The Phillips unit root tests for polynomials of integrated processes

Oliver Stypka, Martin Wagner

SFB
823

The Phillips Unit Root Tests for Polynomials of Integrated Processes

Oliver Stypka
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

Martin Wagner
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany
&
Bank of Slovenia
Ljubljana, Slovenia
&
Institute for Advanced Studies
Vienna, Austria

Abstract

We show that the Phillips (1987) unit root tests have nuisance parameter free limiting distributions when applied to polynomials of integrated processes driven by linear process errors. This substantially generalizes a similar result of Wagner (2012) allowing only for serially uncorrelated errors. The result is based on novel kernel weighted sum limit results involving powers of integrated processes. These results allow us also consider additional modifications of the Phillips (1987) tests applicable to polynomials of integrated processes.

1 Introduction

This paper generalizes the results of an earlier paper of the second author with the same title analyzing the asymptotic behavior of the Phillips (1987) unit root tests applied to polynomials of integrated processes. We show that the Phillips (1987) unit root tests have nuisance parameter free limiting null distributions *also* when applied to polynomials of integrated processes driven by linear process errors and not only in case of serially uncorrelated errors as discussed in Wagner (2012).[1]

---

[1]Wagner (2012) discusses in detail where and why the issue of applying unit root tests to polynomials of integrated processes arises and for sake of brevity we do not repeat this discussion here. Some literatures where the problem arises are the environmental Kuznets curve literature (see, e.g., Grossman and Krueger, 1993, Wagner, 2015), the intensity of use literature (see, e.g., Labson and Crompton, 1993) or the exchange rate target zone literature (see, e.g., Darvas, 2008, Svensson 1992).

The generalization of the result is based on novel limit results for kernel weighted sums involving powers of nonstationary processes derived in the related context of cointegrating polynomial regressions (see Wagner and Hong, 2016) in Stypka *et al.* (2017).

Utilizing the kernel results allows us to consider also two additional modifications – compared to the original Phillips (1987) modifications – of the least squares estimate of the first order serial correlation coefficient and its t-statistic. We thus consider in total six test statistics. The two modifications remove an "additive bias term" present in the limiting distributions of the Phillips (1987) statistics in case of a polynomial degree k larger than one. One of the transformations makes use of Itô's Lemma and simplifies the numerator of the test statistics to the k-th power of a chi-squared random variable with one degree of freedom.

2 The Test Statistics and Their Limiting Distributions

We consider a polynomial transformation, x_t, of an integrated process y_t, i.e.

$$x_t \quad := \quad y_t^k + a_1 y_t^{k-1} + \cdots + a_{k-1} y_t, \tag{1}$$

where

$$y_t \quad = \quad y_{t-1} + u_t, \tag{2}$$

with the details concerning the stationary process u_t given in the following assumption.

Assumption 1 *The process u_t is a linear process, i.e. $u_t = \sum_{j=0}^{\infty} \varepsilon_t$, with $\sum_{j=1}^{\infty} j|c_j| < \infty$, $\sum_{j=0}^{\infty} c_j \neq 0$ and ε_t is a zero mean i.i.d. process with $\mathbb{E}|\varepsilon_0|^p < \infty$ for some $p > \max(8, 4/(1-2b))$ with $0 < b < 1/3$.*

The leading coefficient in (1) can be set equal to one without loss of generality by redefining the variance of ε_t appropriately. For later use we define the variance of u_t as $\sigma_u^2 := \mathbb{E}(u_t^2)$, its long-run variance $\omega_u := \sum_{j=-\infty}^{\infty} \mathbb{E}(u_t u_{t-j})$ and its half long-run variance $\lambda_u := \sum_{j=1}^{\infty} \mathbb{E}(u_t u_{t-j})$.

