Bootstrap Stationarity Test with an Application to Purchasing Power Parity

Emre Aylar

University of Washington

April 2013

Abstract

It is widely known that conventional stationarity tests have significant size distortions in the presence of highly persistent data. In this paper we propose a bootstrap stationarity test that has good size control and also retains power. Our bootstrap procedure consists of two parts; first we utilize a parametric re-sampling scheme that can generate independent bootstrap samples. Our re-sampling scheme also enables us to impose the null constraint on bootstrap samples, which leads to a correctly sized test and good power properties. Second, we choose our test statistic to achieve maximal power while still differentiate the null and the alternative. We compare the empirical size and power performance of our test with existing bootstrap and conventional stationarity tests through Monte-Carlo studies. Simulations demonstrate that our bootstrap test controls size better and has higher power than the competing methods. Finally, we utilize our proposed method to find some evidence against purchasing power parity in real exchange rates of the countries investigated.

1 Introduction

Whether a time series is stationary or it has a unit root has been one of the most important issues in time series econometrics. Unit root tests have been the primary statistical method to differentiate these two alternatives. A testing procedure proposed by Kwiatkowski et al. (1992, KPSS hereafter) takes the stationarity as the null hypothesis instead and complements the unit root tests. Stationarity tests can be useful in cases where unit root tests have size distortion or low power. Moreover, the stationarity null may be more appropriate when testing for some economic theories where the hypothesis to be tested under the null is the one that economic theories suggest. For instance, purchasing power parity predicts that the relative values of currencies fluctuates temporarily around the relative price levels between countries and that in the long run nominal exchange rates converge to these relative prices. Hence, it is more natural to test stationarity of real exchange rates and to look for evidence against mean-reversion to test for purchasing power parity.

This useful stationarity test, however, is known to have significant size distortions in the presence of highly persistent yet stationary data. These size distortions were first documented by KPSS and later Caner and Killian (2001) reported that the size distortions are common for all conventional stationarity tests. The main source of the size distortion of the KPSS test is underestimation of long run variances. Muller (2005), for instance, provides a theoretical explanation of size distortions of the KPSS test in relation to long run variance estimation under local to unity framework.

In this paper, we propose a bootstrap stationarity test that has a good size control over the null parameter space. First we utilize a parametric re-sampling scheme that can generate independent bootstrap samples. This parametric re-sampling scheme, proposed by Kuo and Tsong (2005), allows us to replicate the dependence structure of the original data. In addition, the null constraint can easily be imposed on the generated data with this scheme. The bootstrap tests have correct size given that the dependence structure is replicated successfully and the null constraint is imposed. Second we select our bootstrap test statistic to increase the power of our bootstrap test. This choice of the test statistic is what separates our procedure from other bootstrap stationary tests in the literature. Kuo and Tsong (2005) and Lee and Lee

(2012) choose the Lagrange multiplier test statistic (KPSS test statistic) as their bootstrap test statistic which is slightly different from our choice. Our test statistic has a well defined limit under the null and diverges under the alternative. However, our test statistic diverges at a faster rate than the LM test statistic hence we expect our bootstrap procedure to have better power properties than competing bootstrap tests. Our simulation studies also confirm these findings and the effectiveness of our bootstrap procedure.

The rest of the paper is organized as follows. Section 2 reviews conventional stationarity tests and their size distortions. We introduce our bootstrap procedure in Section 3. Section 4 provides some Monte Carlo studies where we compare the finite sample properties of our bootstrap test with existing bootstrap and conventional stationarity tests. We utilize our bootstrap procedure to test purchasing power parity as an empirical application, and the results are given in Section 5. Section 6 concludes and mentions future research topics.

