A Likelihood Ratio Test of Stationarity Based on a Correlated Unobserved Components Model

James Morley
University of New South Wales
Sydney NSW 2052
Australia

Irina Panovska
Washington University in St. Louis
St. Louis, MO 63130
USA

Tara M. Sinclair[†]
The George Washington University
Washington, DC 20052
USA

September 1, 2011

JEL Classifications: C12, C15, C22

Keywords: Stationarity Test, Likelihood Ratio, Unobserved Components, Parametric Bootstrap, Monte Carlo Simulation, Small-Sample Inference

Abstract

We propose a likelihood ratio (LR) test of stationarity based on a widely-used correlated unobserved components model. We verify the asymptotic distribution and consistency of the LR test, while a bootstrap version of the test is at least first-order accurate. Given empirically-relevant processes estimated from macroeconomic data, Monte Carlo analysis reveals that the bootstrap version of the LR test has better small-sample size control and higher power than commonly used bootstrap Lagrange multiplier (LM) tests, even when the correct parametric structure is specified for the LM test. A key feature of our proposed LR test is its allowance for correlation between permanent and transitory movements in the time series under consideration, which increases the power of the test given the apparent presence of non-zero correlations for many macroeconomic variables. Based on the bootstrap LR test, and in some cases contrary to the bootstrap LM tests, we can reject trend stationarity for U.S. real GDP, the unemployment rate, consumer prices, and payroll employment in favor of nonstationary processes with volatile stochastic trends.

[†] Corresponding author: tsinc@gwu.edu. 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 Institute for International Economic Policy (IIEP) of the Elliott School and the UFF/CCAS fund at George Washington University. We wish to thank Tino Berger, Drew Creal, William Dunsmuir, Neil Ericsson, Gerdie Everaert, Fred Joutz, Maral Kichian, Tom King, Michael McCracken, Michael Owyang, Phil Rothman, Roberto Samaniego, Christoph Schleicher, Herman Stekler, Tatsuma Wada, and participants at the 2011 AMES conference, the 2011 SCE Meetings, the 2011 SNDE meetings, the GWU Institute for Integrating Statistics in Decision Sciences Seminar, and the 2009 Joint Statistical Meetings for helpful discussions and comments. All remaining errors are our own.

Introduction

Beginning in the 1970s, a number of econometric studies suggested that permanent movements in many macroeconomic time series follow a 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 autoregressive unit roots, and that using these series in regression models may lead to spurious inferences about their underlying relationships. Nelson and Plosser (1982) could not reject the autoregressive unit root hypothesis in favor of trend stationarity for 13 out of 14 major U.S. macroeconomic time series using statistical techniques developed by Dickey and Fuller (1979).

One drawback of autoregressive unit root tests is that they can have low power in small samples 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) consider a similar LM-type test, but take a parametric approach to addressing serial correlation.

In this paper, we propose a likelihood ratio (LR) test of stationarity based on a correlated unobserved components model that has previously been applied to many macroeconomic variables in the empirical literature and compare its performance to the widely-used LM tests. Drawing from theoretical results in Davis and Dunsmuir (1996) and Chen, Davis, and Dunsmuir (1996) for a moving-average (MA) unit root test, we verify the asymptotic distribution and consistency of our LR test. Meanwhile, following Gospodinov's (2002) results for an MA(1) model in first differences, a bootstrap version of the LR test of stationarity for our unobserved components model is at least first-order accurate when the bootstrap data are generated under the null hypothesis. Having established the asymptotic validity of our test, we evaluate its small-sample performance given null and alternative processes of the kind estimated for macroeconomic data. Specifically, we estimate null and alternative models for four U.S. macroeconomic time series (real GDP, the unemployment rate, consumer prices, and payroll employment) and make use of the implied data generating processes to simulate data for several Monte Carlo experiments. These experiments reveal that, given the sample sizes and estimated processes for these time series, asymptotic tests are dramatically over-sized, regardless of whether nonparametric or parametric approaches are taken for addressing serial correlation. For the bootstrap versions of the LM tests, the empirical rejection probabilities are closer to the nominal size of the tests, but they still have a tendency to over-reject. Furthermore, the bootstrap LM tests have much lower power against empirically-relevant alternatives than their asymptotic counterparts.[1] By contrast, the bootstrap LR test has excellent small-sample properties, including accurate size and more power than the bootstrap LM tests for empirically-relevant alternatives. When we apply the bootstrap stationarity tests to the four U.S. macroeconomic time series considered in our Monte Carlo analysis, we find that we are able to reject trend stationarity in every case using the bootstrap LR test, while we are unable to reject in some cases using the bootstrap LM tests. Thus, the bootstrap LR test is more informative about the prevalence of stochastic trends in macroeconomic data than its LM counterparts.

[1] Rothman (1997) also finds that the bootstrap version of the KPSS test has low power in the case of an alternative based on estimates for U.S. real per capita GNP.

