A Unit Root Test Using a Fourier Series to Approximate Smooth Breaks

WALTER ENDERS* and JUNSOO LEE**

* Department of Economics, Finance & Legal Studies, University of Alabama, Tuscaloosa, AL 35487-0224, USA (email: wenders@cba.ua.edu)
** Department of Economics, Finance & Legal Studies, University of Alabama, Tuscaloosa, AL 35487-0224, USA (email: jlee@cba.ua.edu).

Abstract

We develop a unit-root test based on a simple variant of Gallant's (1981) flexible Fourier form. The test relies on the fact that a series with several smooth structural breaks can often be approximated using the low frequency components of a Fourier expansion. Hence, it is possible to test for a unit root without having to model the precise form of the break. Our unit root test employing Fourier approximation has good size and power for the types of breaks often used in economic analysis. The appropriate use of the test is illustrated using several interest rate spreads.

JEL Classifications: C12, C22, E17

Keywords: Structural breaks, nonlinear models, Fourier approximation.

I. Introduction

As shown in Perron (1989), traditional unit root tests lose power if structural breaks present in the data-generating process are ignored. If the break date is known, these unit root tests can be modified by including dummy variables to capture changes in the level and trend. Typically, structural breaks in a series are assumed to occur instantaneously and manifest themselves contemporaneously. However, a number of authors have recognized that the effects of structural change on the level or slope of a series can be gradual. For example, Leybourne, Newbold and Vougas (1998) and Kapetanios, Shin and Snell (2003) develop unit-root tests such that the deterministic component of the series is a smooth transition process. To properly use this type of unit-root test, it must be assumed that there is a single gradual break with a known break date and functional form. However, the break dates and the number of breaks are likely to be unknown. The existing literature assumes, *a priori*, the presence of only one or two structural breaks in the series in question. Although it is possible to allow for more breaks, such tests are not powerful, as many parameters need to be estimated. As such, it does not seem fruitful to develop a new test for the purpose of capturing multiple structural breaks with unknown break dates.

The aim of this paper is to develop a unit-root test that can be used in the presence of a small number of smooth breaks in the deterministic components of a series. Specifically, we use a variant of Gallant's (1981) flexible Fourier form to control for the unknown nature of the break(s). We follow Becker, Enders and Lee (2006) and illustrate that the essential characteristics of a series containing a small number of structural breaks can often be captured using the low frequency components of a Fourier approximation. A key feature of the approximation is that we do not need to assume that the break dates, the precise number of

breaks, and/or the exact form of the breaks are known *a priori*. Moreover, the Fourier

approximation can reduce the need to estimate a large number of parameters and, hence, results

in a test with good size and power properties. The test is designed to work when breaks are

gradual and we show that it has good size and power properties in the presence of either LSTAR

(logistic smooth transition autoregressive) or ESTAR (exponential smooth transition

autoregressive) breaks. Nevertheless, we show that our test can perform reasonably well in the

presence of sharp breaks. The appropriate use of our test is illustrated using several interest rate

spreads.

II. Approximating a nonlinear trend with a Fourier series

A simple modification of the Dickey-Fuller (DF) type test is to allow the deterministic

term to be a time-dependent function denoted by $d(t)$:

$$y_t = d(t) + \rho\, y_{t-1} + \gamma \cdot t + \varepsilon_t \tag{1}$$

where ε_t is a stationary disturbance with variance σ_ε^2, and $d(t)$ is a deterministic function of t.

We note that the initial value is assumed to be a fixed value, and ε_t is weakly dependent. If the

functional form of $d(t)$ is known, it is possible to estimate (1) directly and to test the null

hypothesis of a unit root (i.e., $\rho = 1$). When the form of $d(t)$ is unknown, any test for $\rho = 1$ is

problematic if $d(t)$ is mis-specified. Our test is based on the fact that it is often possible to

approximate $d(t)$ using the Fourier expansion:[1]

$$d(t) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sin(2\pi kt / T) + \sum_{k=1}^{n} \beta_k \cos(2\pi kt / T); \ n \leq T / 2 \tag{2}$$

---

[1] As indicated in Becker, Enders and Hurn (2004), structural change can be captured by the relatively low
frequency components of a series since breaks shift the spectral density function towards zero. Becker *et
al*. also show that the higher frequency components of a series are most likely to be associated with
stochastic parameter variation.

where n represents the number of cumulative frequencies contained in the approximation, k represents a particular frequency, and T is the number of observations.[2]

In the absence of a nonlinear trend, all values of $\alpha_k = \beta_k = 0$ so that the standard Dickey-Fuller specification emerges as a special case. There are several reasons why it is inadvisable to use a large value for n. As we demonstrate below, the presence of many frequency components uses degrees of freedom and can lead to an over-fitting problem. As such, we keep the value of n is small so that equation (2) can be viewed as an application of Gallant's (1981) flexible Fourier form (FFF) to modeling $d(t)$. As evidenced by Gallant (1981), Davies (1987), Gallant and Souza (1991), and Bierens (1997), a Fourier approximation using a small number of frequency components can oftentimes capture the essential characteristics of an unknown functional form. Moreover, n should be small since it is important to allow the evolution of the nonlinear trend to be gradual. There is little point in claiming that a series reverts to an arbitrarily evolving mean. Finally, Becker, Enders and Hurn (2004) show that if the number of breaks is unknown, a test for structural change that uses a single trigonometric component can have better power than the well-cited Bai-Perron (1998) test.

One key issue for our test is whether a small number of frequency components can replicate the types of breaks typically seen in economic data. To keep the problem tractable, we begin by considering a Fourier approximation using a single frequency component, so that

$$d(t) \cong \alpha_0 + \alpha_k \sin(2\pi kt/T) + \beta_k \cos(2\pi kt/T), \tag{3}$$

---

[2] When the sample size gets very large, it will be natural to expect that the number of frequencies (n) will also increase accordingly. In the limit, we may let $n = n(T) \to \infty$ as $T \to \infty$. However, as n increases, the tests lose power. As such, in finite samples, it is sufficient to treat n as a finite value ($n \ll T$), and the test depends on n.

where k represents the single frequency selected for the approximation, and α_k and β_k measure the amplitude and displacement of the sinusoidal component of the deterministic term. Thus, even with a single frequency ($n = 1$), we can allow for multiple smooth breaks.

Although (3) is especially suitable to mimicking smooth breaks, the solid lines in Panels 1 to 9 of Figure 1 show nine series for $T = 500$ containing various types of sharp and smooth structural breaks. Panel 1 illustrates a temporary break, which often occurs empirically in macro variables. Panels 2 and 3 allow for breaks in the intercept and slope of a trending series. As indicated by Prodan (2008), the types of sharp breaks illustrated in Panels 1 through 3 are difficult to detect using standard break-testing methodologies. In these cases, using smooth breaks with our Fourier approximation can work better than using dummy variables. Series with smooth breaks are shown in Panels 4 through 9. Panel 4 of the figure shows the following LSTAR break with parameter values $d_1 = 3$, $\varphi = 0.05$, $T = 500$, and $\lambda = 0.5$:

$$y_t = d_1/[\ 1 + \exp(\varphi(t - \lambda T))\ ]. \tag{4}$$

Panel 5 of the figure shows the same LSTAR break setting $\lambda = 0.75$. Panel 6 shows the following ESTAR break for $d_1 = 3$, $\varphi = 0.0003$, $T = 500$, and $\lambda = 0.75$:

$$y_t = d_1[\ 1 - \exp(-\varphi(t - \lambda T)^2)\ ]. \tag{5}$$

Series with multiple smooth breaks are shown in Panels 7 through 9. Of course, the essential features of all nine series are invariant to inverting their magnitudes or to reordering the data from $t = 500$ to $t = 1$.

The dashed line (short dashes) in Panel 1, shows the time path of $y_t = \alpha(t) + \gamma t$ obtained by setting $k = 1$, $\alpha_0 = 1.26$, $\alpha_1 = 1.45$, $\beta_1 = 1.24$, and $\gamma = -0.003$. The values of α_0, α_1, β_1 and γ were selected by setting $k = 1$ and regressing y_t on a constant, a linear time trend t, $\sin(2\pi kt/T)$ and $\cos(2\pi kt/T)$. The sum of squared residuals ($SSR = 302.62$) was about one third of that

5

obtained using only an intercept and trend ($SSR = 896.40$). As illustrated by the long-dashed line in Panel 1, the fit can be improved by introducing a second frequency component in the regression. In terms of equation (2), when we set $n = 2$ (so that we use the frequency components $k = 1$ and $k = 2$) we obtained $SSR = 136.90$.

For our purposes, the precise parameter values for the other panels of the figure are not especially important. The key points illustrated by the nine panels are:

1. A Fourier approximation of structural change using the single frequency $n = 1$ can often serve as a reasonable approximation to breaks of an unknown form. The addition of the second frequency seems to be important if there are several breaks and/or if the breaks are sharp. Moreover, using the specific frequency $k = 1$ often leads to a good approximation to a model with structural change. Since structural breaks shift the spectral density function towards frequency zero, it seems reasonable to use the low frequency components to mimic structural change. In addition, the introduction of a time trend in the estimating equation eliminates the need for the starting and ending values of the approximation to be equal. As such, changes in the intercept and slope of a trending function can also be captured by the approximation.

2. The approximations shown in Panels 1, 3 and 6 are clearly worse than those shown in the other panels. In general, sharp breaks of short duration will not be well-approximated by a few low frequency components.

3. As mentioned above, Leybourne, Newbold and Vougas (1998) and Kapetanios, Shin and Snell (2003) developed tests for a unit root allowing for LSTAR and ESTAR breaks, respectively. However, it is unclear how an applied researcher will know whether to use a

linear test, an LSTAR test, or an ESTAR test. An important feature of the flexible Fourier

form is that it can behave like an LSTAR break or an ESTAR break.

Of course, it is possible to design series with more complicated break patterns than those

shown in Figure 1. However, our experimentation with a large number of break types only

served to reinforce the three points listed above. This is not to say that the use of one or two

frequency components is able to control for all types of nonlinearities in the deterministic

components of a time-series model. Nevertheless, as we show in our Monte Carlo experiments of

Section 4, the approximation does reasonably well for other types of breaks often used in

economic analysis. The issue at hand is not to precisely formulate a model for breaks. Instead,

the goal is to develop a unit-root test controlling for the possibility of smooth breaks of unknown

forms. As such, we now turn to the theory underlying our break test.

