

# On the Behaviour of Phillips-Perron Tests in the Presence of Persistent Cycles

Tomás del Barrio Castro<sup>a</sup>, Paulo M.M. Rodrigues<sup>b</sup> and A.M. Robert Taylor<sup>c</sup>

<sup>a</sup> Department of Applied Economics, University of the Balearic Islands

<sup>b</sup> Banco de Portugal, NOVA School of Business and Economics, Universidade Nova de Lisboa, CEFAGE

<sup>c</sup>Granger Centre for Time Series Econometrics, University of Nottingham

May 2013

## Abstract

In this paper we provide a detailed analysis of the impact of persistent cycles on the well-known semi-parametric unit root tests of Phillips and Perron (1988, *Biometrika* 75, 335–346). It is shown analytically and through Monte Carlo simulations that the presence of complex (near) unit roots can severely bias the size properties of these unit root test procedures.

**Keywords:** Phillips-Perron unit root test, non-stationarity, serial correlation, cyclicity, business cycles.

**JEL:** C12, C22.

## 1 Introduction

Cycles are an important feature of many macroeconomic, financial and other time series. As such, studying their impact on the performance of pre-testing procedures, in particular on the limiting null distributions and finite sample properties of zero frequency unit root test statistics, is of particular relevance. This is undertaken in Castro, Rodrigues and Taylor (2011) [CRT], for the familiar augmented Dickey-Fuller [ADF] tests, the variance ratio test of Breitung (2002) and the M unit root tests of Stock (1999) and Perron and Ng (1996).

This paper contributes to the literature in two ways. First it provides asymptotic and finite sample results for the ordinary least squares [OLS] estimator of the parameter of a nearly integrated first-order autoregressive (AR(1)) model driven by shocks which are generated according to a near integrated cyclical process; that is, a process characterised by a second order autoregressive structure with complex roots in the neighbourhood of unity. A second contribution relates to the discussion of the behaviour of the well known and widely used Phillips-Perron [PP] unit root tests (Phillips, 1987, Phillips and Perron, 1988) in this important context. It will be shown that the presence of persistent cycles can seriously impact upon both the small and large sample properties of these tests.

The remainder of the paper is organized as follows. In section 2 we present our reference time series model which allows for persistent cycles (cyclical near-unit roots) and we briefly

outline the PP unit root tests. Section 3 details the large sample behaviour of the PP unit root statistics and associated tests when persistent cycles are present in the data. In section 4, we report finite sample simulation results relating to the performance of the least squares estimator from a first order autoregression and of the PP tests when persistent cycles are present in the data. Section 5 concludes. All proofs are collected in an mathematical appendix.

## 2 The Model and Unit Root Tests

### 2.1 The Time Series Model

Autoregressive (AR) processes with roots on the complex unit circle are non-stationary and display persistent cyclical behavior similar to that of persistent business cycles (see Bierens, 2001, Allen, 1997). Hence, consider, without loss of generality, a univariate time series  $\{x_t\}$  generated according to an autoregressive process of order 3 [AR(3)], viz.,

$$\vartheta_\phi(L)(1 - \varphi_{0T}L)x_t = \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2), \quad t = 1, 2, \dots, T \quad (1)$$

where  $L$  denotes the usual lag operator. It is assumed throughout that the process is initialised at  $x_{-2} = x_{-1} = x_0 = 0$ , although weakening this assumption to allow these starting values to be of  $o_p(T^{1/2})$  would not change any of the asymptotic results which follow. In (1) the autoregressive polynomial  $\vartheta_\phi(L) = (1 - 2\cos(\phi)\varphi_{\phi T}L + \varphi_{\phi T}^2L^2)$ , with  $\phi \in (0, \pi)$  and  $\varphi_{\phi T} := \exp(c_\phi/T) \simeq (1 + c_\phi/T)$  with  $c_\phi \leq 0$  and fixed. Consequently, when  $c_\phi = 0$  ( $c_\phi < 0$ ),  $\vartheta_\phi(L)$  admits the complex conjugate pair of unit (near-unit) roots,  $\exp(\pm i\phi) \equiv \cos(\phi) \pm i\sin(\phi)$ , at the spectral frequency  $\phi$ . Notice that when  $c_\phi < 0$  a complex conjugate pair of stable roots at frequency  $\phi$  is obtained for any finite  $T$ . The polynomial  $\vartheta_\phi(L)$  generates a persistent cycle of  $2\pi/\phi$  periods. Indeed, and as noted by Díaz-Emparanza (2004), it also generates an aliased cycle of  $2\pi/(2\pi - \phi)$  periods, and as such our analysis in fact covers  $\phi \in (0, 2\pi) - \{\pi\}$ .

In (1) we also allow for a (near) unit root at the zero frequency through the parameter  $\varphi_{0T} := \exp(c_0/T) \simeq (1 + c_0/T)$  with  $c_0 \leq 0$  and fixed; a zero frequency unit (near-unit) root is obtained when  $c_0 = 0$  ( $c_0 < 0$ ). In the case where  $c_0 = c_\phi = 0$ ,  $\{x_t\}$  is integrated of order one at both the zero and  $\phi$  spectral frequencies, denoted  $I_0(1)$  and  $I_\phi(1)$ , respectively. In this case, it follows that  $z_t := \Delta x_t$ , where  $\Delta := (1 - L)$ , will be  $I_\phi(1)$  but  $I_0(0)$ , while  $u_t := \Delta_\phi x_t$ , where  $\Delta_\phi := (1 - 2\cos(\phi)L + L^2)$ , will be  $I_0(1)$  but  $I_\phi(0)$ . Our focus in this paper is on testing the standard zero frequency unit root null hypothesis that  $x_t \sim I_0(1)$ ,  $H_0 : \varphi_{0T} = 1$ , against the alternative that  $x_t \sim I_0(0)$ ,  $H_1 : |\varphi_{0T}| < 1$ , in the case where  $\vartheta_\phi(L)$  admits a pair of near-integrated complex roots at frequency  $\phi$ .

Note that model (1) can be easily extended to allow for deterministic components, weak dependence in  $\{\varepsilon_t\}$ , and unit roots at other cyclical frequencies lying in  $(0, \pi)$  and/or at the Nyquist ( $\pi$ ) frequency, without altering the qualitative conclusions which can be drawn from the analysis of (1). For expositional purposes we will therefore restrict our attention to (1).

## 2.2 The PP Unit Root Test

The use of lag augmentation using lags of the dependent variable in the ADF regression (see, for example, Dickey and Fuller, 1979) is motivated by the need to generate residuals which are free of serial correlation. However, an alternative unit root testing approach that can be used in the context of models with weakly dependent errors is that of Phillips (1987) and Phillips and Perron (1988), known as the Phillips–Perron [PP] unit root tests. In contrast to the ADF approach, the PP tests deal with serial correlation in the errors by employing a nonparametric serial correlation correction factor, which is based on a consistent estimate of the long-run variance of the error process. An in-depth investigation into the behaviour of the PP test when different types of first order AR and MA dependencies are allowed for in the errors is given in Nabeya and Perron (1994), Perron and Ng (1996) and Perron and Ng (1998).