Although much of the econometrics literature on stationarity tests has focused on the LM-type tests, our LR test is motivated by a related literature on likelihood-based inference for an MA(1) model with a root of the MA polynomial close to or equal to 1. An LR test of whether the MA root equals 1 is directly related to stationarity tests due to the equivalence between unobserved components models and autoregressive moving average (ARMA) models for first differences. For example, a random walk plus noise model with normally distributed shocks is equivalent to an MA(1) model in first differences, with the MA root equal to 1 corresponding directly to stationarity for the levels. Davis and Dunsmuir (1996) derive an asymptotic approximation of the distribution of the maximum likelihood estimator and the LR test for an MA(1) model with values of the MA root close to or equal to 1, while Davis, Chen, and Dunsmuir (1996) show that the results generalize to testing the closest MA root for ARMA models. Gospodinov (2002) extends Davis and Dunsmuir's (1996) analysis to show that it is asymptotically valid to bootstrap the LR test for an MA(1) model when imposing the null hypothesis of an MA root equal to 1. We make use of these results in order to verify the asymptotic distribution and consistency of the LR test of stationarity based on a correlated unobserved components model and to consider the asymptotic validity of a bootstrap version of this test.

A key feature of our proposed LR test is its allowance for correlation between permanent and transitory movements in the time series under consideration. This is important because estimates for correlated unobserved components models support non-zero correlations for many macroeconomic variables, meaning that allowing for the correlation increases the power of the LR test by generating higher likelihood values under the alternative.[2] Also, the estimated alternatives when allowing for this correlation often imply large permanent movements in the data over time and are, therefore, far away from the null of stationarity. Thus, we find that the statistical significance of the bootstrap LR test corresponds directly to economic relevance in terms of the importance of stochastic trends in macroeconomic time series.

The remainder of the paper proceeds as follows. In Section 2, we discuss asymptotic and bootstrap tests of stationarity, including both the traditional LM tests and our proposed LR test. In Section 3, we present Monte Carlo analysis of the small-sample size and power performance of the various stationarity tests. In Section 4 we apply the tests to U.S. macroeconomic data. Section 5 concludes.

Section 2: Stationarity Tests

As discussed in KPSS, it is often possible to 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) and LMC (Leybourne and McCabe, 1994), which are both versions of a Lagrange Multiplier (LM) test, and our proposed likelihood ratio (LR) test.

2.1 The LM Statistic

Let $\hat{u}_t$, $t = 1, \ldots, T$, be the estimated residuals from a regression of the time series of interest, y, on an intercept and a time trend. Assuming that that the innovations to the random

[2] See Morley (2007) and Mitra and Sinclair (forthcoming) and references therein for examples of correlated unobserved components models applied to macroeconomic data with significant estimates of non-zero correlation between permanent and transitory movements.

walk component are normally distributed and that the stationary errors are iid N(0, σ_u^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 depends on the partial sum process, S_t, of these residuals, and the estimate of the error variance from the regression, $\hat{\sigma}_u^2$:

$$LM = \sum_{t=1}^{T} S_t^2 / \hat{\sigma}_u^2 \tag{1}$$

The nonstandard asymptotic distribution of the LM statistic can be derived based 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 serial correlation in the error, 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 $\hat{\sigma}_u^2$ with a nonparametric estimator of the "long-run variance" (i.e., 2π times the spectral density of u at frequency zero), which can be denoted as $s^2(l)$:

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

where $s^2(l) = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 + 2T^{-1}\sum_{s=1}^{l} w(s,l)\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}$ and $w(s,l)$ is a weighting function, typically 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 autocorrelation in the errors. In our analysis, 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: $n = \text{int}\left[4 * \left(T/100\right)^{(2/9)}\right]$, where *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, although Müller (2005) considers local-to-unity asymptotics to show that the KPSS tests performs poorly in the presence of high autocorrelation, which is the empirically-relevant context for most macroeconomic data. Caner and Kilian (2001) employ Monte Carlo analysis to show that a parametric bootstrap of the KPSS test reduces small-sample 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 in our Monte Carlo analysis, they also show 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 hereafter) employ a parametric version of the LM test of the null hypothesis of stationarity against the presence of a stochastic trend. 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:

$$y_t^* \equiv y_t - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i}, \tag{3}$$

where the $\hat{\phi}_i$ are the maximum likelihood estimates of ϕ_i from the ARIMA(p, 1, 1) model:

$$\Delta y_t = \delta + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t + \theta u_{t-1} . \tag{4}$$

The ARIMA(p, 1, 1) is the reduced-form representation of the unobserved components model LMC assume under the alternative, which is the local-level model of Harvey (1989). This approach gives consistent estimates of the AR(p) parameters both when the null and the alternative are true.[3] By contrast, if we were to estimate an AR(p) in levels, the estimates would be inconsistent when the alternative is true. In particular, the estimates would capture an autoregressive unit root, rather than converge to their true values, and the test would have little power, as discussed in LMC.

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

$$\text{LMC} = \hat{u}'\mathbf{V}\hat{u} , \tag{5}$$

where $\boldsymbol{V}$ is a T x T matrix with ijth 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, as with the KPSS test, Caner and Kilian (2001) find that a parametric bootstrap version of the LMC test reduces small-sample size distortions compared to the asymptotic test, but the bootstrap version of the test has low power.

[3] 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. Therefore, if we estimate the first difference of a stationary model (i.e. estimating under 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, 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 when the null is true.

2.2 The LR Statistic

Likelihood ratio statistics have been widely used to test for parameter constancy. A stationarity test is an example of a test for parameter constancy with the specific alternative hypothesis of a stochastic trend process. In particular, the alternative can be thought of as a time-varying parameter model in which the long-run 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, has a reduced-form ARIMA(p, 1, p) representation, which follows from the assumption that the long-run 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 in the errors. 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 Monte Carlo analysis, we modify the reduced-form model for the LMC test in order to allow it to capture the full parametric structure under the alternative, while still being consistent under the null. Thus, the differences in performance between the LMC test and the LR test are not due to parametric misspecification, but purely reflect the relative merits of the tests themselves for given data generating processes.

