Helsinki
Center
of
Economic
Research

Discussion Papers

Bootstrap Inference for Stationarity

Biing-Shen Kuo
National Chengchi University, RUESG and HECER

and

Ching-Chuan Tsong
National Chinan University

Discussion Paper No. 50
February 2005

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014
University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 50

Bootstrap Inference for Stationarity *

Abstract

Tests for the stationarity null due to Kwiatkowski et al. (1992) has been an indispensable part of tool kits for empirical time series research. The tests however display considerable size distortions in the presence of highly persistent but stationary processes. Using a local-to-unity framework, the paper offers an asymptotic explanation why the size problem comes into existence. The analysis shows that the tests fail to converge without a renormalization in the parameter space of concern. But it lends limited practical modifications to reducing the size bias, because of an unknown local-to-unity parameter that cannot be consistently estimated. We devise a parametric bootstrap scheme to account for the size distortions instead. Our bootstrap proposal is able to generate independent bootstrap re-samples, regardless of the dependence in the component representation of the considered series. Even in the problematic parameter space, simulations demonstrate that our bootstrap tests exhibit an excellent control over the empirical rejection probabilities, while maintaining a comparable power to the asymptotic counterparts, for both small and moderate sample sizes found in applications.

JEL Classification: C12, C14, C15, C22.

Keywords: stationarity test, size distortion, long-run variance, bootstrap, ARIMA.

Biing-Shen Kuo

Graduate Institute of International Trade
National Chengchi University
Taipei 116
TAIWAN

e-mail: bsku@nccu.edu.tw

Ching-Chuan Tsong

Department of Economics
National Chinan University
Puli 545
TAIWAN

e-mail: canny227@ms17.hinet.net

* This paper was completed while the first author visited the Research Unit for Economic Structure and Growth (RUESG), University of Helsinki in Winter 2004. Kuo thanks the RUESG for the hospitality and research support. Both authors also gratefully acknowledge grants from the National Science Council in Taiwan. Very helpful comments and suggestions from Bruce Hansen, Pentti Saikkonen and seminar participants in Institute of Economics, Academia Sinica and RUESG led to considerable improvements in the presentation.

1 Introduction

Tests for the stationarity null has appeared to be an indispensable part of tool kits when investigating time series property of aggregate variables. Information about stationarity of the observed series coming from evidence with the tests often complements to that from existing unit root tests. On the other hand, the spirit of testing for the stationarity null may be more consistent with classical hypothesis testing where the hypothesis to be tested under the null is the one that researchers believe in when testing for some economic theories. For instance, many international macroeconomists tend to hold the view that relative price levels between countries display at most transitory deviations from purchasing power parity, as a result of market forces. The hypothesis that real exchange rates are mean-reverting is thus natural to be tested under the null, and should not be rejected lightly unless strong evidence against it is established. A partial list for available stationarity tests that possess these features can consist of Kwiatkowski, Phillips, Schmidt and Shin (1992) (short for KPSS, hereafter), Saikkonen and Luukkonen (1993), and Leybourne and McCabe (1994).

For inference on stationarity to be able to be drawn reliably from empirical evidence, test statistics on which statistical decisions are based at least ought to demonstrate a robust control over the rejection frequencies. While a minimal requirement for the test statistics, a satisfactory size control has proved very difficult to meet when the series under test are stationary but highly persistent processes. Caner and Kilian (2001) offer a comprehensive account of the size problem with stationarity tests in the context. Simulations by KPSS (1992) already revealed potential size distortions about their tests. Specifically, there have had considerable over-rejections when the simulated data is drawn from simple autoregressive models with the persistence parameter closer to unit root, for sample sizes that usually encounter in practice. Moreover, an increase in samples does not help reduce but aggravate occurrence of rejections. The latter finding runs counter to the idea of large sample theory on which stationarity tests typically rely: asymptotic approximations yield more accuracy as sample increases. The existence of size problem immediately calls into questions the credibility of empirical evidence with stationarity tests. It is well understood that many observed time series in empirical macroceconomics and international finance often exhibit a strong persistence, and thus fall into the problematic parameter zone. In the presence of spurious rejections, a clear interpretation of rejections by the stationarity tests now turns out to be a formidable task, whether or not the true processes underlying the considered series are stationary.

The purposes of our paper are two fold. The first is to provide a theoretical underpinning for the sources of distorted sizes in stationarity tests. We concentrate on the KPSS tests that have been widely applied in empirical work. The popularity of the test is partly due to a much less computation efforts required by a semiparametric correction for error autocorrelation, compared to the parametric one used by the tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994). The 'semiparametric' correction is accomplished through an estimation of the "long-run variance" accounting for a wide range of short-run dynamics. In a way, the analysis carried out here is parallel to the development in the unit root testing. It is well known that conventional unit root tests, such as Phillips and Perron tests (Phillips and Perron, 1988) and their modified variants (Perron and Ng, 1996), subject to dramatic size distortions when the autoregressive root of the error process is close to the unit circle. Thus, the size problem with the KPSS tests shares a similar nature as that with the aforementioned unit root tests, where the estimation of the long-run variance plays an important role in shaping asymptotic behaviors of either class of tests. Using the local-to-unity framework, developed by Phillips (1987), the simulation evidence about the KPSS tests is able to be reconciled with our analytical results. Of particular concern emerging from our asymptotic analysis is that in the presence of stationary but highly persistent process, the KPSS tests can never converge to any sensible limit distributions without a re-normalization. Precisely, the test statistics diverge to infinity with probability one as samples pass to infinity. This explains why the size performance of the tests are worsened by increasing samples in simulations of Caner and Kilian (2001).

While our analytical results are useful in explaining why the KPSS tests suffer from size distortions, they do not lend practical solutions to reducing the problem. This is because the asymptotics obtained for the re-scaled KPSS tests under the null still depends on unknown local-to-unity coefficient that can not be consistently estimated. Therefore, our results are clearly indicative of the impossibility of mitigating the size distortion based on asymptotic arguments. Recognizing the size problem, some recent empirical studies by Cheung and Chinn (1997), Kuo and Mikkola (1999, 2001), and Caner and Kilian (2001) corrected for the bias by employing size-adjusted finite-sample critical values, in place of asymptotic counterparts. This is a reasonable attempt to correct for the size bias, but is potentially vulnerable to estimation risks. To compute the the size-adjusted critical values, researchers usually start with approximating the data by stationary autoregressive models, and then simulate the finite-sample null distributions by drawing samples from the fitted models as if they were true. It has been shown, however, that autocorrelation estimates tend to be biased downward, especially around the problematic parameter space of consideration (Marriott and Pope, 1954; Shaman and Stine, 1988). On the basis of the resulting critical values, the tests now are likely to overstate evidence in favor of the stationarity null. We conclude that these tests with size-adjusted critical values are incapable of delivering conclusive evidence.

Alternatively, the bootstrap that often provides more accurate approximations than the first-order asymptotic theory may constitute a useful approach to work on to improve inference on stationarity. Our second purpose is thus to develop a bootstrap procedure that can have actual finite sample rejection frequencies closer to asymptotic nominal levels. The development of such bootstrap stationarity tests does not come as straightforward as that of the bootstrap unit root tests. The major difficulty for doing so lies in a lack of a parametric model for bootstrap samples to be independently generated under the *null* of the KPSS tests. This is in contrast to bootstrapping the unit root tests that virtually relies on the estimated Dicky-Fuller regression to generate bootstrap re-samples. The unobserved component model from which the KPSS tests are derived does not directly render the possibility. We resolve the difficulty by making use of the equivalence in second-order moments between the unobserved component model and the parametric ARIMA model (Harvey, 1989). Thus, in estimating the distributions of the KPSS tests, bootstrap re-samples are drawn from the estimated ARIMA(p,1,1) obtained first from a fit to the series under study. It should be emphasized that the bootstrap is to reproduce the behavior of the KPSS tests under the stationarity null, whether or not the observed series comes from the null. This can now be easily ensured by setting the moving-average root equal to one in the fitted ARIMA model when resampling, corresponding to the null hypothesis that the variance of the random-walk equals to zero in the component representation.

Our bootstrap tests for stationarity perform remarkably well. Through simulations, we show that the bootstrap tests are able to have an excellent control over the size for sample sizes found in applications. In most experiments conducted for the parameter space of interest, the empirical rejection frequencies for our bootstrap stationarity tests are close to the nominal levels, in sharp contrast to the asymptotic tests. Furthermore, these results take place at no cost of power loss, where the bootstrap stationarity tests proposed here display a comparable power to or even minor gain over the asymptotic counterparts.

Our inquiries into the size of stationarity tests are not the first in the literature, and some recent studies along the line deserve attentions. Müller (2002) carefully investigates the effects of the choice of the long-run variance estimator on the size performance of the tests. Both his and our analytical work conclude the undesirable property of stationarity tests in the presence of highly persistent but stationary processes, at least 'asymptotically'. For the parametric stationarity tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994), Lanne and Saikkonen (2003) suggest a modification that corrects for the size bias, while Leybourne and McCabe (1999) propose an improved estimator of error variance to increase the power under the alternative. Neither paper considers the bootstrapping as a route to account for the size problem.

The remainder of the paper is organized as follows. Section 2 reviews the test statistics of KPSS. Our analytical results concerning the large sample behaviors of the tests in the presence of highly persistent but stationary processes, together with relevant discussions, are given in Section 3. Section 4 makes it clear the implementation of our bootstrap tests for stationarity, and gauges the empirical performance in terms of size and power. Section 5 re-examines power purchasing parity using both our bootstrap and asymptotic tests for stationarity. Section 6 concludes. All proofs are left to the appendix.