The application of the PP unit root tests is based on the ordinary least squares (OLS) parameter estimate,  $\hat{\alpha}$  from the  $AR(1)$  (pseudo-) equation

$$x_t = \alpha x_{t-1} + u_t. \quad (2)$$

It is straightforward to show, using (A.5) in the Appendix and the results given in Hamilton (1994, p.517), that  $\alpha = 1 + 2c_0 T^{-1}(1 - \cos \phi) + O(T^{-2})$  in (2). Using the estimate  $\hat{\alpha}$ , the PP unit root statistics are then computed as

$$Z_\alpha := T(\hat{\alpha} - 1) - \frac{1}{2}(\hat{\lambda}^2 - s^2) \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1} \quad (3)$$

$$Z_t := \frac{s}{\hat{\lambda}} t_{\hat{\alpha}=1} - \frac{1}{2}(\hat{\lambda}^2 - s^2) \left( \frac{\hat{\lambda}^2}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1/2} \quad (4)$$

where  $t_{\hat{\alpha}=1} := s^{-1}(\hat{\alpha} - 1) \left( \sum_{t=1}^T x_{t-1}^2 \right)^{1/2}$  and  $s^2 := T^{-1} \sum_{t=1}^T \hat{u}_t^2$  and  $\hat{\lambda}^2$  are estimators of the short and long run variances of  $\{u_t\}$ , respectively. Following Perron and Ng (1996), two alternative estimators for the long-run variance may be considered. Firstly, a non-parametric kernel estimator based on the sample autocovariances,  $\hat{\lambda}^2 = s_{WA}^2$ , with  $s_{WA}^2 := \sum_{h=-T+1}^{T-1} \omega(h/m) \hat{\gamma}_h$ ,  $\hat{\gamma}_h := T^{-1} \sum_{t=1}^{T-|h|} \hat{u}_t \hat{u}_{t+|h|}$ , where  $\hat{u}_t$  are the OLS residuals from regressing  $x_t$  on  $x_{t-1}$ , with kernel function  $\omega(\cdot)$  satisfying *e.g.* the general conditions reported in Jansson (2002, Assumption A3) and the bandwidth parameter  $m \in (0, \infty)$  satisfying  $1/m + m^2/T \rightarrow 0$  as  $T \rightarrow \infty$  (which corresponds to Assumption A4 of Jansson, 2002). Secondly, a parametric autoregressive spectral density [ASD] estimator,  $\hat{\lambda}^2 = s_{AR}^2$ , of the form suggested by Berk (1974), where  $s_{AR}^2 := s_k^2 / \left( 1 - \sum_{i=1}^k \hat{\alpha}_i \right)^2$ , and  $s_k^2 := T^{-1} \sum \hat{\varepsilon}_{k,t}^2$ , are computed using the estimates from the ADF-type regression,

$$\Delta x_t = \hat{\rho} x_{t-1} + \sum_{j=1}^k \hat{\alpha}_j \Delta x_{t-j} + \hat{\varepsilon}_{k,t} \quad (5)$$

for  $k = 2$ .

### 3 Asymptotic Distribution Theory

In the case where  $u_t$  in (2) satisfies certain mixing conditions, Phillips (1987) and Phillips and Perron (1988, Theorem 3, p.342) present the limit results of the statistics in (3) and (4) under the unit root null hypothesis,  $\alpha = 1$  in (2). However, where the data are generated according to (1) these mixing conditions are violated. This is because here  $u_t$  in (2) can be shown to be given by  $u_t = \varepsilon_t + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2}$ , where  $\gamma_1 := 2 \cos \phi - T^{-1}(c_0 - 2 \cos \phi(c_0 + c_\phi)) + O(T^{-2})$  and  $\gamma_2 := -1 - T^{-1}(c_0 + 2c_\phi) + O(T^{-2})$ , which admits (near-) unit root behaviour at frequency  $\phi$ . As a consequence the limit results given in Phillips (1987) and Phillips and Perron (1988) are no longer valid. To that end, in Proposition 3.1 we now detail the impact on the asymptotic distributions of the PP tests when the data display persistent cyclical behaviour.

**Proposition 3.1** *Let the time series process  $\{x_t\}$  be generated by (1) with  $\varphi_{0T} := 1 + c_0/T$  and  $\varphi_{\phi T} := 1 + c_\phi/T$ , with  $c_0 \leq 0$ ,  $c_\phi \leq 0$  and fixed. Then for any  $\phi \in (0, \pi)$ , the OLS estimator  $\hat{\alpha}$  from (2) is such that, as  $T \rightarrow \infty$ ,*

$$\hat{\alpha} - 1 \Rightarrow - \frac{(1 - \cos \phi) \int_0^1 \left( \left[ W_{\phi, c_\phi}^\alpha(r) \right]^2 + \left[ W_{\phi, c_\phi}^\beta(r) \right]^2 \right) dr}{2(1 + \cos \phi) \int_0^1 W_{0, c_0}^2(r) dr + \int_0^1 \left( \left[ W_{\phi, c_\phi}^\alpha(r) \right]^2 + \left[ W_{\phi, c_\phi}^\beta(r) \right]^2 \right) dr} \quad (6)$$

where  $W_{0, c_0}(r)$ ,  $W_{\phi, c_\phi}^\alpha(r)$  and  $W_{\phi, c_\phi}^\beta(r)$  are mutually independent standard Ornstein-Uhlenbeck [OU] processes on  $[0, 1]$ .

**Remark 1:** Observing from Tanaka (1996, p.115) and Phillips (1989) that the functionals of the OU processes which appear in the right member of (6) have positive support, and that  $(1 - \cos \phi)$  and  $(1 + \cos \phi)$  both lie in the interval  $(0, 2)$  for all  $\phi \in (0, \pi)$ , it is seen that  $(\hat{\alpha} - 1)$  will take negative values with probability one in the limit. As a result,  $T(\hat{\alpha} - 1)$  will diverge to  $-\infty$  as  $T \rightarrow \infty$ .

**Remark 2:** Noting that the expected value of a squared Brownian motion integrated between zero and one is equal to  $\frac{1}{2}$ , it is seen that in the case where  $c_0 = c_\phi = 0$ ,  $E(\hat{\alpha}) \Rightarrow 1 - \frac{(1 - \cos \phi)}{(1 + \cos \phi) + 1}$ . This result is useful as it tells us that the expected value of  $\hat{\alpha}$  will depend on the frequency  $\phi$  at which the cyclical unit roots occur. In particular, noting that  $\frac{(1 - \cos \phi)}{(1 + \cos \phi) + 1} \rightarrow 0$  as  $\phi \rightarrow 0$  and that  $\frac{(1 - \cos \phi)}{(1 + \cos \phi) + 1} \rightarrow 2$  as  $\phi \rightarrow \pi$ , it follows that  $E(\hat{\alpha})$  will converge to 1 as  $\phi \rightarrow 0$  and to  $-1$  as  $\phi \rightarrow \pi$ . These patterns can clearly be seen in Figure 1 which is discussed in the next section.

Let us now turn our attention to the large sample behaviour of the  $Z_\alpha$  and  $Z_t$  unit root statistics in (3) and (4) when the long-run variance is estimated by either a kernel-based or ASD estimator. These results are established in Theorem 3.1.

**Theorem 3.1** *Let the conditions of Proposition 3.1 hold. Then, for any  $\phi \in (0, \pi)$ , we have that, as  $T \rightarrow \infty$ :*

$$(i) \text{ for } \hat{\lambda}^2 := s_{WA}^2, Z_\alpha \rightarrow -\infty \text{ and } Z_t \rightarrow -\infty;$$

(ii) for  $\hat{\lambda}^2 := s_{AR}^2$ ,  $Z_\alpha \rightarrow \pm\infty$  and  $Z_t \rightarrow \pm\infty$ .