2 Stationarity Tests

In this section we present the idea behind the stationarity tests and their size distortions through KPSS test. There is no loss of generality since other conventional stationarity tests are either parametric version of KPSS (Leybourne and McCabe (1994)) or employing an equivalent model which only transforms the parameter to be tested (Saikkonen and Luukkonen (1993)). Moreover, these tests also share the size problem of the KPSS test. KPSS considers the following unobserved component model as the data generating process

$$y_t = \mu + \beta t + r_t + \epsilon_t$$

where $r_t = r_{t-1} + u_t$ is the random walk component with u_t are iid $(0, \sigma_u^2)$ and ϵ_t is a stationary error. Then the stationarity (null) hypothesis can be formulated as

$$H_0\colon \sigma_u^2 = 0 \ versus \ \ H_1\colon \sigma_u^2 > 0$$

The KPSS test statistic is both the one-sided Lagrange multiplier statistic and locally best invariant test statistic under the normality assumption. However, they modify the test statistic in a nonparametric way to allow for general forms of temporal dependence in stationary error process ϵ_t. The KPSS test statistic is then; $KPSS_i = \hat{\sigma}^{-2} T^{-2} \sum_{t=1}^{T} S_t^2 \quad i = \mu, \tau$ and the partial sum process is defined as $S_t = \sum_{i=1}^{t} e_i$ where e_i is the residual from the regression of y on an intercept for $KPSS_\mu$ and on an intercept and time trend for $KPSS_\tau$. Sum of the squared partial sums is normalized by an estimate of the long run variance $\sigma^2 = lim_{T\to\infty} T^{-1} E(S_t^2)$.

KPSS show that under the null and some regularity conditions, the limiting distributions of the Lagrange multiplier test statistic can be represented as:

$$KPSS_i \xrightarrow{d} \int_0^1 V_i^2(r)dr \quad for\ i = \mu, \tau$$

where $V_\mu(r)$ is a standard Brownian bridge: $V_\mu(r) = W(r) - rW(1)$, and $V_\tau(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(s)ds$ is a second-level Brownian bridge and $W(r)$ is a Wiener process.

KPSS construct a consistent estimator of the long-run variance σ^2, say $\hat{\sigma}^2$, from the residuals e_t ; specifically they use an estimator of the form

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} e_t^2 + 2T^{-1}\sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T} e_t e_{t-s}$$

Here $w(s,l)$ is an optional weighting function that corresponds to the choice of a spectral window. KPSS use the Bartlett window $w(s,l) = 1 - s/(l+1)$ as in Newey and West (1987), which guarantees the nonnegativity of $\hat{\sigma}^2$. For consistency of $\hat{\sigma}^2$, it is necessary that the lag truncation parameter $l \to \infty$ as $T \to \infty$ but at a slower rate; $l = o(T)$. In practice, the lag truncation parameter is chosen as $l = [k(T/100)^{1/4}]$, where k is a constant, and [.] is the largest integer function, following from Schwert (1989).

It is now widely known that the KPSS and other conventional stationarity tests have significant size distortions in the presence of highly persistent but stationary data. It is KPSS that first documented the size problems and later Caner and Killian (2001) provided an extensive simulation study which demonstrates that all conventional stationarity tests share this problem. In order to address this size problem, we provide a simple simulation study in Table 1. Table 1 provides empirical rejection frequencies of the KPSS test at a 5% nominal level with varying degrees of persistence of a simple AR(1) model. We present our results with three different bandwidth numbers to highlight the impact of the choice of lag truncation parameter on the size of the test. To save some space, we will not report the simulations for the tests with an intercept, as they share very similar qualitative outcomes as reported here.

T	ρ	$KPSS_\tau(4)$	$KPSS_\tau(12)$	$KPSS_\tau(18)$
100	0.98	85.72%	41.16%	23.96%
	0.94	80.62%	29.5%	13.96%
	0.9	68.46%	20.6%	9.96%
	0.5	13.68%	5.14%	3.94%
300	0.98	97.76%	71.66%	50.26%
	0.94	89.52%	42.48%	24.82%
	0.9	73.78%	24.48%	14.42%
	0.5	11.64%	5.06%	4.56%
600	0.98	99.22%	79.16%	57.46%
	0.94	89.72%	44%	25.5%
	0.9	69.98%	21.62%	14.48%
	0.5	10.7%	6.02%	5.46%

Notes: 1. The DGP is $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$

2. The rejection frequencies are calculated with 10000 replications

3. $KPSS_\tau(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$

Table 1: Size performance of $KPSS_\tau$ at 5% nominal level

As seen from Table 1, the KPSS test has significant size distortions, ranging from 15% to 99% as opposed to 5% nominal level, when the data is highly persistent yet stationary. Table 1 is also quite informative about the sources of the size distortions of the KPSS test. First, for a given sample size, size distortions get smaller as the bandwidth/lag truncation parameter increases. Thus, one source of size distortions is the underestimation of the long run variance. However, a more accurate estimate of the long run variance is not enough to circumvent the size problem. The asymptotic results of KPSS do not provide a good approximation to the finite sample distribution of the test statistic when the data is highly persistent. This is the second source of the size distortions and it can clearly be seen from Table 1: given a fixed , an increase in sample size does not help alleviate but exacerbate the degree of size distortions. Muller (2005) and Kuo and Tsong (2005) give a theoretical explanation about this surprising finding.