As discussed in the previous literature, the distribution of the LR statistic is nonstandard for tests of parameter constancy 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. However, despite its nonstandard distribution, the LR test has been applied in the literature for tests that the root of an MA lag polynomial for an MA(1) model is close to or equal to 1. To derive the asymptotic approximation of the distribution of the likelihood ratio test statistic for the value of the MA root in this setting, Davis, Chen, and Dunsmuir (1996) make use of the asymptotic approximation of MLE based on local-to-unity analysis for an MA(1) model as follows:

$$\Delta y_t = u_t - \theta u_{t-1}, \tag{6}$$

where $u_t \sim iid(0, \sigma_u^2)$ and $E(u_t^4) < \infty$, with the likelihood ratio statistic given as

$$2\left[l(\theta) - l(\theta = 1)\right] \xrightarrow{d} Z(\tilde{\beta}), \tag{7}$$

where $l(\cdot)$ denotes the log likelihood function, $\beta = T(1 - \theta)$, and

$$Z(\beta) = \sum_{k=1}^{\infty} \frac{\beta^2 \chi_k^2}{\pi^2 k^2 + \beta^2} + \sum_{k=1}^{\infty} \ln\left(\frac{\pi^2 k^2}{\pi^2 k^2 + \beta^2}\right), \tag{8}$$

with $\tilde{\beta}$ being the global maximizer of $Z(\beta)$, $\chi_k \sim$ iid N(0, 1), and $\xrightarrow{d}$ denoting weak convergence on the space of continuous functions [0, ∞).

In determining the asymptotic critical values for this test, we follow Davis and Dunsmuir (1996) and Gospodinov (2002) and consider the local maximizer of $Z(\beta)$, given by $\tilde{\beta}^l = \inf\{\beta \geq 0 : \beta Z'(\beta) = 0 \text{ and } \beta Z''(\beta) + Z'(\beta) < 0\}$. The infinite series is truncated at k = 1000 and $Z'(\beta)$ is computed for a given draw of the χ_k's. If $Z'(0) \leq 0$, set $\tilde{\beta}^l = 0$ for that draw. Otherwise, we find the smallest nonnegative root of $Z'(\beta)$ by grid search. We consider 100,000 replications to obtain the asymptotic distribution, which Davis, Chen, and Dunsmuir (1996) show generalizes to more complicated ARMA processes that correspond to the models considered in this paper, as discussed in further detail below. Table 1 reports the asymptotic critical values for the LR test.

Chen, Davis, and Dunsmuir (1995) establish that the LR test is consistent (i.e., has asymptotic power of 100%) against any fixed alternative. Furthermore, they also show that in small samples, the LBI LM-type tests have slightly higher power (in the third decimal place) only for alternatives that are very close to the stationary null. In terms of a bootstrap version of the LR test for an MA(1) model, Gospodinov (2002) establishes that a bootstrap LR test imposing the null hypothesis of a unit MA root is at least first-order accurate. However, given the nonstandard setting for the test, optimality of the LR test and higher-order accuracy of the bootstrap is more difficult to determine, as discussed by Gospodinov (2002).

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 and has previously been applied to many macroeconomic variables in the empirical literature (see, for example, Morley, 2007, and Mitra and Sinclair, forthcoming). Specifically, we assume that the observed series $\{y_t\}_{t=1}^{T}$ can be decomposed into a random walk with drift and a strictly-stationary AR(p) cycle:

$$y_t = \tau_t + c_t, \quad t = 1,\ldots, T. \tag{9}$$

$$\tau_t = \mu + \tau_{t-1} + \eta_t \,. \tag{10}$$

$$\phi(L) c_t = \varepsilon_t \,, \tag{11}$$

where the roots of $\phi(L)$ lie strictly outside the unit circle, corresponding to stationarity of the cycle component. Following Morley, Nelson, and Zivot (2003), we assume the innovations (η_t, and ε_t) are jointly normally distributed random variables with mean zero and variance-covariance matrix Σ:

$$\begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} \sim N(0,\Sigma), \quad \Sigma = \begin{bmatrix} \omega^2\sigma^2 & \rho\omega\sigma \\ \rho\omega\sigma & \sigma^2 \end{bmatrix},$$

where $\omega \geq 0$ and $\rho \in [-1,1]$.

For this model, trend stationarity is equivalent to the null hypothesis $H_0 : \omega = 0$ versus the composite alternative hypotheses $H_a : \omega > 0$ corresponding to the presence of a stochastic trend. As discussed in Morley, Nelson, and Zivot (2003), the correlated UC model is only identified for AR(p) specifications of the transitory component for which $p \geq 2$. However, assuming this constraint is satisfied, the correlated UC model can be cast into state-space form and the Kalman filter can be applied for maximum likelihood estimation of the parameters for both the restricted and unrestricted models to directly obtain the LR statistic:

$$LR = 2\left(l(\mu,\tilde{\phi},\sigma,\omega,\rho) - l(\mu,\tilde{\phi},\sigma,\omega=0)\right), \tag{12}$$

where $\tilde{\phi}$ denotes the px1 vector of AR parameters. Because $\omega = 0$ lies on the boundary of the parameter space and ρ is not identified under the null, the LR test statistic has a nonstandard distribution.