**Remark 3:** The results in Theorem 3.1 show that the PP tests diverge as the sample size diverges, regardless of whether a kernel-based or ASD long run variance estimator is used. However, while the kernel-based PP unit root statistics always diverge to  $-\infty$  in the presence of persistent cycles, the ASD-based PP statistics diverge to either  $+\infty$  or  $-\infty$ . As a consequence, the kernel-based PP tests will have asymptotic size unity in the presence of persistent cycles, while the ASD-based tests will have asymptotic size of either unity or zero. The behaviour of these statistics is then governed by the large sample behaviour of the long run variance estimate used. To gain more insight into the mechanics observe, using (3) and (6), that

$$\begin{aligned} Z_\alpha &= T(\hat{\rho} - 1) - \frac{1}{2}(\hat{\lambda}^2 - s^2) \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1} \\ &= -TO_p^+(1) - \frac{1}{2} \left( O_p^+((mT)^\delta) - O_p^+(T) \right) / O_p^+(1) \end{aligned} \quad (7)$$

where  $\delta$  is an indicator function such that  $\delta = 1$  when a kernel-based long-run variance estimator is used and  $\delta = 0$  when an ASD based long-run variance estimator is considered, and  $O_p^+(1)$  is used to indicate a quantity that is  $O_p(1)$  and strictly positive in the limit. In the case of the kernel-based estimator,  $\hat{\lambda}^2 := s_{WA}^2$ , it then follows, due to the non-stationarity of  $(1-L)x_t$ , that  $s_{WA}^2 = O_p(mT)$ , where  $m$  is the bandwidth used to compute  $s_{WA}^2$  (see, the proof of Theorem 2 in Taylor 2003), and hence that  $Z_\alpha$  will diverge to  $-\infty$  at the rate  $mT$ , with the same result holding for  $Z_t$ . In the case of the ASD estimator,  $\hat{\lambda}^2 := s_{AR}^2$ , CRT establish the result that  $s_{AR}^2 \xrightarrow{p} \sigma^2 / (4(1 - \cos \phi)^2)$ . Consequently, the  $Z_\alpha$  statistic will now diverge at the slower  $O_p(T)$  rate. Whether it diverges to  $+\infty$  or  $-\infty$  is determined by the relative magnitude of the first and third terms on the right hand side of (7); specifically, for values of  $\phi$  close to zero the term arising from  $s^2$  will dominate, effecting divergence to  $+\infty$ , while otherwise  $Z_\alpha$  will be dominated by  $T(\hat{\rho} - 1)$  and, hence, will diverge to  $-\infty$ . The same holds for the test based on  $Z_t$ . These patterns are again clearly seen in Figure 1 which will be discussed in the next section.

**Remark 4:** It is straightforward to show that the results stated in Proposition 3.1 and Theorem 3.1 continue to hold under weaker linear process conditions on  $\{\varepsilon_t\}$  provided Assumptions 1.1-1.2 of Gregoir (2004), adapted slightly to our situation, are satisfied. Precisely, these conditions entail that  $\varepsilon_t = d(L)e_t$ , where  $(\varepsilon_t, \mathcal{F}_t)$  is a martingale difference sequence, with filtration  $(\mathcal{F}_t)$ , such that  $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2$  and  $\sup_t E(|\varepsilon_t|^{2+\kappa} | \mathcal{F}_{t-1}) < \infty$  a.s. for some  $\kappa > 0$  and where  $d(L) := 1 + \sum_{j=1}^{\infty} d_j z^j$  is such that  $d(z) \neq 0$  for  $z = 0$  and  $z = \phi$  and  $\sum_{j=1}^{\infty} j |d_j| < \infty$ . In such cases the lag truncation order,  $k$ , used in (5) needs to be such that  $1/k + k^3/T \rightarrow 0$  as  $T \rightarrow \infty$ ; see Said and Dickey (1984) and Chang and Park (2002).

## 4 Monte Carlo Experiments

In this section, we investigate the finite sample properties of the PP unit root tests detailed in section 2.2 when (near-) non-stationary cycles are present in the data. All results reported in

this section are based on 5,000 Monte Carlo replications and were programmed using MATLAB. Results are presented for the case of linear de-trended data only; the results for demeaned data are qualitatively similar and are therefore omitted but can be obtained upon request.

The results of the experiments reported in Figures 1-5, relate to data generated according to:

$$\left[1 - \left(1 + \frac{c_0}{T}\right)L\right] \left[1 - 2\cos(\phi)\left(1 + \frac{c_\phi}{T}\right)L + \left(1 + \frac{c_\phi}{T}\right)^2 L^2\right] x_t = \epsilon_t \sim NIID(0, 1) \quad (8)$$

with  $c_0 = 0$ , so that the unit root null hypothesis holds, and with  $\phi \in \{0, 0.1\pi, \dots, \pi\}$  and  $c_\phi \in \{0, -5, -20, -100\}$ . Results are reported for the sample sizes  $T \in \{100, 500\}$ , initialised at  $x_{-2} = x_{-1} = x_0 = 0$ .

**Remark 5:** Notice that although we have included the frequencies  $\phi = 0$  and  $\phi = \pi$  in our Monte Carlo exercise, these were not allowed in the theoretical analysis conducted in the previous section since they lead to second order integration (double unit root behaviour) at the zero and Nyquist frequencies respectively.

We first investigate the finite sample dependence of the OLS estimate,  $\hat{\alpha}$ , from (2) on the frequency  $\phi$  at which a persistent cycle occurs. To that end, in Figure 1, we plot the average (taken across the Monte Carlo replications) value of  $\hat{\alpha}$  against  $\phi$ .

### Insert Figures 1 – 2 about here

Figure 1 clearly highlights the result that for small values of  $c_\phi$  the average of  $\hat{\alpha}$  can vary considerably from 1 in finite samples, mirroring the asymptotic predictions discussed in Remark 2. This is particularly apparent when  $\phi > \frac{\pi}{2}$ ; indeed as  $\phi$  approaches  $\pi$  it clearly converges towards  $-1$ , as predicted by the results in Proposition 3.1 (see Remark 2). For larger values of  $c_\phi$  this effect is attenuated for  $T = 100$  but is still present for  $T = 500$ . As  $\phi$  approaches zero, again the finite samples results would appear to very closely mirror the prediction from the asymptotic theory.

We next investigate the finite sample behaviour of the kernel-based and ASD-based long-run variance estimators used in constructing the PP tests. Results are reported in Figure 2 for both the kernel-based Bartlett (denoted  $s_B^2$ ) and the Quadratic Spectral (denoted  $s_{QS}^2$ ) windows, together with the autoregressive spectral ( $s_{AR}^2$ ) estimator. From the results in Figure 2 we observe that when the constant used to select the kernel window is  $\kappa = 12$ , [the value of this parameter used in the simulations reported in Figures 3-5 below] then for  $c_\phi = 0$  and  $c_\phi = -5$ ,  $s^2 > s_i^2$ ,  $i = AR, B$ , and  $QS$ , and consequently the component  $(s_i^2 - s^2)$  in the PP test statistics will always be negative, thereby effecting a shift to the right of the tests' distributions and consequently leading to an under-rejection of the null hypothesis of a unit root.

