Hitotsubashi Journal of Economics 50 (2009), pp.87-105. © Hitotsubashi University

CONSTRUCTION OF STATIONARITY TESTS WITH LESS SIZE DISTORTIONS

Eiji Kurozumi*

Graduate School of Economics, Hitotsubashi University
Kunitachi, Tokyo 186-8601, Japan
kurozumi@stat.hit-u.ac.jp

Accepted March 2009

Abstract

We propose a (trend) stationarity test with a good finite sample size even when a process is (trend) stationary with strong persistence; this is useful for distinguishing between a (trend) stationary process with strong persistence and a unit root process. It could be considered as a modified version of Leybourne and McCabe's test (1994, LMC), but with a different correction method for serial correlation. A Monte Carlo simulation reveals that in terms of empirical size, our test is closer to the nominal one than the original LMC test and is more powerful than the LMC test with size-adjusted critical values.

Key Words: LM test; stationary; unit root
JEL classification: C12, C22

I. *Introduction*

The discrimination between the unit root and (trend) stationary hypotheses has been one of the primary interests in both theoretical and empirical time series analyses. While various types of unit root tests have been proposed following the seminal work of Dickey and Fuller (1979), the null hypothesis of stationarity is favored in some cases. For example, if we consider purchasing power parity (PPP), the null of PPP against the alternative of no PPP appears to be a natural choice. The PPP hypothesis can be considered to be equivalent to the hypothesis that the real exchange rate is stationary; hence, in this case, the null of stationarity becomes primary interest. Stationarity tests are also used as a complementary tool for unit root tests. For example, if unit root tests reject the null of a unit root for an economic variable while stationarity tests accept the null of stationarity, we can confirm that the economic variable is

* Gauss code provided by Professor Luts Kilian was helpful in conducting our simulation experiments. I am grateful to Professors Hidehiko Ichimura, Atsushi Inoue, Naoto Kunitomo, Pierre Perron, Katsuto Tanaka, Yoshihiro Yajima, and seminar participants held at Hitotsubashi University, Boston University, and the University of Tokyo for helpful comments on the earlier version. All errors are my responsibility. The earlier version of this paper was written while I was a visiting scholar of Boston University. This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology under Grants-in-Aid No. 17203016 and 18730142.

well characterized as a stationary process.

The most widely used stationarity tests are Kwiatkowski et al. (1992) (KPSS) and Leybourne and McCabe (1994) (LMC); the latter has been extended by Leybourne and McCabe (1999) to improve the finite sample power. Both papers consider the local level model and propose the Lagrange multiplier (LM) tests. The difference between these tests is that the KPSS test uses a nonparametric method to correct a series for serial correlation, while the LMC test considers a parametric correction. The advantage of the LMC test is that under the alternative, the test statistic diverges to infinity at a rate faster than that in the KPSS test, thus making it more powerful than the KPSS test. However, this does not necessarily imply that the LMC test is always more favorable than the KPSS test. This is because the assumptions made in KPSS are more general, which makes it applicable to a wide class of processes.

Since the asymptotic null distributions of the KPSS and LMC tests are free of nuisance parameters, we can control the sizes of these tests at least asymptotically. However, according to Caner and Kilian (2001), both the tests suffer from severe size distortions in finite samples when a process is strongly serially correlated. Müller (2005) explains the theoretical reason for the size distortion of the KPSS test, while Lanne and Saikkonen (2003) investigate the sources of the size distortion of the LMC test. The latter paper also proposes a new method to test the null of stationarity. Their method works better than the original LMC test when only a constant is included in a model; however, it still suffers from size distortions when the first-order autocorrelation exceeds 0.9 or when a model exhibits a linear trend. The size distortion problem may be mitigated by using size-adjusted finite sample critical values, as employed by Cheung and Chinn (1997), Rothman (1997), and Kuo and Mikkola (1999) among others; however, Rothman (1997) and Caner and Kilian (2001) point out that the use of size-adjusted critical values reduces the power of the tests, so that they have a tendency to fail to reject the null of stationarity even when a true process has a unit root. Stationarity tests other than the KPSS and LMC tests are proposed by Tanaka (1990), Saikkonen and Luukkonen (1993a, b), Choi (1994), Arellano and Pantula (1995), and Jansson (2004) among others, while tests for parameter constancy may also be considered as stationarity tests. See, for example, Nyblom and Mäkeläinen (1983), Nyblom (1986, 1989), and Nabeya and Tanaka (1988). However, none of these tests appear to be able to overcome the size distortion problem, and they tend to decisively reject the null of stationarity when a process is strongly serially correlated.

As discussed in Caner and Killian (2001) among others, if economic theory holds, some economic variables may be considered to be stationary but strongly serially correlated. Thus, we need to develop stationarity tests that do not suffer from size distortions when a process is strongly serially correlated and that have reasonable power against the alternative. In this paper, we propose a new test that is robust to such a situation. Our test is obtained by modifying the LMC test, and we will show that the size of the modified test is close to the nominal one under the null hypothesis, while its empirical power is considerably greater than that of the original LMC test with size-adjusted critical values. Thus, our test is useful for distinguishing between a serially correlated stationary time series and a unit root process.

The rest of the paper is organized as follows. Section II reviews the LMC test and investigates the sources of its size distortion. We show that the two important sources of the size distortion are the variation in y_t itself and the estimation error of the maximum likelihood estimator (MLE) of the autoregressive (AR) parameter in a model. We propose a new test in Section III and investigate its finite sample property in Section IV. Section V provides empirical examples, and Section VI concludes the paper.

II. *Sources of Size Distortions*

1. Review of the LMC Test

We consider the following local level model as used in LMC:

$$\phi(L)y_t=\mu+e_t,$$
$$e_t=\gamma_t+\varepsilon_t, \qquad \gamma_t=\gamma_{t-1}+v_t, \tag{1}$$

for $t=1,\cdots,T$, where $\phi(L)=1-\phi_1 L-\cdots-\phi_p L^p$ is the p-th order lag polynomial in the lag operator L with all the roots of $\phi(z)=0$ outside the unit circle, $\{\varepsilon_t\}\sim$ $i.i.d.(0,\sigma_\varepsilon^2)$ with $E[\varepsilon_t^4]<\infty$, $\{v_t\}\sim$ $i.i.d.(0,\sigma_v^2)$; $\{\varepsilon_t\}$ and $\{v_t\}$ are independent of each other. We assume that $\gamma_0=0$ because a constant term is included in the model. Since y_t is stationary when $\sigma_v^2=0$ and it is a unit root process when $\sigma_v^2>0$, the testing problem we consider is

$$H_0: \sigma_v^2=0 \quad \text{v.s.} \quad H_1: \sigma_v^2>0. \tag{2}$$

It is well known that the model (1) is equivalent to the ARIMA $(p,1,1)$ model up to the second moment:

$$\phi(L)\Delta y_t=(1-\theta L)u_t, \tag{3}$$

where $\Delta=1-L$, $\{u_t\}\sim$ $i.i.d.(0,\sigma_u^2)$ with $\sigma_u^2=\sigma_\varepsilon^2/\theta$ and $\theta=\{r+2-(r^2+4r)^{1/2}\}/2$ with $r=\sigma_v^2/\sigma_\varepsilon^2$ being a signal-to-noise ratio. In order to identify the model, we assume throughout the paper that $\phi(z)=0$ does not have a root at $z=1/\theta$. Note that θ is equal to 1 under the null of $\sigma_v^2=0$ while $\theta\in(0,1)$ under the alternative. Then, the null hypothesis H_0 may be interpreted as the null of $\theta=1$ for the ARIMA model (3).

As shown in KPSS and LMC, the LM test statistic for the testing problem (2) when $p=0$ is given by $S_T=T^{-2}\sum_{t=1}^{T}(\sum_{j=1}^{T}\hat{e}_j)^2/\hat{\sigma}_e^2$, where $\hat{e}_t$ is the regression residual of y_t on a constant and $\hat{\sigma}_e^2=T^{-1}\sum_{t=1}^{T}\hat{e}_t^2$. KPSS and LMC showed that under H_0,

$$S_T \Rightarrow \int_0^1 V^2(r)\,dr \tag{4}$$

where $\Rightarrow$ denotes weak convergence and $V(r)$ is a standard Brownian bridge. LMC also proved that under H_1, S_T diverges to infinity at the rate of T.