2 Test Statistics

Tests for the stationarity null mounted by Kwiatkowski et al. (1992) is derived from a component model that consists of a deterministic component, a random walk and a stationary error:

$$y_t = \sum_{i=0}^{m} \beta_i t^i + r_t + \epsilon_t, \ t = 1 \dots T, \tag{1}$$

where m could be either 0 or 1 that represents intercept or both intercept and deterministic time trend, respectively; and r_t is a random walk, in which

$$r_t = r_{t-1} + \zeta_t, \tag{2}$$

with fixed initial values r_0 set to zero without loss of generality, and ζ_t being independent stationary process. We assume that the stationary error $\epsilon_t = \sum_{i=0}^{\infty} \psi_i v_{t-i}$ where $\psi_0 = 1$, v_t iid$(0, \sigma_v^2)$ with an unknown distribution F. Let $\Psi(L) = 1 + \sum_{i=1}^{\infty} \psi_i L^i$. Further, $\{\epsilon_t\}$ is assumed to be invertible, i.e. $\Psi(L)$ is non-zero on unit circle, and $\sum_{i=0}^{\infty} i|\psi_i| < \infty$. The class of error processes considered therefore includes the stationary and invertible ARMA process as a special sub-class. Under these assumptions, it is known that $\{\epsilon_t\}$ can have an infinite order autoregressive representation: $\epsilon_t = \sum_{j=1}^{\infty} \phi_j \epsilon_{t-j} + v_t$, where $\Phi(L) = \Psi(L)^{-1} = 1 + \sum_{j=1}^{\infty} \phi_j L^j$.

Whether the series under consideration y_t is stationary however hinges on the variance of random-walk error, σ_ζ^2. Given that ϵ_t is a stationary error, when $\sigma_\zeta^2 > 0$, y_t comes to be stationary only after differencing. Alternatively, the series is stationary around a constant level or a trend, if $\sigma_\zeta^2 = 0$. The hypothesis of interest thus can be formulated as

$$H_0 : \sigma_\zeta^2 = 0 \qquad \text{versus} \qquad H_1 : \sigma_\zeta^2 > 0 \tag{3}$$

The KPSS test is derived based on the Lagrange multiplier (LM) principle. The derivation of the test is equivalent to those considered by Nyblom (1986) and Nabeya and Tanaka (1988) to test for random coefficients. All these statistics are LBI tests and thus possess the optimal property that attains the highest power locally. The calculation of the LM-type test statistics is not as complicated as the derivation. First, regress y_t against an intercept (if $m = 0$), or an intercept and time trend (if $m = 1$), and obtain the residuals, denoted by $\hat{u}_t$. That is, $\hat{u}_t = y_t - \sum_{i=0}^{m} \hat{\beta}_i t^i$ in which $\hat{\beta}_i$ is OLS estimates of β_i. Next, compute the partial sum of the residuals, $S_t = \sum_{i=1}^{t} \hat{u}_i$, and estimate the long-run variance of ϵ_t, based on Newey and West (1987):

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 + 2\frac{1}{T}\sum_{i=1}^{L} w(i, L) \sum_{t=i+1}^{T} \hat{u}_t \hat{u}_{t-i} \tag{4}$$

where $w(i, L) = 1 - i/(1 + L)$ is Bartlett kernel, and L is bandwidth. Here to obtain a consistent estimate for the long-run variance, L needs to be increased as T increases. In practice, many applications choose $L = [k(T/100)]^{1/4}$, where k is constant, and $[\cdot]$ is the

Table 1: Empirical Size performance of $KPSS_\tau(k)$ (no trend)

T	α	$KPSS_\tau(4)$	$KPSS_\tau(8)$	$KPSS_\tau(12)$
300	0.99	0.976	0.882	0.767
	0.98	0.963	0.832	0.698
	0.94	0.846	0.570	0.393
	0.30	0.073	0.062	0.056
600	0.99	0.995	0.957	0.887
	0.98	0.986	0.901	0.774
	0.94	0.847	0.575	0.382
	0.30	0.071	0.056	0.053

Note:
1. The rejection frequency in each entry is calculated based on a DGP $y_t = \alpha y_{t-1} + e_t$, with $e_t \overset{iid}{\sim} N(0,1)$, using asymptotic critical value at 5% nominal level (.146) in 5000 replications.
2. The test statistic $KPSS_\tau(k)$ is, as defined in the text, calculated with a bandwidth number set to $L = [k(T/100)]^{1/4}$.

largest integer function. following from Schwert (1989). We follow the same practice for simulations reported in the paper. The LM test statistic can then be formed by

$$LM = T^{-2}\hat{\sigma}^{-2}\sum_{t=1}^{T} S_t^2$$

We shall denote the test statistic by $KPSS_\mu(k)$, and $KPSS_\tau(k)$, respectively, given $m = 0$ or 1.

Kwiatkowski et al. (1992) establish that under the null and some regularity conditions, the limiting representations of $KPSS_\mu$ and $KPSS_\tau$ can be characterized as:

$$KPSS_\mu \Rightarrow \int_0^1 V_\mu^2(r)dr, \quad KPSS_\tau \Rightarrow \int_0^1 V_\tau^2(r)dr \tag{5}$$

where $\Rightarrow$ denotes weak convergence, $V_\mu(r) = W(r) - rW(1)$ is a standard Brownian bridge, $V_\tau(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(s)ds$, and $W(r)$ is a Wiener process. The tests reject the stationarity null for large values of the statistics by construction. Because these distributions are not standard and free of nuisance parameters, critical values at conventional significance levels needs to be computed via simulations, before the tests can have practical uses.

3 Sources of Size Distortions

It is by now well-documented that the KPSS tests subject to considerable size distortions in the presence of highly persistent but stationary processes. Caner and Kilian (2001) illustrate and re-affirm the points by providing systematic investigations on the tests. Before them, Monte Carlo simulations in KPSS (1992) have revealed potential size problems of their tests. Lee (1996) and Hobjin, Franses and Ooms (1998) focus on the effect of bandwidth selection on both size and power of the tests. Simulations of the sort delivers immediate relevance to interpretations of empirical evidence with the tests. It is not uncommon that aggregate time series that have been most examined are found to be highly persistent.

To appropriately address the size problem, Table 1 replicates partial simulation results reported in Caner and Kilian (2001), following their setup. As will be shown soon, our asymptotic analysis has much to do with the growth rate of the bandwidth. Thus, to permit a clear comparison, our simulations are conducted by considering three different bandwidth numbers respectively. To save space, we will not report the simulations for the tests with an intercept, as they share very similar qualitative outcomes as reported here. Notably, the tests all suffer from very noticeable size bias, between .40 and .98 as opposed to 5% nominal level, when the autoregressive coefficient is close to unit circle. To place an emphasis on the accuracy of asymptotic approximations, we will only report simulations for sample of sizes 300 and 600, considered to be fairly large samples in time series context. We summarize two important observations from the simulations. First, given a fixed sample size, the closer α is to one, the larger size distortions the tests display. What stands out from the simulations is that as α is closer to one, an increase in sample size does not help reduce but aggravate the degree of size distortions for the tests. Second, for fixed values of α, the tests subject to less size distortions, as the bandwidth increases.

Asymptotic theory provided by Kwiatkowski et al. (1992) does not appear to be capable of explaining the aforementioned simulations. It is worthwhile having alternative theoretical explanation why the size problem takes place. We will derive the local asymptotic distributions for the KPSS tests in a local-to-unity framework, following the development in the literature of the unit root testing. The asymptotics obtained from the framework has been found to yield more accurate approximations to the finite-sample distributions, when the autoregressive root in the underlying process is close to unit root (see, for example, Perron and Ng, 1996; Elliott et al., 1996; and Ng and Perron, 2001). It is expected that the finite-sample distributions of the tests can now be better characterized by the local asymptotic representation. Thus, in a local-to-unity setup, we define the data generating process for our analysis as follows.

Definition 1: Let the series under test y_t be generated by:

$$y_t = \sum_{i=0}^{m} \beta_i t^i + u_t, \quad u_t = (1 + c/T)u_{t-1} + \epsilon_t \tag{6}$$

where $c < 0$, $u_0 = 0$ and ϵ_t is mean zero stationary error where $\sigma_\epsilon^2 = \lim_{T\to\infty} T^{-1}E[\sum_{j=1}^{T} \epsilon_j]^2$ is nonzero and finite.

The random walk component is left out, because we are only interested in the asymptotic distributions of the test statistics under the stationarity null. If the autoregressive coefficient is fixed, rather than depending on sample size, the data generating process defined is then one of special cases considered in KPSS that satisfies the assumed strong mixing conditions. Asymptotically, an AR process with fixed coefficient would behave differently from a near-unit-root process defined in (6). In general, the stochastic order of a near-unit-root process is $O_p(\sqrt{T})$ as that of a unit root process. It is this asymptotic property that results in spurious rejections for stationarity tests as observed from the simulations. We are now in a position to state the asymptotics of the tests under the near unit root setup.

Theorem 1 Let y_t be generated as in Definition 1. As $T \to \infty$ and $L = o(T^{1/2})$.[1] Under the null hypothesis that $H_0 : \sigma_\zeta^2 = 0$,

1. If $m = 0$,

$$(\frac{L}{T})KPSS_\mu(k) \implies \frac{\int_0^1 (\int_0^r \bar{J}_c(s)ds)^2 dr}{\int_0^1 \bar{J}_c^2(r)dr} \tag{7}$$

2. If $m = 1$,

$$(\frac{L}{T})KPSS_\tau(k) \implies \frac{\int_0^1 (\int_0^r \tilde{J}_c(s)ds)^2 dr}{\int_0^1 \tilde{J}_c^2(r)dr} \tag{8}$$

[1]The rate is the same as adopted in the asymptotic argument of KPSS, though their simulations consider $L = [k(T/100)]^{1/4}$.

where $\bar{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$, $\tilde{J}_c(r) = J_c(r) + (6r-4)\int_0^1 J_c(s)ds + (6-12r)\int_0^1 sJ_c(s)ds$, and $J_c(r) = \int_0^r e^{c(r-x)}dW(x)$.