Next in Figures 3-5 we report the empirical (null) rejection frequencies of the PP unit root tests for a nominal 5% significance level using the asymptotic critical values appropriate to the case where (near-) non-stationary cycles are not present in the data; that is, from the Dickey-Fuller (1979) distributions (see Hamilton, 1994, Tables B.5 and B.6, pp. 762-763). For

$\hat{\lambda}^2 := s_{WA}^2$ , results are reported for the Bartlett and the Quadratic Spectral kernels, using the data-dependent bandwidth formulations for these kernels suggested in Newey and West (1994, Equations (3.8) to (3.15) and Table 1) with  $k = 12$ . For  $\hat{\lambda}^2 := s_{AR}^2$ , we set the augmentation order in the ADF type regression as  $k = 2$ .

**Insert Figures 3 – 5 about here**

A first observation that can be drawn from the results in Figures 3-5 is that the large sample results given in Theorem 3.1 appear to provide useful predictions for the small sample behaviour of the PP tests in the presence of persistent cycles. From the results in Theorem 3.1 we would anticipate that those tests based on  $s_{WA}^2$  will almost always reject the null hypothesis of a zero frequency unit root, when it is true, when a persistent cyclical component is present in the data, but that the behaviour of the tests based on the ASD estimator,  $s_{AR}^2$ , will depend on the frequency of the persistent cycle. This behaviour becomes more noticeable for the larger sample ( $T = 500$ ) considered and particularly so for the tests based on the Bartlett estimator. From Figures 3-5 we also observe that the tests based on  $s_{AR}^2$  tend to be very undersized for low frequency persistent cycles, with the converse true for high frequency cycles. This is a direct consequence of the results reported in Figures 1-2 where we saw that  $\hat{\alpha} \rightarrow -1$  as  $\phi \rightarrow \pi$ , so that  $T(\hat{\alpha} - 1)$  will dominate the behaviour of the statistics, since  $T(\hat{\alpha} - 1) \rightarrow -\infty$  as  $\phi \rightarrow \pi$ . On the other hand, as  $\phi \rightarrow 0$ , it follows that  $\hat{\alpha} \rightarrow 1$  and the behaviour of the statistics becomes dominated by  $s_{AR}^2$ .

An important point to note in this analysis is that as  $\phi \rightarrow 0$  the cyclical component  $(1 - 2 \cos(\phi)(1 + \frac{c_\phi}{T})L + (1 + \frac{c_\phi}{T})^2 L^2)$  becomes indistinguishable from an I(2) component, as also observed by Bierens (2001). Haldrup and Lildholdt (2002) provide a discussion of the limit and finite sample behaviour of the ADF and PP unit root tests when the data generating mechanism is an I(2) process. Their main conclusions regarding the PP test are consistent with the results observed in Figures 4 and 5 (they only considered kernel based long-run variance estimators in their analysis) when  $\phi$  is in the neighbourhood of zero. In particular, Haldrup and Lildholdt (2002) observe that the densities of the PP tests have only moderate concentration in the negative region; as a result, these tests will very unlikely reject the null hypothesis in favour of stationarity in the I(2) region.

## 5 Conclusions

In this paper we have shown that the Phillips and Perron (1988) [PP] unit root  $t$ - and normalised bias type test statistics have degenerate limiting null distributions in the presence of (near-) non-stationary cycles in the data, yielding tests with an asymptotic size of either one or unity, when (near-) non-stationary cycles are present. The PP tests which employ a kernel-based long run variance estimator will always have asymptotic size of unity in such cases, while the corresponding tests based on an autoregressive spectral density estimator of the long run variance will have size of either unity or zero depending on the frequency at which the non-stationary cycle occurs.

## References

- Allen, D.S. (1997). Filtering permanent cycles with complex unit roots. Working Paper 1997-001A, Federal Reserve Bank of St. Louis (<http://research.stlouisfed.org/wp/1997/97-001.pdf>).
- Berk, K.N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* 2, 389-502.
- Bierens, H.J. (2001). Complex unit roots and business cycles: are they real? *Econometric Theory* 17, 962-983.
- Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics* 108, 343-363.
- Chan, N.H., and C.Z. Wei (1988). Limiting distribution of least squares estimates of unstable autoregressive processes. *The Annals of Statistics* 16, 1, 367-401.
- Chang, Y., and J.Y. Park (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431-448.
- del Barrio Castro, T., P.M.M. Rodrigues and A.M.R. Taylor (2011). The impact of persistent cycles on zero frequency unit root tests. Working Paper 24/2011, Banco de Portugal
- Díaz-Emparanza, I. (2004). A note on the paper by H.J. Bierens: “Complex unit roots and business cycles: are the real?” *Econometric Theory* 20, 636-637.
- 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.
- Haldrup, N. and P. M. Lildholdt (2002). On the Robustness of Unit Root Tests in the Presence of Double Unit Roots. *Journal of Time Series Analysis* 23, 155-171.
- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press: Princeton.
- Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449-59.
- Nabeya, S. and P. Perron (1994). Local asymptotic distribution related to the AR(1) model with dependent errors. *Journal of Econometrics* 62(2), pp. 229-264.
- Newey, W.K., and K.D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-653.
- Perron P., and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435-463.
- Perron P., and Ng, S. (1998). An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* 14(05), pp. 560-603.
- Phillips, P.C.B. (1989). Partially Identified Econometric Models. *Econometric Theory* 5,2,, 181-240.
- Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277-301.
- Phillips, P.C.B., and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika* 75, 335-346.



Said, S.E., and D.A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-607.

Stock, J.H. (1999). A class of tests for integration and cointegration. Engle, R.F. and White, H. (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*. Oxford University Press: Oxford.

Tanaka, K. (1996). *Time Series Analysis, Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons: Chichester.

Taylor, A.M.R. (2003). Robust stationarity tests in seasonal time series processes. *Journal of Business and Economic Statistics*, 21 156-163.

## Appendix

For the purpose of analysing the impact of (near) integrated cycles on the limit distributions of the least-squares estimator of a first order AR model parameter and the PP zero frequency unit root tests discussed in section 2.2 it will prove useful to first consider a frequency specific orthogonal decomposition of  $x_t$ . CRT show that for a time series process  $\{x_t\}$  generated by (1) with  $\varphi_{0T} = 1 + c_0/T$  and  $\varphi_{\phi T} = 1 + c_\phi/T$ , and  $c_0 \leq 0$ ,  $c_\phi \leq 0$  and fixed, for any  $\phi \in (0, \pi)$  the following decomposition holds,

$$x_t = \delta_0 S_{0,c_0}(t) + \mathcal{C}_{\phi,t} + O_p(1) \quad (\text{A.1})$$

where  $\mathcal{C}_{\phi,t} = \frac{\delta_\phi^\alpha}{\sin \phi} \left\{ \sin[\phi(t+1)] S_{\phi,c_\phi}^\alpha(t) - \cos[\phi(t+1)] S_{\phi,c_\phi}^\beta(t) \right\} + \frac{\delta_\phi^\beta}{\sin \phi} \left\{ \cos[\phi(t+1)] S_{\phi,c_\phi}^\alpha(t) + \sin[\phi(t+1)] S_{\phi,c_\phi}^\beta(t) \right\}$ ,  $S_{0,c_0}(t) := \sum_{j=1}^t \varphi_{0T}^{t-j} \varepsilon_j$ ,  $S_{\phi,c_\phi}^\alpha(t) := \sum_{j=1}^t \varphi_{\phi T}^{t-j} \varepsilon_j \cos(j\phi)$ ,  $S_{\phi,c_\phi}^\beta(t) := \sum_{j=1}^t \varphi_{\phi T}^{t-j} \varepsilon_j \sin(j\phi)$ ,  $\delta_0 := 1/2(1 - \cos \phi)$ ,  $\delta_\phi^\alpha := (1 - (2 - \varphi_{0T}) \cos \phi) / 2(1 - \cos \phi)$ , and  $\delta_\phi^\beta := -\sin(\phi) \varphi_{0T} / 2(1 - \cos \phi)$ . To simplify notation, in what follows we define  $\delta_\phi^\alpha := \frac{1}{2} + o_p(T^{-1})$ , and  $\delta_\phi^\beta := -\sin(\phi) / 2(1 - \cos \phi) + o_p(T^{-1})$ .