When $p\geq 1$, y_t must be corrected for serial correlation. KPSS proposed to replace the variance estimator $\hat{\sigma}_e^2$ with a heteroscedasticity and autocorrelation-consistent long-run variance estimator, while LMC considered a parametric approach. According to LMC, we first estimate (3) by using the ML method and obtain $\tilde{\phi}_1,\cdots,\tilde{\phi}_p$ the MLEs of the AR coefficients. Then, by defining

$$\tilde{y}_t=y_t-\tilde{\phi}_1 y_{t-1}-\cdots-\tilde{\phi}_p y_{t-p}, \tag{5}$$

we obtain $\tilde{e}_t$, the regression residual of $\tilde{y}_t$ on a constant. Finally, the LM test statistic, S_T^{LMC}, is constructed by replacing $\hat{e}_t$ in S_T with $\tilde{e}_t$ and it is defined by $S_T^{LMC}=T^{-2}\sum_{t=1}^{T}(\sum_{j=1}^{T}\tilde{e}_j)^2/\hat{\sigma}_e^2$. LMC showed that S_T^{LMC} has the same asymptotic property as S_T.

2. Sources of Size Distortions

It is known that the size of S_T is close to the nominal one when there is no AR structure in y_t ($p=0$) even for small samples such as $T=30$ (see, for example, Table 2 in KPSS). Then, it is natural to conjecture that the size distortion of the LMC test when $p \geq 1$ arises from the estimation error of the AR parameter. Note that $\tilde{y}_t$ is expressed as

$$\tilde{y}_t = \mu + e_t - r_t \quad \text{where} \quad r_t = (\tilde{\phi}_1 - \phi_1) y_{t-1} + \cdots + (\tilde{\phi}_p - \phi_p) y_{t-p}$$

as observed in Lanne and Saikkonen (2003). In this expression, r_t is asymptotically negligible because y_{t-i} is stationary under the null hypothesis while $\tilde{\phi}_i$ is $\sqrt{T}$ consistent for ϕ_i as shown in McCabe and Leybourne (1998). However, since the estimation error of the AR parameter is incorporated in r_t, we can infer that r_t may have a considerable effect on the test statistic in finite samples and that the size distortion of the LMC test is mainly induced by the finite sample behavior of r_t.

In order to further investigate r_t, we consider the AR(1) model here for ease of exposition. In this case, r_t is expressed as $r_t = (\tilde{\phi}_1 - \phi_1) y_{t-1}$; then, the variation in r_t arises from two sources: The estimation error of the MLE of ϕ_1 and the variance of y_t. First, we observe the relation between the variance of y_t and ϕ_1. For a given ϕ_1, the variance of y_t is expressed as $\sigma_\varepsilon^2/(1-\phi_1^2)$ under H_0; then, the variation in r_t, which is caused by y_t, increases as ϕ_1 approaches 1. For example, the variance of y_t is approximately $5\sigma_\varepsilon^2$, $10\sigma_\varepsilon^2$, and $50\sigma_\varepsilon^2$ for $\phi_1 = 0.9$, 0.95, and 0.99, respectively, while the variance for the *i.i.d.* case is σ_ε^2. Then, when ϕ_1 is close to 1, the size distortion of the LMC test is partly induced by large variance of y_t. See also Lanne and Saikkonen (2003).

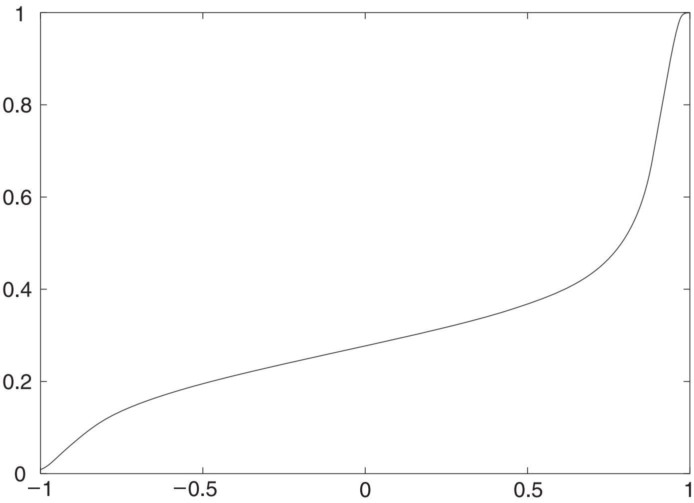
We can also observe that the estimation bias of $\tilde{\phi}_1$ and/or the variation in $(\tilde{\phi}_1 - \phi_1)$ also lead to the large variation in r_t. In order to observe the behavior of $\tilde{\phi}_1$ in finite samples, a simple Monte Carlo simulation is conducted. We consider the following simple data generating process (DGP):

$$y_t = 0.95 y_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim i.i.d.N(0,1), \quad t = 1, \cdots, 100. \tag{6}$$

We fit the ARIMA (1,1,1) model and estimate it by using the GAUSS-ARIMA routine. Section IV contains a detailed explanation of the estimation.

Figure 1 shows the cumulative distribution function (cdf) of $\tilde{\phi}_1$, which is obtained with 50,000 replications. We can observe that the empirical distribution of $\tilde{\phi}_1$ has a thick left-hand tail, which implies that it is severely biased, and the variation in $\tilde{\phi}_1 - \phi_1$ is very large. This poor behavior exhibited by $\tilde{\phi}_1$ appears to be due to the lack of the identification of the model. Note that the likelihood function is known to be relatively flat when the model has an MA unit root. In addition, since the AR coefficient is close to the MA coefficient, both the AR and MA lag polynomials nearly cancel out. As a result, it is difficult to identify the ARIMA model in our situation. In fact, in our simulation, the pair of the estimates of the AR and MA coefficients takes close values such as (0.96, 1) and (0.26, 0.45), even though their variations are large. Undoubtedly, the identification problem is not tackled as the sample sizes increase, but sample sizes such as 100 do not appear to be adequate to overcome this problem.

FIG 1. FINITE SAMPLE DISTRIBUTION OF THE MLE OF THE AR PARAMETER FOR THE ARIMA MODEL



III. *Stationarity Tests with Less Size Distortions*

1. The Variance-based Test

As observed in the previous section, the size distortion of the LMC test is mainly induced by the large variation in r_t, which is due to the large variance of y_t and the poor finite sample behavior of the MLE of the AR parameter. Thus, if we could reduce the variation in r_t from $\tilde{y}_t$, we would be able to construct a stationarity test with less size distortions. Lanne and Saikkonen (2003) proposed to eliminate the large variation in r_t by using the following expression:

$$\tilde{y}_t = \mu + \beta(L)\,\varphi(L)\Delta y_{t-1} + w_t + \beta(L)\,w_{t-1}, \tag{7}$$

where $\beta(L)$ and $\varphi(L)$ are $(p-1)$-th order lag polynomials. Using the residual obtained by the ML estimation of (7), they proposed to construct the LM test statistic S_T^{LS}. This test statistic is shown to be able to reduce the size distortion of the LMC test when the process is moderately serially correlated. In fact, the size of their test is closer to the nominal one as compared with the original LMC test when the AR coefficient is 0.8, as will be seen in Section IV. However, their test still suffers from size distortions when the first order autocorrelation is closer to 1. Hence, although their method works well in some cases, it is not a conclusive solution to the size distortion problem of the stationarity tests.