The limiting representation for the test statistics just derived is a more useful guide to the finite-sample performance of the tests as in Table 1 than given in (5). First, we note that the test statistics under the null has to be re-scaled before having sensible limiting distributions. In other words, under the local-to-unity setup, $KPSS_i = O_p(\frac{T}{L})$, with $i = \mu,\ \tau$. This clearly suggests that without re-normalization, the test statistics are divergent as samples increase, but the bandwidth works in an opposite manner by slowing down the divergent speed of the test statistics. The limit representations are now consistent with the simulation evidence in Table 1, and thus indicative of an inadequacy of the KPSS asymptotic approximations to the small-sample distributions of the tests in the presence of highly persistent but stationary processes.

The sources of the size distortions of the tests can be attributed to two forces. As can be seen from the proofs, the squared partial sums in the numerator of the tests, $\sum S_T^2$, have a stochastic order of $O_p(T^4)$ in the local-to-unity context, as opposed to $O_p(T^2)$ in the standard asymptotics. Further, similar to its asymptotic behavior under the alternative, the estimated long-run variance in the denominator is no longer consistent (i.e. $o_p(1)$) but diverges at a rate of $O_p(TL)$.

The limit results have an important implication for the power of stationarity tests. Intuitively, the test statistics make use of the properties of non-stationary data. Under the alternative of a random walk, the partial sums of the residuals behave as those in spurious regression (see Phillips, 1986), and converge to random variables only after re-normalization. As established in KPSS, the tests are thus consistent under the alternative hypothesis at an order $O_p(\frac{T}{L})$. When the observed series is generated from a highly persistent but stationary process, the partial sums of the residuals resemble those under the alternative in the limit. As a result, the stochastic orders of the tests are the same for under both the alternative of integrated process and the null of highly persistent but stationary process. Based on the standard asymptotic critical values, rejections by the stationarity tests may well result from an underlying process that is either difference stationary, or highly persistent but stationary. In other words, the tests are lack of a discriminatory power between highly persistent but stationary process and an integrated process.

Of much practical relevance to the analytical results is the associated power loss when using the size-corrected critical values. These critical values are calculated by first drawing simulated samples from an estimated autoregressive model fit to the data. The simulated distributions of the test statistics are then constructed based on the pseudo data. The simulated distributions are to mimic the weak limit distributions in (7) or (8). Therefore, as a reflection of the asymptotic counterparts, the simulated test distributions shift more to the right, as the local-to-unity parameter or the persistence rate c is closer to zero. As a result that the simulated test distributions now overlap more with that under the alternative, the tests using the resulting critical values become less capable of detecting against any fixed alternatives. This explains the simulation findings of Caner and Kilian (2001) where the size-adjusted tests experience a sizable power loss, even when there is only a slightly more persistence increase in the null process. For example, their Table 3 shows that the rejection rate decrease from 29% to 20% when the persistence rate increases only by around .02% for a sample of size 100. The gains from using size-controlled critical values do not come without cost.

It is very tempting to suggest practical remedy for size distortions of the tests by making use of the derived limiting representations. There are however difficulties for doing so. As in the unit root testing literature, it proves implausible to consistently estimate the local-to-unity parameter c. Even with a known c, the resulting asymptotic critical values are only useful in situations where highly persistent and stationary series present. When the underlying true processes are not generated from the problematic region, the standard asymptotic counterparts remain applicable. Researchers however can not come to have information concerning the nature of the observed series before testing.

In view of a reduction in the size bias by an increase in the bandwidth, it is a natural question to pose what if the bandwidth increases at a speed higher than $o(\sqrt{T})$. The limiting representations of Theorem 1 in fact hold for $o(T^{1/4}) \leq L \leq o(T)$, not necessarily $o(\sqrt{T})$, where the rate gives the consistency of the long-run variance under the null in the standard asymptotics (see Andrews, 1991). In particular, Müller (2002) reaches a similar conclusion by allowing for the bandwidth grows at $o(T)$. More than that, his asymptotic analysis is much concerned with other important classes of long-run variance estimators that are data-dependent. His theoretical results, explaining simulation evidence in Lee (1996) and Hobjin, Franses and Ooms (1998), all show that it is still difficult to have size of the test under good control in the presence of highly persistent but stationary process, because the limit distributions of the tests under these data-dependent estimators do not come close to those obtained in the standard asymptotics. Recently, Kiefer and Vogelsang (2002) suggest another class of long-run variance estimator using bandwidth equal to sample size. Given our limit representations, by cancelling out the re-scaling factor ($L/T = 1$), it would seem promising that the tests can display satisfactory size behavior when $L = O(T)$. Unfortunately, the tests are always equal to .5 for any data, one of the cases that using large bandwidth does not work. Together, the large-sample analysis appears to be suggestive of alternative approaches to reducing the size distortion in the problematic region for the tests.

4 A Bootstrap Re-sampling Proposal

When the asymptotics fails to yield accurate approximations to the small-sample distributions, researchers often turn to the bootstrap. There have been an increasing interest in applying the method when testing for unit roots (for example, Ferretti and Romo, 1995; Nankervis and Savin, 1996; Psaradakis, 2001; and Park, 2003). The motivation for employing such methods in testing for unit root is as clear as here for stationarity where the asymptotic critical values are not so reliable as they have been promised to be. It has been proven that the bootstrap unit root tests can provide more desirable accuracy in approximations than the asymptotic counterparts, in particular in the presence of negative moving average errors that the asymptotic unit root tests have difficulty to deal with.

Implementation of the bootstrap unit root tests entails generating independent random re-samples from the data for estimating the small-sample distributions of the tests. The procedure is not difficult and is made possible by the use of an autoregressive model on which the unit root test are built. In practice, one starts with fitting an autoregression to the data. The small-sample distribution of the unit root test is estimated by its empirical distribution under sampling from the fitted model. The estimated model is to capture the dependence structure of the underlying data generating process, thereby being able to reduce the DGP to independent random sampling. It is important because whether the bootstrap unit root tests achieve improvements in accuracy, relative to the asymptotic approximations, hinges on if the bootstrap samples can be drawn independently.

The situation is not as straightforward to bootstrap the stationarity tests. Unlike the bootstrap unit root tests that rely on autoregressive models, it is less clear how to generate the independent random re-samples from the 'unobservable' component model as in (1) from which the series under study is generated. Thus, the component model does not lend itself readily to a parametric model to capture the autocorrelation in the series from which re-samples can thus be independently drawn.

The problem with a lack of a suitable parametric model to generate bootstrap samples is in fact more troublesome than it appears. It should never be over-emphasized that the bootstrap is to estimate the *null* finite-sample distributions of the test statistics in the testing context. With the autoregrssion model, the null distributions of the unit root tests to be bootstrapped can be obtained simply by replacing the largest estimated autoregressive root with a unit root in the process of generating bootstrap samples (see references cited above), regardless of whether the studied series is drawn from either the null or the alternative. To place the null constraint into re-sampling schemes is a crucial step to deliver proper size for the test statistics bootstrapped. In the case of the unit root models, Basawa et al. (1991) show that the sampling algorithm without considering a unit root restriction is not valid. The idea of sampling 'restricted' regression errors (under the unit root null) has been also emphasized in Nankervis and Savin (1996). But in the case of bootstrapping the stationarity tests, it might well be an more difficult notion to put into effect to place the restriction of zero random-walk variance, as in (3), without a parametric model for sampling.

We develop a re-sampling scheme that is able to overcome the problems when bootstrapping the stationarity tests. The idea of the sampling proposal is built on the equivalence in second-order moments between an unobservable component model and a parametric ARIMA model (see Harvey, 1989). In other words, the ARIMA representation is a reduced form of the structural component model. For example, for $m = 0$ (models with intercept only), if the regression error ϵ_t and random-walk error ζ_t are iid and independent, the component model in (1) and (2) after differenced is an MA model that can be expressed as $\Delta y_t = (1 - \theta L)\eta_t$, where η_t are iid$(0, \sigma_\eta^2)$ with $\sigma_\eta^2 = \sigma_\epsilon^2/\theta$. The relation of the parameters between the component model and the ARIMA model is found to be

$$\theta = \frac{1}{2}\{\frac{\sigma_\zeta^2}{\sigma_\epsilon^2} + 2 - (\frac{\sigma_\zeta^4}{\sigma_\epsilon^4} + 4\frac{\sigma_\zeta^2}{\sigma_\epsilon^2})^{1/2}\},$$

where $\sigma_\zeta^2/\sigma_\epsilon^2$ is the so-called signal-to-noise ratio. Note that the stationarity null that $\sigma_\zeta^2 = 0$ amounts to $\theta = 1$ in the ARIMA representation, a non-invertible moving average component. Thus testing for the stationarity null based on the component model is equivalent to testing if there is a moving average unit root using the ARIMA model, the idea exploited in constructing the tests proposed by both Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994). Thus, our re-sampling scheme is made available by making use of the corresponding parametric ARIMA model that can reduce the data to independent re-samples. Furthermore, imposing a moving average unit root in the sampling procedure renders it feasible to estimate the bootstrap null distribution of the tests.

We now spell out our re-sampling schemes.

1. Given a sample $\{y_t\}_{t=1}^T$ generated from (1) and (2), fit an ARMA(p,1) to the differenced series $\Delta y_t (= y_t - y_{t-1})$ using the maximum likelihood principle. Specifically, if $m = 0$, the model to be estimated is $\Delta y_t = \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \eta_t - \theta\eta_{t-1}$, while if $m = 1$, $\Delta y_t = \beta + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \eta_t - \theta\eta_{t-1}$. The resulting estimated parameters and residuals are denoted by $\hat{\alpha}_i$, $\hat{\beta}$ (if $m = 1$), $\hat{\theta}$, and $\hat{\eta}_t$.
2. Center the residuals $\hat{\eta}_t$ by $\bar{\eta}_t \equiv \hat{\eta}_t - \frac{1}{T-1}\sum_{t=2}^T \hat{\eta}_t$.
3. Draw a bootstrap sample of size T without replacement from the empirical distribution function of the centered residuals $\{\bar{\eta}_t\}$, and denote it by η_t^*.
4. Set the initials that $y_1^* = y_1, \cdots, y_p^* = y_p$, and generate the bootstrap samples $\{y_t^*\}$ based on the recursive relation that $\Delta y_t^* = \sum_{i=1}^p \hat{\alpha}_i \Delta y_{t-i}^* + \eta_t^* - \eta_{t-1}^*$ $(m = 0)$, or $\Delta y_t^* = \hat{\beta} + \sum_{i=1}^p \hat{\alpha}_i + \eta_t^* - \eta_{t-1}^*$ $(m = 1)$.
5. Calculate $KPSS_\mu(k)$ and $KPSS_\tau(k)$ using $\{y_t^*\}_{t=1}^T$, denoted by $KPSS_\mu^*(k)$ and $KPSS_\tau^*(k)$, respectively.
6. Repeat step 3 to step 4 NB times.
7. Compute the empirical distribution function (edf) for $KPSS_\mu^*(k)$ or $KPSS_\tau^*(k)$, and use the empirical distribution function as an approximation to the cumulative distribution function (cdf) of the bootstrap null distribution for the test statistics.
8. Compute the intended bootstrap critical values, based on the bootstrap null distribution in the preceding step.