The generalisation of the result in (A.1) to lags of  $x_t$  is now provided in Proposition A.1. This result will prove useful in what follows.

**Proposition A.1:** *Let the time series process  $\{x_t\}$  be generated by (1) with  $\varphi_{0T} := 1 + c_0/T$  and  $\varphi_{\phi T} := 1 + c_\phi/T$ , and  $c_0 \leq 0$ ,  $c_\phi \leq 0$  and fixed. Then, for any  $\phi \in (0, \pi)$ ,*

$$x_{t-k} = \delta_0 S_{0,c_0}(t-k) + \mathcal{C}_{\phi,t-k} + O_p(1), \quad k = 1, 2, 3 \quad (\text{A.2})$$

where  $\delta_0 := 1/2(1 - \cos \phi)$ ,  $\mathcal{C}_{\phi,t-k} := \Lambda_{\phi k} \Xi_\phi \mathbf{S}_{\phi,c_\phi}(t-k)$  with  $\Lambda_{\phi,1} := \text{diag}\{- (2 \sin \phi)^{-1}, - [2(1 - \cos \phi)]^{-1}\}$ ,  $\Lambda_{\phi,2} := \text{diag}\left\{ -\frac{(2 \cos \phi + 1)}{2 \sin \phi}, \frac{2 \cos \phi - 1}{2(\cos \phi - 1)} \right\}$ ,  $\Lambda_{\phi,3} := \text{diag}\left\{ -\frac{(4 \cos^2 \phi + 2 \cos \phi - 1)}{2 \sin \phi}, -\frac{(-4 \cos^2 \phi + 2 \cos \phi + 1)}{2(\cos \phi - 1)} \right\}$ ,  $\Xi_\phi := \begin{bmatrix} \sin[(t+1)\phi] & -\cos[(t+1)\phi] \\ \cos[(t+1)\phi] & \sin[(t+1)\phi] \end{bmatrix}$  and  $\mathbf{S}_{\phi,c_\phi}(t-k) := (S_{\phi,c_\phi}^\alpha(t-k), S_{\phi,c_\phi}^\beta(t-k))'$ .

**Proof:** If  $\{x_t\}$  is generated by (1) then using the approximation in CRT it follows that,

$$x_{t-k} = \frac{1}{2(1 - \cos \phi)} \sum_{j=1}^{t-k} \varphi_{0T}^{t-j} \varepsilon_j + \frac{1 - 2 \cos(\phi) + \varphi_{0T} L}{2(1 - \cos \phi) \sin \phi} \sum_{j=1}^{t-k} \sin[\phi(t-k+1-j)] \varphi_{\phi T}^{t-j} \varepsilon_j. \quad (\text{A.3})$$

The first term on the right-hand side of (A.3) is immediately seen to correspond to the zero frequency component,  $\delta_0 S_{0,c_0}(t-k)$ . Turning to the second term, observe that

$$\begin{aligned}
\sum_{j=1}^{t-k} \sin[\phi(t-k+1-j)] \varphi_{\phi T}^{t-j} \varepsilon_j &= \sin[(t-k+1)\phi] \sum_{j=1}^{t-k} \cos(j\phi) \varphi_{\phi T}^{t-j} \varepsilon_j \\
&\quad - \cos[(t-k+1)\phi] \sum_{j=1}^{t-k} \sin(j\phi) \varphi_{\phi T}^{t-j} \varepsilon_j \\
&= \sin[(t-k+1)\phi] S_{\phi,c_\phi}^\alpha(t-k) - \cos[(t-k+1)\phi] S_{\phi,c_\phi}^\beta(t-k).
\end{aligned} \tag{A.4}$$

Consider first the case of  $k=1$ . From (A.3) and (A.4) we obtain that,

$$\begin{aligned}
\mathcal{C}_{\phi,t-1} &= \frac{1-2\cos\phi}{2(1-\cos\phi)} \frac{1}{\sin\phi} \left\{ \sin(\phi t) S_{\phi,c_\phi}^\alpha(t-1) - \cos(\phi t) S_{\phi,c_\phi}^\beta(t-1) \right\} \\
&\quad + \frac{1}{2(1-\cos\phi)} \frac{1}{\sin\phi} \left\{ \sin[\phi(t-1)] S_{\phi,c_\phi}^\alpha(t-2) - \cos[\phi(t-1)] S_{\phi,c_\phi}^\beta(t-2) \right\}.
\end{aligned}$$

Furthermore, using conventional trigonometric identities this can be simplified to,

$$\begin{aligned}
\mathcal{C}_{\phi,t-1} &= -\frac{1}{2\sin\phi} \left\{ \sin[\phi(t+1)] S_\alpha(t-1) - \cos[\phi(t+1)] S_\beta(t-1) \right\} \\
&\quad + \frac{1}{2(\cos\phi-1)} \left\{ \cos[\phi(t+1)] S_\alpha(t-1) + \sin[\phi(t+1)] S_\beta(t-1) \right\} + O_p(1)
\end{aligned}$$

or equivalently as  $\mathcal{C}_{\phi,t-1} = \Lambda_{\phi 1} \Xi_\phi \mathbf{S}_{\phi,c_\phi}(t-1)$ , as defined in Proposition A.1. The results for  $\mathcal{C}_{\phi,t-2}$  and  $\mathcal{C}_{\phi,t-3}$  are obtained along similar lines. ■

**Proof of Proposition 3.1:** Note from (1) with  $\vartheta_\phi(L) = (1-2\cos(\phi)\varphi_{\phi T}L + \varphi_{\phi T}^2 L^2)$ ,  $\varphi_{0T} = (1+c_0/T)$  and  $\varphi_{\phi T} = (1+c_\phi/T)$  that,

$$\begin{aligned}
x_t &= \left[ \left(1 + \frac{c_0}{T}\right) + 2(\cos\phi) \left(1 + \frac{c_\phi}{T}\right) \right] x_{t-1} - \left[ \left(1 + \frac{c_\phi}{T}\right)^2 + 2(\cos\phi) \left(1 + \frac{c_\phi}{T}\right) \left(1 + \frac{c_0}{T}\right) \right] x_{t-2} \\
&\quad + \left(1 + \frac{c_\phi}{T}\right)^2 x_{t-3} + \varepsilon_t
\end{aligned} \tag{A.5}$$

and therefore  $\hat{\alpha} := \left(\sum_{t=1}^T x_{t-1}^2\right)^{-1} \sum_{t=1}^T x_{t-1} x_t$  can be written as a linear combination of three components: the parameters of the process and the first and second order autocorrelations, *i.e.*,

$$\begin{aligned}
\hat{\alpha} &= \left[ \left(1 + \frac{c_0}{T}\right) + 2(\cos\phi) \left(1 + \frac{c_\phi}{T}\right) \right] - \\
&\quad - \left[ \left(1 + \frac{c_\phi}{T}\right)^2 + 2(\cos\phi) \left(1 + \frac{c_\phi}{T}\right) \left(1 + \frac{c_0}{T}\right) \right] \frac{\sum_{t=1}^T x_{t-1} x_{t-2}}{\sum_{t=1}^T x_{t-1}^2} \\
&\quad + \left(1 + \frac{c_\phi}{T}\right)^2 \frac{\sum_{t=1}^T x_{t-1} x_{t-3}}{\sum_{t=1}^T x_{t-1}^2} + \frac{\sum_{t=1}^T x_{t-1} \varepsilon_t}{\sum_{t=1}^T x_{t-1}^2}.
\end{aligned} \tag{A.6}$$

