



## **Bootstrap Tests of Stationarity†**

James Morley  
Department of Economics  
Washington University in St. Louis  
One Brookings Drive  
St. Louis, MO 63130

Tara M. Sinclair  
Department of Economics  
and the Elliott School of International Affairs  
The George Washington University  
2115 G Street NW, Monroe Hall Suite 340  
Washington, DC 20052

RPF Working Paper No. 2008-011  
<http://www.gwu.edu/~forcpgm/2008-011.pdf>

October 24, 2008

RESEARCH PROGRAM ON FORECASTING  
Center of Economic Research  
Department of Economics  
The George Washington University  
Washington, DC 20052  
<http://www.gwu.edu/~forcpgm>

# Bootstrap Tests of Stationarity<sup>†</sup>

James Morley  
Department of Economics  
Washington University in St. Louis  
One Brookings Drive  
St. Louis, MO 63130

Tara M. Sinclair  
Department of Economics  
and the Elliott School of International Affairs  
The George Washington University  
2115 G Street NW, Monroe Hall Suite 340  
Washington, DC 20052

October 21, 2008

JEL Classifications: C12, C15, C22

Keywords: Stationarity Test, Unobserved Components,  
Parametric Bootstrap, Monte Carlo Simulation, Finite Sample Inference

## Abstract

We compare the finite-sample performance of different stationarity tests. Monte Carlo analysis reveals that tests based on Lagrange multiplier (LM) statistics with nonstandard asymptotic distributions reject far more often than their nominal size for trend-stationary processes of the kind estimated for macroeconomic data. Bootstrap versions of these LM tests have empirical rejection probabilities that are closer to nominal size, but they still tend to over-reject. Meanwhile, we find that a bootstrap likelihood ratio (LR) test has very accurate finite-sample size, while at the same time having higher power than the bootstrap LM tests against empirically-relevant nonstationary alternatives. Based on the bootstrap LR test, and in some cases contrary to the bootstrap LM tests, we can reject trend stationarity for US real GDP, the unemployment rate, consumer prices, and payroll employment in favour of unit root processes with large permanent movements.

---

<sup>†</sup> The authors gratefully acknowledge the support of the Murray Weidenbaum Center on the Economy, Government, and Public Policy for this project as well as support from the UFF/CCAS fund at George Washington University. We wish to thank Tino Berger, Drew Creal, Gerdie Everaert, Maral Kichian, Tom King, Michael Owyang, Christoph Schleicher, Tatsuma Wada, and participants in the 14<sup>th</sup> International Conference on Computing in Economics and Finance for helpful discussions and comments. All remaining errors are our own.

## Introduction

Beginning in the 1970s, a number of studies appeared which suggested that permanent movements in many major economic time series follow a variable, or stochastic, trend, instead of a smooth deterministic time trend. Granger and Newbold (1974) were among the first to argue that macroeconomic data as a rule contained stochastic trends, characterized by unit roots, and that using these series in traditional econometric models may lead to spurious regressions. Nelson and Plosser (1982) could not reject the unit root hypothesis in favor of trend stationarity for 13 out of 14 major macroeconomic time series using statistical techniques developed by Dickey and Fuller (1979).

One problem with unit root tests is that they can have low power against estimated trend stationary processes (see, for example, DeJong et al., 1992, and Rudebusch, 1992, 1993). As a consequence, stationarity tests, in which the null hypothesis is level-stationarity or trend-stationarity and the alternative is a nonstationary unobserved components process, have become popular. The most well-known stationarity test is the KPSS test (Kwiatkowski et al. 1992) that has the form of a Lagrange multiplier (LM) test, but has a nonstandard asymptotic distribution. KPSS take a nonparametric approach to addressing any serial correlation in the process under the null hypothesis. Leybourne and McCabe (1994, hereafter LMC) consider a similar LM-type test, but take a parametric approach to addressing serial correlation.

In this paper, we consider the finite-sample properties of stationarity tests given null and alternative processes of the kind estimated for macroeconomic data. In particular, we estimate null and alternative models for four US macroeconomic time series and use the implied data generating processes to simulate data for several Monte Carlo experiments. Our analysis reveals that, given sample sizes and estimated processes, the asymptotic LM tests are dramatically over-

sized, regardless of whether nonparametric or parametric approaches are taken for addressing serial correlation. We also consider bootstrap versions of the LM tests. While the empirical rejection probabilities are closer to the nominal size of the tests, they still have a tendency to over-reject. Meanwhile, the bootstrap LM tests have low power against empirically-relevant alternatives. In addition to asymptotic and bootstrap versions of the LM tests, we propose and consider a bootstrap version of a likelihood ratio (LR) test of stationarity. The LR test statistic does not have a standard asymptotic distribution in this setting for a variety of reasons, including the nonstationary alternative, a pile-up problem, the true parameter being on the boundary of the parameter space, and the presence of a nuisance parameter that is not identified under the null. However, it is relatively straightforward to bootstrap its distribution under the null hypothesis for any given application. It turns out that the bootstrap LR test has excellent finite-sample properties, including accurate size and more power than the bootstrap LM tests for empirically-relevant alternatives. For our application, we apply the bootstrap stationarity tests to macroeconomic data and are able to reject trend stationarity using the bootstrap LR test for US real GDP, the unemployment rate, consumer prices, and payroll employment, while we are unable to reject in some cases using the bootstrap LM tests.

