfied, for example, by the Huber function, which is not differentiable at two points. However, even if ψ does not have derivative at a finite number of points, all the results established below also hold with some complication in the proof. Assumption A2 below states that the error process z_t is a stationary and invertible ARMA.

A2. e_t is a sequence of independent errors with mean zero and finite variance σ^2; all roots of $\phi(L)$ and $\theta(L)$ lie strictly outside the unit circle, and $\phi(L)$ and $\theta(L)$ have no common roots.

Our tests for H_0 based on the M-estimator are $n(\hat{\rho}-1)$ and $\hat{\tau}_S = (\hat{\rho}-1)/\ se(\hat{\rho})$, where $se(\hat{\rho})$ is the standard error of $\hat{\rho}$ and is the square root of the (1,1) element of

$$\hat{\sigma}^2 \left[\sum \dot{\psi}(\hat{e}_t/\hat{\sigma})\hat{e}_{\beta t}\hat{e}'_{\beta t}\right]^{-1} \left[\sum \{\psi(\hat{e}_t/\hat{\sigma})\}^2 \hat{e}_{\beta t}\hat{e}'_{\beta t}\right] \left[\sum \dot{\psi}(\hat{e}_t/\hat{\sigma})\hat{e}_{\beta t}\hat{e}'_{\beta t}\right]^{-1}. \quad (2.3)$$

In (2.3) and in the sequel, $\hat{e}_t$ and $\hat{e}_{\varphi t}$ are the values of $e_t(\varphi)$ and $e_{\varphi t}$ evaluated at $\varphi = \hat{\varphi}$ and $\hat{\sigma}^2$ is an estimator of σ^2 computed from the residuals $\hat{e}_t$. The reason for using the covariance estimator (2.3) is explained after establishing the limiting null distributions of the test statistics in Theorem 1 below.

Theorem 1. *Consider (2.1). Let assumptions A1 and A2 hold. Suppose also that either of the following two conditions hold:*

(a) $\psi(x)$ *is the derivative of a convex function* δ *or*

(b) $A_n(\hat{\beta}-\beta_0) = o_p(1)$ *and* $(\hat{\xi}-\xi_0) = o_p(1)$, *where* $A_n = diag(n^{1/2}, 1, n, \cdots, n^\kappa)$.

For each $\kappa = -1,\ 0, \cdots,$ if $\rho = 1$ and $\beta_0 = \beta_1 = \cdots = \beta_\kappa = 0$, then

$$n(\hat{\rho} - 1) \Rightarrow \{\sigma_\psi/(c\sigma\mu_\psi)\} J' \left\{ \int_0^1 X(r)X'(r)dr \right\}^{-1} \int_0^1 X(r)dW^*(r), \tag{2.4}$$

$$\hat{\tau}_S \Rightarrow \left[J' \{ \int_0^1 X(r)X'(r)dr \}^{-1} J \right]^{-1/2} J' \left\{ \int_0^1 X(r)X'(r)dr \right\}^{-1} \int_0^1 X(r)dW^*(r), \tag{2.5}$$

where $X(r) = (W(r)\ ,\ 1,\ r, \cdots, r^\kappa)'$, W is a standard Brownian motion on $[0,1]$,

$$c = \phi(1)/\ \theta(1),$$

$$W^* = \rho_{e\psi} W + (1 - \rho_{e\psi}^2)^{1/2} W^{**},$$

$\rho_{e\psi} = corr(e_t, \psi(e_t))$, W and W^{**} are two independent standard Brownian motions, $J = (1, 0, \cdots, 0)'$ is a $(\kappa + 2) \times 1$ vector, and $\Rightarrow$ denotes convergence in distribution.

Conditions (a) and (b) in Theorem 1 are similar to those adopted by Knight (1989) and Lucas (1995a) for asymptotic analyses of their robust unit root tests.

From $W^* = \rho_{e\psi} W + (1 - \rho_{e\psi}^2)^{1/2} W^{**}$, we see that the limiting null distribution of $\hat{\tau}_S$ is of the form $\rho_{e\psi} DF + (1 - \rho_{e\psi}^2)^{1/2} N$, where DF is the distribution of τ-statistic of Dickey and Fuller (1979) and N is the standard normal distribution independent of DF. If $\psi(x) = x$ is the Gaussian quasi-score, then $\rho_{e\psi} = 1$ and the limiting null distribution of $\hat{\tau}_S$ is DF. The distribution (2.5) is indexed by one parameter $\rho_{e\psi}$. The percentage points of this distribution is available in Shin and So (1999a, Table 1). Since $\rho_{e\psi}$ can be estimated by the sample correlation $\hat{\rho}_{e\psi}$ of $e_t(\hat{\varphi})$ and $\psi(e_t(\hat{\varphi}))$, we can easily find the critical values from Shin and So (1999a) using $\hat{\rho}_{e\psi}$ instead of $\rho_{e\psi}$.