Proposition 2.1 *The LR statistic in (12) for a correlated UC model in (9)-(11) has the asymptotic distribution given in (7) under the null of stationarity* $H_0 : \omega = 0$ *and the test is consistent at least at rate* $\sqrt{T}$ *for alternatives with a stochastic trend* $H_1 : \omega > 0$.

Proof *See appendix.*

It should be emphasized that allowing for correlation between the permanent and transitory innovations is a crucial feature of our approach. If we had only considered alternatives for which the correlation, ρ, between the permanent and transitory movements was restricted to be 0, 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 Δy). To the extent that this restriction is false, a LR test based on an uncorrelated UC model will have lower power as a result of imposing the restriction. Based on our estimates, the restriction appears to be false for the macroeconomic variables that we consider.

Section 2.3 The Bootstrap Test Procedure

Asymptotic distributions often provide poor approximations to small-sample distributions of test statistics. Thus, the bootstrap can be used to approximate the small-sample distributions of the stationarity tests under consideration, as was done in Rothman (1997) and Caner and Kilian (2001) for the LM statistics.[4] Given that the asymptotic distribution in (7) is pivotal, the first-order accuracy of a bootstrap version of the LR test for UC model follows directly the results in Gospodinov (2002). Thus, consideration of a bootstrap LR test is asymptotically valid. Unfortunately, as discussed by Gospodinov (2002), higher-order accuracy is difficult to determine. However, our Monte Carlo analysis below suggests that the bootstrap test has better empirical size than the asymptotic test in practice.

For 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 of a trend stationary autoregressive process. We also calculate the likelihood value under the alternative, being careful to consider a large number of different starting values

[4] More recently, Cavaliere and Taylor (2005) consider bootstrap versions of the KPSS test that address time-varying second moments.

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 correlation ρ) 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 imposing the null based on the parameters estimated in step 1. Again, this is fully parametric. 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.[5]

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

[5] 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. This maintains the nominal size of a bootstrap test at 5%.

Monte Carlo data, divided by the total number of bootstrap draws (MacKinnon, 2002).

Section 3: Monte Carlo Experiments

Monte Carlo experiments provide the standard way to evaluate the small-sample properties of tests for given data generating processes (DGPs). While both KPSS and LMC use Monte Carlo analysis to evaluate the small-sample properties of their test statistics, their assumed DGPs are highly stylized and do not correspond very closely to estimated processes for most macroeconomic variables. Here, we consider Monte Carlo analysis of the small-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 four macroeconomic data series discussed below. We generate the same number of observations for each of the simulated Monte Carlo samples as we have for the actual corresponding data series. For each of the four trend stationary AR(2) DGPs (the first column of Tables 2a through 2d), we generate 1000 simulated samples and consider empirical rejection probabilities to compare the size of the asymptotic and bootstrap versions of the LR test, KPSS, and LMC. We then use the correlated UC estimates to simulate 1000 data samples under the alternative and consider empirical rejection probabilities to compare the power of the tests.

Section 3.1: The Data Generating Processes

For the empirically-relevant DGPs in our Monte Carlo analysis, we consider parameter values based on estimates for four important U.S. 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 empirical specification and potentially reduces the small-sample distortions of the KPSS test that would occur if we were to use higher frequency data (see Müller, 2005). We consider observations from 1947-2006 for U.S. real GDP and the CPI, from 1948-2006 for the U.S. unemployment rate, and from 1939-2006 for the U.S. total nonfarm payroll employment.[6] For the three monthly series (CPI, unemployment rate, and payroll employment), 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 estimates for 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 2a through 2d present the parameter estimates based on trend stationary and unobserved components models of U.S. 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 autoregressive 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 somewhat across the series.

Section 3.2: Results from the Monte Carlo Experiments

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 below in Section 4. We present the results for the Monte Carlo experiments in Tables 3a through 3d.

As shown in the 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

[6] All data were obtained from the FRED2 database of the Federal Reserve Bank of St. Louis.

stationary processes of the kind estimated for macroeconomic variables. For a nominal 5% test, the bootstrap LR rejection probabilities range from 4.4% to 6.2% under the null hypothesis whereas the bootstrap LM tests are always over-sized, with rejection probabilities ranging from 5.8% to 11.4%. Meanwhile, the asymptotic KPSS test is severely over-sized with rejection probabilities as high as 86.3% under the null. The rejection probabilities of the asymptotic LMC test range widely from 6.9% to 27.1% under the null, while the asymptotic LR test is also over-sized, with rejection probabilities ranging from 12.7% to 29.5% under the null. In all four cases the KPSS was the most severely over-sized. In three of the four cases the asymptotic LMC test had the lowest rejection probability amongst the asymptotic tests. In the fourth case, based on U.S. real GDP parameter estimates, the asymptotic LR test had the lowest rejection probability amongst the asymptotic tests. In terms of the different DGPs, it is perhaps not surprising that the various asymptotic and bootstrap tests have their largest size distortions for the trend stationary DGP based on the CPI data, which is the most persistent process, with AR coefficients summing to 0.998.