A key aspect of our analysis is the explicit consideration of unobserved components models that allow for correlation between permanent and transitory movements in a unit root process. This allowance is important because estimates for such models support non-zero correlations for macroeconomic time series, meaning that allowing for the correlation increases the power of the bootstrap LR test. Meanwhile, the estimated alternatives that are supported by the bootstrap LR test imply the presence of large permanent movements over time. Thus, the

statistical significance of the bootstrap LR test corresponds directly to economic significance in terms of the importance of unit roots in macroeconomic time series.

This paper proceeds as follows. Section 2 discusses asymptotic and bootstrap LM tests of stationarity and proposes a bootstrap LR test. Section 3 presents Monte Carlo analysis of the finite-sample size and power performance of the various stationarity tests, as well as applications of the tests to U.S. macroeconomic data. Section 4 concludes.

## **Section 2: Stationarity Tests**

As discussed in KPSS, we can think about a time series of interest as the sum of a deterministic trend, a random walk component, and a stationary error. In this setting, the test of trend stationarity involves determining whether or not the innovation to the random walk component has zero variance. In this section we focus on three tests of stationarity: KPSS (Kwiatkowski et al., 1992), LMC (Leybourne and McCabe, 1994), and a bootstrap likelihood ratio (LR) test. Both KPSS and LMC use versions of a Lagrange Multiplier (LM) test. We discuss their tests in the following sections, and then we propose the LR test.

### **2.1 The LM Statistic**

Let  $e_t$ ,  $t = 1, \dots, T$ , be the residuals from a regression of the time series of interest,  $y$ , on an intercept and a time trend. Assuming that the innovations to the random walk component are normally distributed and that the stationary errors are iid  $N(0, \sigma_\varepsilon^2)$ , the (one-sided) LM statistic is the locally best invariant (LBI) statistic for the hypothesis that the innovation to the random walk component has a zero variance (Nyblom and Mäkeläinen, 1983; Nabeya and Tanaka, 1988). The statistic is derived from the partial sum process,  $S_t$ , of these residuals, and the estimate of the error variance from the regression,  $\hat{\sigma}_\varepsilon^2$ :

$$LM = \sum_{t=1}^T S_t^2 / \hat{\sigma}_\varepsilon^2 \quad (1)$$

The distribution of the LM statistic depends critically on the assumption of iid errors. However, this assumption is unrealistic for most time series to which a stationarity test would be applied because these series are in general highly dependent over time. To address the error autocorrelation, KPSS take a nonparametric approach, whereas LMC take a parametric approach.

### 2.1.1 KPSS Nonparametric Approach

To allow for general forms of temporal dependence, KPSS modify the LM test statistic by replacing  $\hat{\sigma}_\varepsilon^2$  with a nonparametric estimator of the “long-run variance” (i.e.,  $2\pi$  times the spectral density of  $\varepsilon$  at frequency zero), which can be denoted as  $s^2(l)$ :

$$LM = \sum_{t=1}^T S_t^2 / s^2(l) \quad (2)$$

where  $s^2(l) = T^{-1} \sum_{t=1}^T e_t^2 + 2T^{-1} \sum_{s=1}^l w(s,l) \sum_{t=s+1}^T e_t e_{t-s}$  and  $w(s,l)$  is a weighting function, commonly the Bartlett kernel,  $w(s,l) = 1 - s/(l+1)$ . There is a trade-off between size distortions and test power related to the selection of the lag truncation parameter,  $l$ : the larger the choice of  $l$ , the smaller the size distortion, but the lower the power of the test. Setting  $l$  equal to zero is equivalent to not correcting for error autocorrelation. We use the generalized KPSS test of Hobijn, Franses and Ooms (2004) with the Bartlett kernel, automatic lag selection (following Newey and West, 1994), and initial bandwidth ( $n$ ) as a function of the length of the series:  $n = \text{int} \left[ 4 * (T/100)^{(2/9)} \right]$ , where  $\text{int}$  is a function that takes the integer portion. KPSS derive the asymptotic distribution of their statistic as an integrated Brownian bridge for level stationarity and an integrated second-level Brownian bridge for trend stationarity. Thus, in both cases, the

asymptotic distribution is pivotal. Meanwhile, in terms of finite-sample distributions, Caner and Kilian (2001) use Monte Carlo analysis to show that a parametric bootstrap of the KPSS test reduces size distortions compared to the asymptotic test for stylized stationary processes with similar levels of persistence to those estimated for real exchange rates. However, as we find with our Monte Carlo analysis, they also find that the bootstrap version of the KPSS has very low power against a nonstationary alternative with large permanent movements.