The covariance matrix estimator (2.3), if multiplied by n^2, has weak limit

$$[\sigma_\psi/(c\sigma\mu_\psi)]^2 \left[\int X(r)X'(r)dr \right]^{-1}$$

by Lemma A.1-(i) and (A.1) in the Appendix. By taking the (1,1) element of (2.3) as $\{se(\hat{\rho})\}^2$, we adjust the scale parameter $\sigma_\psi/(c\sigma\mu_\psi)$ which appears in the limiting null distribution (2.4) of $n(\hat{\rho} - 1)$. Thus the limiting null distribution (2.5) of $\hat{\tau}_S$ gets free from the scale parameter. Note that, the estimated covariance matrix (2.3) is based on $\hat{e}_{\beta t}$ instead of $\hat{e}_{\varphi t} = (\hat{e}'_{\beta t}, \hat{e}'_{\xi t})'$. This is to prevent wild estimation of $se(\hat{\rho})$ in case where MA roots are close to the unit root and the matrix $\sum \dot{\psi}(\hat{e}_t)\hat{e}_{\varphi t}\hat{e}'_{\varphi t}$ is close to a singular matrix. This strategy does not change the null asymptotics of the test statistics because the limiting null distribution of $\sum \dot{\psi}(\hat{e}_t)\hat{e}_{\varphi t}\hat{e}'_{\varphi t}$, when suitably normalized, is block diagonal.

Our tests are parametric counterpart of the semiparametric test of Lucas (1995a) based on an M-estimator. The estimate $\hat{\rho}_L$ of unit root of Lucas (1995a) is the first element of $\hat{\beta}_L$ which is a solution to

$$\sum_{t=1}^{n} \psi((y_t - X_t'\beta)/\tilde{\sigma}_z)X_t = 0, \tag{2.6}$$

where $\tilde{\sigma}_z = \text{median}\,|\tilde{z}_t|/0.6745$, and $\tilde{z}_t$ is the residual in the ordinary least squares regression of y_t on X_t. In his model, z_t is a general mixing process satisfying an invariance principle. The limiting null distributions of the statistics $n(\hat{\rho}_L - 1)$ and $\hat{\tau}_L = (\hat{\rho}_L - 1)/se(\hat{\rho}_L)$ depend on nuisance parameters $\bar{\sigma}_{\psi_z\psi_z} = \sum_{j=-\infty}^{\infty} E\{\psi(z_t)\psi(z_{t+j})\}$ and $\bar{\sigma}_{z\psi_z} = \sum_{j=-\infty}^{\infty} E\{z_t\psi(z_{t+j})\}$ as those of the semiparametric tests of Phillips (1987) depend on $\bar{\sigma}_{zz} = \sum_{j=-\infty}^{\infty} E(z_t z_{t+j})$. Lucas (1995a) adopted the semiparametric approach of Phillips (1987) to adjust the nuisance parameters and constructed unit root tests as (3.1) below whose limiting null distributions are free from the nuisance parameters.

3. Numerical Study

Using a Monte-Carlo simulation, we compare finite sample properties of our unit root test $\hat{\tau}_S$ with those of $\hat{\tau}_L$ of Lucas (1995a) for mean-adjusted model. For the data generating process, we consider

$$y_t = \rho y_{t-1} + z_t \ ,$$

where z_t is one of the following : AR(1), $z_t = \phi z_{t-1} + e_t$; MA(1), $z_t = e_t + \theta e_{t-1}$. We select three distributions for the error process e_t : standard normal, $N(0, 1)$ (NR); t-distribution with 4 degrees of freedom (T4); mixed normal (MN), $0.95N(0,1) + 0.05N(0,25)$. For the quasi-score function ψ, we consider the normal score, the Huber score, and the t-score with 3 degrees of freedom given by $\psi_n(x) = x$; $\psi_h(x) = x$ if $|x| \leq 2.5$, $\psi_h = 2.5$ if $x > 2.5$, $\psi_h = -2.5$ if $x < -2.5$; $\psi_t(x) = 4x/(3 + x^2)$.

