Australian School of Business Research Paper No. 2012 ECON 41
First draft October 2012

Australian School of Business Research Paper No. 2012 ECON 41A
May 2013

UNSW Business School Research Paper No. 2012 ECON 41B
This version August 2014

Testing Stationarity with Unobserved Components Models

James Morley
Irina B. Panovska
Tara M. Sinclair

Testing Stationarity with Unobserved Components Models

James Morley
University of New South Wales
Sydney NSW 2052
Australia

Irina B. Panovska
Lehigh University
Bethlehem, PA 18015
USA

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

August 18, 2014

JEL Classifications: C12, C15, C22, E23

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

Abstract

In the aftermath of the global financial crisis, competing measures of the trend in macroeconomic variables such as US real GDP have featured prominently in policy debates. A key question is whether the large shocks to macroeconomic variables will have permanent effects—i.e., in econometric terms, do the data contain stochastic trends? Unobserved components models provide a convenient way to estimate stochastic trends for time series data, with their existence typically motivated by stationarity tests that allow for at most a deterministic trend under the null hypothesis. However, given the small sample sizes available for most macroeconomic variables, standard Lagrange multiplier tests of stationarity will perform poorly when the data are highly persistent. To address this problem, we propose the use of a likelihood ratio test of stationarity based directly on the unobserved components models used in estimation of stochastic trends. We demonstrate that a bootstrap version of this test has far better small-sample properties for empirically-relevant data generating processes than bootstrap versions of the standard Lagrange multiplier tests. An application to US real GDP produces stronger support for the presence of large permanent shocks when using the likelihood ratio test as compared to the standard tests.

[†] 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 two insightful anonymous referees, 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 Greater New York Metropolitan Economics Colloquium, the 2011 SCE Meetings, the 2011 SNDE meetings, the GWU Institute for Integrating Statistics in Decision Sciences Seminar, the 2009 Joint Statistical Meetings, the 2013 Midwest Econometrics Group meetings, and the 2013 NBER-NSF Time Series Conference for helpful discussions and comments. All remaining errors are our own.

Introduction

In the aftermath of the recent global financial crisis, macroeconomists and policymakers are once again debating the relative importance of permanent versus transitory shocks in driving macroeconomic variables. For example, the slow recovery in US real GDP following the Great Recession of 2007-2009 could be due to a lower trend, persistent cyclical weakness, or some blend of the two. The importance of this issue has been highlighted in a recent speech by the Vice Chair of the Federal Reserve, Stanley Fisher, who argues that "[s]eparating out the cyclical from the structural, the temporary from the permanent, impacts of the Great Recession and its aftermath on the macroeconomy is necessary to assessing and calibrating appropriate policies going forward" (Fisher, 2014).

There are many different approaches to trend/cycle decomposition considered in practice (e.g., linear detrending, Hodrick-Prescott filtering, and bandpass filtering). However, assuming a well-specified model, an unobserved components (UC) approach provides a way to estimate stochastic trends in time series data so as to avoid the spurious cycle phenomenon that plagues many of the other methods (e.g., see Nelson and Kang, 1981, Cogley and Nason, 1995, and Murray, 2003). As a result, UC models have become quite popular, especially in macroeconomics.[1] Estimates from these models often imply a large role for permanent shocks in the overall variation of macroeconomic variables, especially when the UC models allow for correlation between permanent and transitory movements (see, for example, Morley, Nelson, and Zivot, 2003, Morley, 2007, Basistha, 2007, and Sinclair, 2009). However, a large point estimate for the variance of permanent shocks may occur even when the true data generating process is

[1] See, inter alia, Harvey (1985), Watson (1986), Clark (1987), Harvey and Jaeger (1993), Kuttner (1994), Proietti (2002), Morley, Nelson, and Zivot (2003), Basistha (2009), Sinclair (2009), Berger and Everaert (2010), Senyuz (2011), Mitra and Sinclair (2012), Bradley et al. (forthcoming), and Ma and Wohar (forthcoming).

stationary or trend stationary (as is argued by Perron and Wada, 2009, for the results in Morley, Nelson, and Zivot, 2003). Thus, it is helpful to motivate the application of a UC model for trend/cycle decomposition by first conducting a stationarity test that allows for at most a deterministic trend under the null hypothesis.

The standard approach to testing stationarity proposed by Kwiatkowski et al. (1992, KPSS hereafter) is to apply a Lagrange multiplier (LM) test for the presence of a random walk component in the residual from a regression of a time series on deterministic terms corresponding to either level or trend stationarity.[2] Calculation of the test statistic is straightforward, as it only requires estimation under the null, but its asymptotic distribution is nonstandard and depends on the deterministic terms allowed for in estimation. KPSS propose accounting for serial correlation in the residuals using the Newey and West (1987) nonparametric estimator of the long-run variance. However, the KPSS test performs poorly in small samples when the data are highly persistent (see Müller, 2005, for an explanation of the poor size and power performance of KPSS-type tests based on local-to-unity asymptotic analysis). Rothman (1997) and Caner and Kilian (2001) use Monte Carlo simulation evidence to show massive size distortions of the KPSS test in small samples given empirically-realistic persistent data generating processes such as might be thought to describe many macroeconomic variables. They find that a bootstrap version of the KPSS test does better in terms of size, but suffers from low power.