The objects of interest in this paper are the asymptotic behavior of the estimated first order serial correlation coefficient, its t-statistic for the null hypothesis that the serial correlation coefficient is equal to one and of the Phillips (1987)-type unit root tests, based on modifications of the coefficient estimate and its t-statistic. All considered quantities are based on the regression

$$x_t \quad = \quad \rho x_{t-1} + v_t, \qquad t = 2, \dots, T. \tag{3}$$

2

We denote the OLS coefficient estimator of ρ from (3) as

$$\hat{\rho} \quad := \quad \frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=2}^{T} x_{t-1}^2} \tag{4}$$

and the t-statistic for the null hypothesis that $\rho = 1$ as

$$t_\rho \quad := \quad \frac{\hat{\rho} - 1}{\sqrt{\hat{\sigma}_v^2 \left(\sum_{t=2}^{T} x_{t-1}^2 \right)^{-1}}}, \tag{5}$$

with $\hat{\sigma}_v^2 := \frac{1}{T} \sum_{t=2}^{T} \hat{v}_t^2$ and $\hat{v}_t := x_t - \hat{\rho} x_{t-1}$. The Phillips (1987) test statistics are given by

$$Z_\rho \quad := \quad T(\hat{\rho} - 1) - \frac{\hat{\lambda}_v}{\frac{1}{T^2} \sum_{t=2}^{T} x_{t-1}^2}, \tag{6}$$

$$Z_t \quad := \quad \sqrt{\frac{\hat{\sigma}_v^2}{\hat{\omega}_v}} t_\rho - \frac{\hat{\lambda}_v}{\sqrt{\hat{\omega}_v \frac{1}{T^2} \sum_{t=2}^{T} x_{t-1}^2}}, \tag{7}$$

with $\hat{\omega}_v := \hat{\sigma}_v^2 + 2 \sum_{h=1}^{M_T} \mathcal{K}\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \hat{v}_t \hat{v}_{t-j}$ and $\hat{\lambda}_v := \frac{1}{2}(\hat{\omega}_v - \hat{\sigma}_v^2)$, where $\mathcal{K}(\cdot)$ is a kernel function and M_T a bandwidth (specified below in Assumption 2).

Because v_t is not a stationary process for $k > 1$, it is clear that $\hat{\sigma}_v^2$, $\hat{\omega}_v$ and $\hat{\lambda}_v$ are not estimates of a variance, long-run variance and half long-run variance. Stypka *et al.* (2017, Theorem 1 and Corollary 1) show in the related context of cointegrating polynomial regressions that this type of kernel weighted sum involving powers of integrated processes (contained in $\hat{v}_t$) converges weakly to limits involving integrals of Brownian motions.[2] These results are key for understanding the asymptotic behavior of the Phillips (1987) unit root tests applied to polynomials of integrated processes. Since our adapted results follow using similar arguments as Stypka *et al.* (2017), we posit the same assumptions on kernel and bandwidth for the kernel weighted sums as in that paper.

Assumption 2 *The kernel function $\mathcal{K}(\cdot)$ satisfies: $\mathcal{K}(0) = 1$, $\mathcal{K}(\cdot)$ is continuous at zero, $\bar{\mathcal{K}}(0) := \sup_{x \geq 0} |\mathcal{K}(x)| < \infty$ and $\int_0^\infty \bar{\mathcal{K}}(x) dx < \infty$, where $\bar{\mathcal{K}}(x) = \sup_{y \geq x} |\mathcal{K}(y)|$. The bandwidth $M_T \to \infty$ fulfills $M_T = O(T^b)$, with the same parameter b as in Assumption 1.*

---

[2] The corresponding results in that paper are conceptually similar to what we need here, but not exactly the same. Stypka *et al.* (2017) consider *static* cointegrating regressions involving powers of integrated processes as regressors, whereas we consider here a *dynamic* time series regression involving polynomials of integrated processes as both dependent variable and (with lag one) as explanatory variable. The structure of the proofs is, however, quite similar with full details available upon request.