Some words about the scheme are worth mentioning. In step 1, based on the equivalence of a reduced form to a component model in second moments, the MA part is to re-parameterize the stationarity property of the data in the ARMA representation. On the other hand, the AR part as an approximation to the assumed infinite-order moving average errors is to capture the dependence structure in data. Of which entertains the highly persistent but stationary processes under consideration. Note that while under the alternative hypothesis, the maximum likelihood estimators of AR and MA coefficients are consistent by the standard asymptotic theory, under the stationarity hypothesis when $\theta = 1$, the consistency holds still following from Potscher (1991). The lag order in general needs to increase with sample size. In practice, the optimal lag order for a time series is usually chosen by some information criteria. We follow this practice in the subsequent Monte-Carlo study.

Centering the residuals in step 2 is justified by two reasons. It not only takes into account that the underlying population distribution has zero expectation, but also works to reduce the downward bias of the autoregression coefficients in small samples (see Horowitz, 2001).

Step 4 is known to be the recursive bootstrap. There have had some comparable re-sampling procedures to the recursive bootstrap in the literature, notably the moving block bootstrap (Künsch, 1989) and stationary bootstrap (Politis and Romano, 1994). These procedures are all capable of reproducing error dependence structure. In contrast to the latter two procedures, the recursive bootstrap is of parametric nature by making use of the autoregression model. Bühlmann (1997) and Horowitz (2001) both emphasize the merit of the use of the recursive bootstrap when the DGP is linear as in our case. It has been found that the recursive bootstrap appears to be the best bootstrap method that provides significant accuracy gains from taking advantage of the knowledge of the linear structure in the DGP. The gains will be embodied in the simulations reported below.

How can the *null* bootstrap distributions of the tests be estimated? Step 4 is the key to yield such estimates. The stationarity null is now conveniently placed into the parametric re-sampling schemes by imposing a moving average unit root, regardless of whether the data is drawn from either the null or the alternative. The next section will show through simulations immediately the empirical relevance for the size performance of the bootstrap tests without imposing the constraint of a moving average unit root when re-sampling.

5 Monte-Carlo Study

This section is devoted to access the finite-sample performance of our bootstrap testing procedures via simulations. The simulation setup under the null is the same as that in Table 1 where the DGP is an AR(1): $y_t = \alpha y_{t-1} + e_t$ with e_t is nid(0,1). Under the alternative, the DGP is assumed to be $y_t = r_t + \epsilon_t$, and $r_t = r_{t-1} + \zeta_t$, with $\epsilon_t \sim \text{nid}(0, 1)$ and $\zeta_t \sim \text{nid}(0, \sigma^2_\zeta)$. To gauge the extent to which the bootstrap tests perform, different degrees of persistence, signal-to-noise ratio and sample sizes are considered by varying α (= {.0,.3,.5,.8,.82,.84,.86,.88,.90,.92,.94,.96,98 }), σ^2_ζ (={ 1, .1, .01, .001,.0001}) and T (={30, 50, 100, 150, 300, 600}). Typically in applications, simple AR processes with an autoregressive parameter greater than .8 are regarded highly persistent. Sample sizes ranging from 30 to 150 are those usually encountered in empirical time series studies. On the other hand, experiments with sample sizes 300 and 600 is to investigate and to represent the performance of the tests in large samples. The rejection rates for both the asymptotic tests and the bootstrap tests are computed and reported at nominal 5% level. The 5% asymptotic critical values for different models are simulated and available in KPSS (.146 for models with intercept only, and .463 for models with both intercept and time trend). Replications for the asymptotic tests are 5,000, while 1,000 for the bootstrap counterparts, with 100 bootstrap re-samples in each replication. A smaller replication number considered for the latter is due to many more computations involved in maximum likelihood estimations.

All the estimations of the autoregressive and moving-average coefficients are carried out by the GAUSS-ARIMA procedure, and initial values of these parameters need to be given prior to estimations. In general, the estimations depend on the choices of initial values, yielding different local maximums of the likelihood for any particular samples. It is particularly sensitive to the choice of initial values for the boundary estimations about the unit root moving average parameter under the null (see Kuo, 1999). Instead of performing the grid search over the parameter spaces as in Leybourne and MaCabe (1994), our strategy in the simulations is to select initial values to be the true values of the parameters set in the DGP. This choice appears to be quite reasonable and less costly in computations, as our experience with different choices of initial values indicates that a global maximum has been most likely to be assured when the initial values are selected to be equal to or close to the true ones. The available distributional results concerning maximum likelihood estimations are mostly built on the assumption of the existence of a unique global maximum over the limit likelihood (see for example, Amemiya, 1985). As a reflection in practice, we consider it very important locating the global maximum of the sample likelihood for correct inferences. To achieve that, it requires experimenting various sets of starting values, though here our simulations skip the searching process by having the 'good' guesses of them.

To compute the asymptotic tests, it entails selecting a bandwidth number. Again as in Table 1, we report the associated results using three different bandwidth numbers. Another practical consideration is how to choose an appropriate autoregression lag length in Δy_t when computing the bootstrap tests. A correct selection of the AR order is an important prerequisite for reproducing samples appropriately. Because researchers generally do not have the luxury of owning information about the DGP *a priori*, the lag order has to be chosen by some data-dependent methods such as AIC and BIC. We employ however only AIC throughout simulations by setting the maximum lag order to be 5, again because of computation considerations. We are specifically interested in examining the effects of the lag length uncertainty. It might well be expected that simulations with BIC produce qualitatively similar results to those with AIC here, given a similar information structure shared by both criterion. Therefore, for each particular replication, the bootstrap re-samples would be generated, based on the chosen order by the criteria as if it is true. Note that three bootstrap test statistics will be computed, using the bandwidth numbers as in computing the asymptotic counterparts. By doing so, we want to investigate whether the finite-sample performance of the bootstrap tests are affected by the choice of bandwidth numbers as the asymptotic counterparts are. Based on the simulation evidence that we will present later, it is quite evident that the finite-sample performance of the bootstrap tests are little sensitive to the choice of bandwidth numbers, to which the asymptotic counterparts are.

The first set of simulation results is associated with the empirical size of both the asymptotic and bootstrap tests. Table 2 presents results for models with intercept, and Table 3 for models with intercept and time trend. It appears from tables that the size of the asymptotic tests is very close to the nominal level for the cases with no autocorrelation ($\alpha = 0$), regardless of sample sizes and bandwidth numbers we considered, implying good asymptotic approximations. The cases with no autocorrelation indeed serve as a benchmark against which how both types of the tests perform in the presence of persistent processes will be examined. The empirical size of the asymptotic tests is no longer closer to the nominal level when the significance of autocorrelation is present. As the persistence in data increases, the over-rejection problem turns increasingly severe. The size-distortion patterns of the asymptotic tests then repeat as documented earlier: increasing in bandwidth reduces the distortion at a slow speed for fixed samples, yet increasing in samples does not help so but deteriorate it for given bandwidth numbers. The distortion has gone even worse for models with both intercept and time trend, due to an efficiency loss from estimating an additional coefficient. The message from the tables is simply that the finite-sample distributions of the tests are badly approximated when the persistence is high. This amounts to saying that the validity of the standard asymptotic approximations to the tests does not hold at all in the presence of high persistence. For the asymptotic analysis to make sense again in the presence of significant persistence, a local-to-unity parameterization and re-scaling are both required, as our preceding analysis has done.

Now turn to the bootstrap tests. There are a few important observations emerging from the tables regarding the tests. Firstly, the bootstrap distributions of the tests appear to bear a great deal closer resemblance to the finite-sample ones, implying the accuracy in the bootstrap critical values. In sharp contrast to the asymptotic counterparts, the bootstrap tests have a predominantly much better control of their sizes. Many instances under investigations for the tests show insignificant difference between the empirical size and the nominal one.

Next, unlike the asymptotic tests, the bootstrap tests display a better size control for models with intercept and time trend than with intercept only. Specifically, when sample size is small, the tests now reject slightly too little and thus are liberal in the presence of high persistence. The under-rejections are very likely to result from the downward bias of autoregressive coefficient estimations. The effect of the bias might be profound. Basically, the shorter the series is, the more persistence displays in data, and the greater is the bias. Consequently, the bootstrap re-samples are generated based on the estimates, smaller than the true ones on average. The resulting bootstrap distribution of the tests are then to the left of what they ought to be, creating under-rejections. Other than this, the control over size by the tests has been very good even in the problematic region for samples of small and moderate size.

Thirdly, the performance of the bootstrap tests appears to be not so much more affected by the choice of bandwidth numbers as the asymptotic ones. Notably, when the generated data is highly persistent, the bootstrap test are quite robust to the choice of bandwidth numbers across two models. The results might seem to be at odds with the intuition that a larger bandwidth number is required for more persistence in data. This does not have be so for the bootstrap tests however. The bootstrap is to estimate the finite-sample distributions of the tests directly. So if the finite-sample distribution of the tests is dependent on some nuisance parameters, as here probably due to not enough bandwidth, the bootstrap gives an estimate of the finite-sample distribution, whether or not the distribution to be estimated is free of nuisance parameters. The large-sample approximation, however, relies on the asymptotic distribution that is free of nuisance parameters. The size robustness to the choice of bandwidth numbers provides additional convenience in applications using the bootstrap tests where whether having a good choice over the bandwidth need not be a major concern.