Instead of eliminating the variation in r_t from $\tilde{y}_t$, we consider constructing a test statistic that remains unaffected by one of the main sources of size distortions, the identification problem. Here, we focus on the estimator of variance σ_u^2 because it behaves relatively well even if some of the characteristic roots of the AR lag polynomial are close to those of the MA lag polynomial. In order to observe this, let us consider a simple AR(1) Gaussian model for ease of exposition. In this case, the vectorized form of (3) becomes $L(\phi)\Delta y = L(\theta)u$, where $\Delta y = [y_1, \Delta y_2, \cdots, \Delta y_T]'$, $u = [u_1, u_2, \cdots, u_T]'$, and $L(a)$ for a given a is defined as

$$L(a) = \begin{bmatrix} 1 & & \mathbf{0} \\ -a & 1 & \\ & \ddots & \ddots \\ \mathbf{0} & & -a \quad 1 \end{bmatrix}.$$

Then, we can see that $\Delta y \sim N(0, \sigma_u^2 \Omega(\phi, \theta))$ where $\Omega(\phi, \theta) = L(\phi)^{-1} L(\theta) L'(\theta) L'^{-1}(\phi)$. Note that under the null hypothesis $\theta = 1$, $\Omega(\phi, \theta) = \Omega(\phi, 1)$ becomes close to an identity matrix when ϕ is close to 1. In this case, the variance matrix of Δy is almost diagonal, $\sigma_u^2 I_T$, which implies that $\{\Delta y_t\}$ seems like an uncorrelated series. Then, we can expect that the realization of Δy_t is also almost uncorrelated. As a result, ϕ and θ would be estimated so that $\Omega(\phi, \theta)$ is close to an identity matrix, which implies that $\Omega(\tilde{\phi}, \tilde{\theta}) \simeq I_T$, irrespective of the difficulty of the identification of ϕ and θ. Since σ_u^2 is estimated by $\tilde{\sigma}_u^2 = T^{-1} \Delta y' \Omega(\tilde{\phi}, \tilde{\theta})^{-1} \Delta y$, we can observe that $\tilde{\sigma}_u^2$ is not considerably affected by the identification problem of ϕ and θ.

In practice, the estimator of variance $\tilde{\sigma}_u^2$ can be constructed by $\tilde{\sigma}_u^2 = T^{-1} \sum_{t=1}^T \tilde{u}_t^2$, where $\tilde{u}_t$ is recursively obtained by

$$\tilde{u}_t = \tilde{\theta} \tilde{u}_{t-1} + \tilde{e}_t - \tilde{e}_{t-1} \quad \text{with} \quad \tilde{u}_1 = \tilde{e}_1 \tag{8}$$

and $\tilde{e}_t$ is the regression residual of $\tilde{y}_t$ on a constant, as defined in Section II.1. The recursive formula (8) is obtained because the ARIMA model (3) is expressed as

$$\begin{aligned} u_t - \theta u_{t-1} &= \phi(L) \Delta y_t \\ &= (y_t - \mu - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p}) \\ &\quad - (y_{t-1} - \mu - \phi_1 y_{t-2} - \cdots - \phi_p y_{t-p-1}) \end{aligned}$$

and the sample analogue of the term within parentheses is given by $\tilde{e}_t$, which is the regression residual of $\tilde{y}_t$ on a constant.

In order to construct the test statistic, we also use the least squares-based estimator of variance of e_t, which is defined as $\hat{\sigma}_e^2 = T^{-1} \sum_{t=1}^T \hat{e}_t^2$; here, $\hat{e}_t$ is the regression residual of y_t on a constant and $y_{t-1}, \cdots, y_{t-p}$,

$$y_t = \hat{\mu} + \hat{\phi}_1 y_{t-1} + \cdots + \hat{\phi}_p \hat{y}_{t-p} + \hat{e}_t. \tag{9}$$

Note that $e_t = \varepsilon_t = u_t$ under the null hypothesis; hence, $Var(e_t) = \sigma_\varepsilon^2 = \sigma_u^2$. The following lemma gives the basic property of the two variance estimators.

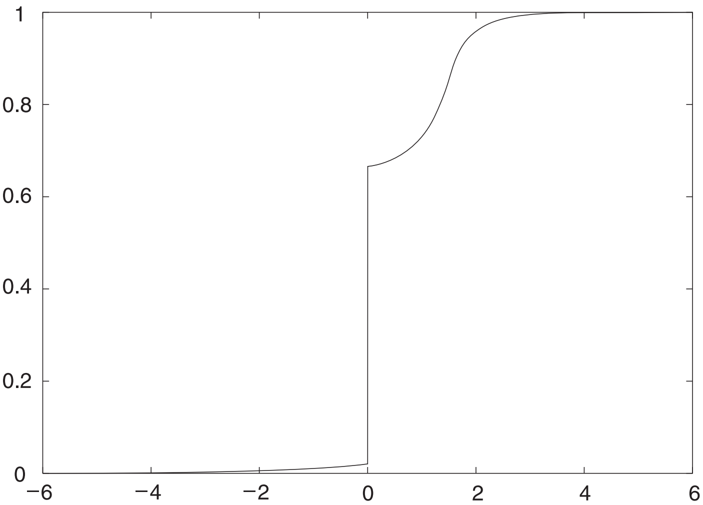
Lemma1 *Under H_0, both $\tilde{\sigma}_u^2$ and $\hat{\sigma}_e^2$ converge in probability to $\sigma_u^2 = \sigma_\varepsilon^2$, while under H_1,*

$$\tilde{\sigma}_u^2 \xrightarrow{p} \sigma_u^2 \quad \text{and} \quad \hat{\sigma}_e^2 \xrightarrow{p} \gamma(0) - \gamma'_{p-1} \Gamma_{p-1}^{-1} \gamma_{p-1},$$

where $\xrightarrow{p}$ denotes convergence in probability, $\gamma(j) = Cov(\Delta y_t, \Delta y_{t-j})$ for a given j, $\gamma_{p-1} = [\gamma(1), \gamma(2), \cdots, \gamma(p-1)]'$ and Γ_{p-1} is a $(p-1)$ square matrix whose (i,j) element consists of $\gamma(i-j)$ for $i, j = 1, \cdots, p-1$. The relation between the two probability limits under H_1 is given by

$$\gamma(0) - \gamma'_{p-1} \Gamma_{p-1}^{-1} \gamma_{p-1} > \sigma_u^2. \tag{10}$$

FIG 2. THE LIMITING NULL DISTRIBUTION OF V_T



This lemma shows that $\tilde{\sigma}_u^2$ is consistent under both the null and the alternative hypothesis, while $\hat{\sigma}_e^2$ is consistent for σ_u^2 only under the null hypothesis. Then, it appears natural to consider constructing a test statistic by taking the difference between these two estimators:

$$V_T = \frac{T(\tilde{\sigma}_u^2 - \hat{\sigma}_e^2)}{\tilde{\sigma}_u^2}.$$

We can see that this test statistic basically has the same structure as the conventional F test statistic. The asymptotic property of this statistic is given by the following theorem.

Theorem 1 (*i*) *Under* H_0,

$$V_T \xrightarrow{d} \eta^2 \int_0^1 \tilde{W}_0^2(s)\,ds - 2\eta \int_0^1 \tilde{W}_0(s)\,dW_0(s)$$
$$+ W_0^2(1) \int_0^1 e^{-2\eta s} ds - 2W_0(1)\tilde{W}_0(1) + \tilde{W}_0^2(1), \qquad (11)$$

where $W_0(r)$ *is a standard Brownian motion,* $\tilde{W}_0(r) = \int_0^r e^{-s\eta}\,dW_0(s)$, *and* $-\eta$ *is the limiting distribution of* $T(\tilde{\theta}-1)$, *which is defined in the Appendix.*
(*ii*) *Under* H_1, $V_T \to -\infty$ *at the rate of T.*

Since V_T diverges to $-\infty$ under the alternative hypothesis, we reject the null hypothesis of stationarity when V_T takes smaller values.

The limiting distribution of V_T is obtained by approximation using 1,000 observations with 50,000 replications, and the approximated limiting cdf is depicted in Figure 2. From the figure, we can see that the limiting cdf is not continuous but has a mass at the origin; $P(V_T < 0)$ is approximately 0.025, and the cdf suddenly increases at 0 from 0.02 to 0.67. The mass of the cdf occurs because of the discontinuity of the limiting distribution of $T(\tilde{\theta}-1)$. More precisely, the distribution of the normalized MLE of θ is known to have a mass at 0 and it is distributed in a unimodal form for the range less than 0, as shown by Cryer and Ledolter (1981), Sargan and Bhargava (1983), Davis and Dunsmuir (1996), and Tanaka (1996). This implies that the limiting distribution of $T(\tilde{\theta}-1)$, $-\eta$, takes a value equal to zero with positive probability. When $\eta=0$, we have $\tilde{W}_0(r)=W_0(r)$ and then $V_T=0$. As a result, V_T becomes equal to zero with positive probability. Due to the mass of the limiting distribution of V_T at 0, we cannot find some critical points such as 5% and 10% points and hence we may not be able to control the size of the test. Therefore, we cannot use V_T as a test statistic.