III. The asymptotic properties of the test statistics

Since a small number of frequency components can mimic a variety of breaks, we begin

our analysis with a data-generating process (DGP) containing only one frequency:

$$y_t = \alpha_0 + \gamma \cdot t + \alpha_k \sin(2\pi kt/T) + \beta_k \cos(2\pi kt/T) + e_t \; ; \; k \leq T/2 \tag{6}$$

$$e_t = \rho \, e_{t-1} + \varepsilon_t . \tag{7}$$

Note that $\rho = 1$ under the null hypothesis of a unit root, and $\rho < 1$ under the alternative

hypothesis. By adopting the above DGP, we are able to show that the asymptotic distribution for

the test of the null hypothesis $\rho = 1$ is invariant to the magnitudes of α_0, α_k, β_k and γ. Our testing

procedure is based on the Lagrange Multiplier (LM) methodology of Schmidt and Phillips

(1992) and Amsler and Lee (1995). Specifically, we employ the LM principle by imposing the

null restriction and estimate the following regression using first-differences:

$$\Delta y_t = \delta_0 + \delta_1 \Delta\sin(2\pi kt/T) + \delta_2 \Delta\cos(2\pi kt/T) + u_t . \tag{8}$$

We denote the estimated coefficients as $\tilde{\delta}_0$, $\tilde{\delta}_1$ and $\tilde{\delta}_2$.[3] Then, we construct a detrended series

using these coefficients as:

$$\tilde{S}_t = y_t - \tilde{\psi} - \tilde{\delta}_0 t - \tilde{\delta}_1 \sin(2\pi kt/T) - \tilde{\delta}_2 \cos(2\pi kt/T), \quad t=2, \dots ,T \tag{9}$$

where $\tilde{\psi} = y_1 - \tilde{\delta}_0 - \tilde{\delta}_1 \sin(2\pi k/T) - \tilde{\delta}_2 \cos(2\pi k/T)$, and y_1 is the first observation of y_t. Once we

subtract $\tilde{\psi}$ from y_t, it follows that $\tilde{S}_1 = 0$. The testing regression is based on the following

regression using the detrended series:

$$\Delta y_t = \phi \tilde{S}_{t-1} + d_0 + d_1 \Delta\sin(2\pi kt/T) + d_2 \Delta\cos(2\pi kt/T) + \varepsilon_t. \tag{10}$$

If y_t is not stationary, it must be the case that $\phi = 0$; hence, the LM test statistic is:

$$\tau_{LM} = t\text{-statistic for the null hypothesis } \phi = 0. \tag{11}$$

To allow for serially correlated as well as heterogeneously distributed innovations, we

assume that the innovations ε_t in the DGP (7) satisfy the following conditions [see Phillips and

Perron (1988, p. 336)]:

Assumption 1. (a) $E(\varepsilon_t) = 0$ for all t; (b) $\sup_t E|\varepsilon_t|^{\delta+\omega} < \infty$ for some $\delta > 2$ and $\omega > 0$; (c) $\sigma^2 =$

$\lim_{T\to\infty} T^{-1} E(S_T^2)$ exists and $\sigma^2 > 0$, where $S_T = \sum_{t=1}^{T} \varepsilon_t$; (d) $\{\varepsilon_t\}_1^{\infty}$ is strong mixing with mixing

numbers α_m that satisfy: $\sum_{1}^{\infty} \alpha_m^{1-2/\delta} < \infty$.

To correct for the effect of autocorrelated errors we augment (10) with lagged values of

$\Delta\tilde{S}_{t-j}, j=1, \dots , p$, so as to ensure that there is no remaining serial correlation in the residuals. To

find the asymptotic distribution of the test statistic, we need to establish:

---

[3] In the earlier version of this paper, we also considered a DF type test. The DF type test can be viewed as a test using the coefficients derived from the regression in levels of y_t on 1, t, $\sin(2\pi kt/T)$ and $\cos(2\pi kt/T)$. It was shown that the LM type tests have better size and power properties than the DF version of the test. We focus on the LM type tests.

Lemma 1: Suppose that y_t is generated by the DGP in (6) and (7) with $\rho = 1$, and one adopts the first step testing regression (8). Then,

$$\sqrt{T}\,(\tilde{\delta}_0 - \delta_0) \to \sigma\, W(1)$$

$$\frac{1}{\sqrt{T}}(\tilde{\delta}_1 - \delta_1) \to \sigma\,[(2\pi k)\textstyle\int_0^1 \cos^2(2\pi kr)dr]^{-1}[W(1) + (2\pi k)\textstyle\int_0^1 \sin(2\pi kr)W(r)dr]$$

$$\frac{1}{\sqrt{T}}(\tilde{\delta}_2 - \delta_2) \to \sigma\,[\textstyle\int_0^1 \sin^2(2\pi kr)dr]^{-1}[\textstyle\int_0^1 \cos(2\pi kr)W(r)dr] \qquad (12)$$

Proof. See the Appendix.

As shown in the Appendix, the asymptotic distribution of τ_{LM} follows from

Theorem 1: Suppose that y_t is generated by the DGP in (6) and (7) with $\rho = 1$, and one adopts the testing regressions (8), (9), and (10). Then, under the null hypothesis:

$$\tau_{LM} \to -0.5\,(\sigma_g/\sigma)\,[\textstyle\int_0^1 \underline{V}(r)^2 dr]^{-1/2} \qquad (13)$$

where $\underline{V}(r)$ is the projection of the process $V(r)$ on the orthogonal complement of the space spanned by the trigonometric function $dz = (1,\ d\sin(2\pi kr),\ d\cos(2\pi kr))'$; *and where* $V(r) =$

$$W(r) - rW(1) - [(2\pi k)\textstyle\int_0^1 \cos^2(2\pi kr)dr]^{-1}[W(1) + (2\pi k)\textstyle\int_0^1 \sin(2\pi kr)W(r)dr]\cdot \sin(2\pi kr) -$$

$[\textstyle\int_0^1 \sin^2(2\pi kr)dr]^{-1}[\textstyle\int_0^1 \cos(2\pi kr)W(r)dr]\cdot \cos(2\pi kr)$, and $W(r)$ is a Wiener process on $r \in [0, 1]$.

The above results show that the asymptotic distribution of τ_{LM} depends on the frequency k, but is invariant to all other parameters in the DGP in (6). Using any values of these parameters will not change the asymptotic distribution of the test. Thus, the tests are pivotal. To obtain critical values via simulations, we employ the DGP in (6) and (7) with $\rho = 1$, $\sigma_\varepsilon^2 = 1$, and use random initial values for ε_0. The critical values of τ_{LM} are reported in Panel a of Table 1 for the sample sizes $T = 100, 200, 500$, and 2500. The critical values were calculated using 100,000

replications for different frequency values of $k = 1$, 2 and 3. Our simulations show that the critical values converge quickly as the sample size increases and that it is reasonable to use simple interpolation to obtain critical values for other sample sizes.

Pre-testing for a nonlinear trend

If a nonlinear trend is absent from the DGP, a standard unit root test without a nonlinear trend will be more powerful than our test. As such, it is possible to increase the power of the unit-root test by pre-testing for a nonlinear trend using a methodology similar to that of Harvey, Leybourne and Xiao (2011). In order to develop a test with a null hypothesis of linearity against the alternative of a nonlinear trend with given frequency k, we can also use the following F-statistic:

$$F(k) = \frac{(SSR_0 - SSR_1(k))/2}{SSR_1(k)/(T-q)} \tag{14}$$

Here, $SSR_1(k)$ denotes the sum of squared residuals (SSR) from equation (10), q is the number of regressors, and SSR_0 denotes the SSR from the regression without the trigonometric terms. The distribution of the F-statistic is non-standard when the unit-root null is imposed on the DGP.[4] Specifically, the asymptotic distribution of $F(k)$ obtained under the unit-root null hypothesis, is given as follows:

Lemma 2: Suppose that y_t is generated by the DGP (6) and (7) with $\rho = 1$, and $\alpha_k = \beta_k = 0$ such that the null implies the absence of the nonlinear functions. Then,

$$F(k) \rightarrow \underset{k}{Max} \frac{1}{8}(\sigma_\varepsilon^2/\sigma^2)[(\int_0^1 \underline{V}(r)^2 dr)^{-1} - (\int_0^1 \underline{V}_0(r)^2 dr)^{-1}] \tag{15}$$

where $\underline{V}_0(r)$ the demeaned Brownian bridge, and $\underline{V}(r)$ is defined in Theorem 1.

---

[4] Harvey *et al.* (2010) considers a pre-test to determine the presence of nonlinear Fourier trends when the order of integration is unknown. Accordingly, their tests can be also used in our framework. Our pre-test is a conservative test as the null of a unit root is imposed on the GDP.

Proof. See the Appendix.

In Panel b of Table 1, we report the simulated critical values of $F(k)$ for the null hypothesis $\delta_1 = \delta_2 = 0$ against the alternative hypothesis of a nonlinear trend with frequency k. If the sample value of $F(k)$ is sufficiently large (so that the null hypothesis $\delta_1 = \delta_2 = 0$ is rejected), one can employ our τ_{LM} statistics using the nonlinear trend Fourier function. If the null is not rejected, it is possible to gain power by using the usual linear unit root tests.

A data-driven method of selecting a single k

A completely agnostic approach to the problem of detecting breaks is to select k using purely statistical means. We follow Davis (1987) by using a grid-search method such that the value $k = \hat{k}$ minimizes the sum of squared residuals from (10). Specifically, for each integer value of k in the interval $1 \leq k \leq kmax$, we estimate (10) and select k from the regression yielding the best fit. We suggest using the integer values 1 through 5 since low frequencies are associated with breaks. As established in Becker, Enders and Hurn (2004), the presence of high frequency components could be due to various forms of stochastic parameter instability. Let $\hat{k}$ denote the value of k that yields the smallest sum of squared residuals. We denote the fact that the critical values of τ_{LM} are a function of $\hat{k}$ by:

$$\tau_{LM} = \tau_{LM}(\hat{k}) \tag{16}$$

Practically speaking, the distribution of the statistic depends on how accurately the Fourier approximation in (6) mimics the actual DGP. As reported in an unpublished appendix available from the authors, Monte Carlo simulations indicate that the frequency is well-estimated. As such, we conjecture that the grid-search procedure yields a consistent estimate of k when the true DGP is given by (6). Hence, to account for the fact that k is an unknown nuisance parameter, we consider the following modification of the F-test given by (14):

11

$$F(\hat{k}) = \underset{k}{Max}\ F(k), \tag{17}$$

where $\hat{k} = \underset{k}{argmax}\ F(k)$. It is obvious that the value of k giving a minimum SSR value will

maximize the F-statistic in (14) such that $\hat{k} = \underset{k}{arginf}\ SSR_1(k)$. We report the critical values of

$F(\hat{k})$ in Panel c of Table 1 labeled 'Critical Values of $F(\hat{k}) = Max\ F(k)$.' To utilize our tests, we

first obtain $\hat{k}$ from (14) by minimizing the SSR and applying the F-test with $F(\hat{k})$ to examine

whether a nonlinear trend exists. If the null of absence of a nonlinear trend is rejected, then we

employ our suggested statistics with a nonlinear Fourier function. If the null is not rejected, we

utilize the usual unit root tests without a nonlinear trend. As we show in Section 4, this version of

our test has very good size and power properties.

The Use of Multiple Frequencies

As suggested by Figure 1, it is often possible to simply impose $n = k = 1$ so as to

approximate a wide variety of smooth breaks. In essence, the method filters out a low frequency

component (such as a break or other form of nonlinearity) that might interfere with the unit-root

test for $\rho = 1$. In this way, (8) can be estimated directly and the estimated coefficients can be

used to construct (9). The t-statistic for the null hypothesis $\phi = 0$ in (10) can be compared to the

critical values reported in the top row of Table 1.