The other parameters are set to $n = 200$; $\rho = 1$, 0.95, 0.90; $\phi = 0$, ± 0.4, ± 0.8; $\theta = 0$, ± 0.4, ± 0.8. Data y_t, $t = 1, \cdots, n$, are simulated starting with $y_t = z_t = 0$ for $t \leq 0$. The normal random numbers e_t are generated by RNNOA, a FORTRAN subroutine in IMSL. The $T4$ random numbers are simulated using the

t-statistic of 5 independent normal numbers. Our unit root test $\hat{\tau}_s = (\hat{\rho}-1)/se(\hat{\rho})$ is constructed from (2.2) with $X_t = (y_{t-1},\ 1)'$ by fitting AR(2) or ARMA(1,1) models to data y_t, $t = 1, \cdots, n$. In computing $\tilde{\sigma}$, we use the residuals $e_t(\tilde{\varphi})$ based on the Gaussian MLE $\tilde{\varphi}$ of φ. The counterpart of Lucas is

$$\hat{\tau}_L = (\sigma_{\psi_z\psi_z}/\bar{\sigma}_{\psi_z\psi_z})(\hat{\rho}_L - 1)/se(\hat{\rho}_L) \\ - \{(\bar{\sigma}_{z\psi_z} - \sigma_{z\psi_z})/(2\bar{\sigma}_{\psi_z\psi_z})\}\left\{n^{-2}\sum(y_{t-1}-\bar{y})^2\right\}^{-1/2}, \qquad (3.1)$$

where $\sigma_{\psi_z\psi_z} = E\{\psi(z_t)\}^2$, $\sigma_{z\psi_z} = E\{z_t\psi(z_t)\}$, $\bar{y} = n^{-1}\sum y_t$, $\hat{\rho}_L$ is computed by solving (2.6) and $se(\hat{\rho}_L)$ is the standard error of $\hat{\rho}_L$ computed from (2.3) with $(X_t,\ z_t,\ \hat{\sigma}_z)$ instead of $(\hat{e}_{\beta t},\ \hat{e}_t,\ \hat{\sigma})$. For adjusting $\sigma_{\psi_z\psi_z}$ and $\sigma_{z\psi_z}$ in (3.1), we use $n^{-1}\sum_{t=1}^n\{\psi(\hat{z}_t)\}^2$ and $n^{-1}\sum_{t=1}^n \hat{z}_t\psi(\hat{z}_t)$, respectively, where $\hat{z}_t = y_t - X_t\hat{\beta}_L$. For adjusting the long-run covariance parameters $\bar{\sigma}_{\psi_z\psi_z}$ and $\bar{\sigma}_{z\psi_z}$ in (3.1), we use the kernel estimate of Andrews (1991) with the Parzen kernel and the "plug in" optimal bandwidth based on the residuals $\hat{z}_t$.

The test statistics $\hat{\tau}_S$ and $\hat{\tau}_L$ are simulated 5,000 times. Their rejection frequencies for testing H_0 against stationary alternative $H_1 : \rho < 1$ are reported. The level of the tests is set to 5%. As critical values of $\hat{\tau}_S$, we use the percentage points from Shin and So (1999a, Table 1). As the critical value of $\hat{\tau}_L$, we use -2.94 from Table 1 of Lucas.

In Table 1, empirical sizes(%) of the two tests are reported. We see that size of our parametric test $\hat{\tau}_S$ is relatively stable for all ranges of ϕ and θ considered here having size 16.2% in the worst case. However, size of the semiparametric test $\hat{\tau}_L$ is very sensitive to the values of ARMA coefficients for all error distributions and quasi-scores. The test $\hat{\tau}_L$ has severe size distortion for negatively autocorrelated errors (ϕ, $\theta = -0.8, -0.4$). When the error process is AR(1) with $\phi = -0.8$ or MA(1) with $\theta = -0.8$, size of Lucas' test $\hat{\tau}_L$ is above 80% for all error distributions and score functions considered here. When the error process z_t is a white noise ($\phi = 0$ or $\theta = 0$) or is positively autocorrelated (ϕ, $\theta = 0.4,\ 0.8$), sizes of the two tests are close to the nominal level 5%.

The null performance of our test $\hat{\tau}_S$ is not sensitive to the types of error distribution and quasi-scores. This is because the limiting null distribution (2.5) of $\hat{\tau}_S$ does not involve any parameters related with error distribution and quasi-score except $\rho_{e\psi}$ in W^*, which was already adjusted by taking the critical values from the distribution indexed by $\rho_{e\psi}$. Also, as the error distribution and quasi-score vary, the null performance of the semiparametric test $\hat{\tau}_L$ remains stable except for few cases of AR(1) with $\phi = -0.4$ and MA(1) with $\theta = -0.4$.