Kuo and Tsong (2005) demonstrate, under the local-to-unity framework, that the KPSS test statistic is $O_p(\frac{T}{l})$. Local to unity framework provides an alternative asymptotics to the distribution of the test statistic for highly persistent data. This alternative explains clearly why the rejection frequencies in Table 1 increase with the sample size; as the sample size increase the KPSS statistic diverges at rate T/l .

3. Bootstrap Procedure

In time series literature, when conventional asymptotics fail to deliver good approximations to the finite sample distributions bootstrap turns out to be quite useful to improve small sample properties. Ferretti and Romo (1996) used bootstrap in unit root testing and provided an alternative to augmented Dickey-Fuller and Phillips-Perron tests in unit root tests with correlated errors. Nankervis and Savin (1996) introduced bootstrap t test in the AR(1) model to fix the size distortions of the conventional t test of the AR parameter. Hansen (1999) provided the grid bootstrap method to construct confidence interval for the largest autoregressive root and achieved uniform coverage probability where the conventional methods failed to do so.

In this paper, we propose a bootstrap stationarity test to deal with the size distortions of the conventional stationarity tests addressed in Section 2. There exits bootstrap stationarity tests by Kuo and Tsang (2005) and Lee and Lee (2012). Our bootstrap test utilizes Kuo and Tsang's resampling scheme, however differs from their test in the choice of the test statistic as will be explained later in this section.

There are two major issues one needs to deal with for bootstrap tests to be successful ; i) generating independent bootstrap resamples, ii) imposing the null constraint on the bootstrap resamples. We overcome these issues by utilizing an ARIMA model which is equivalent (up to second order moments) to the unobserved component model of section 2. Specifically, the unobserved component model $y_t = \beta t + r_t + \epsilon_t$ with $r_t = r_{t-1} + u_t$ is second order equivalent in moments to ARIMA(p,1,1) model;

$$\Phi(L)(1-L)y_t = \beta + (1-\theta L)\eta_t, \;\; \eta_t \overset{iid}{\sim} (0, \sigma_\eta^2)$$

where $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ with roots outside the unit circle. This equivalence is the basis of the stationarity test of Leybourne and McCabe (1994). They utilized above ARIMA(p,1,1) model to take care of the serial correlation and used the residuals to construct the Lagrange multiplier test statistic of KPSS. In our bootstrap resampling scheme, we make use of the parametric ARIMA(p,1,1) model to capture the autocorrelation in the series from which resamples can thus be independently drawn.

We also exploit one-to-one correspondence of model parameters to impose the null constraint on bootstrap resamples. To show this explicitly, let $\lambda = \sigma_u^2/\sigma_\epsilon^2$ then the moving average parameter can be written as $\theta = \frac{1}{2}(\lambda + 2 - (\lambda^2 + 4\lambda)^{1/2})$. Hence, the stationarity null $\sigma_u^2 = 0$ implies $\theta = 1$; a moving average unit root and this idea was exploited in constructing the stationarity test proposed by Saikkonen and Luukkonen (1993). We impose the null constraint by imposing a moving average unit root on generated bootstrap resamples.

Finally, it is the choice of the test statistic that separates our bootstrap test from other bootstrap stationarity tests. Kuo and Tsong (2005) and Lee and Lee (2012) pick the KPSS test statistic, $KPSS = (T^{-2}\sum_{t=1}^{T} S_t^2)/\hat{\sigma}^2$, to bootstrap. On the other hand, we focus only on the numerator; the normalized sum of squared partial sums to gain some power advantages and let's denote it as $NSSPS = T^{-2}\sum_{t=1}^{T} S_t^2$. Under the alternative hypothesis the numerator of the KPSS test statistic is $O_p(T^2)$ and the denominator is $O_p(lT)$. Hence, under the alternative KPSS test statistic diverges at rate T/l whereas the NSSPS test statistic diverges at a faster rate T^2. We then expect this faster divergence rate to translate into a higher power for our bootstrap test than the bootstrap tests using the KPSS test statistic. Moreover, with our choice of the test statistic we also avoid estimating the long run variance in each bootstrap resample. Our simulations indicate that estimation of long run variance in each bootstrap resample deteriorates the size of the bootstrap tests.