We are now ready to state the first result of the paper:

Proposition 1 *Under Assumptions 1 and 2 it holds for $T \to \infty$ that*

$$Z_\rho \Rightarrow \frac{k \int_0^1 W(r)^{2k-1} dW(r) + \binom{k}{2} \int_0^1 W(r)^{2(k-1)} dr}{\int_0^1 W(r)^{2k} dr}, \tag{8}$$

$$Z_t \Rightarrow \frac{\int_0^1 W(r)^{2k-1} dW(r) + \frac{k-1}{2} \int_0^1 W(r)^{2(k-1)} dr}{\sqrt{\int_0^1 W(r)^{2(k-1)} dr \int_0^1 W(r)^{2k} dr}}, \tag{9}$$

with a standard Brownian motion $W(r)$.

Proposition 1 generalizes Wagner (2012, Corollary 1) from the case of u_t being serially uncorrelated to the more general case of u_t being a linear process as specified in Assumption 1. The result shows that the Phillips (1987) unit root tests can be applied to polynomials of integrated processes using the original formulation of Phillips (1987), when the test decisions are based on the appropriate critical values simulated from the asymptotic distributions given in Proposition 1, which have been simulated and tabulated already in Wagner (2012, Table 1) for $k = 1, 2, 3$.

The kernel weighted sum limit results allow us to consider also other modifications than the original Phillips (1987) modifications that also lead to nuisance parameter free limiting distributions. We consider two variants: First, we consider modifications of the coefficient and the t-statistic that remove the additive bias terms, i.e., $\binom{k}{2} \int_0^1 W(r)^{2(k-1)} dr$ and $\frac{k-1}{2} \int_0^1 W(r)^{2(k-1)} dr$, from the numerators of the limiting distributions given in Proposition 1. Second, we consider modifications to the test statistics that allow us to use Itô's Lemma (see, e.g., Theorem 3.3., p. 149 in Karatzas

4

and Shreve, 1991) to simplify the numerators of the test statistics further to $W(1)^{2k}$. This leads to the following four additional test statistics:[3]

$$Z_\rho^* \quad := \quad T(\hat\rho - 1) - \frac{2k\hat\lambda_v + (k-1)\hat\omega_v}{2k\frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2} = Z_\rho - \frac{(k-1)\hat\omega_v}{2k\frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2}, \tag{10}$$

$$Z_t^* \quad := \quad \sqrt{\frac{\hat\sigma_v^2}{\hat\omega_v}}t_\rho - \frac{2k\hat\lambda_v + (k-1)\hat\omega_v}{2k\sqrt{\hat\omega_v \frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2}} = Z_t - \frac{(k-1)\hat\omega_v}{2k\sqrt{\hat\omega_v \frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2}}, \tag{11}$$

$$Z_\rho^{**} \quad := \quad T(\hat\rho - 1) + \frac{\hat\sigma_v^2}{2\frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2}, \tag{12}$$

$$Z_t^{**} \quad := \quad \sqrt{\frac{\hat\sigma_v^2}{\hat\omega_v}}t_\rho + \frac{\hat\sigma_v^2}{2\sqrt{\hat\omega_v \frac{1}{T^2}\sum_{t=2}^T x_{t-1}^2}}. \tag{13}$$

Corollary 1 *Under Assumptions 1 and 2 it holds for $T \to \infty$ that*

$$Z_\rho^* \quad \Rightarrow \quad \frac{k\int_0^1 W(r)^{2k-1}dW(r)}{\int_0^1 W(r)^{2k}dr}, \tag{14}$$

$$Z_t^* \quad \Rightarrow \quad \frac{\int_0^1 W(r)^{2k-1}dW(r)}{\sqrt{\int_0^1 W(r)^{2(k-1)}dr \int_0^1 W(r)^{2k}dr}}, \tag{15}$$

$$Z_\rho^{**} \quad \Rightarrow \quad \frac{W(1)^{2k}}{2\int_0^1 W(r)^{2k}dr}, \tag{16}$$

$$Z_t^{**} \quad \Rightarrow \quad \frac{W(1)^{2k}}{2k\sqrt{\int_0^1 W(r)^{2(k-1)}dr \int_0^1 W(r)^{2k}dr}}. \tag{17}$$