TABLE 1 *Empirical sizes (%) of the mean-adjusted τ-tests for $H_0 : \rho = 1$ against $H_1 : \rho < 1$ based on M-estimators in model $y_t = \beta_0 + \rho y_{t-1} + z_t$*

		$\hat\tau_S$	$\hat\tau_L$	$\hat\tau_S$	$\hat\tau_L$	$\hat\tau_S$	$\hat\tau_L$	$\hat\tau_S$	$\hat\tau_L$	$\hat\tau_S$	$\hat\tau_L$
						$z_t = \phi z_{t-1} + e_t$					
Dist.	ψ	$\phi = 0.8$		$\phi = 0.4$		$\phi = 0.0$		$\phi = -0.4$		$\phi = -0.8$	
NR	ψ_n	7.5	10.6	6.9	3.1	7.1	4.5	7.1	30.7	10.6	94.5
NR	ψ_h	7.6	11.1	7.0	3.4	7.1	4.7	8.1	32.0	10.6	94.2
NR	ψ_t	7.4	9.7	6.3	3.0	7.1	4.7	7.5	26.3	10.9	91.4
T4	ψ_n	4.8	10.5	4.4	3.5	4.9	6.2	5.1	39.6	7.6	96.2
T4	ψ_h	9.2	11.0	8.7	4.1	8.3	9.0	8.5	43.8	12.8	96.5
T4	ψ_t	6.5	9.7	5.5	2.9	6.5	5.0	6.7	30.9	9.1	94.4
MN	ψ_n	4.2	10.4	3.2	2.2	3.9	3.3	4.1	31.8	7.7	96.5
MN	ψ_h	7.4	10.4	6.8	3.6	7.1	5.4	7.5	36.2	10.9	96.6
MN	ψ_t	6.2	8.5	5.6	2.3	6.2	3.1	6.8	20.7	8.9	92.4
						$z_t = e_t + \theta e_{t-1}$					
Dist.	ψ	$\theta = 0.8$		$\theta = 0.4$		$\theta = 0.0$		$\theta = -0.4$		$\theta = -0.8$	
NR	ψ_n	6.2	6.1	6.7	3.9	7.0	4.4	7.9	18.1	15.6	97.2
NR	ψ_h	7.2	6.2	6.4	4.1	7.0	4.2	7.7	20.5	16.1	98.1
NR	ψ_t	6.7	5.6	6.6	3.7	7.1	4.7	7.3	14.0	14.8	91.0
T4	ψ_n	3.6	7.4	4.2	5.4	4.5	6.0	5.4	30.1	10.4	96.1
T4	ψ_h	9.8	8.6	7.7	6.3	8.4	8.4	11.0	38.9	16.2	99.9
T4	ψ_t	7.2	5.9	6.0	3.1	6.4	4.7	6.9	24.3	11.7	96.9
MN	ψ_n	3.8	5.3	3.2	3.1	3.9	3.1	4.1	17.3	10.7	81.2
MN	ψ_h	10.5	6.7	6.6	3.7	7.2	5.0	8.4	31.5	16.0	99.1
MN	ψ_t	7.7	4.4	5.9	2.9	5.8	3.0	6.5	14.1	11.7	90.5

NOTE : *$n = 200$, nominal level = 5%, number of replications = 5,000.*

We next investigate empirical powers of the tests at $\rho = 0.95$, 0.90 given in Table 2.

When ϕ, $\theta = -0.4, -0.8$, direct power comparison of $\hat\tau_L$ and $\hat\tau_S$ is meaningless because sizes of $\hat\tau_L$ and $\hat\tau_S$ substantially different from each other. On the other hand, when ϕ, θ = 0, 0.4, 0.8, power comparison of $\hat\tau_L$ and $\hat\tau_S$ is meaningful because sizes of $\hat\tau_L$ and $\hat\tau_S$ are similar.