In addition to better size properties, we also find higher power using the bootstrap LR test for empirically-relevant nonstationary processes with stochastic trends. In particular, for the power experiments, where the data were simulated under the alternative, the rejection probabilities for the bootstrap LR test are larger than for the bootstrap version of KPSS test in all four cases and larger than for the bootstrap version of LMC test in three of four cases. In the case of the DGP corresponding to the U.S. unemployment rate estimates, the bootstrapped LMC test has slightly higher power (rejection probability of 52.8% for KPSS, as compared to 49.2% for the LR test). However, note that for the unemployment rate DGP under the null, 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 3.3: 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 stochastic trend 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.[7] One reason for this result is that the empirically-relevant and economically-interesting alternatives are in no sense "local" given that the variances of the permanent innovations are large.[8] As discussed in Nyblom (1989), an LR test of parameter constancy can have more power against distant alternatives than the LBI test. Indeed, as mentioned above, Chen, Davis, and Dunsmuir (1995) find that the small-sample power of their LR test for an MA(1) model is higher for all but the closest alternatives for the MA(1) parameter.

It is interesting to note, however, that we find no direct relationship between the small-sample 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.[9] Instead, what appears to matter more is the correlation between the permanent and transitory innovations. We find that the more negative

---

[7] 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.

[8] See Rothman (1997) and Rudebusch (1993) who also discuss the issue of alternatives that are not "local in economic terms" to the null.

[9] Similarly, there is no obvious link between the divergence rates for the LR test discussed in the appendix and the small-sample power results from the Monte Carlo experiment. Specifically, for the DGP based on real GDP data, we have complex roots (off the unit circle) for the implied MA polynomial and imperfect correlation between trend and cycle innovations, corresponding to divergence of the LR test at rate $T^{0.5}$. By contrast, we have real roots and imperfect correlation for the DGP based on the unemployment rate data and perfect correlation for the DGPs based on the CPI and payroll employment data, corresponding to divergence at rate T. Yet, the LR test has somewhat higher small-sample power for the DGP based on real GDP than for the DGPs based on the unemployment rate and the CPI, although the highest small-sample power is for the DGP based on the payroll employment data.

the correlation, the higher the power of the LR test. This is actually 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 nonstationary process with a volatile stochastic trend. Given that a correlation of negative one turns up for two of the four series, it is certainly an empirically-relevant 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 4: Application to Macroeconomic Data

Having considered Monte Carlo analysis to evaluate the small-sample performance of the various stationarity tests for DGPs related to U.S. real GDP, the unemployment rate, the CPI, and payroll employment, we now turn to applying the bootstrap versions of the tests to the actual data. Table 4 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 U.S. real GDP and the unemployment rate, all of the bootstrap tests agree on the rejection of trend stationarity.[10] However, for CPI and payroll employment we find conflicting results. For these

[10] 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 were true. Given that we reject trend stationarity of the unemployment rate for all three tests, this loss of power is not a

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 follow a nonstationary process with a volatile stochastic trend, as one would do using the bootstrap LR test.

Section 5: Conclusions

In this paper, we have verified the validity of asymptotic and bootstrap versions of a likelihood ratio (LR) test for stationarity based on a correlated unobserved components model that has previously been applied to many macroeconomic variables in the empirical literature. We then compared the LR test to asymptotic and bootstrap versions of widely-used Lagrange multiplier (LM) tests of stationarity. For the relatively small sample sizes that are available for macroeconomic time series, our Monte Carlo analysis reveals that the various asymptotic tests of stationarity have huge size distortions given estimated trend-stationary processes. Meanwhile, correcting for these size distortions using bootstrap versions of LM tests results in low power against estimated nonstationary processes with stochastic trends. By contrast, we found that a bootstrap version of our proposed LR test has more accurate small-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"

---

particular concern, especially for the bootstrap LR test that accurately controls size for the null DGP based on the unemployment rate.

(2001, page 641). Thus, having a more powerful test of stationarity in small sample sizes is extremely useful, especially given that the bootstrap LR test leads us to reject trend stationarity in favour of correlated unobserved components processes with large permanent movements for all four of the U.S. macroeconomic time series under consideration. Evidently, the prevalence of stochastic trends in macroeconomic data implied by standard autoregressive unit root tests is confirmed with a powerful version of a stationarity test.

Appendix: Proof of Proposition 2.1

It is straightforward to show that UC model in (9)-(11) is strictly equivalent in moments to a reduced-form ARIMA(p ,1, q) model:

$$\phi(L)(\Delta y_t - \mu) = \phi(L)\eta_t + (1-L)\varepsilon_t = \theta(L)u_t , \tag{A.1}$$

where $u_t \sim N(0,\sigma_u^2)$ and the parameters for the MA polynomial $\theta(L)$ depend on the vector of AR parameters $\tilde{\phi}$, ω, and ρ, with the order of the MA polynomial $q \leq p$. Strict equivalence of the models follows from the normality assumption for the innovations η_t and ε_t in the UC model. However, it should be noted that the results below rely only on second-order equivalence of the models, which would follow from the assumption that the innovations in the UC model and the forecast error u_t in the ARIMA model are iid with finite fourth moments. Also, even though we assume the $p \geq 2$ for identification of the correlated UC model, the results below would hold as long as the process is at least equivalent to a reduced-form ARIMA(0,1,1) process after any cancellation of roots and the specification of an ARIMA model used in estimation under the null and alternative is sufficiently rich enough to capture the true underlying process.[11]

[11] The specific result in terms of the rate of divergence of the test under the alternative hypothesis also requires that the model used in estimation allows for autoregressive dynamics, even if none are present in the true process.