### 2.1.2 Leybourne and McCabe Parametric Approach

Leybourne and McCabe (1994, LMC) employ a parametric version of the LM test of the null hypothesis of stationarity against the presence of a unit root. LMC address serial correlation by assuming an AR( $p$ ) under the null and thus they include  $p$  lagged terms of  $y_t$  in their initial model specification. To obtain their test statistic, LMC construct a series:

$$y_t^* = y_t - \sum_{i=1}^p \phi_i^* y_{t-i}, \quad (3)$$

where the  $\phi_i^*$  are the maximum likelihood estimates of  $\phi_i$  from the fitted ARIMA( $p, 1, 1$ ) model:

$$\Delta y_t = \beta + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \zeta_t + \theta \zeta_{t-1}. \quad (4)$$

The ARIMA( $p, 1, 1$ ) is the reduced form of the unobserved components model LMC assume under the alternative, the local level model of Harvey (1989), which is second-order equivalent in moments to an ARIMA( $p, 1, 1$ ) process. This approach gives consistent estimates of the AR( $p$ ) parameters under the null and under the alternative.<sup>1</sup> By contrast, if we were to estimate

---

<sup>1</sup> McCabe and Leybourne (1998) show that the marginal distribution of the maximum likelihood estimates of AR parameters in the case of an MA unit root is asymptotically the same as the distribution of the maximum likelihood estimates in a pure AR( $p$ ) model. This means that if we estimate the first difference of a stationary model (i.e. estimating the alternative when the null is true), the AR parameter estimates can be used for the null. Meanwhile, for a more complicated alternative, such as the nonstationary unobserved components process considered in this paper,

an  $AR(p)$  in levels, the estimates would be inconsistent under the alternative and the test would have little power, as discussed in LMC.

Similar to KPSS, LMC calculate the residuals,  $e_t$ , from a regression of  $y_t^*$  from equation (3) on an intercept and a time trend. The LMC test statistic is then

$$LMC = e'Ve, \quad (5)$$

where  $V$  is a  $T \times T$  matrix with  $ij$ th element equal to the minimum of  $i$  and  $j$ . LMC derive the asymptotic distributions under level-stationarity and trend-stationarity of standardized versions of (5), which, like the KPSS test, depend on integrated Brownian bridges and are pivotal. Also, similar to the KPSS test, Caner and Kilian (2001) find that a parametric bootstrap version of the LMC test reduces finite-sample size distortions compared to the asymptotic test, but the bootstrap test has low power.

## 2.2 The LR Statistic

Likelihood ratio test statistics have regularly been used for tests of parameter constancy. As discussed extensively in the literature, the distribution of the likelihood ratio test statistic is nonstandard for these tests for a variety of reasons, including that, under the null, variances of time-varying parameters are on the boundary of the parameter space, there may be nuisance parameters that are only identified under the alternative, and because the alternative may be a nonstationary process. Because asymptotic distributions are often unknown, researchers have relied on bootstrapped critical values and p-values. Examples include the bootstrap approach of Feng and McCulloch (1996) applied to finite mixture models. Di Sanzo (forthcoming) also finds

---

it is straightforward to modify the reduced-form model to allow it to capture the full parametric structure under the alternative, while still being consistent under the null.

that a bootstrap-based LR test works well in the case of testing for parameter constancy in the context of Markov switching models.

A stationarity test is an example of a test for parameter constancy with the specific alternative hypothesis of a unit root process. In particular, the alternative can be thought of as a time-varying parameter model in which the unconditional mean follows a random walk. For the LR test proposed here, we follow the parametric approach of LMC. We also assume an  $AR(p)$  under the null. Our alternative, however, is an  $ARIMA(p, 1, p)$  in reduced form, which directly corresponds to the notion that the unconditional mean is a random walk, whereas LMC consider a local level model in which the intercept in an autoregression follows a random walk. If the true process is an unobserved components model, then even accounting for the  $AR(p)$  in constructing the test statistic, there is still a transitory MA component to the residuals. Under the null of (trend) stationarity, our model in first differences is  $ARMA(p, 1)$  with a unit MA root. It should be noted that for our analysis, we modify the reduced-form model for LMC in order to allow it to capture the full parametric structure under the alternative, while still being consistent under the null. Thus, differences in performance for the LMC test versus the bootstrap LR test are not due to any parametric misspecification.

### **Section 2.2.1: The Correlated Unobserved Components (UC) Representation**

In our analysis, we consider a correlated unobserved components (UC) model, which nests the full range of possibilities about the relative importance of permanent and transitory movements in macroeconomic time series.