In this paper, we propose the alternative use of a likelihood ratio (LR) test of stationarity based on a UC model. Although maximum likelihood estimation of the UC model under the

[2] Rothenberg (2000) and Jansson (2004) propose more efficient stationarity tests under fairly general assumptions about the underlying data generating process. However, our focus is on the most commonly-used stationarity tests and in the specific setting of parametric unobserved components models that are widely used to estimate stochastic trends in macroeconomic data.

alternative hypothesis is somewhat more complicated than OLS estimation under the null hypothesis, this is hardly an impediment if the main purpose of conducting the stationarity test is to motivate estimation of a stochastic trend using the UC model in the first place. We establish the validity of our proposed approach by drawing from the theoretical results in Davis, Chen, and Dunsmuir (1995,1996) and Davis and Dunsmuir (1996) for a moving-average (MA) unit root test to verify the asymptotic distribution of the LR test of stationarity based on the UC model. As with the KPSS test, we find Monte Carlo simulation evidence that the LR test is somewhat oversized in small samples for empirically-relevant persistent data generating processes. However, a bootstrap version of the LR test does far better in terms of size and displays higher power than the KPSS test for empirically-relevant alternatives. Furthermore, we show that the improvement in performance of the bootstrap LR test over the bootstrap KPSS test is not just the result of assuming the correct parametric specification for the LR test. Specifically, we also compare the performance to Leybourne and McCabe's (1994, LMC hereafter) version of LM test, also assuming the correct parametric specification when applying this test. The LMC test performs somewhat better than the KPSS test, but its bootstrap version still underperforms the LR test both in terms of size and power.

We apply the various stationarity tests, including the proposed LR test, to postwar quarterly US real GDP assuming trend stationarity under the null hypothesis. Consistent with the power properties found in the Monte Carlo analysis, the bootstrap LM tests do not reject the null, but the bootstrap LR test does reject at the 5% level. We further investigate the sensitivity of our results to the sample period and to allowing for structural breaks. We find that the rejection of the null for postwar quarterly US real GDP is robust for the bootstrap version of the LR test. Thus, we conclude that there is strong evidence for the existence of a stochastic trend in US real

GDP and, according to our UC model estimates, the stochastic trend is responsible for a large portion of the overall fluctuations in real economic activity, including during the Great Recession.

The rest of this paper is organized as follows. In Section 2, we present the correlated UC model of trend/cycle processes and discuss pitfalls with using traditional stationarity tests for such processes given the small samples typically available for macroeconomic variables. In Section 3, we propose the LR test based on a correlated UC model, establish its asymptotic validity, and show that a bootstrap version of the LR test outperforms bootstrap versions of the standard tests in small samples. In Section 4, we apply the various stationarity tests to postwar quarterly US real GDP. Section 5 concludes.

Section 2: UC Models and Traditional Stationarity Tests

A correlated UC model of a trend/cycle process assumes that an observed time series $\{y_t\}_{t=1}^{T}$ can be decomposed into a random walk with drift and a stationary AR(p) cycle:

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

$$\tau_t = \mu + \tau_{t-1} + \eta_t \tag{2}$$

$$\phi(L)c_t = \varepsilon_t, \tag{3}$$

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), the innovations (η_t and ε_t) are assumed to be jointly normally distributed random variables with mean zero and variance-covariance matrix Σ:

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

where $\omega \geq 0$ and $\rho \in [-1,1]$. Restricting $\mu = 0$ and $\omega = 0$ corresponds to level stationarity and $\mu \neq 0$ and $\omega = 0$ corresponds to trend stationarity. A structural break in the trend function corresponds to a break in μ.

There is a vast literature on estimating stochastic trends in time series using UC models of trend/cycle processes. Early examples in a univariate setting include Harvey (1985), Watson (1986), and Clark (1987), all of which impose the correlation $\rho=0$ in estimation. When allowing for a non-zero correlation, Morley, Nelson, and Zivot (2003) find that the estimated variance of permanent shocks for postwar quarterly US real GDP given an AR(2) cycle is much larger than found when imposing a zero correlation, with the estimated correlation being about -0.9. The large estimate for the variance of the permanent shocks can be sensitive to allowing for a structural break in the deterministic trend function (see Perron and Wada, 2009).[3] Meanwhile, Wada (2012) shows that a large magnitude for the estimated correlation should be expected even when the true process is stationary.[4] Thus, the economic significance of large estimates of the variance of permanent shocks and the relevance of a highly negative correlation should be supported by first confirming the statistical significance of the stochastic trend via a stationarity test.

In practice, standard stationarity tests have been shown to behave poorly in small samples when time series data are highly persistent (e.g., Rothman, 1997, and Caner and Kilian, 2001). Müller (2005) provides a theoretical explanation for this poor performance based on local-to-unity asymptotic analysis. He notes that the LM test considered by KPSS concentrates

[3] The findings of a large variance for permanent shocks to real GDP is more robust to allowing for a structural break in the deterministic trend function when considering multivariate UC models (see, for example, Basistha, 2007, and Sinclair, 2009)

[4] Specifically, Wada (2012) shows that the correlation is often 1 or -1 for an estimated correlated UC model when the true process for the observed series is a stationary AR model.

its power on detecting a stochastic trend with a small shock variance compared to a Gaussian white noise error that dominates movements in an observed time series. However, in this case, both the null and the alternative imply a high degree of mean reversion, contrary to the apparently high persistence observed in most macroeconomic variables to which stationarity tests are applied. Meanwhile, depending on the nonparametric correction for serial correlation when estimating the long-run variance, KPSS-type tests fail to control size or are inconsistent when considering local-to-unity asymptotics.

We illustrate the small-sample problems for the KPSS test of trend stationarity given persistent time series processes using a Monte Carlo simulation based on estimated time series models for postwar quarterly US real GDP. For our size experiment, we consider a trend-stationary AR(2) model. For our power experiment, we consider a correlated UC model that allows for a stochastic trend, but nests the trend-stationary AR(2) model when the variance of the trend shocks is zero. The parameters for the data generating processes (DGPs) are reported in Table 1 and correspond to estimates based on US real GDP data for the sample period of 1947Q1-2011Q4.[5] Consistent with the findings in Morley, Nelson, and Zivot (2003) for the same model, but shorter sample period, the estimated variance of permanent shocks for the UC model is large and the estimated correlation between permanent and transitory movements is strongly negative.[6]

Table 2 reports the empirical size and power properties for the KPSS test for trend stationarity (and the LMC and LR tests, discussed in detail below). We consider 500 replications

[5] The data were obtained from the FRED database for the vintage of August 29, 2012. We do not include the last two quarters of the vintage of data because they are based on preliminary estimates that are often heavily revised.