Note that in case $k = 1$ it holds by construction that $Z_\rho^* = Z_\rho$ and $Z_t^* = Z_t$. Also, in case $k = 1$, the Phillips (1987) test statistics can immediately be rewritten to have $W(1)^2$ in the numerator, since, e.g., $Z_\rho = \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2dr} = \frac{W(1)^2-1}{2\int_0^1 W(r)^2dr}$. The corresponding correction term, in case $k = 1$, just adds asymptotically $\frac{1}{2\int_0^1 W(r)^2dr}$. This idea is extended here to $k > 1$ for Z_ρ^{**} and Z_t^{**}. By definition, of course, $W(1)^{2k}$ is the k-th power of a chi-squared random variable with one degree of freedom.

The limiting distributions given in Proposition 1 and Corollary 1 can be simulated and tabulated. Table 1 and Figure 1 display the results for $k = 1, 2, 3$, based on 50,000 replications using standard normally distributed time series of length 1,000 to approximate the functionals of Brownian

---

[3]In our context Itô's Lemma results in $W(1)^{2k} = 2k\int_0^1 W(r)^{2k-1}dW(r) + k(2k-1)\int_0^1 W(r)^{2(k-1)}dr$. Note also that the modification of the OLS estimate and the t-statistic used in Z_ρ^{**} and Z_t^{**} has been used in a different context, for $k = 1$, to develop fixed-b inference for the Phillips and Perron (1988) unit root tests already in Vogelsang and Wagner (2013, p. 617). It has not been used there to consider different unit root tests, however.

k	0.010	0.025	0.050	0.500	0.950	0.975	0.990	Mean	Std.Dev.
Panel A: Coefficient Statistic Z_ρ									
1	-13.698	-10.623	-8.239	-0.885	1.282	1.613	2.023	-1.816	3.204
2	-21.977	-17.163	-13.574	-2.236	2.870	4.233	6.413	-3.369	5.351
3	-36.523	-28.565	-22.674	-4.509	4.174	6.935	11.603	-6.245	8.800
Panel B: t-Statistic Z_t									
1	-2.588	-2.262	-1.965	-0.513	1.285	1.635	2.064	-0.430	0.988
2	-3.350	-2.931	-2.590	-0.946	1.781	2.522	3.434	-0.736	1.343
3	-4.445	-3.890	-3.405	-1.396	1.857	2.951	4.416	-1.176	1.657
Panel C: Coefficient Statistic Z_ρ^*									
1	-13.698	-10.623	-8.239	-0.885	1.282	1.613	2.023	-1.816	3.204
2	-33.215	-26.192	-20.696	-3.788	2.038	3.221	5.089	-5.815	7.559
3	-60.688	-47.690	-38.109	-8.249	2.213	4.638	8.609	-11.805	13.476
Panel D: t-Statistic Z_t^*									
1	-2.588	-2.262	-1.965	-0.513	1.285	1.635	2.064	-0.430	0.988
2	-4.998	-4.389	-3.900	-1.551	1.359	2.121	3.036	-1.445	1.598
3	-7.193	-6.335	-5.635	-2.494	1.120	2.194	3.629	-2.422	2.068
Panel E: Coefficient Statistic Z_ρ^{**}									
1	0.001	0.003	0.013	0.779	2.786	3.420	4.235	0.996	0.937
2	0.000	0.000	0.000	0.657	5.691	7.655	10.533	1.487	2.220
3	0.000	0.000	0.000	0.450	8.791	12.478	18.077	1.971	3.790
Panel F: t-Statistic Z_t^{**}									
1	0.000	0.001	0.005	0.445	1.824	2.133	2.511	0.620	0.604
2	0.000	0.000	0.000	0.267	2.674	3.373	4.301	0.674	0.959
3	0.000	0.000	0.000	0.136	3.104	4.161	5.533	0.671	1.192

Table 1: Simulated percentiles, mean and standard deviation of the limiting distributions given in Proposition 1 and Corollary 1 for $k = 1, 2, 3$.

motions. By definition, the first two panels displaying the values for Z_ρ and Z_t, correspond to Wagner (2012, Table 1).[4]

The distributions of Z_ρ, Z_t, Z_ρ^* and Z_t^* shift to the left with increasing variances as k increases. This behavior is less pronounced for the distributions of Z_ρ^{**} and Z_t^{**}. As can be seen in Figure 1 the fitted densities corresponding to Z_ρ^{**} and Z_t^{**} are extremely steep in the vicinity of zero, which makes the simulation of small quantiles like 1% or 5% typically used for testing more imprecise than for the other distributions. This will be seen to detrimentally effect the performance of these two tests below.