Using results from Chan and Wei (1988), it is straightforward to show that:

$$\begin{aligned}
\frac{1}{T^2} \sum_{t=1}^T x_{t-1} x_{t-1-k} &= \frac{1}{T^2} \sum_{t=1}^T (\mathcal{C}_{0,t-1} + \mathcal{C}_{\phi,t-1}) (\mathcal{C}_{0,t-1-k} + \mathcal{C}_{\phi,t-1-k}) + o_p(1) \\
\Rightarrow \frac{\sigma^2 \int_0^1 W_{0,c_0}^2(r) dr}{4(1-\cos(\phi))^2} &+ \frac{\sigma^2}{4} \left( \frac{\cos(k\phi)}{2(1-\cos(\phi)) \sin^2(\phi)} \right) \int_0^1 \mathcal{W}_{\phi,c_\phi}^{(\alpha\beta)^2} dr, \quad k=0,1,2
\end{aligned} \tag{A.7}$$

where we have defined  $\mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} := [W_{\phi, c_\phi}^\alpha(r)]^2 + [W_{\phi, c_\phi}^\beta(r)]^2$ .

Noting that  $\frac{T^{-2} \sum_{t=1}^T x_{t-1} \varepsilon_t}{T^{-2} \sum_{t=1}^T x_{t-1}^2}$  in (A.6) is  $o_p(1)$ , and using (A.7) it then follows that:

$$\rho_k := \frac{T^{-2} \sum_{t=1}^T x_{t-1} x_{t-1-k}}{T^{-2} \sum_{t=1}^T x_{t-1}^2} \Rightarrow 1 - \frac{(1 - \cos(k\phi)) \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{2 \frac{\sin^2 \phi}{(1 - \cos \phi)} \int_0^1 W_{0, c_0}^2(r)^2 dr + \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}, \quad k = 1, 2. \quad (\text{A.8})$$

Noting that (A.6) can be written as  $\hat{\alpha} = [1 + 2(\cos \phi)] - [1 + 2(\cos \phi)] \rho_1 + \rho_2 + o_p(1)$ , we then have, from the joint convergence result in (A.8), that

$$\begin{aligned} \hat{\alpha} &\Rightarrow (1 + 2 \cos \phi) \frac{(1 - \cos \phi) \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{2 \frac{\sin^2 \phi}{(1 - \cos \phi)} \int_0^1 W_{0, c_0}^2(r)^2 dr + \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr} \\ &\quad + 1 - \frac{(1 - \cos 2\phi) \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{2 \frac{\sin^2 \phi}{(1 - \cos \phi)} \int_0^1 W_{0, c_0}^2(r)^2 dr + \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr} \\ &= 1 + \frac{[(1 + 2 \cos \phi)(1 - \cos \phi) - (1 - \cos 2\phi)] \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{2 \frac{\sin^2 \phi}{(1 - \cos \phi)} \int_0^1 W_{0, c_0}^2(r)^2 dr + \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}. \end{aligned}$$

Noting that  $\cos 2\phi = 2 \cos \phi - 1$  and that  $(\sin^2 \phi) / (1 - \cos \phi) = (1 + \cos \phi)$  this simplifies to the result stated in (6). ■

**Proof of Theorem 3.1:** Consider first the case where kernel based long-run variance estimators are used in the PP test statistics. First from the proof of Theorem 2 in Taylor (2003) it follows that  $\hat{\lambda}^2 = s_{WA}^2 = O_p(mT)$ . Next consider  $s^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2$ . Noting that

$$s^2 = \frac{1}{T} \left( \sum_{t=1}^T \Delta x_t^2 - (\hat{\alpha} - 1) \sum_{t=1}^T x_{t-1} \Delta x_t \right),$$

then using (6), the result from Chan and Wei (1988) that  $T^{-2} \sum_{t=1}^T \Delta x_t^2 \Rightarrow \frac{\sigma^2 \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{4 \sin^2 \phi}$ , and finally that based on Chan and Wei (1988), and the corresponding result that  $T^{-2} \sum_{t=1}^T x_{t-1} \Delta x_t \Rightarrow -\frac{\sigma^2 \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{8 \sin^2 \phi}$ , we obtain that,

$$T^{-1} s^2 \Rightarrow \frac{\sigma^2 \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{4 \sin^2 \phi} \left( \frac{1}{2} + \frac{(1 - \cos \phi) \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr}{2(1 + \cos \phi) \int_0^1 W_{0, c_0}^2(r)^2 dr + \int_0^1 \mathcal{W}_{\phi, c_\phi}^{(\alpha\beta)^2} dr} \right).$$

As a result,  $s^2$  is of  $O_p(T)$ .

For the  $Z_\alpha$  statistic we therefore have that

$$\begin{aligned} Z_\alpha &= T(\hat{\alpha} - 1) - \frac{1}{2} \left( \hat{\lambda}^2 - s^2 \right) \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1} \\ &= -O_p^+(T) - \frac{1}{2} (O_p^+(mT) - O_p^+(T)) / O_p^+(1) \end{aligned}$$

which is therefore seen to diverge to  $-\infty$  at rate  $mT$  as  $T \rightarrow \infty$ , where we have used the result that  $(\hat{\alpha} - 1)$  is always negative (see Remark 1). Regarding the  $Z_t$  statistic, following Ng and Perron (1996), we have that  $Z_t = Z_\alpha \left( \frac{1}{\hat{\lambda}^2 T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{1/2}$  so that,

$$Z_t = -O_p^+(T)/O_p^+ \left( \sqrt{mT} \right) - \frac{1}{2} (O_p^+(mT) - O_p^+(T))/O_p^+ \left( \sqrt{mT} \right)$$

which again diverges to  $-\infty$  as  $T \rightarrow \infty$ .

For the case where a spectral autoregressive long-run variance estimator is used in constructing the test statistics in (3), then since  $\hat{\lambda}^2 = s_{AR}^2 = O_p(1)$ , and  $s^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 = O_p(T)$ , it follows that,

$$\begin{aligned} Z_\alpha &= T(\hat{\alpha} - 1) - \frac{1}{2} \left( \hat{\lambda}^2 - s^2 \right) \left( \frac{1}{T^2} \sum_{t=1}^T x_{t-1}^2 \right)^{-1} \\ &= -TO_p^+(1) - \frac{1}{2} (O_p^+(1) - O_p^+(T))/O_p^+(1) \\ &= \pm O_p^+(T) \end{aligned}$$

and, consequently, diverges to either  $+\infty$  or  $-\infty$  as  $T \rightarrow \infty$ . Similarly for  $Z_t$ , we have that

$$\begin{aligned} Z_t &= -O_p^+(T)/O_p^+(1) - \frac{1}{2} (O_p^+(1) - O_p^+(T))/O_p^+(1) \\ &= \pm O_p^+(T) \end{aligned}$$

which again diverges to either  $+\infty$  or  $-\infty$  as  $T \rightarrow \infty$ . ■

Figure 1: Average AR(1) Parameter Estimate. DGP (8) with  $c_0 = 0$ .