[6] We do not report standard errors for the parameter estimates because Wald-type inferences can be highly misleading in finite samples for UC models given weak identification (see Ma and Nelson, 2012). Instead, in Section 4 below, we consider an LR test of stationarity to evaluate the statistical significance of the variance of permanent shocks.

for our baseline Monte Carlo experiments reported in Table 2. Based on the asymptotic critical value of 0.146 reported in KPSS, the test is severely oversized at a nominal 5% level given a sample size of 260 observations.[7] Similar findings have been noted by Rothman (1997) and Caner and Kilian (2001) in other related contexts of empirically-motivated persistent time series processes. Also established in those studies is an improvement in the size performance when considering bootstrap versions of the KPSS test. We find this improvement for a parametric bootstrap version of the test. However, as in the previous studies, we find that the power drops off dramatically when considering the bootstrap test. Note that, given the computational burden, we consider only up to 199 bootstrap simulations in each Monte Carlo replication.[8] Full details of the parametric bootstrap experiments can be found in the appendix.

One issue to note is that the KPSS test applies a nonparametric correction for serial correlation, even though we are assuming a parametric model for the data. Thus, we also consider the LMC test of trend stationarity that applies a parametric correction for serial correlation. For the LMC test, we estimate the AR parameters using the alternative UC model and apply the estimates to construct residuals that can be used to conduct the same LM test as in KPSS. We note that the UC model estimates of the AR parameters will be consistent under both the null and the alternative. Full details of both LM tests can be found in the appendix. As with the KPSS test, the results in Table 2 make it clear that the LMC test is oversized when based on the asymptotic critical value at a nominal 5% level and given a sample size of 260 observations, although the size distortion is not as severe as for the KPSS test. Also, the parametric bootstrap

[7] Consistent with their asymptotic distributions, we have confirmed that all of the tests considered in this paper have much more accurate size given a sample size of 5000 observations.

[8] To reduce the computational burden we follow the procedure proposed in Davidson and MacKinnon (2000) that stops a given bootstrap experiment at fewer than 199 simulations if the estimated bootstrap *p*-value is significantly smaller or larger than the size at a 5% level. This procedure maintains the nominal size of the bootstrap test at 5%.

version of the LMC test has better size and weaker power, but is not as weak as with the KPSS test. Caner and Kilian (2001) find similar results for the LMC test in their Monte Carlo analysis.

Table 3 presents an additional set of Monte Carlo experiments where we vary the values of individual parameters for the DGP.[9] Table 3A presents the results of size and power experiments when we reduce the values of the AR parameters by 50%. This allows us to explore the role of persistence in the performance of the stationarity tests we are considering. Comparing the size distortions with the baseline case, we see, as expected, that they are generally smaller for the asymptotic tests when the persistence is lower.

Table 3B presents two power experiments where we vary the relative importance of the permanent innovations by changing the value of the ω parameter (holding all other values from the baseline DGP constant). In the first case we reduce the value of ω by 50%. In the second case we reduce it by 90%. The LMC and KPSS tests have improved power when ω is smaller, which is not surprising as they are locally best invariant tests and they maximize the power close to the null.[10]

Table 3C presents four power experiments where we vary the size of the correlation between the permanent and transitory innovations by changing the value of the ρ parameter (holding all other values from the baseline DGP constant). We have considered four cases of correlation in the data generating process: zero correlation ($\rho=0$), low negative correlation ($\rho=0.5\rho_{baseline}$), low positive correlation ($\rho= -0.5\rho_{baseline}$), and high positive correlation ($\rho= -$

[9] For these results, we consider 100 replications as compared to 500 replications for the baseline case. This keeps the computational burden for these additional experiments manageable.

[10] The most notable case is when ω is very small ($\omega= 0.1\omega_{baseline}$). KPSS and LMC should have high power in this case because ω is near the null, and we can see from the Table 3B that they do. The low power of the LR0 test in this case is due two problems. First, UC models where the true correlation is non-zero, but the correlation is restricted to be zero, lead to estimates of the variance of the permanent component that are biased downwards. This problem is discussed in Morley, Nelson, and Zivot (2003) and Oh, Zivot, and Creal (2008). Second, there is a pile-up problem when the true variance is small, but non-zero, whereby the maximum likelihood estimate has non-zero probability of being equal to zero.

$\rho_{baseline}$). When the correlation is large in absolute magnitude in the true DGP, the bootstrap LR test is more powerful than the LR0 test (and also the most powerful test overall). If the correlation in the DGP is zero, then the bootstrap LR0 test is more powerful, as we would expect, because it is always better to impose true values of parameters than to estimate them. But the loss of power from estimating the correlation is minimal and, of course, it would never be known in practice that the correlation is actually zero. For the intermediate correlation cases, both positive and negative, we find that the bootstrap LR test is still the most powerful test.

Section 3: A Likelihood Ratio Test of Stationarity

For the UC model in (1)-(3), stationarity corresponds to the null hypothesis that the variance of permanent shocks is zero, with level stationarity imposed when $\mu = 0$ and trend stationarity considered otherwise.[11] In terms of the model, the null hypothesis is $H_0: \omega = 0$ versus the composite alternative hypotheses of positive variance, $H_a: \omega > 0$, 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:

$$LR = 2(l(\mu_a, \tilde{\phi}_a, \sigma_a, \omega, \rho) - l(\mu_0, \tilde{\phi}_0, \sigma_0, \omega = 0), \quad (4)$$

[11] The distribution of the LR statistic does not depend on whether or not a constant is allowed. We focus here on a trend stationarity test, but we could alternatively think about a test with the null being an autoregressive unit root process such as the well-known test by Dickey and Fuller (1979) and the more recent LR-based tests of Elliott, Rothenberg and Stock (1996) and Jansson and Orregaard Nielsen (2012). Our approach is similar in spirit to that of Jansson and Orregaard Nielsen (2012), but we focus on the null of trend stationarity rather than an autoregressive unit root because that is the appropriate test for the case where a researcher is considering applying a UC model to estimate a stochastic trend if the null is rejected.