Of course, it follows that if the use of $k = 1$ can reasonably represent an unknown

functional form, the use of $k = 1$ *and* $k = 2$ (with $n = 2$) will do even better. Hence, instead

conducting our LM test using a single frequency component, a researcher might want to include

multiple frequencies in the estimating equation. Moreover, since the individual frequency

components are orthogonal to each other, it can be easily shown that the appropriate critical

values will depend only on the particular frequencies included in the estimating equation. Since

12

the researcher has no *a priori* way to select the appropriate frequencies to use so as to mimic the

unknown functional form of $d(t)$, we follow the methodology of Gallant (1981) and Bierens

(1997) and use cumulative frequencies to estimate the unknown functional form. Hence, instead

of conducting our LM test using a single frequency component or an *ad hoc* selection of

frequencies, a researcher might want to include frequencies 1 through n in the estimating

equation. The asymptotic distribution of the test statistic can be modified accordingly to allow

for multiple frequencies. Table 2 reports the critical values of τ_{LM} for $n = 1, 2$ and 3; we denote

these critical values by $\tau_{LM}(n)$ in order to differentiate them from the case when a single

frequency component is used. Hence, for $T = 500$, at the 5% significance level, the critical value

of $\tau_{LM}(3)$ for the null hypothesis that $\rho = 0$ is -5.43 if frequencies 1, 2 and 3 are used in the

estimating equation.

Effect of Ignoring a Nonlinear Trend

One remaining question concerns the effect on the usual LM unit-root tests if a nonlinear

trend in the DGP is ignored. Perron (1989) earlier suggested that there will be a bias against

rejecting a false unit root if an existing structural break is ignored in the usual DF test. We

examine the asymptotic property of the LM tests under this situation with a nonlinear trend.

Lemma 3: Suppose that a nonlinear trend occurs in the data, and the DGP implies (6) and (7)

with $\rho < 1$, but the nonlinear trend is ignored and usual LM tests with a linear trend are

employed. Then, as we prove in the Appendix, the resulting OLS estimate follows:

$$\hat{\phi} \rightarrow \frac{\sigma_\varepsilon^2 (\rho - 1)}{H(k,r)} \tag{18}$$

$$H(k,r) = \sigma_\varepsilon^2 + (1/3)(\varepsilon_\infty^2 + \varepsilon_1 \varepsilon_\infty + \varepsilon_1^2) + \alpha_k^2 \int_0^1 \sin^2(2\pi kr)dr + \beta_k^2 \int_0^1 \cos^2(2\pi kr)dr$$

$$+ \alpha_k^2 - 2\sigma W(1)\left[\alpha_k \int_0^1 r \cdot \sin(2\pi kr)dr + \beta_k \int_0^1 r \cdot \cos(2\pi kr)dr\right];$$

13

and ρ is the true parameter value in the DGP (7).

Looking at the numerator, we note that $\hat{\phi} \to 0$ as $\rho \to 1$. The obvious conclusion is that

the unit-root null will not be rejected for ρ near unity. The power of the test will increase as ρ

$\to 0$, which is also obvious. Second, the denominator gets larger as the magnitude of the

coefficients α_k and β_k increase. Then, we will observe that $\hat{\phi} \to 0$ and this leads to non-rejections

of the null. This result implies that there will be loss of power under the alternative, if the

existing non-trend is ignored. Thus, the loss of power depends on the magnitude of the

coefficients α_k and β_k under the alternative. The loss of power is understood in line with Perron's

(1989) finding that unit-root tests will fail to reject a false unit root if an existing structural break

is ignored. His result generalizes to the failure to control for a nonlinear Fourier break.

IV. The Monte Carlo experiments

In this section, we evaluate the performance of our unit-root test using Monte Carlo

20,000 replications. To correct for the possibility of serial correlation, we augment equation (10)

with lagged values of $\Delta \widetilde{S}_{t-j}$, $j = 1,.., p$, using the general to specific methodology. Starting from a

maximum of $p = 8$, we estimate (10) augmented with a full p lags of $\Delta \widetilde{S}_{t-p}$. If a standard t-test

indicates that the p-th lag is insignificant at the 10% level, p is reduced by one and the procedure

is repeated. To conserve space, we report only results using the critical values for the 5%

significance level.

Size and Power with a single value of k

As a benchmark, we begin by generating the small-sample size and power of the test for

various values of k, α_k and β_k assuming that value of k in the estimating equation is the same as

that in the DGP.[5] Column 4 of Table 3 reports the size of our test for three sets of values for α_k

and β_k, $\rho = 1.0$, and values of $k = 1$, 2 and 3 for a sample size of 100. Consider the case $k = 1$, α_1

$= 0$, and $\beta_1 = 5$. At the 5% significance level, the test rejects the null hypothesis of a unit root

(i.e., $\rho = 1$) in 4.9 % of the 20,000 Monte Carlo replications. Also observe that these results are

insensitive to the actual value of k used in the DGP. We also performed the experiment for $T =$

200, and $T = 500$. The results reported in columns 6 and 8 of the table show that the empirical

size is never less than 4.8% and never more than 5.4%.

As in most unit-root tests, when the sample size is small, the power of the test is low.

Consider Table 3 for the case $k = 1$, $\alpha_1 = 0$, $\beta_1 = 5$, and $\rho = 0.9$. At the 5% significance level, the

test correctly rejects the null hypothesis in only 11.6% of the Monte Carlo replications. We note

that the power improves when a higher value of k is used; in fact, the power can exceed that of a

standard DF test. For instance, when $T = 100$, the power of τ_{LM} is 0.220 and 0.245 for $k = 2$ and

3, respectively. The more important point is that increasing the sample size to $T = 200$ or 500

greatly improves the power of the test, implying that the test is consistent; the power always

exceeds 99% when $\rho = 0.9$ and $T = 500$.

Size and power using an estimated value of k

In this section, we report results where k is estimated from the data. Clearly, the test will

have better performance if the actual value of k in the DGP is used. Our approach is a two-step

procedure in that we first determine if a nonlinear trend exists or not. Specifically, we estimate

(10) and test the null hypothesis $\delta_1 = \delta_2 = 0$ using the values of $F(\hat{k})$ reported in Panel c of Table

1. As such, we use a supremum test $F(\hat{k})$ where $\hat{k} = arginf\ SSR(k)$. If the null hypothesis of a

---

[5] Hence, the entries in the table imply that the researcher 'knows' the correct value of k to include in the
estimating equation. Details of the case where k is estimated along with the other parameters of the model

liner trend is not rejected, we apply the usual linear LM unit-root test. Instead, if the null

hypothesis is rejected, we use our τ_{LM} statistic reported in Table 1. There are several reasonable

ways to determine the most appropriate leg length. Here, the lag determination is done jointly

along the lines suggested by Ng and Perron (1995). Starting from a maximum of $p = 8$ lagged

terms, the procedure looks for significance of the last augmented term. We use the 10%

asymptotic normal value of 1.645 on the t-statistic of the last first-differenced lagged term.

Table 4 reports the size and power of our test for various values of α_k, β_k, and values of k

= 1 through 5 for a sample size of 200. Consider the case $k =1$, $\alpha_k = 0$, and $\beta_k = 5$. At the 5%

significance level, the test rejects the null hypothesis of a unit root (i.e., $\rho = 1$) in 5.6% of the

20,000 Monte Carlo replications. At the 10% significance level, we reject the null hypothesis in

9.9% of the replications. Since the asymptotic critical values are invariant to the magnitudes of

α_k and β_k, it is not surprising to find that that the empirical rejection rates are similar across all

non-zero values of these two parameters. Also observe that these results are insensitive to the

actual value of k used in the DGP. Notice that when $\alpha_k = \beta_k = 0$, the data-generating process

(DGP) is actually linear. Nevertheless, at the 5% and 10% significant levels, the null hypothesis

$\rho = 1$ is rejected in 7.7% and 13.7% of the replications, respectively. As expected, the power of

the test using the estimated value of k is a little lower than the test using the known value of k (as

given in Table 3), but the difference gets smaller when the magnitude of the coefficients gets

large.

Size and Power of the Test with Cumulative Frequencies

The Monte Carlo results reported in Table 5 indicate that our test is correctly sized when

cumulative frequencies are used. The DGP implies (6) and (7), and we used the values $\alpha_k = \beta_k =$

---

are omitted here but they are available from the authors.

0, for $k = 1$, 2, and 3, but the results would be invariant to using other values of α_k and β_k. The first entry in the table indicates that for $T = 100$, and $\rho = 1$, the null hypothesis of a unit root was rejected in 4.9% Monte Carlo replications when the single frequency $k = 1$ was included in the estimating equation. Similarly, for $n = 2$ (so that $k = 1$ *and* $k = 2$ were included in the estimating equation) the simulated size of the test is 0.049. As can be seen from examining the second, fourth, and sixth columns of the table, the test is correctly sized regardless of the value of T and the number of cumulative frequencies included in the estimating equation.

The problem is that the power of the test diminishes rapidly as additional frequency components are added to the estimating equation. In essence, an over-fitting phenomenon occurs when a large number of frequency components are included in the estimating equation. If the number of frequency components increases past $n = 3$, the power of the test deteriorates rapidly even for large T. As such, we recommend that a small value of n, such as $n = 1$ or $n = 2$, be used. We observe that the power is much better for $T = 200$ and for $T = 500$ when n is small. For instance, for $n = 2$ and $\rho = 0.9$, the null hypothesis of a unit root was correctly rejected in 22.1% and 96% of the trials when $T = 200$ and 500, respectively.

Effects of Ignoring Nonlinear Trends

Lemma 3 above indicates that ignoring a *nonlinear* trend affects the performance of a linear unit-root test under the null and alternative hypotheses. Table 6 indicates the magnitude of the size distortion and the loss of power when a standard LM unit-root test is used (so that the nonlinear trend is ignored). When $\rho = 1$, the linear LM test exhibits a serious size distortion regardless of the magnitudes of α_k and β_k, the sample size, and the frequency present in the DGP. Notice that the size of the linear test is usually zero, and is never more than 3.5%, even though we used the 5% critical values to construct the table. Also notice that the power of the linear LM

test is very low, and is usually near zero, whenever $\alpha_k \neq 0$, and/or $\beta_k \neq 0$. This is in contrast to the results shown in Table 3: when the nonlinear trend does not exist in the DGP (i.e., all $\alpha_k = \beta_k = 0$), using trigonometric components does not seriously distort the size of the test and the power can be comparable to the usual unit-root tests.

Effects of Improperly Modeling the Breaks

In a related experiment, we examined the consequences of employing dummy variables to model nonlinear breaks. Specifically, we use Perron's (1997) dummy-endogenous unit-root test in which a dummy variable is used to control for a break occurring at an unknown date. To perform the experiment, we assume that the DGP is given by (6) and (7) and generate 5,000 series with the single frequency component $k = 1$. Although not reported here, Perron's (1997) test results in relatively poor size and power properties.[6] For example, for $k = 1$, $T = 500$, $\alpha_k = 3$, and $\beta_k = 5$, the empirical size of the break test (using the 5% critical value) with dummy variables is 3.6% and the power is only 11.8% (the size-adjusted power is 16.5%). As previously reported in Tables 3 and 4, for these same parameter values, our test is properly sized and has power in excess of 99% regardless of whether $k = 1$ or $n = 2$ is used in the testing regression. The point is that improperly modeling the break can be as problematic as ignoring the break altogether.