Under the null hypothesis $H_0 : \omega = 0$, the implied MA lag order for the corresponding reduced-form ARIMA model is $q = 1$, with the coefficient in the implied MA polynomial $\theta(L) = 1 - \theta L$ restricted to $\theta = 1$. That is, the MA polynomial has a single root equal to 1.

Lemma 1: *Under the alternative hypothesis $H_1 : \omega > 0$, the roots of the MA lag polynomial for the reduced-form ARIMA model in (A.1) corresponding to the UC model in (9)-(11) are strictly different than 1 (although they may be on the unit circle).*

There are two cases to consider for the alternative hypothesis.

Case 1: If the correlation between UC innovations is less than perfect, $\rho \in (-1,1)$, the variance-covariance matrix for the UC model, Σ, is strictly positive definite and invertibility of the MA polynomial $\theta(L)$ follows directly from Theorem 1 in Teräsvirta (1977), which states that the sum of possibly correlated MA processes with positive definite variance-covariance matrix is invertible if and only if the MA polynomials have no common roots of modulus 1. Because the $\phi(L)\eta_t$ and $(1-L)\varepsilon_t$ processes in (A.1) have no common roots of modulus 1 for their lag polynomials due to the stationarity assumption for $\phi(L)$, the MA polynomial $\theta(L)$ is invertible, directly implying that none of its roots is equal to 1.

Case 2: If the correlation between UC innovations is perfect, $\rho = \pm 1$, it implies that $\eta_t = \pm\omega\varepsilon_t$. Thus, the MA polynomial is $\theta(L) = \pm\omega\phi(L) + (1-L)$. Note, then, that an MA root equal to 1 implies that the MA polynomial can be factorized as follows: $\theta(L) = (1-L)\theta^*(L)$, where $\theta^*(L)$ is based on the other roots. It is trivial to show from $\theta(L) = (1-L)\theta^*(L)$ that $\theta(1) = 0$. However, if $\theta(1) = 0$, then $\theta(L) = \pm\omega\phi(L) + (1-L)$ would imply that $\phi(1) = 0$, which contradicts our assumption $\phi(L)$ has roots that are strictly outside the unit circle. Thus, as in the previous case, none of the roots of $\theta(L)$ is equal to 1.

Based on Lemma 1, testing stationarity for the UC model is equivalent to testing whether the corresponding ARIMA(p ,1, q) model has a root equal to 1 for its MA polynomial. In terms of this test, it is again useful to factorize the MA polynomial:

$$\theta(L) = \theta_c(L)\theta^*(L) \tag{A.2}$$

where $\theta_c(L)$ is the factor of the MA polynomial of order one or two with the single root or complex conjugate roots for $\theta(L)$ that are closest to 1 and $\theta^*(L)$ is the residual factor that reflects all of the other roots that are further away from 1. Denoting the root or 2x1 vector of roots closest to 1 as z_c and $\tilde{z}_c$, respectively, with z_c also being the first element of $\tilde{z}_c$, and the vector of all the other roots as $\tilde{z}^*$, the hypotheses $H_0 : \omega = 0$ and $H_1 : \omega > 0$ for the UC model are equivalent to the respective hypotheses $H_0 : z_c = 1$ and $H_1 : z_c \neq 1$ for the ARIMA model.

To impose the null hypothesis for both the UC model and ARIMA model, we can estimate a trend-stationary AR(p) model in levels. Assuming the null hypothesis is true, it is straightforward to show that MLE for the drift, AR parameters, and variance will be consistent for this model. Meanwhile, if we allow for the alternative hypothesis in estimation, consistency of MLE for all of the ARMA model parameters, both under the null and alternative, follows from Pötscher (1991). Focusing on the roots of the MA polynomial and assuming the null hypothesis is true, but allowing for the alternative in estimation, it follows from McCabe and Leybourne (1998) that the implied MLE estimate for z_c will be T-consistent with the Davis and Dunsmuir

(1996) asymptotic distribution given in (7) and the estimates for the elements of $\tilde{z}^*$ will be $\sqrt{T}$-consistent and asymptotically normal.

Conditional on μ, $\tilde{\phi}$, and σ_u which, assuming the null hypothesis is true, will be consistent both when imposing the null and when allowing for the alternative in estimation, as discussed above, the likelihood ratio statistic for testing $H_0: z_c = 1$ vs. $H_1: z_c \neq 1$ for an ARMA model is

$$LR_{z_c=1} = 2\Big((l(z_c) - l(z_c = 1)) + (l(\tilde{z}^* | z_c) - l(\tilde{z}^* = 0 | z_c = 1))\Big) \tag{A.3}$$

Under the null hypothesis, the first term converges to the Davis and Dunsmuir distribution given in (7) as $T \to \infty$. The second term is continuous in the neighbourhood of zero and, from McCabe and Leybourne (1998), is of order $\sqrt{T}$, meaning that it converges in probability to 0 as $T \to \infty$. Thus, given the equivalence of the UC model and the ARIMA model, the LR statistic for testing $H_0: \omega = 0$ vs. $H_1: \omega > 0$ has the asymptotic distribution given in (7) when the null hypothesis is true.