where $\tilde{\phi}$ denotes the *p*x1 vector of AR parameters. Because ω=0 lies on the boundary of the parameter space, the LR test statistic has a nonstandard distribution.[12] The UC model that we consider here is second-order equivalent in moments to an ARIMA(p,1,p*) model under the alternative, and to an ARMA(p,1) model with moving average coefficient on the unit circle under the null. In particular, we can rewrite the model in differences:

$$\Delta y_t = \Delta\tau_t + \Delta c_t \tag{5}$$

$$\Delta y_t = \mu + \eta_t + c_t - c_{t-1} \tag{6}$$

$$\phi(L)(\Delta y_t - \mu) = \phi(\mathrm{L})\eta_t + \varepsilon_t - \varepsilon_{t-1} \tag{7}$$

If ω>0, the right-hand side of equation (7) is an MA process of order smaller than or equal to p. If ω=0, the right hand side of equation (7) is an MA process with a unit root. In Lemma 1 in the appendix, we show that for the empirically popular processes that we consider, the MA coefficient is equal to 1 if and only if ω=0.[13]

In determining the distribution of the LR test statistic, we rely on the theoretical results in Davis, Chen, and Dunsmuir (1995) and Davis and Dunsmuir (1996) for a moving-average unit root test.[14] Specifically, Davis, Chen, and Dunsmuir (1995) make use of the asymptotic

[12] Interestingly, despite its appearance in (4), the correlation parameter ρ does not act as a nuisance parameter for the LR test. This is because the UC model in (1)-(3) is equivalent to a reduced-form ARIMA model. Specifically, assuming a diffuse prior on the initial level of the trend, the likelihoods for the UC model and the reduced-form ARIMA model will be identical, as found in Morley, Nelson, and Zivot (2003). Thus, the likelihood can always be re-parameterized in terms of ARIMA parameters that are identified under the null. We make use of this equivalence in the proof for the distribution of the likelihood ratio statistic.

[13] Oh, Creal, and Zivot (2008) consider a more general UC model where the cycle is allowed to have an MA component. In the case of a more general model, the LR test will still have the asymptotic distribution discussed below, but the estimation will require a two-step approach proposed by Davis, Chen, and Dunsmuir (1996). The differences between the more general model and our model are discussed in the appendix.

[14] We use the approach proposed here because for the particular case when the null is that an MA process has a root that is on the unit circle and the other roots are not (i.e. specifically for the case of Morley, Nelson, and Zivot, 2003), the asymptotic distribution is fully derived by Davis and Dunsmuir. Chernoff (1954), Gouriéroux, Holly, and Monfort (1982) and Andrews (2001) have also studied likelihood ratio tests when parameters are on a boundary.

approximation of MLE based on local-to-unity analysis for an MA(1) model of the first differences as follows:

$$\Delta y_t = u_t - \theta u_{t-1},, \tag{8}$$

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

$$2(l(\theta) - l(\theta = 1) \xrightarrow{d} Z(\tilde{\beta}), \tag{9}$$

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

$$Z(\beta) = \sum_{k=1}^{\infty} \frac{\beta^2 \chi_k^2}{\pi^2 k^2 + \beta^2} + \sum_{k=1}^{\infty} \ln\left(\frac{\pi^2 k^2}{\pi^2 k^2 + \beta^2}\right), \tag{10}$$

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 on [0, ∞).

To obtain 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 $\tilde{\beta}^l = \inf\{\beta \geq 0: \beta Z'(\beta) = 0 \; and \; \beta Z''(\beta) + Z'(\beta) < 0\}$.[15] The infinite series is truncated at k = 1000 and $Z'(\beta)$ is computed for a given draw of the χ_k's. If $Z'(0) \leq 0$, we set $\tilde{\beta}^l$=0 for that draw. Otherwise, we find the smallest nonnegative root of $Z'(\beta)$ by grid search. The asymptotic critical value at 5% for the LR test of a moving-average unit root for an MA(1) model based on 100,000 replications is 1.89.[16]

Davis, Chen, and Dunsmuir (1996) show that the distribution in (6) holds for testing a moving-average unit root for more complicated ARMA models. Given the close relationship

[15] We consider the local maximizer because it is much less computationally involved than the global maximizer. However, as discussed in Davis, Chen, and Dunsmuir (1995), the asymptotic distributions for the LR statistic are very similar for the local and global maximizers.

[16] It is 0.96 for 10% and 4.42 for 1%.

between UC and ARMA models, we build on their result to establish the same asymptotic distribution for a stationarity test based on a UC model and the consistency of the test:

Proposition 1 *Assuming iid innovations with finite fourth moments, the LR statistic for a test of stationarity based on a correlated UC model has the asymptotic distribution given in (6) 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* $H_a: \omega > 0$.

Remark *The proposition follows directly from (i) the second-order equivalence of the UC model to a stationary ARMA model in first differences, (ii) Theorem 4.1 in Davis, Chen, and Dunsmuir (1996), (iii) Theorem 2.1 in Pötscher (1991), and (iv) the theoretical results for MLE of MA roots in McCabe and Leybourne (1998). See appendix for the full proof.*

Meanwhile, in terms of the bootstrap version of the LR test, first-order accuracy follows directly from the equivalence of stationarity to a unit MA root and the more general results in Gospodinov (2002) for a bootstrap LR test given a fixed null about the MA root.[17] Thus, consideration of a bootstrap LR test is also asymptotically valid and, in principle, no worse than considering an LR test based on the asymptotic critical value. Unfortunately, as discussed by Gospodinov (2002), higher-order accuracy is difficult to determine in this setting.