Our bootstrap resampling scheme is described in the following.

1. Given a sample $\{y_t\}_{t=1}^{T}$ generated from the unobserved component model, fit an ARIMA(p,1,1) to the series y_t using the maximum likelihood principle. Specifically, for intercept only, the model to be estimated is $\Delta y_t = \sum_{i=1}^{p} \rho_i \Delta y_{t-1} + \eta_t - \theta\eta_{t-1}$, while for intercept and trend $\Delta y_t = \beta + \sum_{i=1}^{p} \rho_i \Delta y_{t-1} + \eta_t - \theta\eta_{t-1}$. The resulting estimated parameters and residuals are denoted by $\hat{\rho}_i, \hat{\beta}, \hat{\theta}$, and $\hat{\eta}_t$.
2. Center the residuals $\hat{\eta}_t$ by $\bar{\eta}_t = \hat{\eta}_t - \frac{1}{T-1}\sum_{t=2}^{T} \hat{\eta}_t$.
3. Draw a bootstrap sample of size T with 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 using the estimated parameters from step 1 and setting $\theta = 1$ such that $\Delta y_t^* = \sum_{i=1}^{p} \hat{\rho}_i \Delta y_{t-1}^* + \eta_t^* - \eta_{t-1}^*$ (for intercept only) and $\Delta y_t^* = \hat{\beta} + \sum_{i=1}^{p} \hat{\rho}_i \Delta y_{t-1}^* + \eta_t^* - \eta_{t-1}^*$ (for intercept and trend).
5. Calculate $NSSPS_\mu$ and $NSSPS_\tau$ using $\{y_t^*\}_{t=1}^{T}$, denoted by $NSSPS_\mu^*$ and $NSSPS_\tau^*$, respectively
6. Repeat step 3 to step 5 B times
7. Compute the empirical distribution function for $NSSPS_\mu^*$ or $NSSPS_\tau^*$, and use the empirical distribution function as an approximation to the cumulative distribution function 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 remarks are in order here. In step 1, ARIMA(p,1,1) is fit to data to capture the dependence structure of the underlying data generating process. As a result, the residuals are independent and this enables us to generate independent bootstrap samples in steps 3 and 4. The lag order p in general needs to increase with sample size. In practice, the appropriate lag order is usually chosen by some information criteria. We follow this practice in the subsequent Monte Carlo study.

In step 4 we generate our bootstrap samples $\{y_t^*\}$ and two points are worth mentioning. First, in step 4 we replicate the dependence structure of the underlying data generating process on the bootstrap samples $\{y_t^*\}$. This is accomplished by using the parameter estimates from step 1. Second, in step 4 we impose the null (stationarity) constraint on the bootstrap samples by setting $\theta = 1$ regardless of the estimate $\hat{\theta}$ from step 1.

By the bootstrap invariance principle and continuous mapping theorem, the asymptotic limit of the bootstrap version of NSSPS test statistic is equivalent to that of the original NSPSS test statistic and it has the following form under the null

$$NSSPS_i^* \xrightarrow{d} \sigma^2 \int_0^1 V_i^2(r)dr \quad for\ i = \mu, \tau$$

where σ^2 and $V(r)$ are as defined in section 2. The null distribution of the original test statistic is not pivotal and depends on the nuisance parameter σ^2. However, since our bootstrap resampling scheme is able to replicate the autocorrelation structure, the bootstrap test statistic adapts automatically and directly approaches the same limiting distribution. Another example of bootstrap test statistic with non-pivotal limit distribution can be found in Ferretti and Romo (1996).