Suppose we have a series ( $y$ ) which may be represented as the sum of two unobserved components: a “trend” component and a “cycle” component. The “trend” ( $\tau$ ), also called the permanent component, is the steady-state level after removing all temporary movements. The

“cycle” ( $c$ ), also called the transitory component, embodies all temporary movements and is assumed to be stationary:

$$y_t = \tau_t + c_t, \quad t = 1, \dots, T. \quad (6)$$

We begin by assuming that the series is I(1), thus a random walk for each of the trend components allows for permanent movements in the series. It is also possible to allow for a drift ( $\mu$ ) in the trend:

$$\tau_t = \mu + \tau_{t-1} + \eta_t. \quad (7)$$

We model the transitory component as a stationary and invertible ARMA( $p, q$ ) process:

$$\phi_p(L)c_t = \theta_q(L)\varepsilon_t. \quad (8)$$

Following Morley, Nelson, and Zivot (2003), we assume the innovations ( $\eta_t$ , and  $\varepsilon_t$ ) are jointly normally distributed random variables with mean zero and variance-covariance matrix  $\Sigma$ :

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\eta\varepsilon} \\ \sigma_{\eta\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix}.$$

This model also nests the trend stationary null in the case that the variance of the innovation to the permanent component ( $\sigma_\eta^2$ ) is zero (and the covariance,  $\sigma_{\eta\varepsilon}$ , is not identified). As discussed in Morley, Nelson, and Zivot (2003), this UC model is only identified for ARMA( $p, q$ ) specifications of the transitory component for which  $p \geq q + 2$ . This requirement is satisfied in our empirical applications.

Allowing for correlation between the permanent and transitory innovations is the key point of using the correlated UC approach. If we only considered alternatives for which the correlation between the permanent and transitory innovations,  $\sigma_{\eta\varepsilon}$ , is restricted to be zero, we would be placing a strong restriction on the variability of the permanent component (i.e., it can

be no greater than the variability of  $\Delta y$ ). To the extent that this restriction is false, it would reduce the power of our bootstrap LR test. Based on our estimates, the restriction appears to be false for the macroeconomic variables that we consider.

The correlated UC model can be cast into state-space form and we use the Kalman filter for maximum likelihood estimation of the parameters for both the restricted and unrestricted models to directly obtain the likelihood ratio test statistic:

$$LR = -2(LLV_{\sigma_{\eta}=0} - LLV_{unrestricted}), \quad (9)$$

where  $LLV$  is the log of the likelihood value for each model. Unlike for the KPSS and LMC test statistics, the asymptotic distribution of this test statistic is unknown. Therefore in this case we bootstrap its distribution, as discussed next.

### **Section 2.3 The Bootstrap Test Procedure**

Asymptotic distributions often provide poor approximations to finite-sample distributions of test statistics. Given a pivotal asymptotic distribution, such as we have for the LM tests, bootstrap tests can often produce asymptotic refinements such that bootstrap distributions provide a better approximation of finite-sample distributions. Meanwhile, bootstrap distributions can also be useful as a way to approximate unknown asymptotic distributions, such as we have for the LR test, although they do not always provide asymptotic refinements in such settings (see Horowitz, 2001, for a full discussion of bootstrap inference).

If the true data generating process (DGP) for a given series is either a trend stationary AR(p) process or a nonstationary UC process, we are interested in testing whether  $\sigma_{\eta}$ , the standard deviation of the innovation to the permanent component in a correlated UC model, equals zero or if it is positive. If  $\sigma_{\eta}$  equals zero, then  $\sigma_{\eta\varepsilon}$ , the covariance between the permanent

and transitory innovations, is not identified. The distribution of the likelihood ratio test statistic under the null is nonstandard for a variety of reasons, including the fact that the null value of the parameter lies on the boundary of the parameter space and there is an unidentified nuisance parameter under the alternative. However, a parametric bootstrap can be used to approximate the distribution of the LR test statistic. Similarly, for the KPSS and LMC test statistics, the bootstrap can be used to approximate their finite-sample distributions, as was done in Caner and Kilian (2001).

In our analysis, we consider parametric bootstrap tests. Specifically, bootstrap simulated data are based on estimated parameters and distributional assumptions. The full bootstrap testing procedure is given as follows:

- 1) Consistently estimate the parameters under the null. In this case, the model is a trend stationary AR(2) process. We also calculate the likelihood value under the alternative, being careful to consider a large number of different starting values for numerical optimization in order to ensure that we find the global maximum. We then construct the likelihood ratio test statistic for the actual or Monte Carlo data (depending on whether we are using the bootstrap test for actual data or using Monte Carlo simulated data to explore the size and power of the different tests). Note that the test statistic is actually a supLR statistic since we obtain the maximum likelihood estimate of the nuisance parameter (the covariance  $\sigma_{\eta\varepsilon}$ ) under the alternative. We also construct the KPSS statistic and the LMC statistic for the actual or Monte Carlo data, with the appropriate parametric assumption made when constructing the LMC statistic.