The most interesting point is that our test $\hat\tau_S$ is much more powerful than $\hat\tau_L$ for ϕ, $\theta = 0.4$, 0.8 and $\rho = 0.95$, 0.90. For example, for ($\rho = 0.95$, T4, ψ_h), powers of ($\hat\tau_S$, $\hat\tau_L$) are (56.6%, 0.1%) and (71.1%, 31.2%) for $\phi = 0.8$, $\theta = 0.8$, respectively. For $\rho = 0.95$, 0.90 and AR(1) error with $\phi = 0.8$, 0.4, the semiparametric test $\hat\tau_L$ has almost no power or very small power for all error distributions and score functions considered here. On the other hand, our test $\hat\tau_S$ has substantial power for $\rho = 0.95$, 0.90 and AR(1) error with $\phi = 0.8$, 0.4. For $\rho = 0.95$, 0.90 and MA(1) error with $\theta = 0.8$, 0.4, the semiparametric test $\hat\tau_L$ has some power, which

TABLE 2 *Empirical powers (%) of the mean-adjusted τ-tests for $H_0 : \rho = 1$ against $H_1 : \rho < 1$ based on M-estimators in model $y_t = \beta_0 + \rho y_{t-1} + z_t$*

			$\hat{\tau}_S$	$\hat{\tau}_L$	$\hat{\tau}_S$	$\hat{\tau}_L$	$\hat{\tau}_S$	$\hat{\tau}_L$	$\hat{\tau}_S$	$\hat{\tau}_L$	$\hat{\tau}_S$	$\hat{\tau}_L$
			$z_t = \phi z_{t-1} + e_t$									
Dist.	ψ	ρ	$\phi = 0.8$		$\phi = 0.4$		$\phi = 0.0$		$\phi = -0.4$		$\phi = -0.8$	
NR	ψ_n	0.95	25.4	0.0	33.7	1.2	37.5	28.4	42.6	93.7	57.1	100.0
NR	ψ_h		28.2	0.0	36.0	1.6	39.7	31.6	43.6	93.6	59.6	100.0
NR	ψ_t		24.8	0.2	30.5	1.6	34.7	24.2	37.6	88.0	54.3	100.0
T4	ψ_n		30.6	0.0	39.9	4.1	41.7	44.2	47.6	97.4	63.0	100.0
T4	ψ_h		56.6	0.1	68.9	6.7	70.8	62.1	74.8	99.0	83.5	100.0
T4	ψ_t		57.6	1.5	66.7	8.6	70.1	54.9	74.1	97.3	82.6	100.0
MN	ψ_n		30.9	0.0	39.0	2.2	42.7	41.6	49.7	93.5	63.6	100.0
MN	ψ_h		58.5	0.1	71.0	7.7	75.4	69.9	78.8	99.1	85.8	100.0
MN	ψ_t		58.8	2.9	71.1	15.7	72.1	59.7	76.3	96.0	82.8	100.0
NR	ψ_n	0.90	56.4	0.0	79.1	9.5	86.3	76.3	90.4	99.8	96.7	100.0
NR	ψ_h		58.9	0.0	81.6	10.4	86.3	77.4	91.3	99.9	97.6	100.0
NR	ψ_t		49.5	0.1	69.6	9.6	76.2	66.2	80.8	99.6	91.8	100.0
T4	ψ_n		59.2	0.0	81.3	22.9	87.3	88.5	90.7	99.9	96.2	100.0
T4	ψ_h		77.1	0.2	92.6	36.4	96.2	96.6	97.5	100.0	99.3	100.0
T4	ψ_t		75.9	1.3	91.2	34.6	93.9	93.4	95.9	99.9	98.9	100.0
MN	ψ_n		63.3	0.0	83.1	16.3	87.2	84.8	90.5	99.4	95.6	100.0
MN	ψ_h		89.1	0.1	97.3	41.2	98.9	98.1	99.2	100.0	99.8	100.0
MN	ψ_t		84.5	5.2	94.8	44.3	96.8	94.5	97.7	99.8	99.2	100.0
			$z_t = e_t + \theta e_{t-1}$									
Dist.	ψ	ρ	$\theta = 0.8$		$\theta = 0.4$		$\theta = 0.0$		$\theta = -0.4$		$\theta = -0.8$	
NR	ψ_n	0.95	35.0	15.2	35.9	12.8	36.0	23.5	39.0	69.0	47.6	99.8
NR	ψ_h		37.1	17.2	36.6	14.0	38.3	25.2	40.0	73.0	49.4	100.0
NR	ψ_t		32.5	14.2	32.8	12.6	35.2	21.1	36.5	56.0	46.2	99.0
T4	ψ_n		38.5	23.8	41.3	21.8	42.1	39.5	44.6	83.4	51.2	99.4
T4	ψ_h		71.1	31.2	69.3	30.8	70.1	55.9	70.5	96.1	70.7	100.0
T4	ψ_t		71.7	34.2	70.5	29.6	70.7	51.2	72.3	88.0	72.8	99.9
MN	ψ_n		41.6	18.6	42.4	16.2	43.3	37.6	45.0	66.6	54.0	96.7
MN	ψ_h		74.9	42.4	75.1	33.1	75.1	65.3	74.9	94.8	74.1	100.0
MN	ψ_t		73.2	46.8	73.1	38.4	72.5	55.4	73.4	81.0	73.9	99.1
NR	ψ_n	0.90	87.0	29.2	86.9	31.0	84.3	62.9	82.2	92.6	81.7	99.9
NR	ψ_h		88.7	31.0	87.9	34.3	86.4	66.3	84.5	94.8	81.9	100.0
NR	ψ_t		75.9	26.6	75.8	29.0	75.6	53.7	75.0	83.0	76.9	99.8
T4	ψ_n		86.8	40.4	86.7	47.0	87.0	80.2	83.9	94.8	81.6	99.3
T4	ψ_h		96.4	55.9	96.1	60.5	95.8	92.9	93.9	99.9	89.7	100.0
T4	ψ_t		94.6	58.7	94.0	60.7	93.6	87.7	93.3	97.9	89.2	100.0
MN	ψ_n		88.4	28.2	86.9	36.4	86.9	77.0	84.6	85.6	83.4	97.4
MN	ψ_h		99.0	67.6	98.7	66.3	98.4	95.2	97.9	99.4	94.8	100.0
MN	ψ_t		97.4	71.6	96.8	68.2	96.2	89.9	96.2	94.8	93.5	100.0