Robustness of Fourier Tests in the Presence of Various Breaks

Although our test works well in the presence of trigonometric components in the DGP, it has reasonable size and power properties in the presence of other types of breaks. To determine the how our unit-root test works in the presence of different types of breaks, we examined the

---

[6] The complete results are available from the authors on request.

empirical performance of the size and power of the test in the presence of the types of breaks

shown in the nine panels of Figure 1:

$$y_t = I_t d_1 + (1 - I_t)d_2 + e_t \text{ where } I_t = 1 \text{ if } t \leq T/2 \text{ and } I_t = 0 \text{ otherwise.} \qquad (19)$$

$$y_t = I_t d_1 + (1 - I_t)d_2 + e_t \text{ where } I_t = 0 \text{ if } T/4 < t \leq 3T/4 \text{ and } I_t = 1 \text{ otherwise.} \qquad (20)$$

$$y_t = I_t(d_1 + t/T) + (1 - I_t)(d_2 + d_3 t/T) + e_t \text{ where } I_t = 1 \text{ if } t \leq T/2 \text{ and } I_t = 0 \text{ otherwise.} \qquad (21)$$

$$y_t = I_t(d_1 + t/T) + (1 - I_t)(d_2 + d_3 t/T) + e_t \text{ where } I_t = 0 \text{ if } T/4 < t \leq 3T/4 \text{ and } I_t = 1 \text{ otherwise.} \qquad (22)$$

$$y_t = d_1/[1 + \exp(\varphi(t - \lambda T))] + e_t. \qquad (4)'$$

$$y_t = d_1[1 - \exp(-\varphi(t - \lambda T)^2)] + e_t. \qquad (5)'$$

where $e_t = \rho e_{t-1} + v_t$. Columns 5 through 8 of Table 7 report the size and power of our unit root

test with Fourier functions when we used $k = 1$ in the estimating equation. For example, for

break type 1 (equation 19), $\rho = 1$, and $T = 200$, the entries in column 5 show that the null

hypothesis of a unit root was rejected in 5.3% of the simulated series for $d_1 = 3$, and in 5.2% of

the series for $d_1 = 6$. If you read down columns 5 and 7, it should be clear that the size of the test

is approximately 5% for all of the simulated breaks for $T = 200$ and $T = 500$, respectively.[7] This

is an encouraging result since our unit-root test is fairly robust to the DGP with structural

changes in level and trend shifts. We noted in the earlier experiment that the converse did not

hold; the dummy-endogenous tests have poor size and power in the presence of a Fourier

function in the DGP. In contrast, column 6 indicates that the power of our test for break type 1 is

0.354 for $T = 200$ and $\rho = 0.9$. For $T = 500$, Column 7 shows that the size remains good and the

power improves substantially. In fact, the power is always in excess of 96.8%. Columns 9

---

[7] To save space, we do not report values for $T = 100$. The power of our test and of the dummy-
endogenous test is relatively low in the case. For example, in equation (19) with $d_1 = 3$, the power of our

through 12 of Table 7 report the size and power when we used the cumulative method for $n = 2$. In all instances, both the size and power of the test using $k = 1$ is superior to that using the cumulative method.

The last four columns of the table report the size and power of the dummy-endogenous break tests using the estimated break dates. Of course, this test should have better performance than ours since it is designed for the type of sharp breaks used in the simulations. Nevertheless, it does not perform substantially better than the Fourier test. In fact, columns 13 and 15 of the table show that the size distortion of the test can be noticeable. This occurs since the dummy-endogenous tests are not invariant to the magnitude of breaks. With non-zero breaks under the null, they tend to exhibit over-rejections that increase as the magnitude of breaks increases. Once we corrected for the size distortion, the size-adjusted power of the dummy-endogenous test (shown in the second row of each cell in columns 13 and 15) is comparable to that of our test.

As reported in Table 8, little of substance is changed if we repeat the simulation exercise for smooth LSTAR and ESTAR breaks given by (4) and (5). As listed in the table, break types 1 and 2 are the LSTAR breaks shown in Panels 4 and 5 of Figure 1. Type 3 is for the offsetting LSTAR breaks shown in Panel 7 and Type 4 is for the ESTAR break shown in Panel 6. As should be clear from the table, the size of the test is quite good, indicating that the Fourier tests are robust to the DGP with smooth transitions. The power is low for $T = 100$ but is excellent for $T = 500$. The power of the test using $n = 1$ exceeds that for $n = 2$ even for the large sample size.

V. The term structure of interest rates

We now use the test to examine the term structure of interest rates. A number of papers, including Enders and Granger (1998) and Shin and Lee (2001), suggest that interest rate spreads

---

test was 0.114 and the power of the dummy-endogenous break test was 0.125. Further details are

20

should be stationary such that the adjustment towards the long-run equilibrium follows a threshold process. To explore this possibility we obtained monthly values of the 3-month T-bill rate and the 1-year (R1) and 3-year (R3) rates on U.S. government securities over the 1990:1 through 2003:11 period. The time paths of the R1 – R3 spread is shown in Figure 2; the time paths of the other spreads are similar. It seems that the spread is subject to several structural breaks. Thus, usual unit root tests using one or two breaks can pose a problem. The 3-year rate was substantially above R1 throughout the early 1990s. However, in the 1995 - 2000 period, the spreads between R3 and R1 declined substantially. In mid-2001, R3 did not decline as rapidly as the other two rates so that the spread became large.

It is plausible to argue that persistence of the magnitudes of the gaps is due to several structural smooth breaks in the equilibrium level of the spread. To employ our test, we estimated each spread as a nonlinear Fourier process in the form of (6) with $k = 1$. We used the estimated coefficients to construct $\tilde{S}_t$ and $\tilde{\psi}$ as in (9) and estimated an equation in the form of (10) with lagged values of $\Delta \tilde{S}_{t-i}$ to correct for serial correlation. The sample values of the τ_{LM}-statistics for the null hypothesis $\phi = 0$ are as follows:

R-long	R-short	τ_{LM}	k	p	DF	TAR
R1	T-bill	−5.50	1	12	−3.52	0.243
R3	T-bill	−5.76	1	11	−3.27	0.917
R3	R1	−5.05	1	11	−2.72	0.702

As reported in Table 1, the critical values of τ_{LM} for $T = 100$ are −4.69, −4.11 and −3.82 at the 1%, 5% and 10% significance levels, respectively. As such, if we wanted to use the 1% significance level, it would be possible to conclude that the spreads are all stationary. The situation can be quite different when we do not use a trigonometric component to control for

available from the authors on request.

nonlinearities. If we use the (R1 – T-Bill)$_t$ spread, the linear Dickey-Fuller test (see the next-to-last column of the table labeled 'DF') indicates that the *t*-statistic for the null hypothesis is –3.52. Using the standard Dickey-Fuller distribution, we can just reject the null hypothesis of a unit root at the 1% significance level (the critical value is –3.51). However, it is not possible to reject the unit-root hypothesis at the 1% level for the (R3 – *T*-bill)$_t$ spread or at the 5% level for the (R3 – R1)$_t$ spread.

For our purposes, the key issue is how the spreads adjust over time. In particular, we wanted to examine whether the interest rate spreads follow a threshold process. When we apply Hansen's (1997) test for threshold behavior, we are unable to reject the null hypothesis of no threshold at any conventional significance level. We used 1000 Monte Carlo replications to obtain the appropriate critical values for the test. The *prob*-values (listed under 'TAR' in the table above) for the three spreads are 0.243, 0.917 and 0.702, respectively. Hence, it is unlikely that our Fourier approximation detects any type of threshold behavior.

To obtain a better understanding of the nature of the Fourier adjustment process, the solid line in Figure 2 shows the time path of the difference between the 3-year rate and 1-year rate. One can clearly see that the spread was much higher in the early 1990s and in 2002 than in the intervening years. We estimated the following model (with *t*-statistics in parentheses):

$$\Delta y_t = -0.223\ y_{t\text{-}1} + 0.160 + 0.080\ \sin(2\pi kt/T) + 0.069\ \cos(2\pi kt/T) + \Sigma\ \beta_i \Delta \tilde{S}_{t-i} \qquad (23)$$
$$(-5.38)\qquad (5.29)\quad (3.97)\qquad\qquad\quad (4.16)$$

where $k = 1$, and y_t is the gap between 3-year rate and the 1-year rate.

The time-varying mean of the $\{y_t\}$ sequence can be obtained by dividing the Fourier intercept [i.e., $0.160 + 0.080\ \sin(2\pi kt/T) + 0.069\ \cos(2\pi kt/T)$] by one minus the sum of the autoregressive coefficients ($= 0.232$). You can clearly see by the smooth line in Figure 2, the

time-varying mean mimics the fact that the spread was much higher in the early 1990s and in the later part of the sample than in the intervening years. Given that the spreads are stationary, we also examined the breaks that are identified by the Bai and Perron (1998) procedure. Allowing for a maximum of five breaks with a minimum break-size of 12 months, the BIC selected three breaks. The first occurs at 1994:11, the second at 2000:12, and the third at 2002:6. The dashed line in Figure 2 shows the breakpoints and the four sub-period means of the (R3 – R1) spread. Notice that these sharp breaks are similar to the Fourier intercept. The difference, of course, is that the Fourier intercept is smooth and that the Bai-Perron procedure does not embody a unit-root test.

Although there are fewer than 200 observations, we wanted to determine how the test performs when $n = 2$ (so that trigonometric terms for $k = 1$ and $k = 2$ were included in the estimating equations). It turns out that nothing of consequence is altered when we use a second frequency component to capture the behavior of $\alpha(t)$. In fact, the time path of the intercept using $n = 2$ almost exactly coincides with the Fourier intercept shown in Figure 2. Consider:

R-long	R-short	τ_{LM}	n	p
R1	T-bill	−5.49	2	12
R3	T-bill	−6.27	2	11
R3	R1	−5.81	2	11

Table 2 indicates that the critical values of $\tau_{LM}(2)$ for $T = 100$ are −5.49, −4.92 and −4.62 at the 1%, 5% and 10% significance levels, respectively. As such, the null hypothesis of a unit root can be rejected at the 5% significance level for the (R1 – T-Bill)$_t$ spread and, since there are more than $T = 100$ sample observations, at the 1% level for the other two spreads.

VI. Summary and conclusion

The paper develops a unit-root test that can be used to control for several smooth breaks of an unknown functional form. The test is based on the fact that a small number of low frequency components of a Fourier approximation can often capture a process of gradual structural change. Nevertheless, the test can often capture sharp breaks and other types of nonlinear trends that we examined. Since the test can mimic a small number of smooth breaks using a parsimonious number of parameters, it can avoid the loss of power that can occur in unit root tests using a large number of estimated dummy variables. Our test reaffirms the well-known result that interest rate spreads are stationary even though there have been long-swings in the value of the spread.