- 2) Simulate bootstrap data based on the parameters estimated under the null in step 1. We simulate bootstrap data a maximum of 499 times for each bootstrap test in our applications, while we do so 199 times for each bootstrap test in our Monte Carlo exercises.<sup>2</sup>
- 3) For each bootstrap data series, estimate both the null and alternative models. For the alternative models we consider a large number of starting values for numerical optimization in order to ensure that we obtain the global maximum.
- 4) For each bootstrap data series, construct bootstrap draws of the test statistics based on the estimates from step 3.
- 5) Calculate a bootstrapped p-value as the number of bootstrap draws of a given test statistic that are greater than the test statistic found from the actual or Monte Carlo data, divided by the total number of bootstrap draws (MacKinnon, 2002).

### **Section 3: Empirical Analysis**

We apply the various stationarity tests to four important US macroeconomic time series. For ease of modeling, all series have been transformed to the quarterly frequency which allows us to use an AR(2) transitory component as a reasonable specification in all cases. We consider observations from 1947.1 through 2006.4 for US real GDP and the CPI, from 1948.1 through 2006.4 for the US unemployment rate, and from 1939.1 through 2006.4 for the US total nonfarm payroll employment.<sup>3</sup> For the three monthly series (CPI, unemployment rate, and employment),

---

<sup>2</sup> For the Monte Carlo experiments, we use the procedure proposed in Davidson and MacKinnon (2000) and consider fewer than 199 draws in a given bootstrap experiment if the estimated p-value from the bootstrap is significantly smaller or larger than the size at a 5% level.

<sup>3</sup> All data were obtained from the FRED2 database of the Federal Reserve Bank of St. Louis.

quarterly averages of the data are used. We further transform the data by taking 100 times the natural log of each series except for the unemployment rate, which is modeled directly in levels.

The four series provide a range of different empirically-relevant processes with which to illustrate the relative effectiveness of the different stationarity tests in our Monte Carlo experiments. Tables 1a through 1d present the parameter estimates based on trend stationary and unobserved components models of US Real GDP, the unemployment rate, the CPI, and payroll employment, respectively. For all four series under the null, the autoregressive dynamics are highly persistent, which is exactly the setting where standard unit root tests have low power and we would like to use a stationarity test. Under the alternative, the permanent movements are large, but their relative importance and the correlations with transitory movements vary across the series.

### **Section 3.1: The Monte Carlo Experiments**

Monte Carlo experiments provide the standard way to evaluate the finite-sample properties of tests for given data generating processes (DGPs). While both KPSS and LMC use Monte Carlo analysis to evaluate the finite sample properties of their test statistics, their assumed DGPs are highly stylized and do not appear to correspond very closely to estimated processes for macroeconomic variables. Here, we consider Monte Carlo analysis of the finite-sample size and power properties of stationarity tests for empirically-relevant DGPs.

For our experiments, we simulate data based on our estimates of the null and alternative models for each of the four data series discussed above. We generate the same number of observations for each of the simulated Monte Carlo samples as we have for the actual data series. For each of the four trend stationary AR(2) DGPs (the first column of Tables 1a through 1d), we generate 1000 simulated samples and consider empirical rejection probabilities to compare the

size of our bootstrap test with both the asymptotic and bootstrap versions of KPSS and LMC. We then use the nonstationary correlated UC estimates to simulate 1000 data samples under the alternative and consider empirical rejection probabilities to compare the power of the tests. For each simulated Monte Carlo sample, we follow the full bootstrap procedure outlined in Section 2.3 that we also apply to the actual macroeconomic data later in Section 3.3. We present the results for the Monte Carlo experiments in Tables 2a through 2d.

As is shown in these tables, we find that empirical rejection probabilities for the bootstrap LR test are much closer than those for the other tests to the nominal size of the test for trend stationary processes of the kind estimated for macroeconomic variables. For a nominal 5% test, the bootstrap LR size ranges from 4.5% to 6.2% whereas the bootstrap LM tests are always oversized and range from 5.8% to 11.4%. Meanwhile, the asymptotic KPSS test is severely oversized with an empirical rejection probability of as much as 86.3% with data generated under the null. The empirical rejection probability of the asymptotic LMC test ranges widely from 6.9% to 27.1%.

We also find higher power using the bootstrap LR test for empirically-relevant unit root processes. In particular, for the power experiments with data simulated under the alternative, the empirical rejection probabilities for the LR statistic are larger than for the bootstrap version of KPSS in all four cases and larger than for the bootstrap version of LMC in three of four cases. In the case of the DGP corresponding to the US unemployment rate estimates, the bootstrapped LMC test has slightly higher power (52.8% for KPSS as compared to 49.2% for the LR test). However, in that case, the size of the bootstrapped LMC test is not very well controlled at 7.3% compared to the bootstrap LR size of 4.9% for a nominal size of 5%. Thus, the bootstrap LR test always performs best in terms of size-adjusted power.