Lastly, the use of AIC appears to reduce the risk of an uncertainty in selecting the appropriate AR order in Δy_t, to a large extent. Our simulations show that the correct order, equal to one in the setup, can be picked up most frequently, though there are a few instances where the estimations are overfitted. Overall, the bootstrap tests exhibit a very satisfactory control over size, and thus subject little to size distortions.

It is important that the good empirical size performance of the bootstrap tests does not come at the cost of power loss. We now examine the empirical power performance of the bootstrap tests. Table 4 summarizes the results for the empirical power. As a benchmark, we also report the empirical power of the asymptotic tests. Note the empirical power reported for the asymptotic ones has been adjusted for the size distortions, the so-called size-adjusted power. It is infeasible, because the finite-sample critical values of the asymptotic tests under the null are generally unknown, and need to be computed case by case. As seen clearly from the table, the empirical power of the bootstrap test is comparable to, or is slightly higher than that of the asymptotic counterparts. The power superiority of the bootstrap test to the asymptotic counterparts is to a very minor extent (between 1% and 2%), as a result of sampling results. Furthermore, it is a numerical reflection of test consistency when observing the empirical power of the bootstrap tests increase as sample size increases for fixed bandwidth, and signal-to-noise ratio. It is also obvious to find that the power of the bootstrap tests increases as the signal-to-noise ratio increases. Although it remains unknown whether the power of the bootstrap tests will depend on the bandwidth as the asymptotic ones do, it is still possible to assess to what extent the bootstrap will depend on the bandwidth via simulations. Bear in mind that the errors are iid, and it needs very small bandwidth in computations. The results show that the traditional wisdom applies again: choosing larger bandwidth costs the power of the bootstrap tests in a quite rapid speed. It implies that the choice of bandwidth for the bootstrap tests requires cautions in the sense that it may not affect the size but the power of the tests.

6 An Application to Purchasing Power Parity Debate

In this section we apply the bootstrap tests to the real exchange rates in the post-Bretton Woods period. This is to illustrate how the bootstrap tests can perform in applications. Of particular concern is to seek stationarity, or mean-reverting property in the real exchange rates that corresponds to the notion of long-run purchasing power parity. The importance of the parity comes from that it is the cornerstone assumption underlying many open macroeconomic models. Testing for the parity is equivalent to search for empirical supports for implications of the theory of concern. Thus, tests for stationarity of the real exchange rates under the null well serve the purpose. There has never been lack of empirical efforts on whether long-run purchasing power parity holds at all in the literature. To our knowledge, testing for purchasing power parity using a bootstrap version of the KPSS tests has not been attempted yet. We shall apply the tests to a panel of real exchange rates in the post-Bretton Woods. The use of this panel can be motivated for several reasons. Using the same panel samples, the findings here based on the bootstrap tests can be compared fairly to earlier work based on the asymptotic counterparts (for example, Culver and Papell, 1999; Caner and Kilian, 2001). The real exchange rates time series over the panel are not long spanned, and tend to be better characterized by a highly persistent autoregressive process This is exactly where the asymptotic tests may have difficulty in having the size under good control. Moreover, the panel has observations of the real exchange rates available on the both monthly and quarterly sampling frequencies. This allows our theoretical results derived in the local-to-unity context to have something to say about the empirical evidence.

The real exchange rates are constructed from the consumer price index series and the exchange rate series for the price of U.S. dollars in respective currency. Data is obtained from the IMF publication, *International Financial Statistics*. Monthly data is available for the following 18 countries over the period 1973.1-1998.12: Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States. Quarterly data is available over 1973.I - 1998.IV for Australia, Ireland, and New Zealand, in addition to the same 18 countries as in monthly data.

Table 5 presents the results of applying both the asymptotic tests and the bootstrap tests to each real exchange rate time series on both sampling frequencies. For all calculations, the bandwidth number is set to $L = [12(T/100)]^{1/4}$, following Caner and Kilian (2001). To demonstrate how the size bias is related to the degrees of persistence in the real exchange rates, we also report the largest autoregressive root from fitting ARMA(p,1) to the differenced series from which the bootstrap re-samples are generated. Again the autoregressive lag order is chosen by the AIC as in the Monte-Carlo study. More than half of the chosen autoregressive order from the data is one, and some are distributed evenly over other orders up to 5. The table also reports the bootstrap critical values at 5% and 10% levels, giving an idea how the bootstrap critical values might deviate from the asymptotic counterparts for each series.

The results with the asymptotic tests are not much different from those presented in Caner and Kilian (2001) using data up to 1997.4 only. At monthly frequency, the asymptotic test now rejects the null of stationarity for 9 out of 17 countries, 8 of them same and 1 additional. Using quarterly data, we have the same rejections by the test for 4 countries, except Sweden. Thus, we come to observe that more rejections appear to take place for higher frequencies, as our theoretical results predict. This is to say that many of them are spurious rejections. The supposition appears to be reasonable because many of the rejections indeed come from series having the largest autoregressive root in the problematic zone of significance persistence. In other words, once applying the bootstrap tests to the data, rejections will be much fewer, given that the tests can much correct for the size distortions. This is the case here. The bootstrap test at monthly frequency, now only rejects for 6 of 17 countries, instead of 9 by the asymptotic one.[2] So there are 3 countries (Austria, Canada, and Spain) for which the

[2]Sweden is included in these 6 countires. However note that it is Sweden only among others where the

asymptotic test reject but not the bootstrap test. The corresponding largest roots estimated for these 3 countries are greater than .84, suggesting spurious rejections by the asymptotic test. At the same time, it shows that correction for the size bias from the bootstrap test is at work for these 3 countries. On the other hand, for 3 of these 6 countries for which both two versions of the test reject (Italy, Japan and Portugal), the largest autoregressive root estimated are found between .3 and .6, away from the problematic region. Rejections like this may well be considered as a more conclusive evidence against long-run purchasing power parity, implying a prevailing random-walk component. Using quarterly data, the bootstrap test now reject for 3 of 20 countries, rather than 4 by the asymptotic one. Having the roots lying below .7, rejections by both versions of the test is very likely to be strong evidence against long-run purchasing power parity for all these 3 countries. The root found for Ireland is around .82, indicating again a spurious rejection due to the size problem with the asymptotic test. It explains already why the asymptotic test rejects but not the bootstrap one. Summing up, the use of the bootstrap test apparently lends more credible supports for purchasing power parity in the long-run. Many previous rejections by the asymptotic test could just originate from the size distortions, and could not constitute strong evidence against the mean-reverting property in real exchange rates.

7 Concluding Remarks

The paper begins with two goals. The first is to provide a large-sample explanation for the considerable size distortions associated with the KPSS tests in the presence of highly persistent but stationary processes. Using a local-to-unity approach, the tests are found to be unable to weakly converge to some sensible distributions without a re-scaling. The asymptotic distributions derived in the problematic region are dramatically different from the standard counterparts, consequently yielding the size bias. Not only do the results explain the sources of the size distortions, but also deliver practical implications. Of much relevance to the power of the tests is that in the presence of highly persistent but stationary processes, the tests possess the same stochastic order under both the null and the alternative, and thus are lack of capability of discriminating between the hypotheses. Related to the foregone is that the results are useful to explain why employing the size-adjusted tests may be at cost

bootstrap test rejects, but the asymptotic not.

of a power loss, as Caner and Kilian (2001) documented. Although the tests are ideal to apply in some cases, the analysis however is indicative that to correct for the size bias in applications where many time series could be highly persistent in nature, either relying on the asymptotic approximations or employing the size-adjusted critical values is unlikely to be useful or promising.

The second, and more important goal of the paper is to seek an alternative strategy to correct for the distortions, given the fact that the information drawn from the tests has been considered important and valuable. We appeal to the bootstrap version of the tests. We overcome the difficulty in reproducing independent bootstrap re-samples, due to a lack of appropriate parametric model that corresponds to the component representation of the series under test. It is done simply by utilizing an equivalence relation between component models and parametric ARIMA models. With the parametric models, it becomes feasible and straightforward to place the null constraint of stationarity when generating the re-samples, whether or not the series come from the null or the alternative. The procedure is crucial for the bootstrap tests to work properly. Our simulation evidence lends excellent credence to the use of our bootstrap tests in applications. Compared to the asymptotic counterparts, the bootstrap tests are far less prone to the size distortions, particularly as far as the problematic area is concerned. For sample of sizes encountered in applications, the empirical size of the bootstrap tests has proven to be nearly away from the nominal one. This suggests the accuracy of the bootstrap critical values. One further merit of the bootstrap tests is their insensitivity to the choice of bandwidth numbers, affecting much the accuracy of the asymptotic approximations.

Our bootstrap scheme is not limited to apply only to the KPSS tests. Other stationarity tests proposed, including parametric tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994), have been complained by suffering the similar size distortions problem as the KPSS tests experience. Applying the bootstrap scheme to account for the size bias associated with the parametric tests for stationarity is more straightforward. This is simply because to draw independent bootstrap re-samples, the bootstrap scheme counts on an ARIMA model with an moving average unit root from which these parametric tests are derived. Some complications, though not difficult, would be involved in applying the bootstrap scheme to the test proposed by Choi (1994), a modified version of the KPSS tests.

Some research as suggested has been under way, and will be reported by the authors in the future.

We nevertheless make no claim of the bootstrap consistency for our bootstrap proposal, not to mention the asymptotic refinement. Theoretical work along the line is to justify the validity and superiority of our bootstrap algorithm. This appears very difficult because the scheme is non-linear in nature due to the ARIMA estimation. It remains unknown whether the Edgeworth expansion is available for the model under study. A good point for the task to start with might have been given by Park (2003) that focuses on the unit root testing. We leave it for future research.