TABLE 1
Critical Values of τ_{LM}

k	$T = 100$			$T = 200$			$T = 500$			$T = 2500$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel a: Critical Values of τ_{LM}												
1	-4.69	-4.10	-3.82	-4.61	-4.07	-3.79	-4.57	-4.05	-3.78	-4.56	-4.03	-3.77
2	-4.25	-3.57	-3.23	-4.18	-3.55	-3.23	-4.13	-3.54	-3.22	-4.15	-3.54	-3.22
3	-3.98	-3.31	-2.96	-3.94	-3.30	-2.98	-3.94	-3.31	-2.98	-3.94	-3.30	-2.98
4	-3.85	-3.18	-2.86	-3.80	-3.18	-2.88	-3.81	-3.19	-2.88	-3.80	-3.19	-2.88
5	-3.75	-3.11	-2.81	-3.73	-3.12	-2.83	-3.75	-3.14	-2.83	-3.74	-3.13	-2.83
Panel b: Critical Values of $F(k)$												
1	7.08	8.48	11.52	6.93	8.28	11.10	6.85	8.15	10.81	6.79	8.07	10.69
2	3.71	4.90	7.72	3.73	4.86	7.51	3.74	4.84	7.45	3.71	4.83	7.32
3	2.87	3.78	5.99	2.90	3.80	5.89	2.93	3.82	5.97	2.94	3.83	5.95
4	2.61	3.41	5.36	2.63	3.44	5.36	2.65	3.45	5.36	2.65	3.46	5.37
5	2.47	3.26	5.16	2.48	3.25	5.09	2.53	3.28	5.09	2.52	3.29	5.03
Panel c: Critical Values of $F(\hat{k}) = Max\ F(k)$												
	7.50	8.80	11.79	7.34	8.60	11.32	7.24	8.45	11.03	7.18	8.37	10.90

TABLE 2
Critical Values of $\tau_{LM}(n)$ Using Cumulated Frequencies

n	$T = 100$			$T = 200$			$T = 500$			$T = 2500$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	-4.69	-4.10	-3.82	-4.61	-4.07	-3.78	-4.57	-4.05	-3.78	-4.56	-4.03	-3.77
2	-5.49	-4.90	-4.61	-5.37	-4.84	-4.57	-5.30	-4.79	-4.52	-5.28	-4.77	-4.51
3	-6.18	-5.59	-5.29	-6.04	-5.48	-5.21	-5.95	-5.42	-5.16	-5.89	-5.38	-5.13

TABLE 3
Finite Sample Performance of τ_{LM}

DGP			T = 100		T = 200		T = 500	
k	α_k	β_k	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$
1	0	5	0.049	0.116	0.054	0.395	0.050	0.997
	3	0	0.049	0.119	0.053	0.396	0.049	0.996
	3	5	0.052	0.119	0.053	0.396	0.050	0.996
2	0	5	0.050	0.220	0.054	0.624	0.047	0.998
	3	0	0.050	0.216	0.054	0.624	0.052	0.997
	3	5	0.049	0.209	0.054	0.623	0.050	0.998
3	0	5	0.049	0.245	0.050	0.719	0.050	0.999
	3	0	0.049	0.247	0.049	0.720	0.051	0.999
	3	5	0.050	0.243	0.050	0.719	0.050	0.999

Note: Entries in the table were constructed using the 5% critical values from Table 1.

TABLE 4
Finite Sample Performance of τ_{LM} Using the $F(\hat{k})$-test

DGP			Size: $\rho = 1$		Power: $\rho = 0.9$	
k	α_k	β_k	5%	10%	5%	10%
1	0	0	0.077	0.137	0.756	0.886
	0	5	0.056	0.099	0.284	0.309
	3	0	0.067	0.118	0.277	0.396
	3	5	0.054	0.093	0.327	0.391
2	0	0	0.075	0.129	0.759	0.896
	0	5	0.040	0.076	0.543	0.657
	3	0	0.037	0.067	0.287	0.408
	3	5	0.049	0.090	0.593	0.755
3	0	0	0.077	0.134	0.754	0.887
	0	5	0.056	0.095	0.697	0.841
	3	0	0.029	0.058	0.428	0.567
	3	5	0.048	0.093	0.694	0.851

Note: Simulations are reported for $T = 200$.

TABLE 5
Size and Power of the $\tau_{LM}(n)$ test (Cumulative Frequencies)

n	$T = 100$		$T = 200$		$T = 500$	
	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$
1	0.049	0.117	0.049	0.386	0.050	0.997
2	0.049	0.081	0.049	0.221	0.051	0.957
3	0.048	0.068	0.051	0.153	0.051	0.845

Note: Entries in the table were constructed using the 5% critical values from Table 2.

TABLE 6
Effects of Ignoring Nonlinear Trends

DGP			$T = 100$		$T = 500$	
k	α_k	β_k	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$
1	3	3	0.010	0.003	0.035	0.548
	3	5	0.002	0.000	0.028	0.029
	5	3	0.003	0.000	0.027	0.066
2	3	3	0.000	0.000	0.018	0.541
	3	5	0.000	0.000	0.008	0.015
	5	3	0.000	0.000	0.007	0.022
3	3	3	0.000	0.000	0.009	0.549
	3	5	0.000	0.000	0.002	0.012
	5	3	0.000	0.000	0.002	0.016

Note: Entries in the table were constructed using the 5% critical values from a traditional linear LM unit root test.

TABLE 7
Effects of Level and Trend Shifts

Breaks	d_1	d_2	d_3	Fourier Approximation: $n=1$				Fourier Approximation: $n=2$				Dummy-Endogenous Break Tests			
				$T=200$		$T=500$		$T=200$		$T=500$		$T=200$		$T=500$	
				$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$	$\rho=1.0$	$\rho=0.9$
Level															
Type 1 (Eqn. 19)	3	0		0.053	0.354	0.051	0.993	0.057	0.222	0.048	0.937	0.040	0.253	0.057	0.925
													0.301		0.917
	6	0		0.052	0.266	0.050	0.968	0.056	0.181	0.047	0.879	0.072	0.419	0.069	0.953
													0.333		0.935
Type 2 (Eqn. 20)	3	0		0.059	0.362	0.048	0.993	0.056	0.194	0.048	0.903	0.037	0.257	0.055	0.913
													0.300		0.908
	6	0		0.059	0.282	0.052	0.970	0.044	0.108	0.045	0.716	0.089	0.357	0.067	0.877
													0.224		0.821
Trend															
Type 3 (Eqn. 21)	0	0	0.2	0.054	0.394	0.051	0.996	0.058	0.236	0.048	0.951	0.034	0.144	0.065	0.827
													0.197		0.781
	0	0	0.4	0.054	0.393	0.051	0.996	0.058	0.236	0.048	0.950	0.078	0.150	0.075	0.591
													0.111		0.511
Type 4 (Eqn. 22)	0	0	0.2	0.053	0.386	0.052	0.996	0.058	0.235	0.049	0.950	0.032	0.132	0.054	0.842
													0.180		0.838
	0	0	0.4	0.053	0.384	0.052	0.996	0.058	0.234	0.049	0.949	0.035	0.066	0.047	0.568
													0.088		0.577

Note: For Type 1 and 2 breaks in the DGP, the endogenous break tests with level shifts were used. For Type 3 and 4 breaks in the DGP, the endogenous break tests with trend shifts were used. The values at the second rows denote size-adjusted powers of the relevant tests.

TABLE 8
Effects of Smooth Transition Shifts

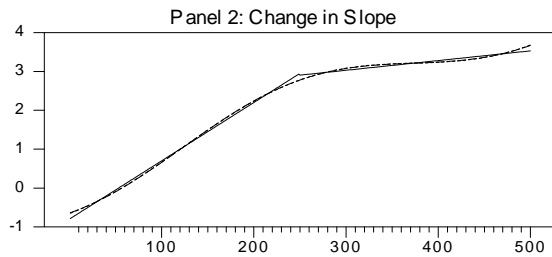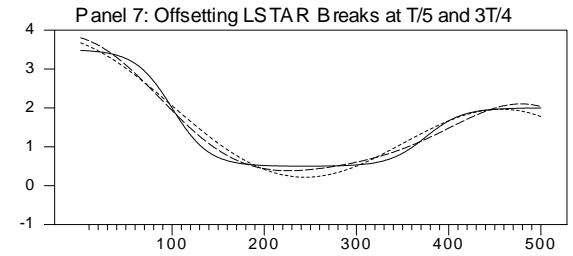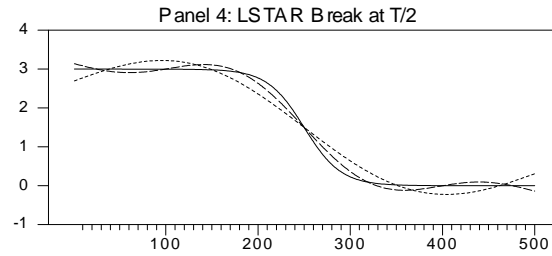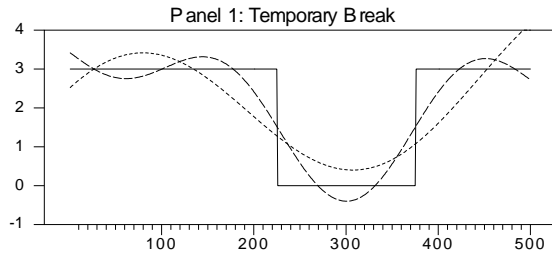| Breaks	φ	d_1	Fourier Approximation: $n = 1$						Fourier Approximation: $n = 2$					
			$T = 100$		$T = 200$		$T = 500$		$T = 100$		$T = 200$		$T = 500$	
			$\rho = 1.0$	$\rho = 0.9$	$\rho = 1.0$	$\rho = 0.9$	$\rho = 1.0$	$\rho = 0.9$	$\rho = 1.0$	$\rho = 0.9$	$\rho = 1.0$	$\rho = 0.9$	$\rho = 1.0$	$\rho = 0.9$
Type 1	0.05	3	0.054	0.122	0.054	0.395	0.051	0.997	0.045	0.075	0.058	0.238	0.048	0.952
LSTAR		6	0.054	0.121	0.054	0.392	0.051	0.996	0.044	0.075	0.058	0.238	0.048	0.952
$\lambda = 0.5$	0.10	3	0.054	0.122	0.054	0.395	0.051	0.997	0.045	0.074	0.058	0.238	0.048	0.952
		6	0.054	0.122	0.054	0.395	0.051	0.997	0.045	0.075	0.058	0.238	0.048	0.952
Type 2	0.05	3	0.053	0.122	0.054	0.395	0.051	0.996	0.045	0.075	0.058	0.238	0.048	0.952
LSTAR		6	0.053	0.120	0.053	0.391	0.050	0.996	0.044	0.075	0.058	0.238	0.048	0.952
$\lambda = 0.75$	0.10	3	0.053	0.122	0.054	0.395	0.051	0.997	0.044	0.075	0.058	0.239	0.048	0.952
		6	0.053	0.121	0.053	0.394	0.051	0.997	0.044	0.075	0.058	0.239	0.048	0.952
Type 3	0.05	3	0.053	0.123	0.054	0.396	0.051	0.997	0.045	0.075	0.058	0.238	0.048	0.952
Offsetting		6	0.054	0.122	0.055	0.396	0.051	0.997	0.045	0.075	0.058	0.238	0.048	0.952
LSTAR	0.10	3	0.054	0.122	0.054	0.396	0.051	0.997	0.046	0.075	0.059	0.238	0.048	0.952
		6	0.054	0.122	0.055	0.395	0.051	0.996	0.046	0.074	0.058	0.238	0.048	0.952
Type 4	0.05	3	0.053	0.121	0.054	0.393	0.051	0.997	0.044	0.074	0.058	0.239	0.048	0.952
ESTAR		6	0.051	0.117	0.053	0.385	0.051	0.996	0.044	0.073	0.058	0.236	0.048	0.951
	0.10	3	0.053	0.121	0.054	0.393	0.051	0.997	0.044	0.074	0.058	0.239	0.048	0.952
		6	0.051	0.117	0.053	0.385	0.051	0.996	0.044	0.073	0.058	0.236	0.048	0.951

Note: Entries in the table were constructed using the appropriate 5% critical value from Table 1 or from Table 2.