When the alternative hypothesis is true, the estimates for $\tilde{\phi}$ are no longer consistent when imposing the null in estimation, as discussed in Leybourne and McCabe (1994). In this case, imposing the null is equivalent to estimation of a trend-stationary AR(*p*) model in levels when there is an autoregressive unit root. Thus, following the Phillips (1987), the implied MLE for $\phi(1)$ when imposing the null converges arbitrarily close to 0 at rate *T*, even though the true $\phi(1)$ is strictly not equal to 0. By contrast, from Pötscher (1991), the implied MLE for $\phi(1)$ when allowing for the alternative is consistent at rate $\sqrt{T}$. Thus, based on the differences in estimates for $\tilde{\phi}$ alone, the LR statistic for testing stationarity will diverge at rate $\sqrt{T}$.

For some alternative DGPs, the LR statistic will diverge at a faster rate than $\sqrt{T}$. There are four cases to consider.

Case 1: If the correlation between UC innovations is less than perfect, $\rho \in (-1,1)$ and the MA polynomial $\theta_c(L)$ is of order 1, the first term of the LR statistic in (A.3) diverges at rate *T*, following Davis, Chen, and Dunsmuir (1996). The second term diverges at rate $\sqrt{T}$ given the $\sqrt{T}$-consistency of the roots of $\theta^*(L)$, $\tilde{z}^*$, which follows from the invertibility of $\theta(L)$ due to Theorem 1 in Teräsvirta (1977) and the consistency results for ARMA models in Pötscher (1991). Thus, in this case, the overall LR statistic in (A.3) diverges at rate *T*.

Case 2: If the correlation between UC innovations is less than perfect, $\rho \in (-1,1)$ and the MA polynomial $\theta_c(L)$ is of order 2 (i.e., the roots closest to 1 are complex conjugates), the LR statistic in (A.3) is modified as follows:

$$LR_{z_c=1} = 2\Big((l(\tilde{z}_c) - l(\tilde{z}_c = (1,0)')) + (l(\tilde{z}^* \,|\, \tilde{z}_c) - l(\tilde{z}^* = 0 \,|\, \tilde{z}_c = (1,0)'))\Big). \tag{A.4}$$

Because the MLE for the MA parameters are $\sqrt{T}$-consistent when allowing for the alternative, again following from the invertibility of $\theta(L)$ due to Theorem 1 in Teräsvirta (1977) and the consistency results for ARMA models in Pötscher (1991), the LR statistic diverges at rate $\sqrt{T}$ in this case.

Case 3: If the correlation between UC innovations is perfect, $\rho = \pm 1$, and the MA polynomial $\theta_c(L)$ is of order 1, we have a similar result to Case 1. Denoting the vector of roots of $\theta(L)$ as $\tilde{z}$, we have two subcases to consider. First, if all of the roots $\tilde{z}$ are strictly off the unit circle, then we have the same result as in Case 1 that the LR statistic diverges at rate *T*. However, if some of the roots $\tilde{z}$ lie on the unit circle, the estimates are consistent following Pötscher

(1991), but at an unknown rate. If the second term in (A.3) diverges at a faster rate than *T*, then the LR statistic will diverge at a faster rate. Thus, in this case, the overall LR statistic diverges at least at rate *T*.

Case 4: If the correlation between UC innovations is perfect, $\rho = \pm 1$, and the MA polynomial $\theta_c(L)$ is of order 2, we have a similar result to Case 2. If all of the roots $\tilde{z}$ are strictly off the unit circle, then we have the same result as in Case 2 that the LR statistic in (A.4) diverges at rate $\sqrt{T}$. However, if some of the roots $\tilde{z}$ lie on the unit circle, the estimates are again consistent at an unknown rate. Thus, in this case, based on the differences in the estimates for $\tilde{\phi}$, the LR statistic diverges at least at rate $\sqrt{T}$.

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.

Cavaliere, G. and A. M. R. Taylor (2005). "Stationarity Tests Under Time-Varying Second Moments," *Econometric Theory*, 21(06): 1112-1129.

Chen, M., R. A. Davis, and W. T. M. Dunsmuir (1995). "Inference for MA(1) Processes with a Root on or Near the Unit Root Circle," *Probability and Mathematical Statistics*, 5: 227-242.

Davidson, R. and J. G. MacKinnon (2000). "Bootstrap Tests: How Many Bootstraps?" *Econometric Reviews* 19: 55-68.

Davis, R. A., Chen, M., and W. T. M. Dunsmuir (1996). "Inference for Seasonal Moving Average Models with a Unit Root," Athens Conference on Applied Probability and Time Series Analysis: Volume II Time Series Analysis in Memory of E. J. Hannan, Eds. P. M. Robinson and M. Rosenbaltt, Lecture Notes in Statistics No. 115, Springer Verlag, New York: 160-176.

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

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.

Gospodivov, N. (2002). "Bootstrap-Based Inference in Models with a Nearly Noninvertible Moving Average Component." *Journal of Business and Economic Statistics* 20: 254-268.

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.

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.

Mitra, S., and T. M. Sinclair (forthcoming). “Output Fluctuations in the G-7: An Unobserved Components Approach.” *Macroeconomic Dynamics*.

Morley, J. C. (2007). “The Slow Adjustment of Aggregate Consumption to Permanent Income.” *Journal of Money, Credit, and Banking* 39: 615-638.

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.

Müller, U. K. (2005). “Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series.” *Journal of Econometrics* 128: 195-213.

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.