Given the different variants of Phillips-type tests, the next question is which one to use. We address this question by plotting local asymptotic power (LAP) type curves. To be precise, we simulate data according to

$$y_t = \left(1 - \frac{c}{T}\right) y_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T, \tag{18}$$

where $y_0 = 0$, the ε_t are i.i.d. standard normally distributed, $T = 1000$ and we consider a grid of 41 values of $c \in [0, 80]$ with mesh two. For each value of c we calculate 50,000 test statistics. The corresponding empirical rejection probabilities are then plotted as a function of c and we refer

---

[4]`MATLAB` code to simulate these distributions is available upon request. The densities plotted in Figure 1 are estimated using Gaussian kernels with bandwidth according to the rule of thumb of Silverman (1986).
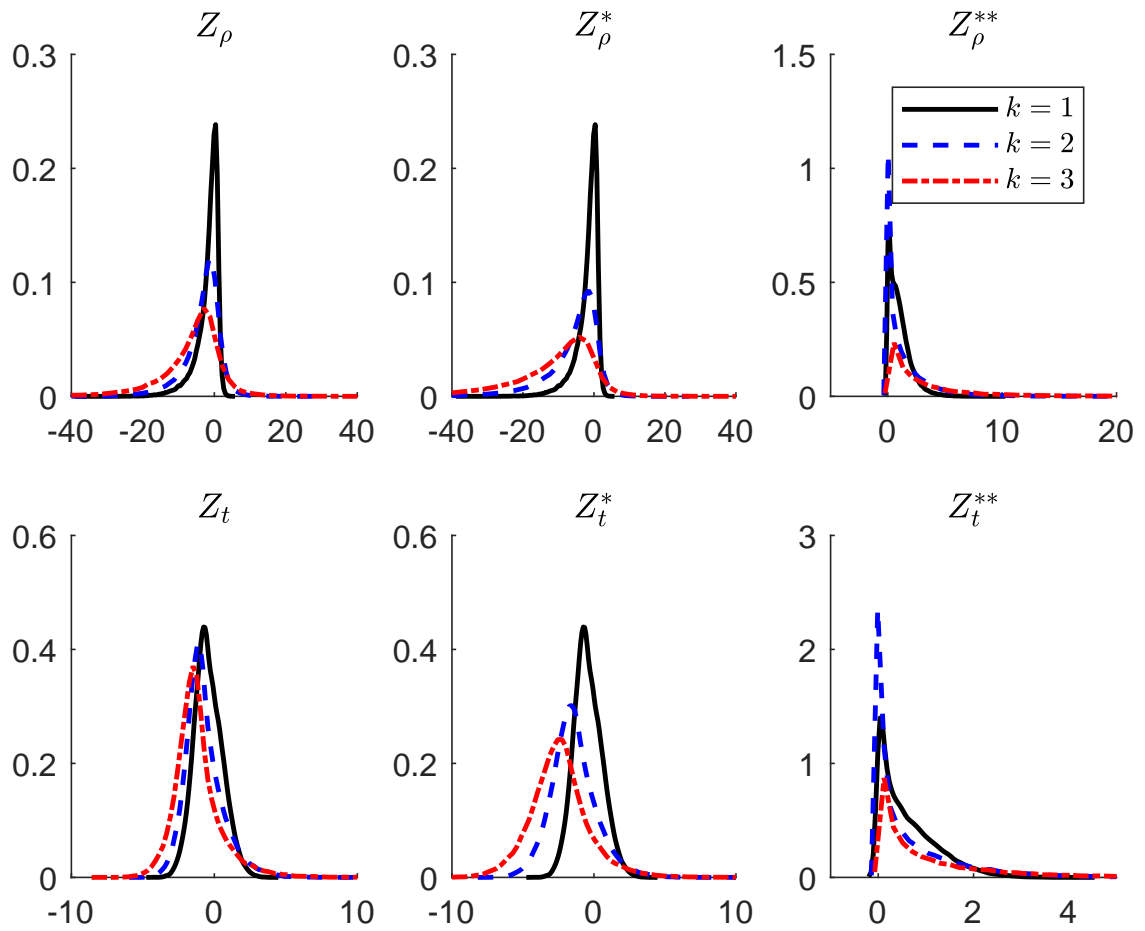
Figure 1: Densities corresponding to the limiting distributions given in Proposition 1 and Corollary 1. The upper graphs display the densities corresponding to the coefficient statistics and the lower graphs display the densities corresponding to the t-statistics.