Returning to Tables 2 and 3, we find that the LR test is oversized in small samples when based on the asymptotic critical value at a nominal 5% level and given a sample size of 260 observations, with the size distortion similar to the LMC test, but not as severe as for the KPSS test. The parametric bootstrap version of the LR test is correctly sized, with the key result being that the drop off in power is not nearly as dramatic as for the LM tests.

[17] Admittedly, the model is MA(1) with a unit root only for the special case considered here. Morley (2011) shows, however, that more general unobserved components models with correlated components that nest the model in Morley, Nelson, and Zivot (2003), such as those discussed in Oh, Creal, and Zivot (2008) and Proietti (2008), have the same implications in terms of the volatility of the stochastic trend.

It should be emphasized that allowing for correlation between the permanent and transitory movements is important for the power of our LR test. In Tables 2 and 3 we also report the LR0 test where the correlation, ρ, was restricted to be 0 in estimation. This places a strong restriction on the estimated variability of the permanent component (specifically, that it can be no greater than the variability of Δy_t). To the extent that this restriction is false, as it is for the DGP considered in our baseline power experiment, the LR0 test based on an uncorrelated UC model has, by construction, lower power as a result of imposing the restriction.[18]

Comparing across all of the experiments reported in Tables 2 and 3, we can see that our proposed bootstrap LR test performs well in all cases. Our test particularly outperforms the other tests in the empirically-relevant baseline case where the DGP was based on estimates for US real GDP, which are discussed in detail next.

Section 4: Application to US Real GDP

Having considered Monte Carlo analysis to evaluate the small-sample performance of the various stationarity tests for DGPs based on estimates for US real GDP, we now turn to applying the tests to the actual data. We first present unit root tests for the actual data in Table 4. We report both the traditional augmented Dickey Fuller test (Dickey and Fuller, 1979, ADF) and the more recent LR-based test of Elliott, Rothenberg and Stock (1996, ERS). Not surprisingly, we fail to reject the presence of a unit root. However, failing to reject a unit root does not confirm its existence. Therefore, we move on to focus on stationarity tests.

[18] Note that the size of the asymptotic LR test is very similar when imposing zero correlation in estimating the alternative model. Meanwhile, the LM tests only require estimation under the null. So, by construction, they are not affected by the consideration of a non-zero correlation between permanent and transitory movements under the alternative.

Table 5 reports the results of applying the bootstrap versions of the stationarity tests to the actual data, beginning with the same 1947Q1-2011Q4 sample period that provided estimates for the DGPs considered in our Monte Carlo analysis. We consider bootstrap tests based on 4999 simulations. For this sample period, both bootstrap KPSS and LMC tests fail to reject the null of a trend-stationary AR(2) process in favor of the correlated UC process.[19] Conversely, the more powerful bootstrap LR test rejects the null hypothesis at the 5% level.

The 1947-2011 period includes the Great Recession near the end of the sample. Because the Great Recession corresponded to a large decline in the level of real GDP and its long-term implications remain unresolved, the rejection of stationarity could be driven by the inclusion of this (possibly incomplete) episode in the sample. Therefore, we also consider a pre-crisis sample period of 1947Q1-2006Q4 that ends just before the Great Recession. For the pre-crisis sample, all three tests reject the trend-stationary null at the 5% level. Interestingly, the test statistics are all (at least slightly) lower for the pre-crisis sample. However, the bootstrap critical values are also lower in all three cases, related to the fact that the estimated trend-stationary AR(2) model for the pre-crisis data implies less persistence (the sum of the AR coefficients is 0.97 instead of 0.99). To the extent that the data display more mean reversion, we would expect the traditional stationarity tests to perform somewhat better, including in terms of power, than when the data are highly persistent (again, see Müller, 2005, on this point).

The third case that we consider addresses Perron and Wada's (2009) concern that the rejection of trend stationarity for postwar US real GDP may be due to the exclusion of known

[19] Based on the asymptotic critical values of the tests, both the KPSS and the LMC reject the null hypothesis. But, as found in the Monte Carlo analysis, the asymptotic versions of these tests are massively oversized in finite samples. Therefore, inference should be based on the bootstrap versions of these tests given their better size properties.

structural breaks.[20] A practical advantage of the bootstrap tests is that they can automatically accommodate for known structural breaks in the trend function or in the error variance (while the asymptotic critical values depend on such breaks).[21] For our application, we first adjust the data based on the 1947Q1-2011Q4 data in order to take into account two structural breaks in the growth rate: a break in the mean in 1973Q1 (Perron, 1989, and Perron and Wada, 2009) and a break in the variance in 1984Q1 (Kim and Nelson, 1999, and McConnell Perez-Quiros, 2000). After modelling and removing the known breaks from the data,[22] we conduct the same empirical analysis as previously and find that all three of the test statistics are smaller than before, consistent with Perron and Wada's (2009) supposition. The bootstrap LR test, however, still rejects the trend-stationary null, whereas the other two tests fail to reject the null. This result again illustrates the power benefits of the bootstrap LR test compared to the bootstrap versions of the LM tests of stationarity.

Section 5: Conclusions

Properly separating trend and cycle movements in macroeconomic variables is important for policy analysis, forecasting, and testing between competing theories. An important first step then in conducting empirical analysis of macroeconomic data is to test for the existence of stochastic trends with a stationarity test. We have investigated the small-sample properties of stationarity tests when the data are highly persistent and can be captured by an unobserved components (UC) model. Monte Carlo analysis confirms that standard asymptotic tests display

[20] Due to considerable complication of the asymptotic analysis, we leave consideration of an unknown number of structural breaks at unknown breakdates for future research.
[21] We are thankful to an anonymous referee for pointing this out to us.
[22] Specifically, we standardize the growth rates allowing for the breaks in mean and variance and then reconstruct the level of real GDP based on the standardized growth series. This approach is equivalent to modeling the known structural breaks in the UC model for both the estimates and the bootstrap.

severe small-sample size distortions in this setting, while bootstrap versions of these tests suffer from weak power. We propose the alternative use of a likelihood ratio test of stationarity based on the UC model and demonstrate the superior power properties of a bootstrap version of this test. An application to postwar US real GDP supports the existence of a stochastic trend that is responsible for a large portion of the overall fluctuations in real economic activity, even when excluding the recent Great Recession or allowing for structural breaks in the mean and variance of the growth rate.