NOTE : $n = 200$, *nominal level* $= 5\%$, *number of replications* $= 5,000$.

is greatly dominated by the power of the parametric test $\hat{\tau}_S$. Power advantage of $\hat{\tau}_S$ over $\hat{\tau}_L$ for positively autocorrelated errors (ϕ, $\theta = 0.4, 0.8$) is more conspicuous for AR(1) error than for MA(1) error. This is because AR(1) error with $\phi = 0.4$, 0.8 has stronger positive autocorrelation that MA(1) error with $\theta = 0.4$, 0.8. When z_t is white noise, empirical powers of $\hat{\tau}_S$ are slightly larger than those of $\hat{\tau}_L$. However, this small power advantage of $\hat{\tau}_S$ dues to the slightly higher sizes as shown in Table 1 and may not be a real advantage.

When error is not normal (T4, MN), power gains of the tests based on the nonnormal quasi-score functions (ψ_h, ψ_t) over those based on the normal score (ψ_n) are obvious uniformly over all parameters considered here. For example, for error T4 with $\rho = 0.95$ and $\theta = 0.8$, the tests $\hat{\tau}_S$ and $\hat{\tau}_L$ have powers (41.6%, 74.9%, 73.2%) and (18.6%, 42.4%, 46.8%), respectively, for the quasi-scores (ψ_n, ψ_h, ψ_t). When error is normal, power loss due to using the false score ψ_h or ψ_t instead of the true score ψ_n is not significant for all ϕ and θ considered here. For example, for the normal error NR with $\rho = 0.95$ and $\theta = 0.8$, the tests $\hat{\tau}_S$ and $\hat{\tau}_L$ have powers (35.0%, 37.1%, 32.5%) and (15.2%, 17.2%, 14.2%), respectively, for the quasi-scores (ψ_n, ψ_h, ψ_t).

From this Monte-Carlo study, we conclude that, when the error process z_t is AR(1) or MA(1), $\hat{\tau}_S$ has more stable size than $\hat{\tau}_L$ for a wider range of the parameters of z_t, especially for negatively autocorrelated errors. Also, when the error process has positive autocorrelation, $\hat{\tau}_S$ has higher power than $\hat{\tau}_L$. This size and power advantage of $\hat{\tau}_S$ over $\hat{\tau}_L$ seems to be uniform for a wide range of error distributions and quasi-score functions.