### **Section 3.2: Why does the power of the tests vary across different DGPs?**

While the LM tests are LBI, meaning that they have the highest asymptotic power against “local” alternatives (i.e., alternatives in which the variance of the shocks to the random walk component is small, even asymptotically under the thought experiment of letting the variance shrink with the sample size), they clearly do not have the highest power against the alternatives in our Monte Carlo power experiments.<sup>4</sup> One reason for this result is that the alternatives are in no sense “local” given that the variances of the permanent shocks are large. As discussed in Nyblom (1989), a likelihood ratio test of parameter constancy can have more power against distant alternatives than the LBI test. However, it is interesting to note that there is no direct relationship between the power of the tests and the signal-to-noise ratios for the DGPs as measured by the relative variances of the permanent and transitory innovations. Instead, what appears to matter more is the correlation between the permanent and transitory innovations, with the more negative the correlations, the more the LR test has power. This is related to a signal-to-noise issue in the following sense: when the correlation is exactly negative one, as it is for the DGPs based on the CPI and payroll employment, there is no independent transitory shock, but only transitory movements due to a slow adjustment of the process to permanent shocks. Thus, the “signal-to-independent-noise” ratio is infinite. In such cases, the LR test has very high power, while the LM-based tests do not. Evidently, the LM-based tests are more easily fooled into thinking the whole process is trend stationary with highly persistent autoregressive dynamics, rather than a unit root process with large permanent movements. Given that a correlation of negative one turns up for two of the four series, it is certainly an empirically-relevant

---

<sup>4</sup> Bailey and Taylor (2002) show that if the cycle and trend innovations are contemporaneously correlated, then the test statistic used by KPSS is still the LBI test for a null of stationarity.

phenomenon. Meanwhile, in the case of the unemployment rate, which has the lowest correlation (in absolute value) and, therefore, the lowest signal-to-independent-noise ratio, the various tests have more similar power.

### **Section 3.3: Applications**

Having used Monte Carlo analysis to evaluate the finite-sample performance of the various stationarity tests for DGPs related to US Real GDP, the unemployment rate, the CPI, and payroll employment, we now turn to applying the bootstrap tests to the actual data. Table 3 presents the results. In all four cases we can reject the null of a trend stationary AR(2) process in favor of the correlated UC process using the bootstrap LR test. For US real GDP and the unemployment rate, all of the bootstrap tests agree on the rejection of trend stationarity.<sup>5</sup> However, for CPI and payroll employment we find conflicting results. For these two series, which were the two series with an estimated perfect negative correlation between permanent and transitory innovations, the bootstrap versions of KPSS and LMC do not reject the null, while the bootstrap LR test does. The estimates under the alternative for CPI and payroll employment correspond to large permanent movements and no independent transitory shocks. Thus, to conclude that they are stationary (as one would do using the bootstrap LM tests) would result in quite different long-horizon forecasts and economic implications than to conclude that they have a unit root (as one would do using the bootstrap LR test).

---

<sup>5</sup> It is possible that level stationarity is a more appropriate null hypothesis for the unemployment rate. However, allowing for trend stationarity should only serve to diminish the power of our tests if level stationarity is true. Given that we reject trend stationarity of the unemployment rate for all three tests, this loss of power is not a particular concern, especially for the bootstrap LR test that accurately controls size (0.049) for the null DGP based on the unemployment rate.

## Section 4: Conclusions

For the relatively small sample sizes that are available for macroeconomic time series, Monte Carlo analysis reveals that asymptotic tests of stationarity have large size distortions given estimated trend-stationary processes. Meanwhile, correcting for these size distortions using bootstrap versions of Lagrange multiplier (LM) tests results in low power against estimated nonstationary processes. By contrast, we find that a bootstrap likelihood ratio (LR) test has more accurate finite-sample size and higher power than bootstrap LM tests. As discussed by Caner and Kilian, “we learn very little from conducting tests with size-corrected critical values except in the rare case of a rejection of stationarity” (2001, page 641). Thus, having a more powerful test of stationarity is extremely useful, especially given that the bootstrap LR test allows us to reject trend stationarity in favour of correlated unobserved components processes with large permanent movements for all four of the macroeconomic time series under consideration.

One potential drawback of Monte Carlo analysis of finite-sample size and power for competing tests is uncertainty about how general the results are for other data generating processes (DGPs). However, it should be noted that advances in computational power make it feasible to consider Monte Carlo analysis in any given empirical setting. Thus, a practitioner can do what we have done in this paper by determining the finite-sample properties of the various asymptotic and bootstrap tests for relevant DGPs prior to conducting the tests on the actual data. Our main contribution, then, is to suggest the inclusion of a bootstrap LR test of stationarity in the set of possible tests that a practitioner should consider. For the DGPs that we consider, the bootstrap LR test has the best finite-sample performance in terms of size-adjusted power, leading us to focus on it in our applications to macroeconomic data. In other settings, Monte Carlo

analysis will reveal which test has the best size control and power. However, given that we have considered a range of different processes for macroeconomic data and the bootstrap LR test was the best performing in every case, it is likely to have broad applicability as the most powerful stationarity test in many other applications.