References

Andrews, D. W. K. (2001). "Testing when a parameter is on the boundary of the maintained hypothesis." *Econometrica* 69: 683-734.

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.

Basistha, A. (2007). "Trend-Cycle Correlation, Drift Break and the Estimation of Trend and Cycle in Canadian GDP." *Canadian Journal of Economics* 40: 584-606.

Basistha, A. (2009). "Hours per Capita and Productivity: Evidence from Correlated Unobserved Components Model." *Journal of Applied Econometrics* 24: 187-206.

Berger, T. and G. Everaert (2010). "Labour Taxes and Unemployment Evidence from a Panel Unobserved Component Model." *Journal of Economic Dynamics and Control* 34: 354-364.

Bradley, M., D. Jansen, and T.M. Sinclair (forthcoming). "How Well Does "Core" Inflation Capture Permanent Price Changes?" *Macroeconomic Dynamics,* forthcoming.

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.

Chernoff, H. (1954). "On the distribution of the likelihood ratio." *The Annals of Mathematical Statistics* 25: 573-8.

Clark, P. K. (1987). "The Cyclical Component of U.S. Economic Activity." *Quarterly Journal of Economics* 102: 797-814.

Cogley, T. and J. M. Nason (1995). "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control* 19: 253-278.

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

Davis, R. A., M. Chen, 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.

Davis, R. A., Chen, M., and W. T. M. Dunsmuir (1996). "Inference for Seasonal Moving Average Models with a Unit Root," in P. M. Robinson and M. Rosenbaltt (Eds.), *Athens Conference on Applied Probability and Time Series Analysis: Volume II Time Series Analysis in*

Memory of E. J. Hannan, Lecture Notes in Statistics No. 115, Springer Verlag, New York: 160-176.

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

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.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). "Efficient tests for an autoregressive unit root." *Econometrica* 64: 813-836.

Fisher, S. (2014). The Great Recession: Moving Ahead. Speech at the "The Great Recession—Moving Ahead," a Conference Sponsored by the Swedish Ministry of Finance, Stockholm, Sweden. August 11, 2014.
http://www.federalreserve.gov/newsevents/speech/fischer20140811a.htm

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

Gouriéroux, C., A. Holly, and A. Monfort (1982). "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." *Econometrica* 50: 63-80.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge, U.K.: Cambridge University Press.

Harvey, A. C. and A. Jaeger. (1993). "Detrending, Stylized Faces and the Business Cycle." *Journal of Applied Econometrics* 8: 231-247.

Harvey, A. C. and T. M. Trimbur (2003). "General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series." *Review of Economics and Statistics* 85: 244-255.

Hobijn, B., P. H. Franses and M. Ooms (2004). "Generalizations of the KPSS-Test for Stationarity." *Statistica Neerlandica* 58: 483-502.

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

Jansson, M. and M. Orregaard Nielsen (2012). "Nearly efficient likelihood ratio tests of the unit root hypothesis." *Econometrica* 80: 2321-2332.

Kim, C.-J. and C. R. Nelson (1999). "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle." *Review of Economic and Statistics* 81: 608-616.

Kuttner, K. N. (1994). "Estimating Potential Output as a Latent Variable." *Journal of Business & Economic Statistics* 12: 361-368.

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: 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.

McConnell, M. M. and G. Perez-Quiros (2000). "Output Fluctuations in the United States: What Has Changed since the Early 1980s? *American Economic Review* 90:1464-1476.

Ma, J. and C. R. Nelson (2012). "The Superiority of the LM Test in a Class of Models where the Wald Test Performs Poorly; Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components," Working Paper.

Ma, J. and M. E. Wohar (forthcoming). "An Unobserved Components Model That Yields Business and Medium-Run Cycles." *Journal of Money, Credit, and Banking*, forthcoming.

Mitra, S., and T. M. Sinclair (2012). "Output Fluctuations in the G-7: An Unobserved Components Approach." *Macroeconomic Dynamics* 16: 396-422.

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. (2011). "The Two Interpretations of the Beveridge–Nelson Decomposition." *Macroeconomic Dynamics*, 15, pp 419-439.

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

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

Murray, C. J. (2003). "Cyclical Properties of Baxter-King Filtered Time Series." *Review of Economics and Statistics* 85: 472-476.

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 H. Kang (1981). "Spurious Periodicity in Inappropriately Detrended Time Series." *Econometrica* 49: 741-751.

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

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

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

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.

Oh, K. H., E. W. Zivot, and D. D. Creal (2008). "The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics." *Journal of Econometrics* 146: 207-219.

Perron, P. (1989). "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis." *Econometrica* 57: 1361-1401.

Perron, P. and T. Wada (2009). "Let's Take a Break: Trends and Cycles in US Real GDP." *Journal of Monetary Economics* 56: 749-765.

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.

Proietti, T. (2002). "Forecasting With Structural Time Series Models," in M. P. Clements and D. F. Hendry (Eds.), *A Companion to Economic Forecasting*, Oxford, Blackwell Publishers: 105-132.

Proietti, T. (2006). "Trend-cycle decompositions with correlated components." *Econometric Reviews* 25: 61-84.

Rothenberg, T. (2000). "Testing for Unit Roots in AR and MA Models," in P. Marriott and M. Salmon (Eds.), *Applications of Differential Geometry to Econometrics*, Cambridge University Press, Cambridge: 281-293.

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

Senyuz, Z. (2011). "Factor Analysis of Permanent and Transitory Dynamics of the US Economy and the Stock Market." *Journal of Applied Econometrics* 26: 975-998.