Appendix : Proofs

Lemma A.1. *Under assumptions A1-A2, we have, uniformly in ξ,*

(i) $n^{-1}A_n^{-1}\sum e_{\beta t}e'_{\beta t}A_n^{-1} \Rightarrow B\int_0^1 X(r)X'(r)drB$,

(ii) $n^{-1}A_n^{-1}\sum e_{\beta t}e'_{\xi t} = o_p(1)$,

(iii) $n^{-1}\sum e_{\xi t}e'_{\xi t} - \sigma^2\Gamma(\xi) = o_p(1)$,

(iv) $n^{-1/2}A_n^{-1}\sum \psi(e_t)e_{\beta t} \Rightarrow -\sigma_\psi B\int_0^1 X(r)dW^*(r)$,

(v) $n^{-1/2}\sum \psi(e_t)e_{\xi t} \Rightarrow \sigma_\psi N_{p+q}$,

where $B = diag(c\sigma, 1, 1, \cdots, 1)$, and N_{p+q} is a $(p+q)$-variate normal random vector having mean zero and covariance matrix $\sigma^2\Gamma(\xi)$, the Fisher information *matrix of* ξ.

Proof. The proof is similar to that of Lemma A.2 of Shin and Lee (1999) and is omitted. □

Proof of Theorem 1 under (a)

Let $D_n = diag(n^{1/2}A_n, n^{1/2}I_{p+q})$ and $Z_n(u) = \sum_t[\delta(e_t(\varphi_0 + D_n^{-1}u)) - \delta(e_t)]$. Then, $\hat{u} = D_n(\hat{\varphi} - \varphi_0) = [n^{1/2}A_n(\hat{\beta} - \beta_0)', n^{1/2}(\hat{\xi} - \xi_0)']'$ is the unique solution to the M-equation

$$D_n \partial Z_n(u)/\partial u = \sum_t \psi(e_t(\varphi_0 + D_n^{-1}u))e_{\varphi t}(\varphi_0 + D_n^{-1}u) = 0.$$

A Taylor expansion around $u = 0$ together with $e_t(\varphi_0) = e_t$ gives

$$Z_n(u) = u'D_n^{-1}\sum_t \psi(e_t)e_{\varphi t} + 0.5u'D_n^{-1}\sum_t \left\{\dot{\psi}(e_t^*)e_{\varphi t}^* e_{\varphi t}^{*\prime} + \psi(e_t^*)e_{\varphi\varphi t}^*\right\} D_n^{-1}u,$$

where $e_{\varphi\varphi t} = \partial^2 e_t(\varphi)/\partial\varphi\partial\varphi'$, e_t^*, $e_{\varphi t}^*$, and $e_{\varphi\varphi t}^*$ are evaluated at $\varphi = \varphi^*$ which is between φ_0 and $(\varphi_0 + D_n^{-1}u)$. According to the same argument for (A.8)–(A.9) of Shin and Lee (1999)

$$u'D_n^{-1}\sum_t \left\{\dot{\psi}(e_t^*)e_{\varphi t}^* e_{\varphi t}^{*\prime} - E\{\dot{\psi}(e_t)\}e_{\varphi t}e_{\varphi t}'\right\} D_n^{-1}u = o_p(1), \qquad \text{(A.1)}$$

$$u'D_n^{-1}\sum_t \psi(e_t^*)e_{\varphi\varphi t}^* D_n^{-1}u = O_p(n^{-1/2})$$

uniformly over u in compact sets. Therefore,

$$Z_n(u) = u'D_n^{-1}\sum_t \psi(e_t)e_{\varphi t} + 0.5u'D_n^{-1}\sum_t \dot{\psi}(e_t)e_{\varphi t}e_{\varphi t}'D_n^{-1}u + o_p(1)$$

and by Lemma A.1, $Z_n(u) \Rightarrow Z(u)$, where

$$\begin{aligned} Z(u) &= u'\sigma_\psi \left[-B\int_0^1 X'(r)dW^*(r),\ N_{p+q}\right]' \\ &\quad + 0.5\mu_\psi u' diag\left[B\int_0^1 X(r)X'(r)drB,\ \sigma^2\Gamma(\xi)\right] u. \end{aligned}$$

Since $Z_n(u)$ is convex and $Z(u)$ has unique minimizer, Lemma A of Knight (1989) implies that the sequence of minimizers of $Z_n(u)$ converges in distribution to the minimizer of $Z(u)$. Now, setting the derivative of $Z(u)$ to zero, we get the limiting distribution of $n(\hat{\rho} - 1)$. Limiting distribution of $\hat{\tau}_S$ follows from that of $n(\hat{\rho}-1)$ and the fact that (2.3), if multiplied by n^2, has weak limit $[\sigma_\psi/(c\sigma\mu_\psi)]^2[\int X(s)X'(s)ds]^{-1}$ by Lemma A.1-(i) and (A.1). □

Proof of Theorem 1 under (b)