Appendix: Mathematical Proofs

We shall only present proofs for the case with $m = 1$, because that with $m = 0$ follows very easily. Under the assumptions in Theorem 1, it can be established that

$$\left[\begin{array}{c} \frac{1}{\sqrt{T}}(\hat{\beta}_0 - \beta_0) \\ \sqrt{T}(\hat{\beta}_1 - \beta_1) \end{array}\right] \Longrightarrow \sigma_\epsilon(\int_0^1 g(r)g(r)'dr)^{-1}(\int_0^1 g(r)J_c(r)dr)$$

where $g(r) = [1, r]'$. Denote $A = diag(1, \frac{1}{T})$ and recall that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimates from a regression of y_t on an intercept and a time trend. Given that u_t is a near-unit-root process, we can obtain

$$\begin{aligned}
(\frac{1}{\sqrt{T}})A^{-1}\left[\begin{array}{c} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{array}\right] &= \left\{T^{-1}\sum_{t=1}^{T} A \left[\begin{array}{c} 1 \\ t \end{array}\right]\left[\begin{array}{cc} 1 & t \end{array}\right] A\right\}^{-1} \frac{1}{T}\sum_{t=1}^{T} A \left[\begin{array}{c} 1 \\ t \end{array}\right](\frac{1}{\sqrt{T}})u_t \\
&\Longrightarrow \sigma_\epsilon(\int_0^1 g(r)g(r)'dr)^{-1}(\int_0^1 g(r)J_c(r)dr)
\end{aligned}$$

Now, let $r \in [0, 1]$ and denote $[Tr]$ the largest integer parts of Tr. The partial sum process can be defined as $S_{[Tr]} \equiv \sum_{t=1}^{[Tr]} \hat{u}_t$, where $\hat{u}_t \equiv y_t - (\hat{\beta}_0 + \hat{\beta}_1 t)$ are the OLS residulas. Some calculations can give

$$\begin{aligned}
T^{-3/2}S_{[Tr]} &= T^{-3/2}\sum_{t=1}^{[Tr]} \hat{u}_t \\
&= T^{-3/2}\sum_{t=1}^{[Tr]}\{u_t - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)t\} \\
&= T^{-3/2}\sum_{t=1}^{[Tr]} u_t - \frac{1}{T}\sum_{t=1}^{[Tr]}\left[\begin{array}{cc} 1 & \frac{t}{T} \end{array}\right]\left[\begin{array}{c} \frac{1}{\sqrt{T}}(\hat{\beta}_0 - \beta_0) \\ \sqrt{T}(\hat{\beta}_1 - \beta_1) \end{array}\right]
\end{aligned}$$

$$
\begin{aligned}
&\Longrightarrow \sigma_\epsilon \int_0^r \{J_c(s) - g(s)'(\int_0^1 g(r)g(r)'dr)^{-1}(\int_0^1 g(r)J_c(r)dr)\}ds \\
&\equiv \sigma_\epsilon \int_0^r \tilde{J}_c(s)ds
\end{aligned}
$$

As a result,

$$
\frac{1}{T^4}\sum_{t=1}^{T} S_t^2 = \frac{1}{T}\sum_{t=1}^{T}(\frac{S_t}{T^{3/2}})^2 \Longrightarrow \sigma_\epsilon^2 \int_0^1 (\int_0^r \tilde{J}_c(s)ds)^2 dr
$$

On the other hand, given the definition of $\hat{u}_t$, using the same line of arguments, it can be shown that $\frac{1}{\sqrt{T}}\hat{u}_{[Tr]} \Longrightarrow \sigma_\epsilon \tilde{J}_c(r)(= J_c(r) + (6r-4)\int_0^1 J_c(s)ds + (6-12r)\int_0^1 sJ_c(s)ds)$. Following from the definition of the long-run variance in (4) and the arguments in KPSS (page 168) gives

$$
\begin{aligned}
(\frac{1}{TL})\hat{\sigma}^2 &\Longrightarrow (\int_{-1}^1 (1-|x|)dx)(\sigma_\epsilon^2 \int_0^1 \tilde{J}_c^2(r)dr) \\
&= \sigma_\epsilon^2 \int_0^1 \tilde{J}_c^2(r)dr
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(\frac{L}{T})KPSS_\tau(k) &= \frac{\frac{1}{T^4}\sum_{t=1}^T S_t^2}{\frac{1}{TL}\hat{\sigma}^2} \\
&\Longrightarrow \frac{\int_0^1(\int_0^r \tilde{J}_c(s)ds)^2 dr}{\int_0^1 \tilde{J}_c^2(r)dr}
\end{aligned}
$$

The intended results are thus established.

References

[1] Amemiya, T. (1985), *Advanced Econometrics*, Cambridge: Harvard University Press.

[2] Andrew, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[3] Basawa, I. V., A. K. Mallik, W. P. McCormick and R. L. Taylor (1991), "Bootstrapping Unstable First Order Autoregressive Process," *Annals of Statistics,* 19, 1098-1101.

[4] Bühlmann, P. (1997), "Sieve Bootstrap for Time Series," *Bernoulli,* 3, 123-148.

[5] Caner, M. and L., Kilian (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, 639-657.

[6] Cheung, Y.-W. and Chinn, M. (1997), "Furhter Investigation of the Uncertain Unit Root in GNP," *Journal of Business and Economic Statistics* 16, 349-356.

[7] Choi, I. (1994), "Residual-Based Tests for the Null of Stationariy with Applications to U.S. Macroeconomic Time Series," *Econometric Theory* 10, 720-746.

[8] Culver, S.E., and D.H. Papell (1999), "Long-Run Purchasing Power Parity with Short-Run Data: Evidence with a Null Hypothesis of Stationarity," *Journal of International Money and Finance* 18, 751-768.

[9] Elliott, G., T. J. Rothenberg and J. H. Stock (1996), "Efficient Tests for An Autoregressive Unit Root," *Econometrica* 64, 813-836.

[10] Engel, C. (2000), "Long-Run PPP May Not Hold After All," *Journal of International Economics* 51, 243-273.

[11] Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge University Press.

[12] Hobjin, B., P.H. Franses, and M. Ooms (1998), "Generalizations of the KPSS-test for Stationarity," Discussion Paper 9802, Econometric Institute, Erasmus University Rotterdam.

[13] Horowitz, L. (2001), "The Bootstrap," in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics*, vol. 5, 3159 - 3228, Amsterdam: Elsevier.

[14] Kiefer, N. and T. Volgelsang (2002), "Heteroskedasticity-Autocorrelated Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory* 18, 1350-1366.

[15] Kuo, B.-S. (1999), "Asymptotics of ML Estimator for Regression Model with a Stochastic Trend Component," *Econometric Theory* 15, 24-49.

[16] Kuo, B.-S. and A. Mikkola (1999), "Re-examining Long-run Purchasing Power Parity," *Journal of International Money and Finance* 18, 251-266.

[17] Kuo, B.-S. and A. Mikkola (2001), "How Sure Are We About PPP? Panel Evidence with the Null of Stationary Real Exchange Rates," *Journal of Money, Credit, and Banking*, 33, 767-789.

[18] Künsch, H. R. (1989), "The Jackknife And the Bootstrap for General Stationay Observations," *Annals of Statistics,* 17, 1217-1241.

[19] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992), "Testing the Null of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics* 54, 159-178.

[20] Lee, J. (1996), "On the Power of Stationarity Tests Using Optimal Bandwith Estimates," *Economics Letters* 51, 131-137.

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

[22] Leybourne, S. and B. McCabe (1999), "Modified Stationarity Tests with Data Dependent Model Selection Rules," *Journal of Business and Economic Statistics* 17, 264-270.

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

[24] Marriott, F. H. C., and J. A. Pope (1954), "Bias in the Estimation of Autocorrelations," *Biometrika* 74, 390-402.

[25] Müller, U. K. (2002), "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," Discussion paper No. 2002-26, Department of Economics, Unversity of St. Gallen.

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

[27] Nankervis, J. and N. Savin (1996), "The Level and Power of the Bootstrap t Test in the AR(1) Model With Trend," *Journal of Business and Economic Statistics* 14, 161-168.

[28] Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

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

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

[31] Park, J.Y. (2003), "Bootstrap Unit Root Tests," *Econometrica* 71/6, 1845-1895.

[32] Perron, P. and S. Ng (1996), "Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," *Review of Economic Studies* 63, 435-465.

[33] Phillips, P. C. B. (1987), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika,* 74, 535-547.

[34] Phillips, P. C. B. and P. Perron (1988), "Testing for a unit root in a time series regression," *Biometrika,* 75, 335-346.

[35] Politis, D. N. and J. P. Romano (1993), "The Stationary Bootstrap," *Journal of the American Statistical Association* 83, 1303-1313.

[36] Potscher, B.M. (1991), "Non-invertibility and Pseudo-Maximum Likelihood Estimation of Mis-specified ARMA Model," *Econometric Theory* 7, 435-449.

[37] Psaradakis, Z. (2001), "Bootstrap Tests for an Autoregressive Unit Root in the Presence of Weakly Dependent Errors," *Journal of Time Series Analysis* 22, 577-594.

[38] Saikkonen, P. and R. Luukkonen (1993), "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association* 188, 596-601.

[39] Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics* 7, 147-160.

[40] Shaman, P. and R. A. Stine (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association* 83, 842-848.