## References

- Bailey, R. W. and A. M. R. Taylor (2002). "An Optimal Test Against a Random Walk Component in a Non-Orthogonal Unobserved Components Model." *Econometrics Journal* 5: 520-532.
- 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.
- Davidson, R. and J. G. MacKinnon (2000). "Bootstrap Tests: How Many Bootstraps?" *Econometric Reviews* 19: 55-68.
- DeJong, D., J. Nankervis, N. E. Savin and C. H. Whiteman (1992). "Integration Versus Trend-Stationarity in Time Series." *Econometrica* 60: 423-433.
- Dickey, D. A. and W. A. Fuller (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427-431.
- Di Sanzo, Silvestro (forthcoming). "Testing for Linearity in Markov Switching Models: A Bootstrap Approach." *Statistical Methods & Applications*.
- Feng, Z. D. and C. E. McCulloch (1996). "Using Bootstrap Likelihood Ratios in Finite Mixture Models." *Journal of the Royal Statistical Society. Series B (Methodological)*. 58(3): 609-617.
- Granger, C. W. J. and P. Newbold (1974). "Spurious Regressions in Econometrics." *Journal of Econometrics* 2: 111-120.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge, U.K.: Cambridge University Press.
- Hobijn, B., P. H. Franses and M. Ooms (2004). "Generalizations of the KPSS-Test for Stationarity" *Statistica Neerlandica* 58(4): 483-502.
- Horowitz, J. L. (2001). "The Bootstrap," in *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E.E. Leamer, eds., Elsevier Science B.V., Ch. 52, pp. 3159-3228.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992). "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root." *Journal of Econometrics* 54: 159-178.
- Leybourne, S. J, and B. P. M. McCabe (1994). "A Consistent Test for a Unit Root." *Journal of Business & Economic Statistics* 12(2): 157-166.
- MacKinnon, J. (2002). "Bootstrap Inference in Econometrics." *Canadian Journal of Economics* 35: 615-645.

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

Morley, J. C., C. R. Nelson, and E. Zivot (2003). "Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different?" *The Review of Economics and Statistics* 85(2): 235-243.

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: 218-35.

Nelson, C. R. and C. I. Plosser (1982). "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics* 10: 139-162.

Newey, W.K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61: 631-653.

Nyblom, J. (1989). "Testing for the Constancy of Parameters Over Time." *Journal of the American Statistical Association* 84(405): 223-230.

Nyblom, J. and T. Mäkeläinen (1983). "Comparison of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model." *Journal of the American Statistical Association* 78: 856-64.

Rudebusch, G. D. (1992). "Trends and Random Walks in Macroeconomic Time Series: A Re-Examination." *International Economic Review* 33: 661-680.

Rudebusch, G. D. (1993). "The Uncertain Unit Root in Real GNP." *The American Economic Review* 83: 264-272.

**Table 1: Parameter Estimates  
Used for the Monte Carlo Simulations and Empirical Results**

Table 1a: Quarterly Real GDP 1947.1 - 2006.4

Description		AR(2)	UC
<b>Log Likelihood</b>	LLV	-319.274948	-316.048545
<b>S.D. of Permanent Innovation</b>	$\sigma_{\eta}$	Restricted to be 0	1.114937 (0.126771)
<b>S.D. of Temporary Innovation</b>	$\sigma_{\varepsilon}$	0.916809 (0.041781)	0.560207 (0.126678)
<b>Correlation btwn. Innovations</b>	$\sigma_{\eta\varepsilon}$	Restricted to be 0	-0.94432 (0.005852)
<b>Drift</b>	$\mu$	0.829816 (0.022145)	0.825633 (0.031162)
<b>1<sup>st</sup> AR parameter</b>	$\varphi_1$	1.316875 (0.060587)	1.363392 (0.098793)
<b>2<sup>nd</sup> AR parameter</b>	$\varphi_2$	-0.34609 (0.061048)	-0.77877 (0.107044)

Table 1b: Quarterly average unemployment rate 1948.1-2006.4 (rate)

Description		AR(2)	UC
<b>Log Likelihood</b>	LLV	-50.400775	-47.877176
<b>S.D. of Permanent Innovation</b>	$\sigma_{\eta}$	Restricted to be 0	0.21258 (0.109434)
<b>S.D. of Temporary Innovation</b>	$\sigma_{\varepsilon}$	0.297635 (0.013732)	0.385414 (0.053168)
<b>Correlation btwn. Innovations</b>	$\sigma_{\eta\varepsilon}$	Restricted to be 0	-0.763463 (0.120433)
<b>Drift</b>	$\mu$	0.005202 (0.004702)	0.002737 (0.013801)
<b>1<sup>st</sup> AR parameter</b>	$\varphi_1$	1.585140 (0.049453)	1.480788 (0.074246)
<b>2<sup>nd</sup> AR parameter</b>	$\varphi_2$	-0.641610 (0.049619)	-0.58358 (0.074527)