Sinclair, T. M. (2009). "The Relationships between Permanent and Transitory Movements in U.S. Output and the Unemployment Rate." *Journal of Money, Credit, and Banking* 41: 529-542.

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

Wada, T. (2012) "On the Correlations of Trend-Cycle Errors." *Economics Letters* 116: 396-400.

Watson, M. W. (1986). "Univariate Detrending Methods with Stochastic Trends." *Journal of Monetary Economics* 18: 1-27.

Table 1: Parameters for Monte Carlo Simulations

Description		AR(2)	UC
S.D. of Permanent Innovations	$\omega\sigma$	Restricted to be 0	1.23
S.D. of Temporary Innovations	σ	0.92	0.81
Correlation btwn. Innovations	ρ	---	-0.93
Drift	μ	0.80	0.78
1st AR parameter	φ_1	1.37	1.27
2nd AR parameter	φ_2	-0.38	-0.66

Note: Parameters are based on estimates from 100*ln of Quarterly Real GDP 1947Q1-2011Q4.

Table 2: Baseline Monte Carlo Results

Results Based on Simulated Data with Parameters from Table 1

Nominal Size 5%	**Asymptotic**	**Bootstrap**
KPSS	78.3%	7.1%
LMC	33.3%	6.1%
LR0	29.0%	5.2%
LR	25.9%	5.4%
Power	**Asymptotic**	**Bootstrap**
KPSS	91.2%	22.4%
LMC	82.2%	46.4%
LR0	52.0%	39.0%
LR	88.6%	76.0%

Note: Sample size is 260 observations and we consider 500 replications.

Table 3: Additional Monte Carlo Results

Table 3A: Results Based on Simulated Data with AR Parameters Reduced by 50%

Nominal Size 5%	**Asymptotic**	**Bootstrap**
KPSS	37%	7%
LMC	38%	3%
LR0	1%	2%
LR	30%	6%
Power	**Asymptotic**	**Bootstrap**
KPSS	100%	70%
LMC	83%	38%
LR0	10%	14%
LR	93%	64%

Note: Sample size is 260 observations and we consider 100 replications.

Table 3B: Additional Power Experiments
Changing the Relative Importance of the Permanent Innovations (ω)

	$\omega = 0.5\omega_{baseline}$		$\omega = 0.1\omega_{baseline}$	
Power	**Asymptotic**	**Bootstrap**	**Asymptotic**	**Bootstrap**
KPSS	44%	58%	84%	89%
LMC	59%	63%	100%	100%
LR0	75%	77%	47%	59%
LR	54%	59%	100%	100%

Note: Sample size is 260 observations and we consider 100 replications. All parameters are the same as the baseline reported in Table 1 except for restrictions noted above each column of results.

Table 3C: Additional Power Experiments
Changing the Correlation Between the Innovations (ρ)

	$\rho = 0$		$\rho = 0.5\rho_{baseline}$		$\rho = -0.5\rho_{baseline}$		$\rho = -\rho_{baseline}$	
Power	**Asymptotic**	**Bootstrap**	**Asymptotic**	**Bootstrap**	**Asymptotic**	**Bootstrap**	**Asymptotic**	**Bootstrap**
KPSS	68%	28%	82%	22%	99%	81%	98%	98%
LMC	77%	56%	94%	52%	81%	48%	82%	49%
LR0	89%	58%	68%	50%	91%	41%	94%	43%
LR	74%	56%	89%	79%	95%	94%	96%	98%

Note: Sample size is 260 observations and we consider 100 replications. All parameters are the same as the baseline reported in Table 1 except for restrictions noted above each column of results.

Table 4: Unit Root Tests for our Empirical Example

Data Series	ADF Statistic	ERS Statistic
Real GDP 1947Q1 – 2011Q4	-1.71 (0.75)	19.90 (>0.10)

Note: p-values reported in parentheses. For the ERS statistic, the bound on the p-value is based on the critical value for the test at a 10% level. Tests are conducted in EViews and lag selection is based on AIC.

Table 5: Empirical Results

Data Series	KPSS Statistic	LMC Statistic	LR0 Statistic	LR Statistic
Real GDP 1947Q1 – 2011Q4	0.36 (0.14)	3.33 (0.07)	1.29 (0.21)	**7.45 (0.02)**
Real GDP 1947Q1 – 2006Q4	**0.35 (0.04)**	**2.82 (0.02)**	**2.96 (<0.01)**	**5.54 (0.03)**
Real GDP 1947Q1 – 2011Q4 drift & var break	0.17 (0.28)	1.65 (0.09)	**1.98 (0.01)**	**3.49 (0.048)**

Note: Statistics in bold represent rejection of the null at the 5% level. Bootstrapped p-values reported in parentheses for all tests.

Appendix A: Bootstrap Procedure

Given our focus on testing stationarity with UC models, we consider parametric bootstrap tests. Specifically, simulated data are based on estimated parameters and distributional assumptions of the models. The full bootstrap testing procedure is given as follows:

1) Consistently estimate the parameters of the assumed autoregressive process under the null of trend stationarity and obtain the likelihood value. We also calculate the likelihood value under the alternative of the specified unobserved components process, 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). 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 model and parameters estimated in step 1.[23] Again, this is fully parametric. We consider 4999 bootstrap simulations in our applications, while we do up

[23] We also considered a modified bootstrap procedure proposed by an anonymous referee where we imposed the AR parameters estimated from the alterative and setting to zero the variance of the stochastic trend for constructing the bootstrap samples. As hypothesized by the referee, this modification worked best in the vicinity of the null. In that case, however, the other tests reported in Tables 2 and 3 also performed well.

to 199 bootstrap simulations in each Monte Carlo replication due to the computational burden.

3) For each bootstrap simulation, 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 simulation, 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).