2. The Modified Test Using Information from V_T

Instead of using V_T as a test statistic, we regard it as an indicator to show whether the process appears stationary or nonstationary. In other words, the process may be considered to be stationary if $V_T \geq c$ for a given value of c and nonstationary otherwise. From Figure 2, the natural candidate for the value of c may be 0. In this case, $\lim_{T\to\infty} P(V_T \geq 0)$ is approximately 0.98 under the null hypothesis, while it is zero under the alternative. Then, the process is asymptotically correctly specified as stationary with probability 0.98, while it is consistently identified as a nonstationary process under the alternative.

Using the indicator function $I(\cdot)$, we propose to use the following estimator of the AR parameter in order to correct y_t for serial correlation:

$$\phi_i^* = \hat{\phi}_i I(V_T \geq 0) + \tilde{\phi}_i I(V_T < 0)$$

for $i=1,\cdots,p$, where $\hat{\phi}_i$ is the least squares estimator (LSE) of ϕ_i for the levels AR model (9), while $\tilde{\phi}_i$ is the MLE of the ARIMA model (3). In other words, ϕ_i^* is equal to the LSE of ϕ_i when $V_T \geq 0$, and it becomes equal to the MLE when $V_T < 0$. This estimator is motivated from the fact that the LSE does not suffer from the identification problem, so that the finite sample behavior of the LSE is better than the MLE under the null hypothesis when some of the roots of $\phi(z)=0$ are close to unity. Therefore, as far as the size of the test is concerned, it is more plausible to use the LSE of ϕ_i to correct y_t for serial correlation. However, it is known that the test statistic based on the LSE is inconsistent; hence, in order for the test to be consistent, serial correlation should be corrected not by the LSE but by the MLE of ϕ_i under the alternative. From the asymptotic distributional property, we can expect that V_T tends to take positive values under the null hypothesis even in finite samples and then ϕ_i^* becomes equal to $\hat{\phi}_i$ with high probability. Thus, the size of the stationarity test would be improved using ϕ_i^*. On the other hand, V_T tends to take negative values under the alternative from Theorem 1 (ii), which implies that it is often the case that ϕ_i^* equals $\tilde{\phi}_i$, so that the test with ϕ_i^* is expected to have non trivial power.

It is also possible to construct an estimator in a more general form such that

$$\phi_i^+ = \hat{\phi}_i f(V_T;\vartheta) + \hat{\phi}_i(1-f(V_T;\vartheta))$$

where $f(\cdot;\vartheta)$ is a function of V_T that depends on the parameter ϑ. For example, we may choose a smooth transition type function as $f(\cdot;\vartheta)$. In this paper, we chose $f(V_T;\vartheta)=I(V_T \geq 0)$ just because it is a very simple form and can be easily applied in practice.

Using ϕ_i^*, we correct y_t for serial correlation and modify the LM test statistic as follows:

$$S_T^* = \frac{1}{T^2} \frac{\sum_{t=1}^{T}(\sum_{j=1}^{t} e_j^*)^2}{\sigma_e^{*2}},$$

where e_t^* is the regression residual of $y_t^* = y_t - \phi_1^* y_{t-1} - \cdots - \phi_p^* y_{t-p}$ on a constant and $\sigma_e^{*2} = T^{-1} \sum_{t=1}^{T} e_t^{*2}$. We term the test based on S_T^* as the modified LM test. The asymptotic property of S_T^* is given in the following theorem.

Theorem 2 *S_T^* has the same limiting distribution as the LMC test statistic under the null hypothesis, while S_T^* is $O_p(T)$ under the alternative.*

From this theorem, we can see that our test has the same asymptotic property as the original LMC test. However, as shown in the next section, our test performs better than the LMC test in finite samples when the process is strongly serially correlated.

3. Extension to the trend stationary process

The modified LM test that is investigated in the previous subsection can be extended to the test for the null hypothesis of trend stationarity. The model is expressed as

$$\phi(L) y_t = \mu_0 + \mu_1 t + e_t, \qquad e_t = \gamma_t + \varepsilon_t, \qquad \gamma_t = \gamma_{t-1} + v_t,$$

and (3) becomes

$$\phi(L) \Delta y_t = \mu_1 + (1 - \theta L) u_t. \tag{12}$$

As in the previous subsections, we estimate (12) by using the ML method and obtain the estimated residual $\tilde{e}_t$ by regressing $\tilde{y}_t = y_t - \tilde{\phi}_1 y_{t-1} - \cdots - \tilde{\phi}_p y_{t-p}$ on a constant and a linear trend. We construct $\tilde{\sigma}_u^2$ in exactly the same way as in the previous subsection by using $\tilde{u}_t$, which is recursively obtained by (8). Further, we calculate $\hat{\sigma}_e^2$ from the regression residual of y_t on a constant, a linear trend, and $y_{t-1}, \cdots, y_{t-p}$. Then, we can construct V_T for the trend stationarity case. Once again, our preliminary simulation unfortunately shows that the limiting distribution of V_T has a mass at 0; therefore, we cannot use V_T as a test statistic. Hence, we construct ϕ_i^* as in the previous subsection and obtain the test statistic S_T^* from e_t^*, the regression residual of $y_t^* = y_t - \phi_1^* y_{t-1} - \cdots - \phi_p^* y_{t-p}$ on a constant and a linear trend.

IV. *Finite Sample Properties*

In this section, we investigate the finite sample property of the modified LM test that is considered in the previous section by using a Monte Carlo simulation. The data are generated according to the following system:

$$y_t = \phi y_{t-1} + \mu_0 + \mu_1 t + \gamma_t + \varepsilon_t, \quad \gamma_t = \gamma_{t-1} + v_t,$$

where $\{\varepsilon_t\} \sim i.i.d.N(0,1)$, $\{v_t\} \sim i.i.d.N(0, \sigma_v^2)$, and they are independent of each other. We set $\mu_0 = \mu_1 = 0$, $\gamma_0 = 0$, $\phi = 0.8$, 0.9, 0.95, 0.99, $\sigma_v^2 = 0$, 0.01, 0.1, 1, and the sample sizes are 100 and 200 (the first 100 observations are discarded). The level of significance is 0.05, and the number

TABLE 1. SIZE AND POWER OF THE STATIONARITY TESTS (Constant Mean)

		$T=100$			$T=200$		
ρ	ϕ	S_T^{LMC}	S_T^{LS}	S_T^*	S_T^{LMC}	S_T^{LS}	S_T^*
(a) size of the tests							
0	0.8	0.148	0.081	0.045	0.099	0.051	0.051
0	0.9	0.321	0.244	0.062	0.181	0.113	0.066
0	0.95	0.502	0.460	0.098	0.345	0.276	0.053
0	0.99	0.710	0.673	0.130	0.710	0.664	0.122
(b) nominal power of the tests							
0.01		0.613	0.500	0.110	0.755	0.635	0.267
0.1	0.8	0.831	0.778	0.273	0.887	0.829	0.311
1		0.990	0.963	0.846	1.000	0.996	0.909
0.01		0.730	0.686	0.128	0.765	0.721	0.112
0.1	0.9	0.925	0.836	0.598	0.992	0.963	0.775
1		0.999	0.974	0.947	1.000	0.995	0.983
0.01		0.757	0.680	0.209	0.829	0.767	0.334
0.1	0.95	0.950	0.838	0.733	0.999	0.981	0.935
1		0.994	0.962	0.975	1.000	0.997	0.998
0.01		0.717	0.590	0.284	0.881	0.763	0.622
0.1	0.99	0.952	0.734	0.786	0.993	0.954	0.981
1		0.991	0.891	0.985	1.000	0.989	1.000
(c) size-adjusted power of the tests							
0.01		0.484	0.434	0.116	0.715	0.634	0.265
0.1	0.8	0.719	0.715	0.278	0.867	0.829	0.309
1		0.919	0.937	0.850	0.999	0.996	0.909
0.01		0.401	0.407	0.098	0.693	0.685	0.100
0.1	0.9	0.362	0.262	0.565	0.908	0.915	0.769
1		0.608	0.529	0.935	0.976	0.985	0.981
0.01		0.216	0.193	0.072	0.304	0.251	0.332
0.1	0.95	0.132	0.031	0.449	0.509	0.334	0.933
1		0.370	0.112	0.762	0.663	0.583	0.998
0.01		0.022	0.017	0.089	0.010	0.004	0.263
0.1	0.99	0.019	0.001	0.506	0.046	0.000	0.745
1		0.136	0.001	0.768	0.193	0.004	0.855

of replications is 1,000. For the estimation of the ARIMA model, the GAUSS-ARIMA routine is used throughout the simulation. As in LMC and Caner and Kilian (2001), the likelihood function is evaluated for a grid of initial values for θ ranging from 0 to 1 in increments of 0.05, with the initial value of ϕ being fixed at $\theta-0.1$, in addition to the default values given by the GAUSS-ARIMA routine. Furthermore, as recommended by Lanne and Saikkonen (2003), we estimate the AR parameter by the generalized least squares method for each starting value of θ, and using a given θ and the estimate of ϕ as the initial values, we maximize the likelihood function based on the Kalman filter algorithm by using the GAUSS-Optmum routine. The estimates obtained by this procedure are also used as the initial values for the GAUSS-ARIMA routine.