4. Simulation Studies

In this section we investigate the finite sample performance of our bootstrap NSSPS test and compare it to asymptotic and bootstrap KPSS test. First, we look the empirical size performance of competing tests. The simulation setup under the null is the same as that in Table 1 where the data generating process is an AR(1): $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$. We consider two different sample sizes: T=100 and T=300 are approximately the sample sizes encountered in practice. To observe the impact of persistence of the data on the size performance we vary parameter ρ; $\rho = \{0, 0.3, 0.5, 0.8, 0.9, 0.92, 0.94\ 0.96\ 0.98\}$. As it is obvious from this set of values, we are mainly interested in the size performances for highly persistent yet stationary data. In estimating the long run variance for the KPSS test, sample size dependent lag truncation parameters are considered such that $l = [k(T/100)^{\ 1/4}]$.

In our bootstrap resampling scheme, we apply the AIC rule to choose the autoregressive lag order. Other lag order selections, as long as they satisfy the growth condition, may be considered in practice. The number of replications for the asymptotic tests are 10,000, while 1,000 replications are performed for the bootstrap tests with 100 bootstrap resamples in each replication. The rejection rates for all tests are computed and reported at nominal 5% level.

Table 2 presents the size performances of the tests for models with intercept. To save some space, we will not report the simulation results of the tests for models with intercept and time trend, as they share very similar qualitative outcomes as reported here. The main conclusion from Table 2 is that our bootstrap test, $NSSPS$, controls size better than both the

T	$\boldsymbol{\rho}$	$KPSS_{\mu}(4)$		$KPSS_{\mu}(12)$		$NSSPS_{\mu}$
		asym.	*boot.*	*asym.*	*boot.*	*boot.*
100	0.98	81.9%	19.6%	47.3%	15.3%	12.0%
	0.96	73.4%	15.1%	36.1%	10.2%	10.4%
	0.94	66.1%	11.8%	29.3%	9.3%	9.0%
	0.92	58.7%	11.3%	23.3%	8.1%	8.1%
	0.9	53.3%	11.1%	20.1%	7.4%	7.9%
	0.8	32.7%	8.9%	9.5%	5.1%	6.8%
	0.5	12.1%	5.5%	4.7%	5.0%	7.3%
	0.3	7.2%	6.0%	4.2%	5.8%	6.1%
	0	4.1%	4.6%	3.3%	5.7%	6.3%
300	0.98	92.1%	21.7%	57.6%	17.6%	10.9%
	0.96	81.8%	12.9%	40.7%	9.7%	5.3%
	0.94	70.9%	10.3%	29.9%	8.4%	4.3%
	0.92	59.7%	9.2%	23.9%	7.2%	4.7%
	0.9	52.1%	8.1%	18.7%	6.5%	5.0%
	0.8	28.5%	6.0%	10%	5.7%	5.7%
	0.5	9.3%	5.9%	5.7%	5.6%	6.0%
	0.3	7.4%	5.8%	4.8%	5.5%	5.7%
	0	4.7%	5.2%	3.8%	5.2%	5.5%

Notes: 1. The DGP is $y_t = \rho y_{t-1} + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} N(0,1)$

2. $KPSS_{\mu}(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$

3. *boot.* and *asym.* denote the bootstrap test and asymptotic test respectively

Table 2: Size performances of asymptotic and bootstrap stationarity tests at 5% nominal level

asymptotic and bootstrap KPSS tests. The size distortions of the asymptotic KPSS test is similar to the size distortions in Table 1. The size distortions increase with sample size and persistence of the data, and decreases with the bandwidth. The bootstrap KPSS test reduces size distortions and has empirical size close to nominal level. Contrary to the asymptotic counterpart, size distortions of the bootstrap KPSS decrease with sample size. Increasing bandwidth improves the empirical size performance of the bootstrap KPSS test as well, however a larger bandwidth also leads to a reduction in power, as in Table 3, so there is a size-power tradeoff in the choice of bandwidth for the bootstrap KPSS test.