Figure 1: Sharp, ESTAR and LSTAR Breaks

Sharp Breaks

One Break

Two Breaks

Panel 1: Temporary Break

Panel 4: LSTAR Break at T/2

Panel 7: Offsetting LSTAR Breaks at T/5 and 3T/4

Panel 2: Change in Slope

Panel 5: LSTAR Break at 3T/4

Panel 8: Reinforcing LSTAR Breaks at T/5 and 3T/4

Panel 3: Change in Level and Slope

Panel 6: ESTAR Break at 3T/4

Panel 9: ESTAR Breaks at T/5 and 3T/4
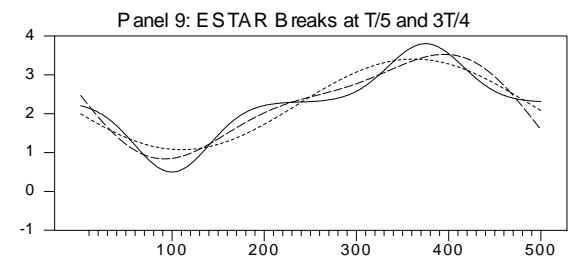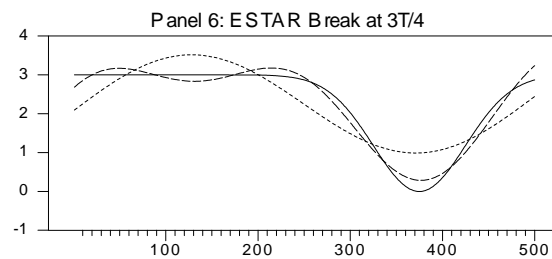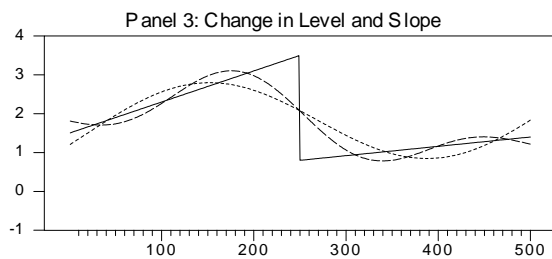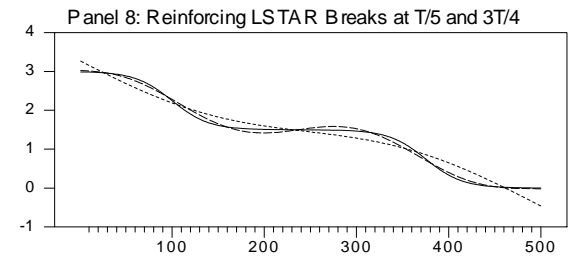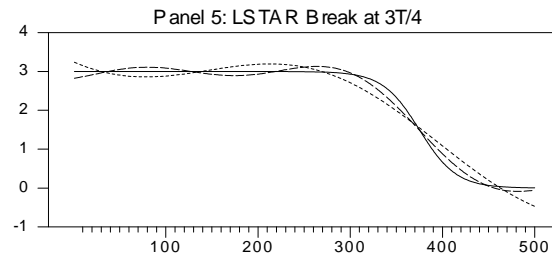
Series: ____   1-Frequency: _ _ _   2-Frequencies: __ __

Figure 2: The R3 - R1 Spread and the Two Intercepts

———— SPREAD ———— Fourier – – – – Bai-Perron
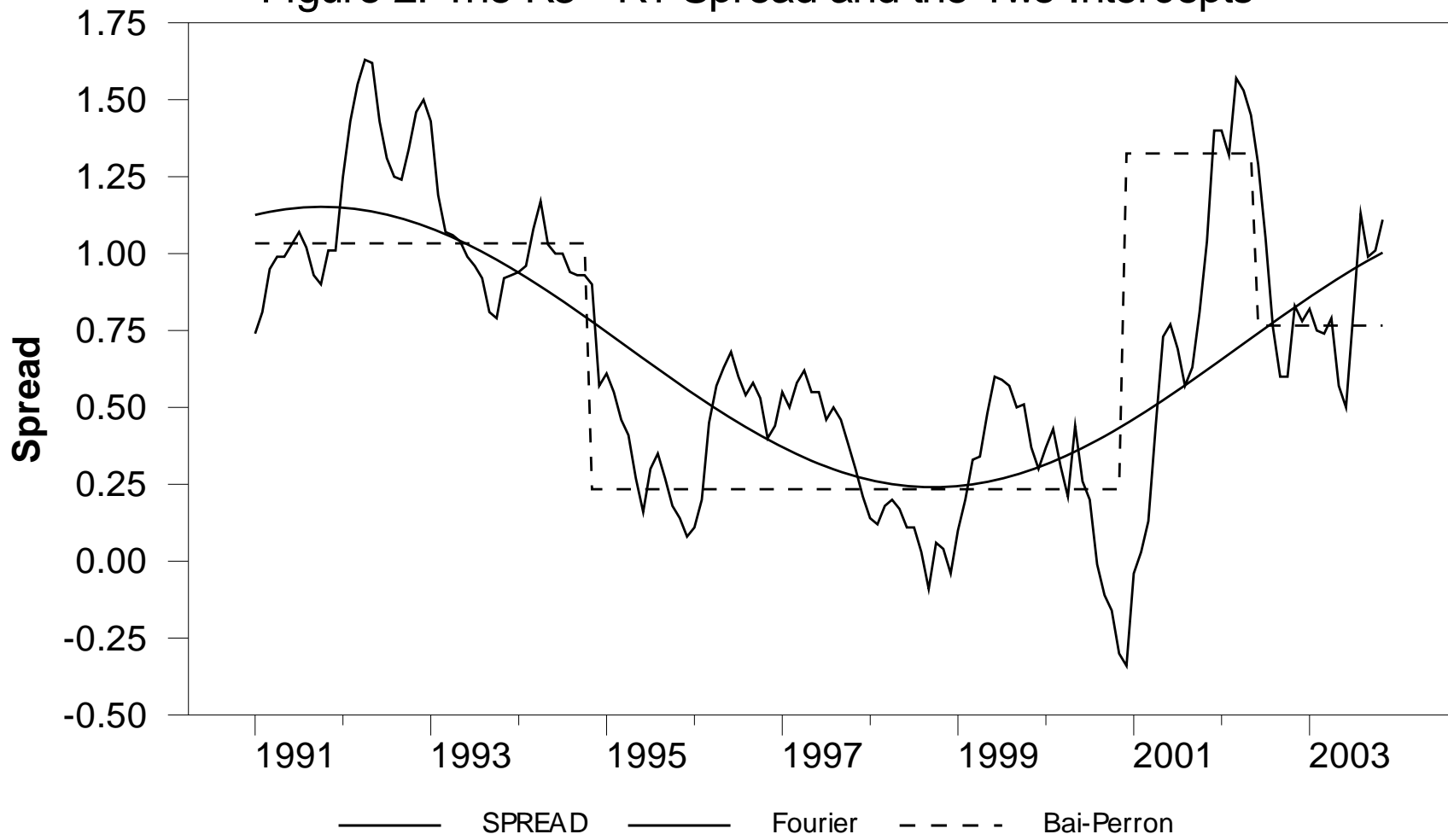
References

Amsler, C. and Lee, J. (1995). 'An LM test for a unit root in the presence of a structural change', *Econometric Theory*, Vol. 11, pp. 359 – 368.

Bai, J. and P. Perron (1998) . 'Estimating and testing linear models with multiple structural changes, *Econometrica*, Vol. 66, pp. 47 – 78.

Becker, R., W. Enders, and S. Hurn (2004). 'A general test for time dependence in parameters, *Journal of Applied Econometrics*, Vol. 19, pp. 899-906.

Becker, R., W. Enders, and J. Lee (2006). 'A stationary test with an unknown number of smooth breaks. *Journal of Time Series Analysis*, Vol. 27, pp. 381 – 409.

Bierens, H. (1997). 'Testing for a unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate, *Journal of Econometrics*, Vol. 81, pp. 29 – 64.

Davies, R. B. (1987). 'Hypothesis testing when a nuisance parameter is only identified under the Alternative', *Biometrika*, Vol. 47, pp. 33 – 43.

Enders, W. and C. Granger (1998). 'Unit-toot tests and asymmetric adjustment with an example using the term structure of interest rates', *Journal of Business and Economic Statistics*, Vol. 16, pp. 304 – 311.

Gallant, A. R. (1981). 'On the bias in flexible functional forms and an essentially unbiased form: the flexible Fourier form', *Journal of Econometrics*, Vol. 15, pp. 211 – 245.

Gallant, R. and G. Souza (1991). 'On the asymptotic normality of Fourier flexible form estimates', *Journal of Econometrics*, Vol. 50, pp. 329 – 353.

Hansen, B. (1997). 'Inference in TAR models', *Studies in Nonlinear Dynamics and Econometrics* 2, Article 1.

Harvey, D., S. Leybourne, and L. Xiao. (2010). 'Testing for nonlinear deterministic components when the order of integration is unknown', *Journal of Time Series Analysis*, Vol. 35, pp. 379 – 91.

Kapetanios, G., Y. Shin and A. Snell (2003). 'Testing for a unit root in the nonlinear STAR framework', *Journal of Econometrics*, Vol. 112, pp. 359 – 379.

Leybourne, S., P. Newbold and D. Vougas (1998). 'Unit roots and smooth transitions', *Journal of Time Series Analysis*, Vol. 19, pp. 83–97.

Perron, P. (1989). 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica*, Vol. 57, pp. 1361 – 1401.

Perron, P. (1997). 'Further evidence on breaking trend functions in macroeconomic variables', *Journal of Econometrics*, Vol. 80, pp. 355–385.

Phillips, P.C.B. and P. Perron (1988). 'Testing for a unit root in time series regression', *Biometrika*, Vol. 75, pp. 335–346.

Prodan, R. (2008). ' Potential pitfalls in determining multiple structural changes with an application to purchasing power parity', *Journal of Business and Economics Statistics*, Vol. 26, pp. 50-65.

Schmidt, P. and P. Phillips (1992). 'LM Tests for a unit root in the presence of deterministic Trends', *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 257–287.

Shin, D. and O. Lee, 2001, Tests for asymmetry in possibly nonstationary time series data', *Journal of Business and Economic Statistics*, Vol. 19, pp. 233 – 44.