Table 1 reports the simulation results of the LMC test, the Lanne and Saikkonen (LS) test, and the modified LM test when only a constant is included in the model. As is observed in

TABLE 2. SIZE AND POWER OF THE STATIONARITY TESTS (Linear Trend)

		$T=100$			$T=200$		
ρ	ϕ	S_T^{LMC}	S_T^{LS}	S_T^{*}	S_T^{LMC}	S_T^{LS}	S_T^{*}
(a) size of the tests							
0	0.8	0.079	0.082	0.046	0.055	0.049	0.043
0	0.9	0.183	0.207	0.067	0.093	0.086	0.043
0	0.95	0.317	0.341	0.127	0.191	0.216	0.063
0	0.99	0.453	0.470	0.136	0.477	0.491	0.146
(b) nominal power of the tests							
0.01		0.231	0.240	0.099	0.442	0.416	0.185
0.1	0.8	0.681	0.670	0.282	0.807	0.793	0.357
1		0.997	0.981	0.903	1.000	1.000	0.979
0.01		0.437	0.458	0.139	0.512	0.513	0.129
0.1	0.9	0.921	0.870	0.624	0.994	0.984	0.852
1		0.999	0.991	0.975	1.000	1.000	0.998
0.01		0.645	0.634	0.256	0.820	0.785	0.400
0.1	0.95	0.954	0.906	0.760	1.000	0.989	0.967
1		0.997	0.981	0.987	1.000	0.999	0.999
0.01		0.728	0.676	0.311	0.888	0.829	0.640
0.1	0.99	0.942	0.857	0.782	0.998	0.983	0.987
1		1.000	0.954	0.992	1.000	0.995	1.000
(c) size-adjusted power of the tests							
0.01		0.187	0.185	0.114	0.427	0.420	0.211
0.1	0.8	0.629	0.628	0.298	0.802	0.794	0.375
1		0.968	0.964	0.907	1.000	1.000	0.979
0.01		0.235	0.255	0.122	0.481	0.486	0.137
0.1	0.9	0.284	0.216	0.608	0.957	0.976	0.855
1		0.515	0.410	0.975	0.996	0.999	0.998
0.01		0.183	0.165	0.152	0.277	0.244	0.363
0.1	0.95	0.072	0.033	0.693	0.313	0.083	0.964
1		0.282	0.043	0.980	0.541	0.362	0.998
0.01		0.062	0.057	0.188	0.023	0.016	0.515
0.1	0.99	0.011	0.001	0.729	0.024	0.000	0.976
1		0.105	0.002	0.989	0.161	0.004	1.000

panel (a), the size of the LMC test is greater than the nominal size, 0.05; it is 14.8% for $\phi=0.8$ and it becomes greater than 50% for $\phi\geq 0.95$. The size of the LS test is relatively close to the nominal size for $\phi=0.8$, but it suffers from size distortions for $\phi\geq 0.9$. On the other hand, the modified LM test proposed in this paper has good finite sample size; the size of the modified test is close to 5% for $\phi\leq 0.9$ when $T=100$ and for $\phi\leq 0.95$ when $T=200$. Although none of the tests have size close to 0.05 for $\phi=0.99$, the modification proposed in this paper improves the finite sample performance of the stationarity tests under H_0.

Panel (b) reports the nominal powers of the three tests. Although the LMC and LS tests appear to be more powerful than the modified LM test, the powers of the former tests are mainly due to large size distortions.

Further, we investigate the size-adjusted power of the tests; the results of this investigation are summarized in panel (c). For $\phi=0.8$, the size-adjusted powers of the LMC and LS tests are higher than that of the modified LM test. However, for $\phi \geq 0.9$, the modified LM test is more powerful than the other two tests except in the case where the signal-to-noise ratio is very small, $\rho = \sigma_v^2 = 0.01$. The size-adjusted powers of the LMC and LS tests are very low for $\phi \geq 0.95$, which implies that if we use size-adjusted critical values for these tests, we rarely reject the null hypothesis of stationarity. This result is consistent with those found in Rothman (1997) and Caner and Kilian (2001).

When a linear trend is included, the size distortions of the tests are not as severe as those in the non-trending case. Overall, the relative performance of the tests is similar to the case in which only a constant is included in the model.

V. *Empirical Examples*

In this section, we apply the stationarity test proposed in the previous section to two macroeconomic data sets. The first data set comprises of the monthly yen/dollar real exchange rate measured in logarithms from January 1973 to December 2004. The sample size T is 384. The nominal exchange rate is obtained from the Bank of Japan, while the US and Japanese consumer price indices (CPI) are taken from the IMF's *International Financial Statistics*.

The results are summarized in Table 3. Only a constant is included in the model because the real exchange rate is not a trending series. The lag length p is selected using the modified Akaike information criterion proposed by Ng and Perron (2001), which is robust to the existence of a large negative MA root. Since the first-order autocorrelation of the demeaned series is 0.993, we expect that the LMC and LS tests suffer from size distortions. In fact, these two tests reject the null of stationarity at the 1% significance level, while the modified LM test statistic is insignificant. Further, we check the hypothesis in the opposite direction using the ADF-GLS test proposed by Elliott, Rothenberg, and Stock (1996), and the null of a unit root is not rejected by this test. This is consistent with the results obtained by the LMC and LS tests, but the rejection of the latter two tests might be due to size distortions. Since the null of stationarity cannot be rejected by the modified LM test, we should not conclude that the PPP hypothesis does not hold for the yen/dollar real exchange rate.

The second data set includes the annualized quarterly inflation rates calculated from the CPI for industrial and developing countries, which are obtained from the IMF's *International Financial Statistics*. The inflation rates are calculated by taking the logged differences, and they range from the first quarter in 1969 to the third quarter in 2004 ($T=143$). Again, these two series are highly persistent and then we need to carefully interpret the results of the stationarity tests. For industrial countries, the ADF-GLS test cannot reject the null of a unit root, while all

TABLE 3. EMPIRICAL RESULTS

Series	T	p	Corr	Stationarity tests			Unit root tests
				S_T^{LMC}	S_T^{LS}	S_T^*	ADF-GLS
Yen-dollar	384	2	0.993	17.114***	14.450***	0.119	−0.459
CPI(industrial)	143	9	0.983	10.353***	10.502***	0.414*	−0.477
CPI(developing)	143	5	0.957	2.629***	2.133***	0.343	−0.803

Notes: The symbols *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

the three stationarity tests reject the null hypothesis, although the modified LM test rejects the hypothesis of stationarity only at the 10% significance level. Then, judging from these results, the CPI for industrial countries is well characterized as a unit root process.

On the other hand, the results of the CPI for developing countries are mixed. The LMC and LS tests reject the null hypothesis but the modified LM test supports the stationarity of the CPI; on the other hand, the ADF-GLS test does not reject the unit root hypothesis. Thus, in this example, we cannot conclude whether the CPI for developing countries is well characterized as a unit root or stationary process.