Phillips, P. C. B. (1987). “Time Series Regression with a Unit Root,” *Econometrica* 55: 277-301.

Pötscher, B. M. (1991). “Noninvertibility and Psuedo-Maximum Likelihood Estimation of Misspecified ARMA Models,” *Econometric Theory* 7: 435-449.

Rothman, P. (1997). “More Uncertainty about the Unit Root in U.S. Real GNP.” *Journal of Macroeconomics* 19(4): 771-780.

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.

Teräsvirta, T. (1977). “The Invertibility of Sums of Discrete MA and ARMA Processes,” Scandinavian Journal of Statistics 4(4): 165-170.

Table 1: Asymptotic Critical Values for the LR Test of Stationarity

	10%	5%	1%
Critical Value	0.96	1.89	4.42

Table 2: Parameter Estimates
Used for the Monte Carlo Simulations and Empirical Results

Table 2a: Quarterly Real GDP 1947.1 - 2006.4

Description		**AR(2)**	**UC**
Log Likelihood	LLV	-319.275	-316.049
S.D. of Permanent Innovation	σ_η	Restricted to be 0	1.115 (0.127)
S.D. of Temporary Innovation	σ_ε	0.917 (0.042)	0.560 (0.127)
Correlation btwn. Innovations	$\sigma_{\eta\varepsilon}$	-	-0.944 (0.006)
Drift	μ	0.830 (0.022)	0.826 (0.031)
1st AR parameter	φ_1	1.317 (0.061)	1.363 (0.099)
2nd AR parameter	φ_2	-0.346 (0.061)	-0.779 (0.107)

Table 2b: Quarterly average unemployment rate 1948.1-2006.4

Description		**AR(2)**	**UC**
Log Likelihood	LLV	-50.401	-47.877
S.D. of Permanent Innovation	σ_η	Restricted to be 0	0.212 (0.109)
S.D. of Temporary Innovation	σ_ε	0.298 (0.014)	0.385 (0.053)
Correlation btwn. Innovations	$\sigma_{\eta\varepsilon}$	-	-0.763 (0.120)
Drift	μ	0.005 (0.005)	0.003 (0.014)
1st AR parameter	φ_1	1.585 (0.049)	1.481 (0.074)
2nd AR parameter	φ_2	-0.642 (0.050)	-0.584 (0.075)

Table 2c: Quarterly average of monthly CPI Index 1947.1 - 2006.4

Description		AR(2)	UC
Log Likelihood	LLV	-186.780	-180.649
S.D. of Permanent Innovation	σ_η	Restricted to be 0	2.836 (0.219)
S.D. of Temporary Innovation	σ_ε	0.528 (0.024)	2.577 (0.217)
Correlation btwn. Innovations	$\sigma_{\eta\varepsilon}$	-	-1.000 (-)
Drift	μ	0.924 (0.083)	0.933 (0.189)
1st AR parameter	φ_1	1.762 (0.042)	0.778 (0.021)
2nd AR parameter	φ_2	-0.763 (0.042)	0.100 (0.021)

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

Table 2d: Quarterly average payroll employment 1939.1-2006.4

Description		AR(2)	UC
Log Likelihood	LLV	-248.895	-226.851
S.D. of Permanent Innovation	σ_η	Restricted to be 0	1.066 (0.003)
S.D. of Temporary Innovation	σ_ε	0.600 (0.026)	1.373 (0.010)
Correlation btwn. Innovations	$\sigma_{\eta\varepsilon}$	-	-1.000 (-)
Drift	μ	0.536 (0.016)	0.602 (0.065)
1st AR parameter	φ_1	1.734 (0.040)	1.522 (0.028)
2nd AR parameter	φ_2	-0.760 (0.041)	-0.525 (0.031)

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

Table 3: Monte Carlo Results

Table 3a: Based on U.S. Real GDP Parameter Estimates

Nominal Size 5%	Asymptotic	Bootstrap
KPSS	62.6%	5.8%
LMC	25.2%	9.0%
LR	18.4%	4.8%
Power	**Asymptotic**	**Bootstrap**
KPSS	84.4%	19.8%
LMC	95.6%	51.4%
LR	93.7%	77.0%

Table 3b: Based on Unemployment Rate Parameter Estimates

Nominal Size 5%	Asymptotic	Bootstrap
KPSS	17.9%	6.4%
LMC	9.2%	7.3%
LR	0.3%	4.8%
Power	**Asymptotic**	**Bootstrap**
KPSS	56.9%	45.9%
LMC	60.5%	52.8%
LR	16.3%	49.2%

Table 3c: Based on CPI Parameter Estimates

Nominal Size 5%	**Asymptotic**	**Bootstrap**
KPSS	86.3%	11.3%
LMC	27.1%	11.4%
LR	3.6%	3.4%
Power	**Asymptotic**	**Bootstrap**
KPSS	90.5%	10.4%
LMC	93.8%	24.5%
LR	55.7%	63.2%

Table 3d: Based on Payroll Employment Parameter Estimates

Nominal Size 5%	**Asymptotic**	**Bootstrap**
KPSS	23.8%	6.2%
LMC	6.9%	7.1%
LR	0.6%	4.2%
Power	**Asymptotic**	**Bootstrap**
KPSS	85.9%	21.7%
LMC	59.0%	21.7%
LR	100%	100.0%

Table 4: Empirical Application Results

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

*Reject the null of trend stationarity at 5%.