Table 1c: Quarterly average of monthly CPI Index 1947.1 - 2006.4

Description		AR(2)	UC
<b>Log Likelihood</b>	LLV	-186.779607	-180.649386
<b>S.D. of Permanent Innovation</b>	$\sigma_{\eta}$	Restricted to be 0	2.835569 (0.219249)
<b>S.D. of Temporary Innovation</b>	$\sigma_{\varepsilon}$	0.527621 (0.024138)	2.576874 (0.217162)
<b>Correlation btwn. Innovations</b>	$\sigma_{\eta\varepsilon}$	Restricted to be 0	-1.00 (-)
<b>Drift</b>	$\mu$	0.924399 (0.083369)	0.932901 (0.188636)
<b>1<sup>st</sup> AR parameter</b>	$\varphi_1$	1.761617 (0.041782)	0.777762 (0.020947)
<b>2<sup>nd</sup> AR parameter</b>	$\varphi_2$	-0.763240 (0.041753)	0.100151 (0.021188)

Note: Standard errors in this table are based on fixing the correlation at its MLE.

Table 1d: Quarterly average employment 1939.1-2006.4

Description		AR(2)	UC
<b>Log Likelihood</b>	LLV	-248.895314	-226.851103
<b>S.D. of Permanent Innovation</b>	$\sigma_{\eta}$	Restricted to be 0	1.065988 (0.002730)
<b>S.D. of Temporary Innovation</b>	$\sigma_{\varepsilon}$	0.600347 (0.025843)	1.372891 (0.010493)
<b>Correlation btwn. Innovations</b>	$\sigma_{\eta\varepsilon}$	Restricted to be 0	-1.00 (-)
<b>Drift</b>	$\mu$	0.536330 (0.016391)	0.601739 (0.065365)
<b>1<sup>st</sup> AR parameter</b>	$\varphi_1$	1.733809 (0.040087)	1.521692 (0.027518)
<b>2<sup>nd</sup> AR parameter</b>	$\varphi_2$	-0.760276 (0.041434)	-0.52481 (0.031349)

Note: Standard errors in this table are based on fixing the correlation at its MLE.

## Table 2: Monte Carlo Results

Table 2a: Based on US Real GDP Parameter Estimates

Nominal Size 5%	Asymptotic	Bootstrap
KPSS	62.6%	5.8%
LMC	25.2%	9.0%
LR	N/A	5.4%
Power	Asymptotic	Bootstrap
KPSS	84.4%	19.8%
LMC	95.6%	51.4%
LR	N/A	79.2%

Table 2b: Based on Unemployment Rate Parameter Estimates

Nominal Size 5%	Asymptotic	Bootstrap
KPSS	17.9%	6.4%
LMC	9.2%	7.3%
LR	N/A	4.9%
Power	Asymptotic	Bootstrap
KPSS	56.9%	45.9%
LMC	60.5%	52.8%
LR	N/A	49.2%

Table 2c: Based on CPI Parameter Estimates

<b>Nominal Size 5%</b>	<b>Asymptotic</b>	<b>Bootstrap</b>
KPSS	86.3%	11.3%
LMC	27.1%	11.4%
LR	N/A	6.2%
<b>Power</b>	<b>Asymptotic</b>	<b>Bootstrap</b>
KPSS	90.6%	10.6%
LMC	94.1%	24.9%
LR	N/A	68.5%

Table 2d: Based on Employment Parameter Estimates

<b>Nominal Size 5%</b>	<b>Asymptotic</b>	<b>Bootstrap</b>
KPSS	23.8%	6.2%
LMC	6.9%	7.1%
LR	N/A	4.4%
<b>Power</b>	<b>Asymptotic</b>	<b>Bootstrap</b>
KPSS	85.9%	21.7%
LMC	59.0%	21.7%
LR	N/A	100.0%

**Table 3: Empirical Application Results**

<b>Data Series</b>	<b>KPSS Statistic (Bootstrapped P-value)</b>	<b>LMC Statistic (Bootstrapped P-value)</b>	<b>LR Statistic (Bootstrapped P-value)</b>
<b>Real GDP</b>	0.360 (0.026)*	2.994 (0.020)*	6.453 (0.034)*
<b>Unemployment Rate</b>	0.237 (0.008)*	0.507 (0.016)*	5.047 (0.030)*
<b>CPI</b>	0.258 (0.447)	2.497 (0.136)	12.260 (0.026)*
<b>Employment</b>	0.184 (0.116)	0.085 (0.295)	44.088 (<0.001)*

\*Reject the null of trend stationarity at 5%.