Note that the condition (b) is $n^{-1/2}D_n(\hat{\varphi}-\varphi_0)=o_p(1)$. Taylor expansion of the M-equation yields

$$0=D_n^{-1}\sum_t \psi(e_t)e_{\varphi t}+D_n^{-1}\sum_t \dot{\psi}(e_t)e_{\varphi t}e'_{\varphi t}D_n^{-1}D_n(\hat{\varphi}-\varphi_0)+R_n,$$

where

$$\begin{aligned} R_n &= (r_{1n}+r_{2n}+r_{3n})n^{-1/2}D_n(\hat{\varphi}-\varphi_0),\\ r_{1n} &= n^{1/2}D_n^{-1}\sum_t\{\dot{\psi}(e_t)-\dot{\psi}(e_t^*)\}e^*_{\varphi t}e^{*\prime}_{\varphi t}D_n^{-1},\\ r_{2n} &= n^{1/2}D_n^{-1}\sum_t \dot{\psi}(e_t)\{e_{\varphi t}e'_{\varphi t}-e^*_{\varphi t}e^{*\prime}_{\varphi t}\}D_n^{-1},\\ r_{3n} &= n^{-1/2}D_n^{-1}\sum_t \psi(e_t^*)e^*_{\varphi\varphi t}D_n^{-1}, \end{aligned}$$

and $e^*_{\varphi t}$ and $e^*_{\varphi\varphi t}$ are evaluated at $\varphi=\varphi^*$ which lies between φ_0 and $\hat{\varphi}$. By similar arguments leading to (A.10) and (A.9) of Shin and Lee (1999), $r_{1n}=O_p(1)$ and $r_{3n}=O_p(1)$. Since $(\varphi^*-\varphi_0)=o_p(n^{1/2}D_n^{-1})$, we have $r_{2n}=O_p(1)$. Therefore, $R_n=O_p(1)o_p(1)=o_p(1)$. Now, the result follows from Lemma A.1 and

$$D_n(\hat{\varphi}-\varphi_0)=-\left[D_n^{-1}\sum_t \dot{\psi}(e_t)e_{\varphi t}e'_{\varphi t}D_n^{-1}+o_p(1)\right]^{-1}D_n^{-1}\sum_t \psi(e_t)e_{\varphi t}.$$

□

References

Ahn, S. K. (1993). "Some tests for unit roots in autoregressive-integrated-moving average models with deterministic trends", *Biometrika*, **80**, 855–868.

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, **59**, 817–858.

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

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

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

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

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

Martin, R. D. and Yohai, V. J. (1985). "Robustness in time series and estimating ARMA models", In *Handbook of Statistics 5* (P. R. Krshnaiah, ed.), 119–155, North-Holland, Amsterdam.

Mittnik, S. and Rachev, S. T. (1993). "Modeling asset returns with alternative stable distributions", *Econometric Reviews*, **12**, 261–330.

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. and Hall, A. (1991). "Testing for unit roots in autoregressive moving average models: an instrumental variable approach", *Journal of Econometrics*, **48**, 325–353.

Phillips, P. C. B. (1987). "Time series regression with a unit root", *Econometrica*, **55**, 277–301.

Said, E. S. (1991). "Unit roots test for time series data with a linear time trend", *Journal of Econometrics*, **47**, 285-303.

Said, E. S. and Dickey, D. A. (1985). "Hypothesis testing in ARIMA$(p, 1, q)$ models", *Journal of the American Statistical Association*, **80**, 369–374.

Schwert, G. W. (1989). "Tests for unit roots : A Monte Carlo investigation", *Journal of Business & Economic Statistics*, **7**, 147–159.

Shin, D. W. and Fuller, W. A. (1998). "Unit root tests based on unconditional maximum likelihood estimation for the autoregressive moving average", *Journal of Time Series Analysis*, **19**, 591–599.

Shin, D. W. and Lee, O. (1999). "*M*-estimation for regressions with integrated regressors and ARMA errors", Technical report, Ewha University, Seoul.

Shin, D. W. and So, B. S. (1999a). "Unit root tests based on adaptive maximum likelihood estimation", *Econometric Theory*, **15**, 1–23.

Shin, D. W. and So, B. S. (1999b). "Normal tests for unit root tests in autoregressive processes with infinite variances or finite variances", *Statistics & Probability Letters*, **44**, 387–397.

Yap, S. F. and Reinsel, G. C. (1995). "Results on estimation and testing for unit root in the nonstationary autoregressive moving-average model", *Journal of Time Series Analysis*, **16**, 339–353.