VI. *Conclusion*

In this paper, we investigate how to construct stationarity tests with less size distortions. First, we propose to construct a variance-based test that is not affected by the identification problem. However, we find that we cannot control the size of the test because the pdf of the test statistic has a mass. Instead, we regard the variance-based test statistic as an indicator to show whether the process appears as a stationary or a unit root process. We propose the correction of serial correlation using either the least squares estimator of the AR parameter for the levels AR model or MLE of the AR parameter for the ARIMA model, depending on the indicator based on V_T. A Monte Carlo simulation shows that the modified LM test performs fairly well in finite samples. Although our simulation settings are limited, it appears that the modified LM test proposed in this paper is useful in distinguishing between a stationary process with strong persistence and a unit root process.

Appendix: Mathematical Proofs

Lemma A.1 *Under H_0, for $0 \leq r \leq 1$,*

$$T(\tilde{\theta}-1) \Rightarrow -\eta, \tag{A.1}$$

$$\frac{1}{\sigma_\varepsilon\sqrt{T}}\sum_{t=1}^{[Tr]}\varepsilon_t \Rightarrow W_0(r), \tag{A.2}$$

$$\frac{1}{\sigma_\varepsilon\sqrt{T}}\sum_{t=1}^{[Tr]}\tilde{\theta}^{\,t-j}\varepsilon_j \Rightarrow \int_0^r e^{-s\eta}dW_0(s) \equiv \tilde{W}_0(r), \tag{A.3}$$

$$\frac{1}{\sigma_\varepsilon^2 T}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1}\tilde{\theta}^{\,t-j-1}\varepsilon_j\right)\varepsilon_t \Rightarrow \int_0^1 \tilde{W}_0(s)\,dW_0(s), \tag{A.4}$$

where $[a]$ denotes the largest integer $\leq a$ and η is a global maximizer of

$$Z(\eta^*) = \sum_{k=1}^{\infty}\frac{\eta^{*2}}{k^2\pi^2+\eta^{*2}}W_k^2(1) + \sum_{k=1}^{\infty}\ln\left(\frac{k^2\pi^2}{k^2\pi^2+\eta^{*2}}\right)$$

with $\{W_k(r)\}_{k=0}^{\infty}$ as sequence of independent standard Brownian motions.

Proof of Lemma A.1: McCabe and Leybourne (1998) showed that $\tilde{\theta}$ has the same limiting property as the MLE for the case where the AR parameter is known. Then, $T(\tilde{\theta}-1)$ has the same limiting distribution as the normalized MLE for the ARIMA (0, 1, 1) model, and the limiting distribution of the latter is given by Theorem 2.2 in Davis and Dunsmuir (1996).

(A.2) holds by the functional central limit theorem. From Proposition A2 in Davis and Dunsmuir (1996), we can observe that $W_k(r)$ for $k\geq 1$ are the limits of

$$U_k,T=\sqrt{\frac{2}{T+1}}\sum_{j=1}^{T}\frac{\varepsilon_j}{\sigma_\varepsilon}\cos\left(\frac{jk\pi}{T+1}\right)+o_p(1)$$

and the joint convergence and independence of $W_0(r)$ and $W_k(r)$ for $k\geq 1$ are established by Theorem 2.2 in Chan and Wei (1988).

From (A.1) and the continuous mapping theorem, we have

$$\tilde{\theta}^{[Tr]}=\left(1+\frac{T(\tilde{\theta}-1)}{T}\right)^{[Tr]}\Rightarrow e^{-r\eta}.$$

Since η is a function of $\{W_k(r)\}_{k=1}^{\infty}$ by definition, it is independent of $W_0(k)$. This implies that $\tilde{\theta}$ is asymptotically independent of the partial sum of ε_t and hence we obtain (A.3) and (A.4) by using Theorem 4.1 in Hansen (1992).

A.1 Proof of Lemma 1

We first consider the probability limit of $\tilde{\sigma}_u^2$. Under the null hypothesis, y_t can be expressed as

$$y_t=\phi(L)^{-1}(\mu+\varepsilon_t)=c+x_t,\quad \text{where}\quad c=\phi(1)^{-1}\mu\quad \text{and}\quad x_t=\phi^{-1}(L)\varepsilon_t.$$

We can also see that $\tilde{e}_t=\tilde{y}_t-\bar{\tilde{y}}$, where $\bar{\tilde{y}}_t=T^{-1}\sum_{t=1}^{T}\tilde{y}_t$, because $\tilde{e}_t$ is the regression residual of $\tilde{y}_t$ on a constant. Since $\Delta y_t=\Delta x_t$ and $\tilde{y}_t=\mu+\varepsilon_t-(\tilde{\phi}_1-\phi_1)y_{t-1}-\cdots-(\tilde{\phi}_p-\phi_p)y_{t-p}$, equation (8) becomes

$$\begin{aligned}\tilde{u}_t-\tilde{\theta}\tilde{u}_{t-1} &= \tilde{e}_t-\tilde{e}_{t-1}\\ &= \tilde{y}_t-\tilde{y}_{t-1}\\ &= \Delta\varepsilon_t-(\tilde{\phi}_1-\phi_1)\Delta y_{t-1}-\cdots-(\tilde{\phi}_p-\phi_p)\Delta y_{t-p}\\ &= \Delta\varepsilon_t-(\tilde{\phi}-\phi)'\Delta z_{t-1}\end{aligned}$$

for $t\geq 2$, where $\tilde{\phi}=[\tilde{\phi}_1,\cdots,\tilde{\phi}_p]'$, $\phi=[\phi_1,\cdots,\phi_p]'$, and $z_{t-1}=[x_{t-1},\cdots,x_{t-p}]'$ with initial values $\tilde{u}_1=\tilde{e}_1$. This equation is expressed in the matrix form as

$$L(\tilde{\theta})\tilde{u}=L(1)\varepsilon-L(1)z(\tilde{\phi}-\phi)+r^*, \tag{A.5}$$

where $\tilde{u}=[\tilde{u}_1,\cdots,\tilde{u}_T]'$, $\varepsilon=[\varepsilon_1,\cdots,\varepsilon_T]'$, $z=[z_0,\cdots,z_{T-1}]'$, and $r^*=[r_1^*,0,\cdots,0]$ with $r_1^*=\tilde{e}_1-\varepsilon_1+(\tilde{\phi}-\phi)'z_0$. Thus, $\tilde{u}=L^{-1}(\tilde{\theta})L(1)\varepsilon-L^{-1}(\tilde{\theta})L(1)z(\tilde{\phi}-\phi)+L^{-1}(\tilde{\theta})r^*$ and we can observe that $\tilde{u}_t$ is expressed as

$$\tilde{u}_t=\varepsilon_t+(\tilde{\theta}-1)\sum_{j=1}^{t-1}\tilde{\theta}^{t-j-1}\varepsilon_j-(\tilde{\phi}-\phi)'z_{t-1}-(\tilde{\theta}-1)(\tilde{\phi}-\phi)'\sum_{j=1}^{t-1}\tilde{\theta}^{t-j-1}z_{j-1}+\tilde{\theta}^{t-1}r_1^* \tag{A.6}$$

for $t\geq 2$ and $\tilde{u}_1=\tilde{e}_1$. Here, note that $\tilde{e}_t=\varepsilon_t-\bar{\varepsilon}-(\tilde{\phi}-\phi)'(z_{t-1}-\bar{z})$, where $\bar{\varepsilon}$ and $\bar{z}$ are the sample means of ε_t and z_{t-1}, respectively. Then, since $(\tilde{\phi}-\phi)=O_p(T^{-1/2})$ as shown by

McCabe and Leybourne (1998), we can observe that $r_1^* = -\bar{\varepsilon} + O_p(T^{-1})$. In addition, since $\sum_{j=1}^{t-1} \tilde{\theta}^{t-j-1} z_{j-1}$ is shown to be $O_p(T^{-1/2})$ in the same way as (A.3), the second-last term in (A.6) is $O_p(T^{-1})$. Using these results and Lemma A.1, we have

$$\begin{aligned}\sum_{t=1}^{T} \tilde{u}_t^2 = &\sum_{t=1}^{T} \varepsilon_t^2 + (\tilde{\theta}-1)^2 \sum_{t=2}^{T} \left(\sum_{j=1}^{t-1} \tilde{\theta}^{t-j-1} \varepsilon_j \right)^2 + (\tilde{\phi}-\phi)' \sum_{t=2}^{T} z_{t-1} z'_{t-1} (\tilde{\phi}-\phi) \\ &+ \bar{\varepsilon}^2 \sum_{t=1}^{T} \tilde{\theta}^{2(t-1)} + 2(\tilde{\theta}-1) \sum_{t=2}^{T} \left(\sum_{j=1}^{t-1} \tilde{\theta}^{t-j-1} \varepsilon_j \right) \varepsilon_t - 2(\tilde{\phi}-\phi)' \sum_{t=2}^{T} z_{t-1} \varepsilon_t \\ &- 2\bar{\varepsilon} \sum_{t=1}^{T} \tilde{\theta}^{t-1} \varepsilon_t - 2\bar{\varepsilon}(\tilde{\theta}-1) \sum_{t=1}^{T} \tilde{\theta}^{t-1} \left(\sum_{j=1}^{t-1} \tilde{\theta}^{t-j-1} \varepsilon_j \right) + O_p(T^{-1/2}). \qquad \text{(A.7)}\end{aligned}$$

Since the first term on the right-hand side of (A.7) dominates the other terms, we can observe that $\tilde{\sigma}_u^2$ is consistent under the null hypothesis.