Appendix B: The LM Statistics

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 walk component are normally distributed and that the stationary errors are iid $N(0, \sigma_u^2)$, the one-sided LM statistic is the locally best invariant (LBI) statistic for the hypothesis that the innovations to the random walk component have a zero variance (Nyblom and Mäkeläinen, 1983; Nyblom, 1986; Nabeya and Tanaka, 1988; Bailey and Taylor, 2002). The statistic depends on the partial sum process, S_t, of these residuals, and the estimate of the error variance from the regression, $\sigma^2{}_u$:

$$LM = \sum_{t=1}^{T} S_t^2 / \hat{\sigma}_u^2 \qquad \text{(B.1)}$$

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.

B.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 $s^2(l)$:

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

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:$n = int[4 * \left(\frac{T}{100}\right)^{\frac{2}{9}}]$, 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.

B.2 LMC Parametric Approach

LMC employ a parametric version of the LM test of the null hypothesis of stationarity against the presence of a stochastic trend. They 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, they construct the series:

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

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

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

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.[24] 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

$$LMC = \hat{u}' V \hat{u}, \tag{B.5}$$

[24] 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.

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 (B.5), which, like the KPSS test, depend on integrated Brownian bridges and are pivotal.

Appendix C: Proof of Proposition 1

Taking first differences of the UC model in (1)-(3), it is straightforward to show that is strictly equivalent in moments to a reduced-form ARIMA(p,1, q) model:

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

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 $\widetilde{\phi}$, ω, and ρ, with the order of the MA polynomial $q \leq p$ (see Morley, Nelson, and Zivot, 2003, on this equivalence). Strict equivalence of the models follows from the normality assumption for the innovations η_t and ε_t in the UC model, as outlined in equations (5)-(7). However, the results for the likelihood ratio test rely only on second-order equivalence of the models, which would follow from the more general 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 $p \geq 2$ for identification of the correlated UC model, the results for the likelihood ratio test will 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.[25] As discussed in the main text, the equivalence of the UC model to the ARIMA model also explains why the correlation ρ does not act as an unidentified nuisance parameter in terms of the distribution of the likelihood ratio statistic under the null hypothesis. Specifically, as we make use of below, the likelihood for the UC model can be re-parameterized in terms of ARIMA parameters that are identified under the null hypothesis.

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 $H_a: \omega > 0$, the roots of the MA lag polynomial for the reduced-form ARIMA model in (C.1) corresponding to the UC model in (1)-(3) 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 (C.1) have no common roots of modulus 1 for their lag polynomials due to

[25] 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.

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 that $\phi(L)$ 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:

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

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 the 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$\tilde{z}_c^*$, the hypotheses H_0: $\omega = 0$ and H_a: $\omega > 0$ for the UC model are equivalent to the respective hypotheses H_0: $z_c = 0$ and H_1: $z_c \neq 0$ 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 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 and related to the approach taken in Davis, Chen, and Dunsmuir (1996), the likelihood ratio statistic for testing H_0: $z_c = 0$ vs. H_1: $z_c \neq 0$ for an ARMA model is

$$LR_{z_c=1} = 2(\big(l(z_c) - l(z_c = 1)\big) + \big(l(\widetilde{z^*}|z_c) - l(\widetilde{z^*} = 0|z_c = 1)\big) \quad \text{(C.3)}$$

Under the null hypothesis, the first term converges to the Davis and Dunsmuir distribution given in (6) as $T \to \infty$. The second term is continuous in the neighborhood of zero and, from McCabe and Leybourne (1998), is of order $\sqrt{T}$-, meaning that it converges 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_a: $\omega > 0$ has the asymptotic distribution given in (6) 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 (C.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 (C.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 (C.3) is modified as follows:

$$LR_{z_c=1} = 2\left(\left(l(\tilde{z}_c) - l(\tilde{z}_c = (1,0)')\right) + \left(l(\widetilde{z^*}|z_c) - l(\widetilde{z^*} = 0|\tilde{z}_c = (1,0)')\right)\right) \qquad \text{(C.4)}$$

Because the MLE for the MA parameters are $\sqrt{T}$-consistent when allowing for the alternative, again following from the invertibility of $\theta(L)$ directly from 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 (C.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 (C.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 $\tilde{z}$.

All of the derivations above use the assumption that the cycle is an AR(p) process. If we consider the more general model introduced by Oh, Creal, and Zivot (2008), where $\phi(L)c_t = (1 + \theta_\varepsilon)\varepsilon_t$. The correlation between the trend and the cyclical shocks in this model is only identified if θ_ε is known (in our model, we considered the empirically popular restricted case with $\theta_\varepsilon = 0$). As shown by OCZ, the variance of the permanent shock does not depend on the correlation ρ, so any tests that are based on ω will not depend on the correlation. However, it is important to note that in this case the model does not reduce to an ARMA(2,1) with a unit root under the null, but to an ARMA(2,2)

with a single unit root. To see this, if we add an MA component to equation (3), equation (7) becomes

$$\phi(L)(\Delta y_t - \mu) = \phi(\mathrm{L})\eta_t + (1 + \theta_\varepsilon)(1 - L)\varepsilon_t.\text{C.5}$$

However, it is important to note that from Theorem 1 in Teräsvirta (1977), under the alternative, the MA component will have roots that are strictly different from 1 (but may be 1 in modulus). Under the null, the MA component will have one root equal to 1 and a root that is not on the unit circle. Under the assumption that $\theta_v < 1$, under the null, the MA component will have a root exactly equal to one if and only if $\omega = 0$. Imposing $\omega = 0$ imposes that exactly one root of the MA coefficient is equal to 1. This is exactly the case considered by Davis, Chen, and Dunsmuir (1996), who show that the likelihood ratio statistics for the null where we impose that exactly one root is equal to one versus the alternative where the roots are unrestricted still follows the DD(1995) distribution (and the estimates for the MA and AR coefficients will still be consistent). In this case, the LR tests discussed in equations (C.3 through C.4 can be directly replaced by the adapted version of the DD(1995) test. It is, however, important to note that if the true θ_v is not equal to zero, and it is close to -1, this may lead to size distortions in finite samples.