Our bootstrap NSSPS test displays size control close to that of the bootstrap KPSS with the larger bandwidth when the sample size is 100 but it clearly outperforms all other alternatives when the sample size increases to 300. The same bootstrap resampling scheme is used for both KPSS and our NSSPS test, then what accounts for the better performance of NSSPS over KPSS? We believe that NSSPS has better size control than KPSS because estimation of long run variance in each bootstrap resample deteriorates the size of the KPSS tests. Especially for highly persistent data, the estimate of the long run variance is quite noisy and this noisy estimate is the source of the different size performances of bootstrap NSPSS and KPSS. Finally, we explain the size distortions of our bootstrap test when $\rho \geq 0.95$ for $T = 100$ and $\rho \geq 0.975$ for $T = 300$ in Table 2. These are due to imprecise estimates of the ARIMA model in step 1 of our resampling scheme. The estimates of the autoregressive and moving average parameters are not accurate due to near cancellation effect when ρ is very close to unity and this is the source of aforementioned size distortions of NSSPS in Table 2.

T	$\lambda = \sigma_u^2/\sigma_\epsilon^2$	$KPSS_\mu(4)$		$KPSS_\mu(12)$		$NSSPS_\mu$
		$asym.$	$boot.$	$asym.$	$boot.$	$boot.$
100	1	82.1%	82.0%	61.2%	60.7%	84.1%
	0.1	75.9%	73.2%	58.6%	59.3%	76.3%
	0.01	50.6%	56.3%	42.6%	46.9%	63.1%
	0.001	14.6%	17.6%	13.4%	13.2%	20.5%
	0.0001	5.7%	8.5%	5.8%	6.8%	8.0%
300	1	96.2%	97.2%	81.6%	83.3%	99.8%
	0.1	95.7%	94.7%	80.9%	81.2%	94.9%
	0.01	88.0%	88.4%	75.3%	77.1%	94.5%
	0.001	55.4%	57.2%	50.8%	52.0%	61.3%
	0.0001	15.2%	17.2%	14.8%	15.3%	18.1%

Notes: 1. The DGP is $y_t = r_t + \epsilon_t$, and $r_t = r_{t-1} + u_t$, with $\epsilon_t \overset{iid}{\sim} N(0,1)$ and $u_t \overset{iid}{\sim} N(0, \sigma_u^2)$
2. $KPSS_\mu(k)$ is calculated with a bandwidth number set to $l = [k(T/100)^{1/4}]$
3. *boot.* and *asym.* denote the bootstrap test and asymptotic test respectively
4. The figures reported for the asymptotic test are the size-adjusted empirical power

Table 3: Power performances of asymptotic and bootstrap stationarity tests at 5% nominal level

Table 3 presents the empirical power performances of the tests for models with intercept only. The data generating process used in this simulation is the unobserved component model of section 2 with $\lambda = \sigma_u^2/\sigma_\epsilon^2$ is defined as the signal-to-noise ratio. The autoregressive lag order in our resampling scheme is again selected by AIC. The number of replications for the asymptotic tests are 10,000, while 1,000 replications are performed for the bootstrap tests with 100 bootstrap resamples in each replication. The rejection rates for all tests are computed and reported at nominal 5% level.

As we argued in Section 3, our choice of the test statistic results in superior power of our NSSPS test to the bootstrap KPSS test. When we compare our bootstrap NSSPS test to the bootstrap KPSS, we see significantly higher power for the NSSPS relative to the KPSS with a large bandwidth number. The difference is less pronounced for the KPSS with a smaller bandwidth. However, it is important to keep in mind that the bootstrap NSSPS controls size much better than the bootstrap KPSS with small bandwidth. Thus, simulation results from Table 2 and Table 3 combined clearly favors our bootstrap NSSPS test over the bootstrap KPSS test. We also report the size-adjusted powers of the asymptotic KPSS test and see that the size improvement of our bootstrap test does not come at the expense of power loss.

5. Empirical Results

We apply our proposed bootstrap test to a set of real exchange rate data to test purchasing power parity. We seek stationarity, or mean-reverting property in the real exchange rates that corresponds to the notion of long-run purchasing power parity. Even though stationarity in the real exchange rates is not enough to prove purchasing power parity (PPP), rejection of stationarity is strong evidence against it. The stationarity test maintains what economic theory predicts under the null and rejects the theory only when there is strong evidence against it. However, asymptotic stationarity tests may result in spurious rejections of PPP due to size distortions underlined in section 2. As Table 1 suggests, the asymptotic KPSS test rejects the stationarity null too often when the data is indeed stationary. Moreover, real exchange rate data is characterized by a highly persistent autoregressive process and this is exactly where the size distortions are most severe. In order to have a more reliable result we test PPP with the bootstrap NSSPS test developed in section 3.