Table 2 Empirical Size Performance of $KPSS_{\mu}$ (with intercept)

		$KPSS_{\mu}(4)$		$KPSS_{\mu}(8)$		$KPSS_{\mu}(12)$			$KPSS_{\mu}(4)$		$KPSS_{\mu}(8)$		$KPSS_{\mu}(12)$	
T	α	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*	T	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*
30	0.980	0.573	0.031	0.304	0.030	0.033	0.037	50	0.610	0.038	0.425	0.039	0.206	0.041
	0.960	0.516	0.036	0.241	0.040	0.022	0.038		0.544	0.029	0.347	0.031	0.130	0.027
	0.940	0.473	0.040	0.191	0.039	0.015	0.034		0.483	0.032	0.296	0.032	0.099	0.036
	0.920	0.437	0.035	0.161	0.037	0.012	0.029		0.451	0.036	0.256	0.031	0.083	0.023
	0.900	0.405	0.033	0.139	0.033	0.009	0.033		0.417	0.025	0.222	0.024	0.071	0.030
	0.880	0.379	0.035	0.122	0.036	0.007	0.037		0.379	0.048	0.196	0.041	0.062	0.041
	0.860	0.355	0.028	0.107	0.026	0.006	0.033		0.343	0.048	0.173	0.042	0.053	0.047
	0.840	0.329	0.038	0.095	0.040	0.006	0.040		0.312	0.059	0.157	0.062	0.049	0.057
	0.820	0.307	0.046	0.087	0.042	0.005	0.049		0.285	0.050	0.142	0.049	0.044	0.050
	0.800	0.288	0.036	0.080	0.039	0.005	0.036		0.261	0.051	0.129	0.051	0.042	0.048
	0.500	0.113	0.054	0.041	0.063	0.006	0.053		0.099	0.068	0.055	0.064	0.026	0.062
	0.300	0.067	0.053	0.031	0.062	0.005	0.050		0.068	0.056	0.041	0.062	0.020	0.063
	0.000	0.035	0.062	0.021	0.058	0.003	0.050		0.041	0.059	0.030	0.062	0.013	0.058
100	0.980	0.711	0.031	0.495	0.029	0.366	0.026	150	0.810	0.030	0.601	0.033	0.442	0.028
	0.960	0.623	0.036	0.409	0.035	0.274	0.031		0.718	0.041	0.485	0.038	0.337	0.034
	0.940	0.554	0.045	0.338	0.043	0.224	0.049		0.628	0.043	0.402	0.039	0.254	0.038
	0.920	0.494	0.047	0.291	0.044	0.185	0.047		0.551	0.039	0.331	0.041	0.200	0.045
	0.900	0.438	0.044	0.251	0.045	0.156	0.043		0.481	0.041	0.270	0.037	0.163	0.044
	0.880	0.386	0.051	0.218	0.049	0.132	0.050		0.420	0.055	0.227	0.051	0.140	0.055
	0.860	0.346	0.050	0.190	0.052	0.114	0.052		0.369	0.048	0.194	0.047	0.125	0.046
	0.840	0.312	0.048	0.167	0.053	0.103	0.056		0.329	0.044	0.169	0.041	0.111	0.037
	0.820	0.285	0.070	0.151	0.063	0.091	0.063		0.290	0.046	0.151	0.043	0.099	0.042
	0.800	0.263	0.047	0.136	0.052	0.084	0.059		0.257	0.049	0.139	0.046	0.092	0.045
	0.500	0.097	0.058	0.061	0.066	0.043	0.067		0.095	0.050	0.069	0.051	0.052	0.059
	0.300	0.062	0.067	0.047	0.062	0.037	0.062		0.069	0.061	0.055	0.062	0.044	0.067
	0.000	0.042	0.061	0.036	0.064	0.031	0.059		0.048	0.053	0.046	0.052	0.037	0.057
300	0.980	0.868	0.032	0.682	0.031	0.532	0.034	600	0.907	0.051	0.725	0.054	0.578	0.054
	0.960	0.750	0.042	0.515	0.042	0.375	0.045		0.752	0.049	0.511	0.050	0.370	0.048
	0.940	0.634	0.049	0.393	0.049	0.278	0.050		0.613	0.049	0.371	0.048	0.258	0.051
	0.920	0.535	0.046	0.312	0.047	0.213	0.049		0.503	0.054	0.282	0.051	0.197	0.052
	0.900	0.452	0.039	0.257	0.040	0.176	0.040		0.417	0.048	0.232	0.047	0.161	0.047
	0.880	0.387	0.047	0.218	0.047	0.152	0.046		0.353	0.050	0.194	0.051	0.136	0.051
	0.860	0.341	0.043	0.188	0.046	0.134	0.049		0.302	0.050	0.165	0.052	0.120	0.050
	0.840	0.300	0.044	0.168	0.043	0.119	0.040		0.260	0.053	0.149	0.053	0.110	0.052
	0.820	0.267	0.045	0.151	0.045	0.109	0.046		0.231	0.049	0.134	0.050	0.101	0.050
	0.800	0.242	0.048	0.136	0.048	0.097	0.049		0.208	0.052	0.123	0.053	0.094	0.058
	0.500	0.089	0.045	0.069	0.045	0.058	0.047		0.086	0.057	0.067	0.061	0.059	0.060
	0.300	0.068	0.075	0.056	0.076	0.050	0.077		0.069	0.044	0.060	0.049	0.054	0.047
	0.000	0.050	0.050	0.047	0.048	0.044	0.055		0.053	0.054	0.052	0.058	0.050	0.056

Notes: The DGP is $y_t = \alpha y_{t-1} + e_t$ with $e_t \overset{iid}{\sim} N(0,1)$. $KPSS_{\mu}(k)$ and $KPSS_{\tau}(k)$ are as defined in the text, respectively. *boot.* denotes the bootstrap test, *asym.* the asymptotic test, and T sample sizes. The figures reported are the rejection frequencies at 5% nominal significance level, based on 5,000 replications for the asymptotic tests, and 1,000 for the bootstrap tests with 100 re-samples. The asymptotic critical values for 5% level is .463 for $KPSS_{\mu}(k)$ and .146 for $KPSS_{\tau}(k)$.

Table 3 Empirical Size Performance of $KPSS_\tau$ (with intercept and time trend)

		$KPSS_\tau(4)$		$KPSS_\tau(8)$		$KPSS_\tau(12)$			$KPSS_\tau(4)$		$KPSS_\tau(8)$		$KPSS_\tau(12)$	
T	α	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*	T	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*
30	0.980	0.497	0.048	0.203	0.060	0.141	0.057	50	0.612	0.072	0.367	0.072	0.169	0.066
	0.960	0.485	0.070	0.194	0.068	0.140	0.058		0.587	0.063	0.338	0.064	0.146	0.063
	0.940	0.468	0.057	0.180	0.050	0.135	0.059		0.550	0.060	0.301	0.065	0.124	0.059
	0.920	0.448	0.060	0.162	0.061	0.133	0.054		0.514	0.053	0.268	0.056	0.105	0.057
	0.900	0.429	0.051	0.148	0.038	0.129	0.051		0.474	0.057	0.244	0.053	0.090	0.059
	0.880	0.404	0.051	0.134	0.049	0.129	0.066		0.441	0.051	0.215	0.055	0.081	0.050
	0.860	0.379	0.044	0.123	0.043	0.129	0.047		0.407	0.041	0.192	0.047	0.071	0.051
	0.840	0.363	0.039	0.112	0.043	0.128	0.054		0.374	0.049	0.171	0.051	0.067	0.047
	0.820	0.337	0.058	0.103	0.041	0.127	0.052		0.347	0.039	0.154	0.039	0.064	0.038
	0.800	0.320	0.050	0.094	0.053	0.130	0.058		0.316	0.055	0.141	0.050	0.061	0.059
	0.500	0.122	0.050	0.051	0.060	0.185	0.061		0.115	0.051	0.058	0.044	0.050	0.058
	0.300	0.072	0.061	0.043	0.060	0.213	0.059		0.067	0.051	0.044	0.057	0.050	0.056
	0.000	0.041	0.056	0.036	0.053	0.246	0.056		0.040	0.062	0.034	0.058	0.046	0.048
100	0.980	0.802	0.056	0.547	0.056	0.384	0.053	150	0.906	0.042	0.714	0.042	0.475	0.039
	0.960	0.750	0.043	0.471	0.045	0.322	0.046		0.854	0.055	0.613	0.054	0.383	0.050
	0.940	0.696	0.055	0.409	0.052	0.266	0.057		0.787	0.051	0.519	0.052	0.304	0.055
	0.920	0.632	0.045	0.350	0.048	0.215	0.046		0.718	0.054	0.435	0.057	0.248	0.053
	0.900	0.574	0.047	0.301	0.049	0.181	0.047		0.655	0.057	0.372	0.057	0.201	0.054
	0.880	0.520	0.049	0.259	0.048	0.154	0.050		0.590	0.056	0.316	0.057	0.170	0.056
	0.860	0.472	0.055	0.225	0.051	0.135	0.056		0.531	0.055	0.274	0.053	0.148	0.053
	0.840	0.422	0.051	0.197	0.053	0.117	0.054		0.470	0.055	0.236	0.056	0.128	0.058
	0.820	0.379	0.057	0.175	0.052	0.105	0.054		0.423	0.054	0.208	0.059	0.113	0.057
	0.800	0.344	0.057	0.159	0.056	0.097	0.052		0.377	0.055	0.182	0.056	0.103	0.055
	0.500	0.110	0.046	0.068	0.048	0.048	0.048		0.112	0.063	0.073	0.059	0.054	0.064
	0.300	0.070	0.075	0.053	0.067	0.041	0.063		0.072	0.051	0.057	0.051	0.046	0.067
	0.000	0.044	0.063	0.040	0.065	0.036	0.063		0.045	0.059	0.044	0.068	0.041	0.070
300	0.980	0.964	0.035	0.820	0.035	0.670	0.032	600	0.986	0.053	0.894	0.052	0.753	0.051
	0.960	0.908	0.043	0.682	0.047	0.500	0.046		0.930	0.060	0.713	0.063	0.526	0.064
	0.940	0.830	0.039	0.550	0.039	0.376	0.040		0.837	0.059	0.557	0.065	0.375	0.066
	0.920	0.748	0.049	0.442	0.046	0.290	0.045		0.730	0.062	0.432	0.060	0.279	0.058
	0.900	0.652	0.049	0.361	0.049	0.231	0.045		0.634	0.067	0.342	0.065	0.217	0.063
	0.880	0.570	0.054	0.301	0.053	0.191	0.054		0.544	0.065	0.280	0.067	0.174	0.066
	0.860	0.494	0.051	0.253	0.056	0.166	0.053		0.468	0.062	0.231	0.063	0.154	0.062
	0.840	0.435	0.057	0.220	0.058	0.143	0.057		0.401	0.061	0.197	0.064	0.136	0.062
	0.820	0.383	0.059	0.197	0.060	0.127	0.062		0.356	0.063	0.172	0.062	0.121	0.061
	0.800	0.340	0.047	0.176	0.057	0.115	0.056		0.310	0.055	0.157	0.058	0.111	0.056
	0.500	0.111	0.050	0.077	0.047	0.065	0.049		0.104	0.063	0.074	0.063	0.065	0.061
	0.300	0.076	0.068	0.062	0.066	0.058	0.070		0.075	0.064	0.063	0.060	0.059	0.058
	0.000	0.055	0.059	0.052	0.060	0.050	0.062		0.054	0.070	0.053	0.063	0.052	0.064

Notes: The DGP is $y_t = \alpha y_{t-1} + e_t$ with $e_t \overset{iid}{\sim} N(0,1)$. $KPSS_\mu(k)$ and $KPSS_\tau(k)$ are as defined in the text, respectively. *boot.* denotes the bootstrap test, *asym.* the asymptotic test, and T sample sizes. The figures reported are the rejection frequencies at 5% nominal significance level, based on 5,000 replications for the asymptotic tests, and 1,000 for the bootstrap tests with 100 re-samples. The asymptotic critical values for 5% level is .463 for $KPSS_\mu(k)$ and .146 for $KPSS_\tau(k)$.