APPENDIX

Proof of Lemma 1

For the estimated coefficients from the first step regression in (8), we let $\delta = (\delta_0, \delta_1, \delta_2)'$, and $\Delta Z_t = [1, \Delta\sin(2\pi kt/T), \Delta\cos(2\pi kt/T)]'$, and $D_T = diag[\sqrt{T}, 1/\sqrt{T}, 1/\sqrt{T}]$ to have

$$D_T(\tilde{\delta} - \delta) = D_T(\Delta Z'\Delta Z)^{-1}\Delta Z'u = [D_T^{-1}(\Delta Z'\Delta Z)D_T^{-1}]^{-1}\cdot D_T^{-1}\Delta Z'u \qquad (A.1)$$

where $\Delta Z = (\Delta Z_2, .., \Delta Z_T)'$ and $u = (u_2, .., u_T)'$. First, it is easy to show that

$$D_T^{-1}(\Delta Z'\Delta Z)D_T^{-1} = diag\Big[\frac{T\text{-}1}{T}, T\sum_{t=2}^{T}\Delta\sin^2(2\pi kt/T), T\sum_{t=2}^{T}\Delta\cos^2(2\pi kt/T)\Big], \qquad (A.2)$$

for which all off-diagonal terms are zero due to the orthogonality property

$$\sum_{t=2}^{T}\Delta\sin(2\pi kt/T)\Delta\cos(2\pi kt/T) = 0, \text{ and } \sum_{t=2}^{T}\Delta\sin(2\pi kt/T) = \sum_{t=2}^{T}\Delta\cos(2\pi kt/T) = 0.$$

Since $\Delta\sin(2\pi kt/T) = (2\pi k/T)\cos(2\pi kt/T) + o_p(1)$, and $\Delta\cos(2\pi kt/T) = -(2\pi k/T)\sin(2\pi kt/T) + o_p(1)$,

$$T\sum_{t=2}^{T}\Delta\sin^2(2\pi kt/T) \rightarrow (2\pi k)^2\int_0^1\cos^2(2\pi kr)dr \qquad (A.3)$$

$$T\sum_{t=2}^{T}\Delta\cos^2(2\pi kt/T) \rightarrow (2\pi k)^2\int_0^1\sin^2(2\pi kr)dr \qquad (A.4)$$

Second, we have

$$D_T^{-1}\Delta Z'u = \Big[\frac{1}{\sqrt{T}}\sum_{t=2}^{T}u_t, \sqrt{T}\sum_{t=2}^{T}u_t\Delta\sin(2\pi kt/T), \sqrt{T}\sum_{t=2}^{T}u_t\Delta\cos(2\pi kt/T)\Big]'.$$

We note that $\frac{1}{\sqrt{T}}\sum_{t=2}^{T}u_t \rightarrow \sigma W(1)$, which is a standard result. For the second and third terms, we

need to utilize the following asymptotics in Proposition 1. Then, combining the above results in (A.2) ~ (A.4), we can obtain the results in Lemma 1. □

Proposition 1

$$\sqrt{T}\sum_{t=2}^{T}u_t\Delta\sin(2\pi kt/T) \rightarrow \sigma(2\pi k)[W(1) + (2\pi k)\int_0^1\sin(2\pi kr)W(r)dr] \qquad (A.5)$$

$$\sqrt{T}\sum_{t=2}^{T}u_t\Delta\cos(2\pi kt/T) \rightarrow \sigma(2\pi k)^2[\int_0^1\cos(2\pi kr)W(r)dr] \qquad (A.6)$$

Proof: We employ the result in Bierens (1994, Lemma 9.6.3): $\sum_{t=2}^{T}F(t/T)u_t = F(1)S_T(1) - \int_0^1 f(r)S_T(r)dr$, where $f(r)$ is $F'(r)$. For (A.5), we choose $F(x) = \cos(2\pi kt/T)$. Then, we can show $F(1)S_T(1) - \int_0^1 f(r)S_T(r)dr = \sigma[W(1) + (2\pi k)\int_0^1\sin(2\pi kr)W(r)dr]$. For (A.6), we choose $F(x) = \sin(2\pi kt/T)$, and follow the similar procedure to obtain the desired result. □

Proof of Theorem 1

We let $S_t = \sum_{j=2}^{t} \varepsilon_j$ and $[rT]$ be the integer part of rT, $r \in [0,1]$. Then, it is easy to show

that the expression $\tilde{S}_t$ in (8) can be given as:

$$\frac{1}{\sqrt{T}}\tilde{S}_{[rT]} = \frac{1}{\sqrt{T}}S_{[rT]} - \frac{1}{\sqrt{T}}(\tilde{\delta}_0 - \delta_0)\, rT - \frac{1}{\sqrt{T}}(\tilde{\delta}_1 - \delta_1)\sin(2\pi k rT/T) - \frac{1}{\sqrt{T}}(\tilde{\delta}_2 - \delta_2)\cos(2\pi k rT/T)$$

$$\rightarrow \sigma\, V(r) = \sigma\left\{ W(r) - rW(1) - [(2\pi k)\int_0^1 \cos^2(2\pi kr)dr]^{-1}[W(1) + (2\pi k)\int_0^1 \sin(2\pi kr)W(r)dr]\right.$$

$$\sin(2\pi kr) - [\int_0^1 \sin^2(2\pi kr)dr]^{-1}[\int_0^1 \cos(2\pi kr)W(r)dr]\cdot\cos(2\pi kr)\biggr\} \tag{A.7}$$

Now, from the second step regression (8), we obtain:

$$\tilde{\phi} = (\tilde{S}_1' M_{\Delta Z} \tilde{S}_1)^{-1}(\tilde{S}_1' M_{\Delta Z}\, \Delta y), \tag{A.8}$$

where $\tilde{S}_1 = (\tilde{S}_1,.., \tilde{S}_{T-1})'$, $\Delta Z = (\Delta Z_2,..,\Delta Z_T)'$, $\Delta y = (\Delta y_2,..,\Delta y_T)'$, and $M_{\Delta Z} = I - \Delta Z(\Delta Z'\Delta Z)^{-1}\Delta Z'$. From

the results in (A.7), we have:

$$T^{-2}\,\tilde{S}_1' M_{\Delta Z}\tilde{S}_1 \rightarrow \sigma^2\int_0^1 \underline{V}(r)^2 dr, \tag{A.9}$$

where $\underline{V}(r)$ is the projection of the process $V(r)$ on the orthogonal complement of the space

spanned by $dz = (1,\, d\sin(2\pi kr),\, d\cos(2\pi kr))'$ where $r \in [0, 1]$. That is,

$$\underline{V}(r) = V(r) - \tilde{\delta}'dz,$$

$$\tilde{\delta} = \underset{\delta}{argmin}\, \int_0^1 (V(r) - \delta'dz)^2 dr.$$

Following SP, we can similarly show that for the second term in (A.8):

$$\frac{1}{T}\tilde{S}_1' M_{\Delta Z}\Delta y = \frac{1}{T}\tilde{S}_1' M_{\Delta Z}\,\varepsilon = \frac{1}{T}\tilde{S}_1'\,\underline{\varepsilon} \rightarrow -0.5\sigma_\varepsilon^2\,, \tag{A.10}$$

where $\underline{\varepsilon} = M_{\Delta Z}\,\varepsilon$. Theorem 1 is thus proved by combining the results in (A.9) - (A.10). □

Proof of Lemma 2

We now obtain the asymptotic distribution of the F-test in (14) which is based on the testing

regression for the LM type statistic. First, we examine SSR_0 which is obtained from the

(restricted) regression with a linear trend for the usual LM statistic.

$$\Delta y_t = \hat{\phi}\hat{y}_{t-1} + \hat{d}_0 + \hat{u}_t,$$

where $\hat{y}_{t-1} = (y_t - y_1) - \hat{\gamma}(t-1)$ and $\hat{\gamma} = (1/T)\sum \Delta y_t$. Then, under the null of a unit root:

$$SSR_0 = \sum \hat{u}_t^2 = \sum (\varepsilon_t - \hat{d}_0 - \hat{\phi}\hat{y}_{t-1})^2 = \sum [(\varepsilon_t - \bar{\varepsilon}) - \hat{\phi}(\hat{y}_{t-1} - \bar{y}_1)]^2$$

$$= \sum (\varepsilon_t - \bar{\varepsilon})^2 + \hat{\phi}^2 \sum (\hat{y}_{t-1} - \bar{y}_1)^2 - 2\hat{\phi} \sum (\hat{y}_{t-1} - \bar{y}_1)(\varepsilon_t - \bar{\varepsilon}) \qquad (A.11)$$

where $\bar{\varepsilon} = \dfrac{1}{T} \sum \varepsilon_t$, and $\bar{y}_1 = \dfrac{1}{T\text{-}1} \sum \hat{y}_{t-1}$. The first term will be cancelled with the same term that

appears in SSR_1 under the null of linearity. The second term can be written as

$$(T\hat{\phi})^2 T^{-2} \sum (\hat{y}_{t-1} - \bar{y}_1)^2 \rightarrow [-\tfrac{1}{2}(\sigma_\varepsilon^2/\sigma^2) \, (\int_0^1 \underline{V}_0(r)^2 dr)^{-1}]^2 (\sigma^2 \int_0^1 \underline{V}_0(r)^2 dr)$$

$$= \tfrac{1}{4}(\sigma_\varepsilon^4/\sigma^2) \, (\int_0^1 \underline{V}_0(r)^2 dr)^{-1} \qquad (A.12)$$

where $\underline{V}_0(r)$ is the demeaned Brownian bridge. The third term in (A.11) is shown to follow

$$- 2(T\hat{\phi}) \cdot (1/T) \sum (\hat{y}_{t-1} - \bar{y}_1)(\varepsilon_t - \bar{\varepsilon}) \rightarrow -2 \left[-\tfrac{1}{2}(\sigma_\varepsilon^2/\sigma^2) \, (\int_0^1 \underline{V}_0(r)^2 dr)^{-1}\right] (-\tfrac{1}{2}\sigma_\varepsilon^2)$$

$$= -\tfrac{1}{2}(\sigma_\varepsilon^4/\sigma^2) \, (\int_0^1 \underline{V}_0(r)^2 dr)^{-1} \qquad (A.13)$$

Next, SSR_1 is similarly obtained from the unrestricted regression (10).

$$\Delta y_t = \tilde{\phi} \, \tilde{y}_{t-1} + \tilde{d}_0 + \tilde{d}_1 \Delta\sin(2\pi kt/T) + \tilde{d}_2 \Delta\cos(2\pi kt/T) + \tilde{u}_t.$$

where $\tilde{y}_{t-1}$ is defined in (9a). Then, we get under the null of a unit root:

$$SSR_1 = \sum \tilde{u}_t^2 = \sum (\varepsilon_t - \tilde{d}_0 - \tilde{d}_1 \Delta\sin(2\pi kt/T) - \tilde{d}_2 \Delta\cos(2\pi kt/T) - \tilde{\phi} \, \tilde{y}_{t-1})^2$$

$$= \sum (\varepsilon_t - \tilde{\phi} \, \tilde{y}_{t-1})^2 = \sum \underline{\varepsilon}_t^2 + \tilde{\phi}^2 \sum \tilde{y}_{t-1}^2 - 2\tilde{\phi} \sum \tilde{y}_{t-1} \underline{\varepsilon}_t \qquad (A.14)$$

where $\underline{\varepsilon}_t$ is the element of $\underline{\varepsilon} = M_{AZ} \, \varepsilon$, which is given in (A.10), and $\tilde{y}_{t-1}$ is the element of $\tilde{y}_1 =$

$M_{AZ} \, \tilde{y}_1$ with $\tilde{y}_1 = (\tilde{y}_1, ..., \tilde{y}_{T-1})'$. The first term in (A.14) is cancelled with the similar term in SSR_0

under the null of linearity. The second and third terms in (A.14) follow the same asymptotics as

in (A.12) and (A.13), except that $\underline{V}_0(r)$ is replaced with $\underline{V}(r)$, which is defined in (A.9).