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*. Quarterly data is available for the following 9 developed countries over the period 1973.I-2012.IV (post-Bretton Woods period):

Australia, Canada, Denmark, Japan, Norway, New Zealand, Sweden, Switzerland, and United Kingdom.

Country	$KPSS_{\mu}(12)$ $asym.$	p	$NSSPS_{\mu}$ $boot.$
Australia	0.3008	1	0.075
Canada	**0.3625***	2	$\mathbf{0.047}^{\dagger\dagger}$
Denmark	0.2353	4	0.048
Japan	**0.609****	5	$\mathbf{0.194}^{\dagger}$
Norway	0.0968	1	0.011
New Zealand	0.2907	1	0.065
Sweden	**0.6196****	4	0.179
Switzerland	**0.4641****	1	0.091
United Kingdom	0.2218	1	0.007

Notes: 1. Quarterly data is available over 1973.I-2012.IV
2. p is the lag order of the ARIMA(p,1,1) fitted to the data and is chosen by AIC
3. $KPSS_{\mu}(12)$ is calculated with a bandwidth number set to $l = [12(T/100)^{1/4}]$
4. **(*) represents a rejection at 5% (10%) level using the asymptotic test, and $^{\dagger\dagger}(^{\dagger})$ using the bootstrap test
5. The figures below $NSSPS_{\mu}$ is calculated with the data, and critical values (not reported here) are computed using 5000 replications with 1000 resamples

Table 4: Test results for purchasing power parity

Table 4 reports the stationarity test results for the real exchange rates of 9 countries. Since the real exchange rates do not exhibit trending behavior, models with an intercept only are considered in this study. First, let's look at the results of the asymptotic KPSS test. The critical values for the asymptotic KPSS test can be found in KPSS (1992) and these are 0.463 and 0.347 at the 5% and 10% significance levels respectively. Based on these critical values, the asymptotic KPSS test rejects purchasing power parity for Japan, Sweden and Switzerland at 5% significance level and for Canada at 10%. Thus, the asymptotic KPSS test finds evidence against purchasing power parity for almost half of the countries studied.

We finally report the results of our bootstrap test. Due to the size distortions of the asymptotic KPSS test, some of the rejections mentioned in previous paragraph may be spurious. This is also suggested by the results of our bootstrap stationarity test. The rejections for Sweden and Switzerland are overturned by the bootstrap test, whereas for Canada and Japan purchasing power parity is still rejected by the bootstrap method. Hence, our bootstrap test provides more evidence for the existence of PPP than the asymptotic test. Since the size and power performance of the bootstrap test surpasses that of the asymptotic test, the bootstrap test offers a more reliable conclusion.

6. Conclusion

We propose a bootstrap stationarity test to reduce possible size distortions of the asymptotic stationarity test. Our bootstrap test utilizes the resampling scheme propped by Kuo and Tsong (2005). We show how the resampling scheme generates independent bootstrap resamples and imposes the stationarity constraint on the generated resamples. Next, we select our bootstrap test statistic to improve upon the existing bootstrap methods and we demonstrate how the faster divergence rate of our test statistic under the alternative translates into higher power.

Monte Carlo studies show that our bootstrap test not only offers higher power but also controls the size better than other bootstrap tests. Finally, we apply our test to real exchange rates to test purchasing power parity. We find less evidence against PPP than what the asymptotic tests provide.

References

Caner, M. and L. Killian (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

Ferretti, N. and J. Romo (1996), "Unit Root Bootstrap Tests for AR(1) Models," *Biometrika* 83, 849-860

Hansen B.E. (1999), "The Grid Bootstrap and The Autoregressive Model," *The Review of Economics and Statistics* 81, 594-607

Kuo, B.-S. and C.-C. Tsong (2005), "Bootstrap Inference for Stationarity," *HECER Discussion Papers* 50, 1-31

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

Lee, J. and Y.I. Lee (2012), "Size Improvement of the KPSS Test Using Sieve Bootstraps," *Economics Letters* 116, 483-486

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

Muller, U. (2005), "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series" *Journal of Econometrics* 128, 195-213

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

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

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

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