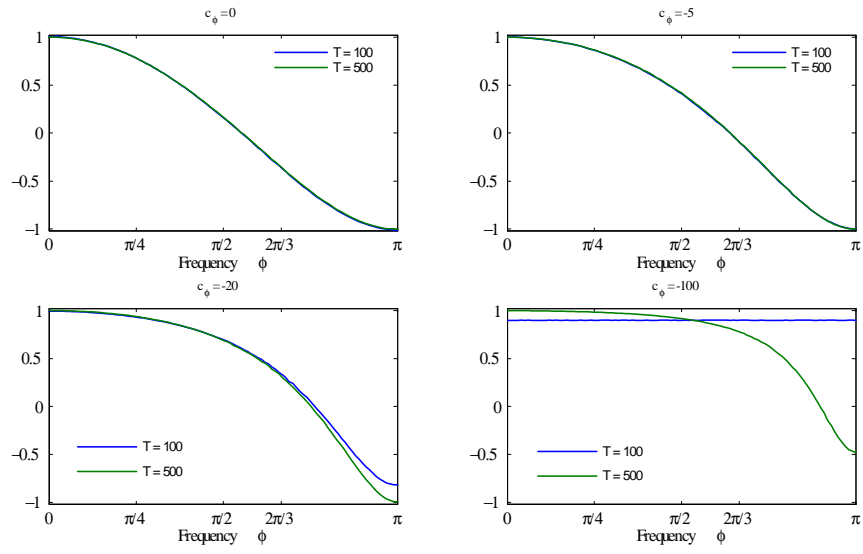
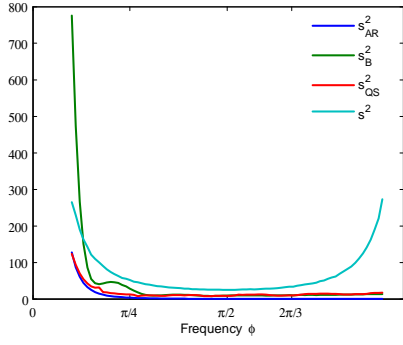
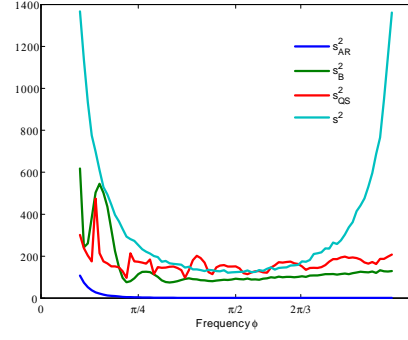


Figure 2: Average Long Run Variance Estimates. DGP (8) with  $c_0 = 0$

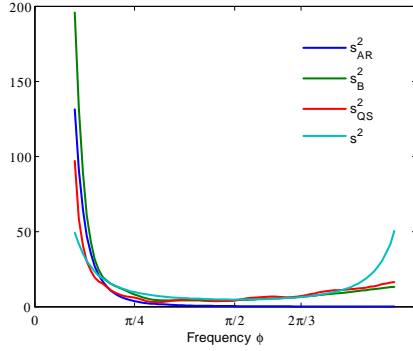
$T = 100, c_\phi = 0$



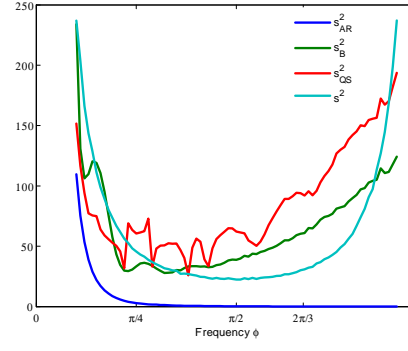
$T = 500, c_\phi = 0$



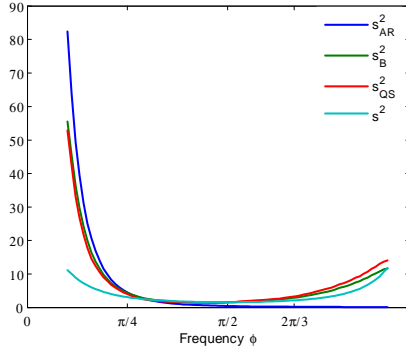
$T = 100, c_\phi = -5$



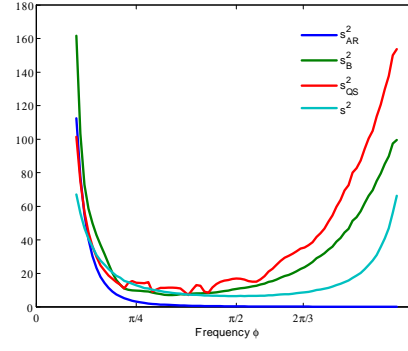
$T = 500, c_\phi = -5$



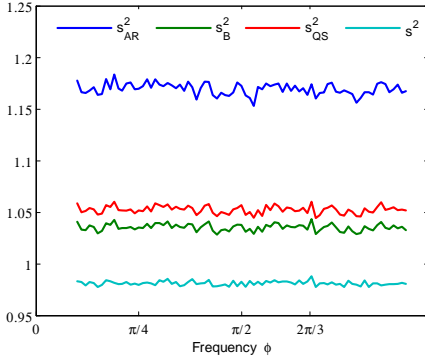
$T = 100, c_\phi = -20$



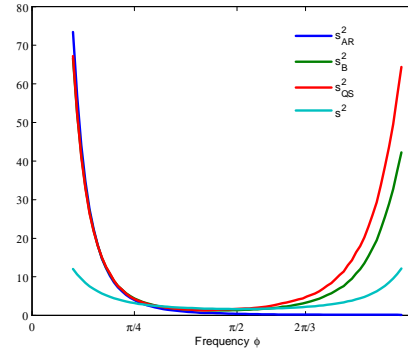
$T = 500, c_\phi = -20$



$T = 100, c_\phi = -100$



$T = 500, c_\phi = -100$



Note:  $\kappa$  is the constant used to determine the kernel window, i.e. for the Bartlett,  $m = \lceil (\kappa T/100)^{2/9} \rceil$  and for the Quadratic Spectral kernel  $m = \lceil \kappa(T/100)^{2/25} \rceil$ . The results in this table are computed with  $\kappa = 12$ .

Figure 3: Empirical Rejection Frequencies of the PP Tests using an Autoregressive Spectral Density Estimator of the Long Run Variance. DGP (8) with  $c_0 = 0$ .