Table 4 Empirical Power Performance of KPSS Tests

Panel A: $KPSS_\mu$ (with intercept)

		$KPSS_\mu(4)$		$KPSS_\mu(8)$		$KPSS_\mu(12)$			$KPSS_\mu(4)$		$KPSS_\mu(8)$		$KPSS_\mu(12)$	
T	α	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*	T	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*
30	1	0.647	0.635	0.487	0.495	0.380	0.392	50	0.704	0.725	0.596	0.607	0.488	0.496
	0.1	0.442	0.454	0.346	0.372	0.272	0.283		0.594	0.609	0.520	0.498	0.423	0.397
	0.01	0.141	0.148	0.119	0.126	0.106	0.113		0.257	0.281	0.233	0.247	0.198	0.228
	0.001	0.062	0.051	0.057	0.059	0.062	0.072		0.075	0.091	0.076	0.084	0.072	0.080
	0.0001	0.051	0.050	0.051	0.056	0.055	0.062		0.052	0.054	0.056	0.056	0.055	0.059
100	1	0.821	0.829	0.690	0.687	0.612	0.605	150	0.914	0.919	0.792	0.836	0.710	0.723
	0.1	0.759	0.792	0.654	0.696	0.586	0.631		0.881	0.887	0.765	0.792	0.693	0.689
	0.01	0.506	0.528	0.465	0.478	0.426	0.438		0.681	0.693	0.618	0.639	0.573	0.582
	0.001	0.146	0.159	0.137	0.148	0.134	0.143		0.278	0.297	0.262	0.273	0.248	0.264
	0.0001	0.057	0.084	0.057	0.079	0.058	0.077		0.077	0.079	0.072	0.078	0.074	0.082
300	1	0.962	0.960	0.893	0.900	0.816	0.825	600	0.992	0.984	0.959	0.956	0.919	0.924
	0.1	0.957	0.958	0.886	0.897	0.809	0.832		0.990	0.989	0.958	0.965	0.917	0.932
	0.01	0.880	0.889	0.818	0.836	0.753	0.782		0.972	0.970	0.935	0.928	0.896	0.898
	0.001	0.554	0.543	0.532	0.524	0.508	0.496		0.818	0.824	0.783	0.795	0.751	0.765
	0.0001	0.152	0.178	0.150	0.174	0.148	0.175		0.368	0.396	0.362	0.384	0.355	0.379

Panel B: $KPSS_\tau$ (with intercept and time trend)

		$KPSS_\tau(4)$		$KPSS_\tau(8)$		$KPSS_\tau(12)$			$KPSS_\tau(4)$		$KPSS_\tau(8)$		$KPSS_\tau(12)$	
T	σ^2_ζ	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*	T	*asym.*	*boot.*	*asym.*	*boot.*	*asym.*	*boot.*
30	1	0.463	0.486	0.232	0.206	0.011	0.026	50	0.612	0.622	0.398	0.434	0.181	0.219
	0.1	0.213	0.238	0.132	0.152	0.026	0.033		0.389	0.398	0.281	0.305	0.135	0.168
	0.01	0.074	0.096	0.062	0.078	0.044	0.056		0.110	0.117	0.097	0.110	0.072	0.075
	0.001	0.051	0.070	0.052	0.066	0.047	0.063		0.058	0.070	0.056	0.058	0.052	0.066
	0.0001	0.050	0.057	0.051	0.062	0.047	0.063		0.053	0.063	0.051	0.051	0.049	0.069
100	1	0.810	0.803	0.599	0.625	0.461	0.497	150	0.917	0.890	0.767	0.758	0.619	0.592
	0.1	0.689	0.686	0.524	0.534	0.416	0.435		0.853	0.863	0.714	0.739	0.571	0.599
	0.01	0.289	0.291	0.245	0.253	0.217	0.213		0.486	0.527	0.417	0.465	0.361	0.384
	0.001	0.073	0.083	0.072	0.086	0.071	0.076		0.114	0.128	0.108	0.129	0.101	0.119
	0.0001	0.051	0.068	0.050	0.066	0.054	0.058		0.053	0.061	0.052	0.053	0.053	0.056
300	1	0.980	0.965	0.906	0.900	0.807	0.805	600	0.998	0.985	0.976	0.962	0.936	0.931
	0.1	0.966	0.952	0.886	0.887	0.790	0.809		0.996	0.991	0.974	0.970	0.933	0.924
	0.01	0.814	0.803	0.732	0.734	0.654	0.660		0.970	0.978	0.937	0.947	0.891	0.896
	0.001	0.287	0.324	0.271	0.304	0.250	0.273		0.683	0.657	0.644	0.618	0.602	0.583
	0.0001	0.071	0.092	0.071	0.086	0.068	0.087		0.172	0.175	0.166	0.175	0.156	0.172

Notes: Under the alternative, the DGP is $y_t = r_t + \epsilon_t$, and $r_t = r_{t-1} + \zeta_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and $\zeta_t \overset{iid}{\sim} N(0,\sigma^2_\zeta)$. $KPSS_\mu(k)$ and $KPSS_\tau(k)$ are as defined in the text, respectively. *boot.* denotes the bootstrap test, *asym.* the asymptotic test, and T sample sizes. The figures reported for the bootstrap tests are the rejection frequencies at 5% nominal significance level, based on 1,000 replications with 100 re-samples. Those reported for the asymptotic counterparts are the size-adjusted empirical power.

Table 5 Testing for Purchasing Power Parity:
Asymptotic vs. Bootstrap KPSS Tests

	Monthly data				Quarterly Data			
			boot. cv.				boot. cv.	
Country	KPSS	$\hat{\alpha}$	5%	10%	KPSS	$\hat{\alpha}$	5%	10%
Australia	n/a	n/a	n/a	n/a	$0.514^{**\ddagger}$	0.018	0.401	0.350
Austria	0.439^{*}	0.835	0.785	0.678	0.225	0.436	0.399	0.322
Belgium	0.224	0.898	0.929	0.824	0.118	0.851	0.572	0.504
Canada	0.924^{**}	0.960	1.252	1.057	$0.463^{**\dagger}$	0.707	0.470	0.394
Denmark	0.276	0.785	0.540	0.475	0.147	0.644	0.502	0.404
Finland	0.119	0.882	0.853	0.756	0.070	0.836	0.474	0.404
France	0.196	0.777	0.562	0.499	0.110	0.298	0.409	0.341
Germany	0.237	0.818	0.764	0.623	0.128	0.861	0.708	0.641
Greece	$0.355^{*\ddagger}$	0.937	0.302	0.235	0.178	0.856	0.555	0.509
Ireland	n/a	n/a	n/a	n/a	0.391^{*}	0.818	0.674	0.560
Italy	$0.367^{*\dagger}$	0.688	0.380	0.285	0.204	0.038	0.400	0.336
Japan	$1.315^{**\ddagger}$	0.641	0.626	0.483	$0.629^{**\ddagger}$	0.017	0.416	0.344
Netherlands	0.221	0.983	1.684	1.658	0.120	0.863	0.432	0.356
Norway	0.156	0.745	0.382	0.306	0.089	0.536	0.385	0.322
New Zealand	n/a	n/a	n/a	n/a	0.158	0.758	0.444	0.346
Portugal	$0.459^{*\dagger}$	0.344	0.477	0.382	0.227	0.736	0.572	0.481
Spain	0.395^{*}	0.923	0.962	0.886	0.209	0.722	0.493	0.419
Sweden	$0.296^{\dagger}$	0.835	0.333	0.260	0.158	0.928	0.824	0.789
Switzerland	$0.608^{**\ddagger}$	0.865	0.327	0.252	0.313	0.365	0.405	0.326
United Kingdom	0.410^{*}	0.970	1.434	1.262	0.254	0.834	0.371	0.310

Notes: Data is obtained from the IMF publication, *International Financial Statistics*. The real exchange rates are constructed from the consumer price index series and the exchange rate series for the price of U.S. dollars in respective currency. Monthly data is available over the period 1973.1-1998.12, and quarterly data is available over 1973.I - 1998.IV. For all calculations, the bandwidth number is set to $L = [12(T/100)]^{1/4}$. $\hat{\alpha}$ denotes The largest autoregressive root from fitting ARMA(p,1) to the differenced series from which the bootstrap re-samples are generated. the autoregressive lag order is chosen by the AIC. 'boot. cv.' is short for bootstrap critical values. The asymptotic critical value for KPSS test is 0.463 (0.347) at the 5 (10)% significance level. $**$ ($*$) represents a rejection at 5 (10)% level using the asymptotic critical values, and ‡ (†) using the bootstrap asymptotic critical values.