Under the alternative, it is well known that both $(\tilde{\theta}-\theta)$ and $(\tilde{\phi}-\phi)$ are $O_p(T^{-1/2})$; subsequently, the consistency of $\tilde{\sigma}_u^2$ is proved in a similar manner.

Next, we investigate the variance estimator $\hat{\sigma}_e^2$, which is estimated by the least squares method for the levels AR model. The consistency of $\hat{\sigma}_e^2$ under the null hypothesis is a well known result. In order to show the probability limit of $\hat{\sigma}_e^2$ under the alternative, we first express (1) as

$$\Delta y_t = \mu + (\rho-1) y_{t-1} + \psi' w_{t-1} + e_t,$$

where $\rho = \phi_1 + \cdots + \phi_p$, $\psi = [\psi_1, \cdots, \psi_{p-1}]'$ with $\psi_j = -\sum_{i=j+1}^{p} \phi_i$ for $j = 1, \cdots, p-1$, and $w_{t-1} = [\Delta y_{t-1}, \cdots, \Delta y_{t-p+1}]'$. By standard linear regression algebra, the normalized least squares estimator is expressed as

$$\begin{bmatrix} T(\hat{\rho}-1) \\ \hat{\psi} \end{bmatrix} = \begin{bmatrix} T^{-2}\sum_{t=1}^{T} \hat{y}_{t-1}^2 & T^{-1}\sum_{t=1}^{T} \hat{y}_{t-1}\hat{w}'_{t-1} \\ T^{-2}\sum_{t=1}^{T} \hat{y}_{t-1}\hat{w}_{t-1} & T^{-1}\sum_{t=1}^{T} \hat{w}_{t-1}\hat{w}'_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\sum_{t=1}^{T} \hat{y}_{t-1}\Delta\hat{y}_t \\ T^{-1}\sum_{t=1}^{T} \hat{w}_{t-1}\Delta\hat{y}_t \end{bmatrix},$$

where $\hat{y}_{t-1}$, $\hat{w}_{t-1}$, and $\Delta\hat{y}_t$ are the regression residuals of y_{t-1}, w_{t-1}, and Δy_t on a constant. Since y_t is a unit root process while Δy_t is stationary, we can see that $T^{-2}\sum_{t=1}^{T} \hat{y}_{t-1}^2 = O_p(1)$, $T^{-1}\sum_{t=1}^{T} \hat{y}_{t-1}\hat{w}'_{t-1} = O_p(1)$, $T^{-1}\sum_{t=1}^{T} \hat{w}_{t-1}\hat{w}'_{t-1} = O_p(1)$, $T^{-1}\sum_{t=1}^{T} \hat{y}_{t-1}\Delta\hat{y}_t = O_p(1)$, and $T^{-1}\sum_{t=1}^{T} \hat{w}_{t-1}\Delta\hat{y}_t = O_p(1)$. Then, the first matrix on the right-hand side of the above equation takes the form of an asymptotic upper block diagonal matrix, so that

$$T(\hat{\rho}-1) = O_p(1) \quad \text{and} \quad \hat{\psi} = \left(T^{-1} \sum_{t=1}^{T} \hat{w}_{t-1}\hat{w}'_{t-1} \right)^{-1} T^{-1} \sum_{t=1}^{T} \hat{w}_{t-1}\Delta\hat{y}_t + O_p(T^{-1}). \qquad \text{(A.8)}$$

The regression residual $\hat{e}_t$ is then expressed as

$$\begin{aligned}\hat{e}_t &= \Delta\hat{y}_t - \hat{\psi}'\hat{w}_{t-1} - (\hat{\rho}-1)\hat{y}_{t-1} \\ &= \Delta\hat{y}_t - \hat{\psi}'\hat{w}_{t-1} + O_p(T^{-1/2}),\end{aligned}$$

where the last equality holds because $(\hat{\rho}-1) = O_p(T^{-1})$ from (A.8) and $\hat{y}_{[Tr]} = O_p(T^{1/2})$ for $0 \le r \le 1$. Then, we have

$$\begin{aligned}\hat{\sigma}_e^2 &= T^{-1}\sum_{t=1}^{T}\Delta\hat{y}_t^2-2\hat{\phi}'T^{-1}\sum_{t=1}^{T}\hat{w}_{t-1}\Delta\hat{y}_t+\hat{\phi}'T^{-1}\sum_{t=1}^{T}\hat{w}_{t-1}\hat{w}_{t-1}'\hat{\phi}+o_p(1)\\ &= T^{-1}\sum_{t=1}^{T}\Delta\hat{y}_t^2 \qquad\qquad (A.9)\\ &\quad-\left(T^{-1}\sum_{t=1}^{T}\hat{w}_{t-1}'\Delta\hat{y}_t\right)\left(T^{-1}\sum_{t=1}^{T}\hat{w}_{t-1}\hat{w}'_{t-1}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T}\hat{w}_{t-1}\Delta\hat{y}_t\right)+o_p(1)\\ &\xrightarrow{p} \gamma(0)-\gamma'_{p-1}\Gamma_{p-1}\gamma_{p-1}, \qquad\qquad (A.10)\end{aligned}$$

where the second equality is established using (A.8) and the convergence in probability holds because Δy_t is a stationary process with mean zero.

Note that (A.9) is asymptotically equivalent to the normalized sum of squared residuals that are obtained from the regression of $\Delta\hat{y}_t$ on $\Delta\hat{y}_{t-1},\cdots,\Delta\hat{y}_{t-p+1}$. Since $\Delta\hat{y}_t$ is expressed in an AR(∞) form as

$$(1-\theta L)^{-1}\phi(L)\Delta\hat{y}_t=\hat{u}_t$$

from expression (3), we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{e}_t^2\geq\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{t,\bar{p}}^2 \qquad (A.11)$$

for any $\bar{p}\geq p$, where $\hat{e}_{t,\,p}$ is the regression residual of $\Delta\hat{y}_t$ on $\Delta\hat{y}_{t-1},\cdots,\Delta\hat{y}_{t-\bar{p}}$. Note that the left-hand side of (A.11) converges in probability to (A.10), while the right-hand side of (A.11) converges in probability to σ_u^2 when $\bar{p}\to\infty$ at a suitable rate as $T\to\infty$. Then, we have the inequality (10). Further, we note that the limit of $\hat{\sigma}_e^2$ can be considered as the mean squared error of the best linear prediction (see, for example, Brockwell and Davis, 1991), which coincides with innovation variance σ_u^2 if and only if Δy_t is expressed as an AR($p-1$) process. Since it is assumed that $\phi(z)=0$ does not have a root at $z=1/\theta$, (A.11) asymptotically becomes a strict inequality.□