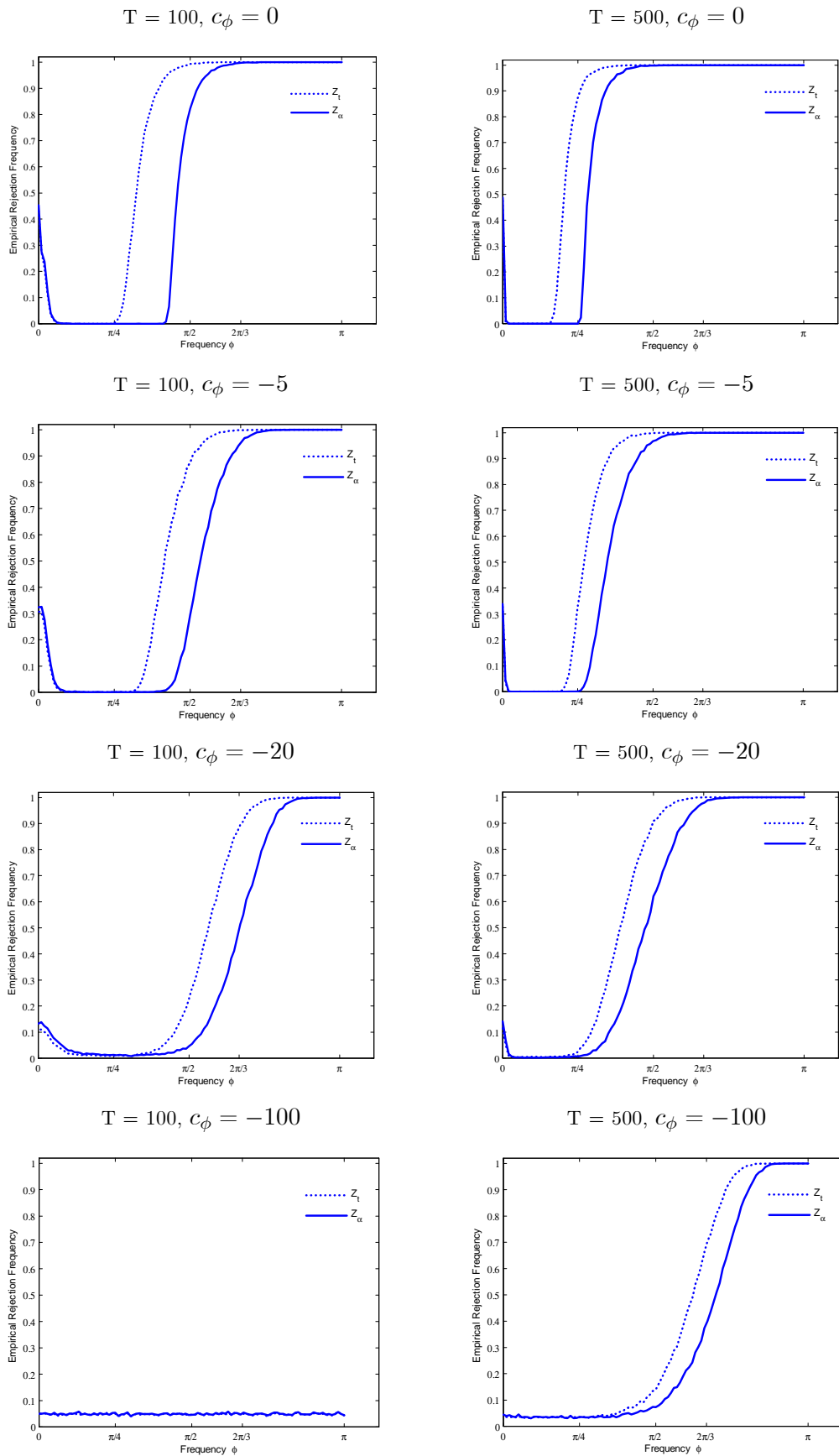


Figure 4: Empirical Rejection Frequencies of the PP Tests using a Kernel-based Long Run Variance Estimator (Bartlett). DGP (8) with  $c_0 = 0$ .

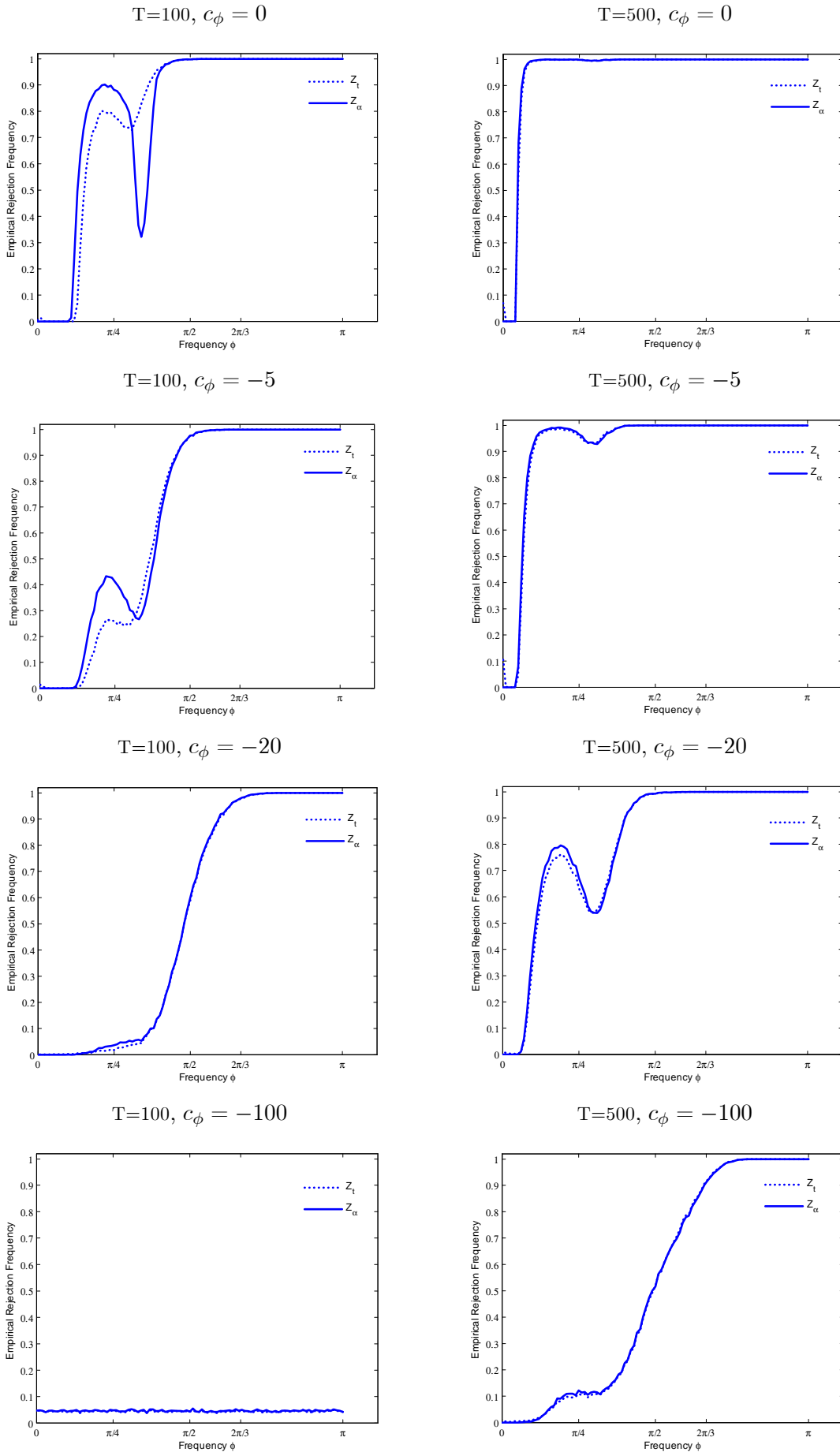




Figure 5: Empirical Rejection Frequencies of the PP Tests using a Kernel-based Long Run Variance Estimator (Quadratic Spectral). DGP (8) with  $c_0 = 0$ .