Finally, the denominator of the F-statistic is given as

$$\frac{1}{T\text{-}q} \sum \tilde{u}_t^2 + \frac{1}{T\text{-}q} O_p(1) \rightarrow \sigma_\varepsilon^2 \qquad (A.15)$$

where the $O_p(1)$ terms of the above expression are the same terms in (A.14). The asymptotic

distribution of the F-statistic is given by collecting terms in (A.12) through (A.15).

$$F(k) \rightarrow \tfrac{1}{8}(\sigma_\varepsilon^2/\sigma^2)[(\int_0^1 \underline{V}(r)^2 dr)^{-1} - (\int_0^1 \underline{V}_0(r)^2 dr)^{-1}] \qquad \square$$

Proof of Lemma 3

The DGP implies (6), which includes the nonlinear trigonometric terms, but they are ignored in the testing regressions. Thus, the first step regression is $\Delta y_t = \delta_0 + u_t$. Then, we can show that the OLS estimate of δ_0 is given as

$$\hat{\delta}_0 = \text{mean of } \Delta y = \delta_0 + \overline{\Delta e} \tag{A.16}$$

Then, we construct a detrended series using $\hat{\delta}_0$ as:

$$\hat{S}_t = y_t - y_1 - \tilde{\delta}_0(t\text{-}1) = y_t - y_1 - \delta_0(t\text{-}1) - (\tilde{\delta}_0 - \delta_0)(t\text{-}1)$$
$$= (e_t - e_1) + a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T) - a_1\sin(2\pi k/T) - a_2\cos(2\pi k/T) - (t\text{-}1)\overline{\Delta e}.$$

The second step regression involves $\Delta y_t = \phi \hat{S}_{t\text{-}1} + c + u_t$. For simplicity, we ignore the constant term, which is 0 in the population. Then, for the denominator of $\hat{\phi}$, we have:

$$\frac{1}{T}\sum_{t=2}^{T}\hat{S}_{t\text{-}1}^2 \approx \frac{1}{T}\sum_{t=2}^{T}\hat{S}_t^2 = \frac{1}{T}\sum_{t=2}^{T}\big[(e_t - e_1) + a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T) - a_1\sin(2\pi k/T)$$
$$- a_2\cos(2\pi k/T) - (t\text{-}1)\overline{\Delta e}\,\big]^2.$$

This can be expressed as:

$$\frac{1}{T}\sum_{t=2}^{T}[(e_t - e_1) - (t\text{-}1)\overline{\Delta e}]^2 + \frac{1}{T}\sum_{t=2}^{T}[a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T)]^2$$
$$+ \frac{1}{T}\sum_{t=2}^{T}[-a_1\sin(2\pi k/T) - a_2\cos(2\pi k/T)]^2 + 2\frac{1}{T}\sum_{t=2}^{T}[(e_t - e_1) - (t\text{-}1)\overline{\Delta e}][a_1\sin(2\pi kt/T)$$
$$+ a_2\cos(2\pi kt/T)] + 2\frac{1}{T}\sum_{t=2}^{T}[(e_t - e_1) - (t\text{-}1)\overline{\Delta e}][-a_1\sin(2\pi k/T) - a_2\cos(2\pi k/T)]$$
$$+ 2\frac{1}{T}\sum_{t=2}^{T}[a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T)][-a_1\sin(2\pi k/T) - a_2\cos(2\pi k/T)] \tag{A.17}$$

For each term, we can show:

$$\frac{1}{T}\sum_{t=2}^{T}[(e_t - e_1) - (t\text{-}1)\overline{\Delta e}]^2 \rightarrow \sigma_\varepsilon^2 + (1/3)(\varepsilon_\infty^2 + \varepsilon_1\varepsilon_\infty + \varepsilon_1^2) \tag{A.17a}$$

$$\frac{1}{T}\sum_{t=2}^{T}[a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T)]^2 \rightarrow a_1^2\int_0^1\sin^2(2\pi kr)dr + a_2^2\int_0^1\cos^2(2\pi kr)dr \tag{A.17b}$$

$$\frac{1}{T}\sum_{t=2}^{T}[-a_1\sin(2\pi k/T) - a_2\cos(2\pi k/T)]^2 \rightarrow a_2^2 \tag{A.17c}$$

$$2\frac{1}{T}\sum_{t=2}^{T}[(e_t - e_1) - (t\text{-}1)\overline{\Delta e}][a_1\sin(2\pi kt/T) + a_2\cos(2\pi kt/T)]$$
$$= 2\frac{1}{T}\sum_{t=2}^{T}(e_t - e_1)a_1\sin(2\pi kt/T) + 2\frac{1}{T}\sum_{t=2}^{T}(e_t - e_1)a_2\cos(2\pi kt/T)$$
$$- 2\frac{1}{T}\sum_{t=2}^{T}(t\text{-}1)\overline{\Delta e}\cdot a_1\sin(2\pi kt/T) - 2\frac{1}{T}\sum_{t=2}^{T}(t\text{-}1)\overline{\Delta e}\cdot a_2\cos(2\pi kt/T)$$

37

$$\rightarrow 0 + 0 - 2\sigma W(1)a_1 \int_0^1 r \cdot \sin(2\pi kr)dr - 2\sigma W(1)a_2 \int_0^1 r \cdot \cos(2\pi kr)dr \qquad \text{(A.17d)}$$

In the above, the first and the second terms in (A.17d) are degenerate since $\dfrac{1}{\sqrt{T}}\sum\limits_{t=2}^{T} e_t \sin(2\pi kt/T)$

$\rightarrow \sigma\int_0^1 \cos(2\pi kr)W(r)dr$, and $\dfrac{1}{\sqrt{T}}\sum\limits_{t=2}^{T} e_t \cos(2\pi kt/T) \rightarrow \sigma W(1) + \sigma(2\pi k)\int_0^1 \sin(2\pi kr)W(r)dr$. The

results for the third and fourth terms follow, since $T\overline{\Delta e} \rightarrow \sigma W(1)$, $\dfrac{1}{T}\sum\limits_{t=2}^{T} (t/T)\sin(2\pi kt/T) \rightarrow \sigma\int_0^1 r$

$\sin(2\pi kr)dr$, and $\dfrac{1}{T}\sum\limits_{t=2}^{T} (t/T)\cos(2\pi kt/T) \rightarrow \sigma\int_0^1 r \cos(2\pi kr)dr$. Thus, the asymptotic distribution of

the numerator of $\hat{\phi}$ is obtained by collecting the terms in (A.17a) - (A.17d).

Next, we examine the numerator of $\hat{\phi}$. Since $\Delta\hat{S}_t = \Delta e_t - \overline{\Delta e} + a_1\Delta\sin(2\pi kt/T) + a_2\Delta\cos(2\pi kt/T)$, we have:

$$\frac{1}{T}\sum_{t=2}^{T} \Delta\hat{S}_t\hat{S}_{t-1} = \frac{1}{T}\sum_{t=2}^{T} \left[(\Delta e_t - \overline{\Delta e} + a_1\Delta \sin(2\pi kt/T) + a_2\Delta \cos(2\pi kt/T)\right]\left[(e_t - e_1)\right.$$

$$+ a_1 \sin(2\pi kt/T) + a_2 \cos(2\pi kt/T) - a_1 \sin(2\pi k/T) - a_2 \cos(2\pi k/T) - (t-1)\overline{\Delta e}\right]$$

$$= \frac{1}{T}\sum_{t=2}^{T} (e_t - \overline{\Delta e}) \left[(e_{t-1} - e_1) - (t-2)\overline{\Delta e}\right] + \frac{1}{T}\sum_{t=2}^{T} (\Delta e_t - \overline{\Delta e})\left[a_1 \sin(2\pi kt/T) + a_2 \cos(2\pi kt/T)\right.$$

$$\left. - a_1 \sin(2\pi k/T) - a_2 \cos(2\pi k/T)\right] + \frac{1}{T}\sum_{t=2}^{T} \left[a_1\Delta \sin(2\pi kt/T) + a_2\Delta\cos(2\pi kt/T)\right]\left[a_1 \sin(2\pi kt/T)\right.$$

$$\left. + a_2 \cos(2\pi kt/T) - a_1 \sin(2\pi k/T) - a_2 \cos(2\pi k/T)\right] + \frac{1}{T}\sum_{t=2}^{T} \left[a_1\Delta\sin(2\pi kt/T)\right.$$

$$\left. + a_2\Delta\cos(2\pi kt/T)\right] \left[(e_{t-1} - e_1) - (t-2)\overline{\Delta e}\right] \qquad \text{(A.18)}$$

The first term in the last equation in (A.18) follows:

$$\frac{1}{T}\sum_{t=2}^{T} (\Delta e_t - \overline{\Delta e}) \left[(e_{t-1} - e_1) - (t-2)\overline{\Delta e}\right] = \frac{1}{T}\sum_{t=2}^{T} \Delta e_t(e_{t-1} - e_1) - \frac{1}{T}\sum_{t=2}^{T} \overline{\Delta e}(e_{t-1} - e_1)$$

$$+ \frac{1}{T}\sum_{t=2}^{T} \Delta e_t (t-2)\overline{\Delta e} + \frac{1}{T}\sum_{t=2}^{T} \overline{\Delta e}(t-2)\overline{\Delta e} \rightarrow \sigma_\varepsilon^2(\rho - 1) + o_p(1)$$

For the above result, we note $\dfrac{1}{T}\sum\limits_{t=2}^{T} \Delta e_t(e_{t-1} - e_1) \approx \dfrac{1}{T}\sum\limits_{t=2}^{T} (e_{t-1} - e_1)e_{t-1}$ since $e_1\dfrac{1}{T}\sum\limits_{t=2}^{T} \Delta e_t = e_1\overline{\Delta e}$

$\rightarrow 0$. Also, $\dfrac{1}{T}\sum\limits_{t=2}^{T} (e_{t-1} - e_1)e_{t-1} \rightarrow \gamma_1 - \sigma_\varepsilon^2 = \sigma_\varepsilon^2(\rho - 1)$ where ρ is the AR coefficient in the DGP

(7). It can be easily seen that all remaining terms are degenerate. The second term in the last

equation in (A.14) can be shown as $o_p(1)$ by employing the results that $\dfrac{1}{\sqrt{T}}\sum\limits_{t=2}^{T} e_t \sin(2\pi kt/T) \rightarrow$

$\sigma\int_0^1 \cos(2\pi kr)W(r)dr$, and $\dfrac{1}{\sqrt{T}}\sum\limits_{t=2}^{T} e_t \sin(2\pi kt/T) \rightarrow \sigma\left[W(1)+(2\pi k)\int_0^1 \cdot\sin(2\pi kr)W(r)dr\right.$. The third

and fourth terms in the last equation in (A.18) can be shown as $o_p(1)$. Thus,

38

$$\frac{1}{T}\sum_{t=2}^{T}\Delta\hat{S}_t\hat{S}_{t-1} \to \sigma_\varepsilon^2(\rho - 1) \qquad\qquad (A.19)$$

The result in Lemma 3 follows by combining the results in (A.17) and (A.19). □