A.2 Proof of Theorem 1

Note that under the null hypothesis,

$$\begin{aligned}y_t&=\mu+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\varepsilon_t\\ &=\mu^*+\phi' z_{t-1}+\varepsilon_t,\end{aligned}$$

where $\mu^*=\mu+c\rho$. Then, $\hat{e}_t$ is numerically equal to the regression residual of y_t on a constant and z_{t-1}. By denoting the LSEs of μ^* and ϕ as $\hat{\mu}^*$ and $\hat{\phi}$, we have

$$\hat{e}_t=\varepsilon_t-(\hat{\mu}^*-\mu)-(\hat{\phi}-\phi)'z_{t-1}.$$

Using this expression, we can observe that

$$\sum_{t=1}^{T}\hat{e}_t^2=\sum_{t=1}^{T}\varepsilon_t^2+T(\hat{\mu}^*-\mu^*)^2+(\hat{\phi}-\phi)'\sum_{t=1}^{T}z_{t-1}z'_{t-1}(\hat{\phi}-\phi)$$

$$-2(\hat{\mu}^*-\mu^*)\sum_{t=1}^{T}\varepsilon_t-2(\hat{\phi}-\phi)'\sum_{t=1}^{T}z_{t-1}\varepsilon_t+O_p(T^{-1/2}). \qquad \text{(A.12)}$$

The limiting distribution of the test statistic is obtained by taking the difference between (A.7) and (A.12). Here, note that $\sqrt{T}(\tilde{\phi}-\phi)$ has the same limiting distribution of the MLE of ϕ for a known θ as proved by McCabe and Leybourne (1998), which implies that $\tilde{\phi}$ has the same asymptotic property as the MLE of ϕ for the levels AR(p) model. Then, we can see that $\sqrt{T}$ $(\tilde{\phi}-\phi)$ has the same limiting distribution as $\sqrt{T}(\hat{\phi}-\phi)$. Hence, the terms associated with $\tilde{\phi}$ and $\hat{\phi}$ are canceled out from the difference between (A.7) and (A.12), so that

$$\begin{aligned}
T(\tilde{\sigma}_u^2-\hat{\sigma}_e^2)&=\sum_{t=1}^{T}\tilde{u}_t^2-\sum_{t=1}^{T}\hat{\varepsilon}_t^2\\
&=(\tilde{\theta}-1)^2\sum_{t=2}^{T}\left(\sum_{j=1}^{t-1}\tilde{\theta}^{\,t-j-1}\varepsilon_j\right)^2+\bar{\varepsilon}^2\sum_{t=1}^{T}\tilde{\theta}^{2(t-1)}+2(\tilde{\theta}-1)\sum_{t=2}^{T}\left(\sum_{j=1}^{t-1}\tilde{\theta}^{\,t-j-1}\varepsilon_j\right)\varepsilon_t\\
&\quad+2\bar{\varepsilon}\sum_{t=1}^{T}\tilde{\theta}^{\,t-1}\varepsilon_t+2\bar{\varepsilon}(\tilde{\theta}-1)\sum_{t=2}^{T}\tilde{\theta}^{\,t-1}\left(\sum_{j=1}^{t-1}\tilde{\theta}^{\,t-j-1}\varepsilon_j\right)\\
&\quad-T(\hat{\mu}^*-\mu^*)^2+2(\hat{\mu}^*-\mu^*)\sum_{t=1}^{T}\varepsilon_t+o_p(1).
\end{aligned}$$

Since it is shown from standard econometric theory that $T^{1/2}(\hat{\mu}^*-\mu^*)=T^{-1/2}\sum_{t=1}^{T}\varepsilon_t+O_p(T^{-1/2})$, we have, using Lemma A.1,

$$\begin{aligned}
V_T&=\frac{1}{\tilde{\sigma}_u^2}\left(\sum_{t=1}^{T}\tilde{u}_t^2-\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)\\
&\xrightarrow{d}\ \eta^2\int_0^1\tilde{W}_0^2(s)\,ds-2\eta\int_0^1\tilde{W}_0(s)\,dW_0(s)\\
&\qquad+W_0^2(1)\int_0^1 e^{-2\eta s}ds-2W_0(1)\tilde{W}_0(1)+W_0^2(1).
\end{aligned}$$

(ii) is obtained from the result in Lemma 1.□

A.3 Proof of Theorem 2

Note that the $\sqrt{T}$ consistency of the LSE of ϕ is a well known result, while McCabe and Leybourne (1998) showed that the MLE of ϕ is $\sqrt{T}$ consistent. As a result, we can see that ϕ^* is also $\sqrt{T}$ consistent. Following LMC's demonstration that the $\sqrt{T}$ consistency of the AR parameter is sufficient for the test statistic to have the same limiting distribution as S_T under H_0, we have the first result.

Regarding to the second result, we have the asymptotic inequality given by (10) under the alternative; thus, ϕ^* is equal to $\tilde{\phi}$ with a probability approaching 1. On the observation that $\tilde{\phi}$ is $\sqrt{T}$ consistent under the alternative, we have $S_T^*=O_p(T)$ as shown in LMC.□

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, pp.817-858.

Arellano, C., and Pantula, S. G. (1995), "Testing for Trend Stationarity Versus Difference Stationarity," *Journal of Time Series Analysis*, 16, pp.147-164.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Caner, M., and Kilian, L. (2001), "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate," *Journal of International Money and Finance*, 20, pp.639-657.

Chan, N. H., and Wei, C. Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, pp.367-401.

Cheung, Y. W., and Chinn, M. D. (1997), "Further Investigation of the Uncertain Unit Root in GNP," *Journal of Business and Economic Statistics*, 15, pp.68-73.

Choi, I. (1994), "Residual-Based Tests for the Null of Stationarity With Applications to U.S. Macroeconomic Time Series," *Econometric Theory*, 10, pp.720-746.

Cryer, J. D., and Ledolter, J. (1981), "Small-Sample Properties of the Maximum Likelihood Estimator in the First-Order Moving Average Model," *Biometrika*, 68, pp.691-694.

Davis, R. A., and Dunsmuir, W. T. M. (1996), "Maximum Likelihood Estimation for MA(1) Processes With a Root on or Near the Unit Circle," *Econometric Theory*, 12, pp.1-29.

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, pp.427-431.

Elliott, G., Rothenberg, T. J., and Stock, J. (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, pp.813-836.

Hansen, B. E. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, pp.489-500.

Jansson, M. (2004), "Stationarity Testing With Covariates," *Econometric Theory*, 20, pp.56-94.

Kuo, B. S., and Mikkola, A. (1999), "Re-Examining Long-Run Purchasing Power Parity," *Journal of International Money and Finance*, 18, pp.251-266.

Kwiatkowski, D., Phillips, P. C. B, Schmidt, P., and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, pp.159-178.

Lanne, M., and Saikkonen, P. (2003), "Reducing Size Distortions of Parametric Stationarity Tests," *Journal of Time Series Analysis*, 24, pp.423-439.

Leybourne, S. J., and McCabe, B. P. M. (1994), "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, pp.157-166.

Leybourne, S. J., and McCabe, B. P. M. (1999), "Modified Stationarity Tests With Data-Dependent Model-Selection Rules," *Journal of Business and Economic Statistics*, 17, pp.264-270.

McCabe, B. P. M., and Leybourne, S. J. (1998), "On Estimating an ARMA Model With an MA Unit Root," *Econometric Theory*, 14, pp.326-338.

Müller, U. K. (2005), "Size and Power of Tests of Stationarity in Highly Autocorrelated Time

Series," *Journal of Econometrics*, 128, pp.195-213.

Nabeya, S., and Tanaka, T. (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random Walk Alternative," *Annals of Statistics*, 16, pp.218-235.

Ng, S., and Perron, P. (2001), "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power," *Econometrica*, 69, pp.1519-1554.

Nyblom, J. (1986), "Testing for Deterministic Linear Trend in Time Series," *Journal of the American Statistical Association*, 81, pp.545-549.

Nyblom, J. (1989), "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, pp.223-230.

Nyblom, J., and Mäkeläinen, T. (1983), "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association*, 78, pp.856-864.

Rothman, P. (1997), "More Uncertainty About the Unit Root in U.S. Real GNP," *Journal of Macroeconomics*, 19, pp.771-780.

Saikkonen, P., and Luukkonen, R. (1993a), "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association*, 88, pp.596-601.

Saikkonen, P., and Luukkonen, R. (1993b), "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models," *Econometric Theory*, 9, pp.343-362.

Sargan, J. D., and Bhargava, A. (1983), "Maximum Likelihood Estimation of Regression Models With First Order Moving Average Errors When the Root Lies on the Unit Circle," *Econometrica*, 51, pp.799-820.

Tanaka, K. (1990), "Testing for a Moving Average Unit Root," *Econometric Theory* 6, pp.433-444.

Tanaka, K. (1996), *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, New York: Wiley.