to these results as "empirical" LAP. In all simulations long-run and half-long run variances are estimated using the Bartlett kernel and the Andrews (1991) bandwidth rule.

A process y_t as given in (18) is called near integrated by, e.g., Phillips (1987) when $c > 0$. Considering near integrated alternatives is the basis for local asymptotic power analysis in unit root and cointegration analysis. Here we just perform this type of LAP curve simulations without providing fully fledged formal derivations of local asymptotic power curves. Performing these derivations in full detail is beyond the scope of this paper as they would require one to generalize a number of results from the integrated to the near integrated case. Back-of-the-envelope calculations indicate that *usual* local asymptotic power results involving Ornstein-Uhlenbeck in place of Wiener processes can be obtained, and the corresponding results are displayed in Figure 3 in the Appendix.[5]

Figure 2 carries two messages. First, empirical LAP is virtually identical and marginally decreasing with increasing k for the original Phillips statistics Z_ρ and Z_t and the Z_ρ^*- and Z_t^*-modifications. Empirical LAP is very low – in fact lower than size for $k = 2, 3$ – for Z_ρ^{**} and Z_t^{**}. This effect is partly driven by the shape of the tests' distributions discussed above that render decisions very sensitive. To assess the uncertainty introduced by simulating LAP empirically it is informative to consider "analytical" LAP as displayed in Figure 3 in the Appendix, based on the mentioned back-of-the-envelope calculations. This figure shows a "more typical" LAP shape also for Z_ρ^{**} and Z_t^{**}. Nevertheless, also LAP based on asymptotic results shows that these two tests have distinctly lower power than the other four tests.

3 A Small Simulation Assessment

To assess the finite sample performance of the different test variants we simulate data generated according to:

$$y_t \;=\; y_{t-1} + u_t, \tag{19}$$

$$u_t \;=\; \gamma u_{t-1} + \varepsilon_t, \tag{20}$$

---

[5]In a bit more detail, the following results need to be generalized from the integrated to the near integrated case: First, the functional central limit theorem involving powers of integrated processes and the stationary error term, $T^{-\frac{k+1}{2}} \sum_{t=2}^T y_{t-1}^k u_t \Rightarrow \int_0^1 B_u(r)^k dB_u(r) + \lambda_u k \int_0^1 B_u(r)^{k-1} dr$, needs to be extended to the case when y_t is near integrated. The corresponding limit is key for the limit of $T(\hat{\rho} - 1)$. Second, limit results for kernel weighted sums involving $\hat{v}_t = \Delta y_t^k - (\hat{\rho} - 1)y_{t-1}^k$ need to be extended from the integrated to the near integrated case. For Z_ρ this leads, e.g., to $Z_\rho \Rightarrow -ck + \frac{k \int_0^1 J_c(r)^{2k-1} dW(r) + \binom{k}{2} \int_0^1 J_c(r)^{2(k-1)} dr}{\int_0^1 J_c(r)^{2k} dr}$ in case of near integrated alternatives where as usual $J_c(r) := \int_0^r e^{-c(r-s)} dW(s)$.
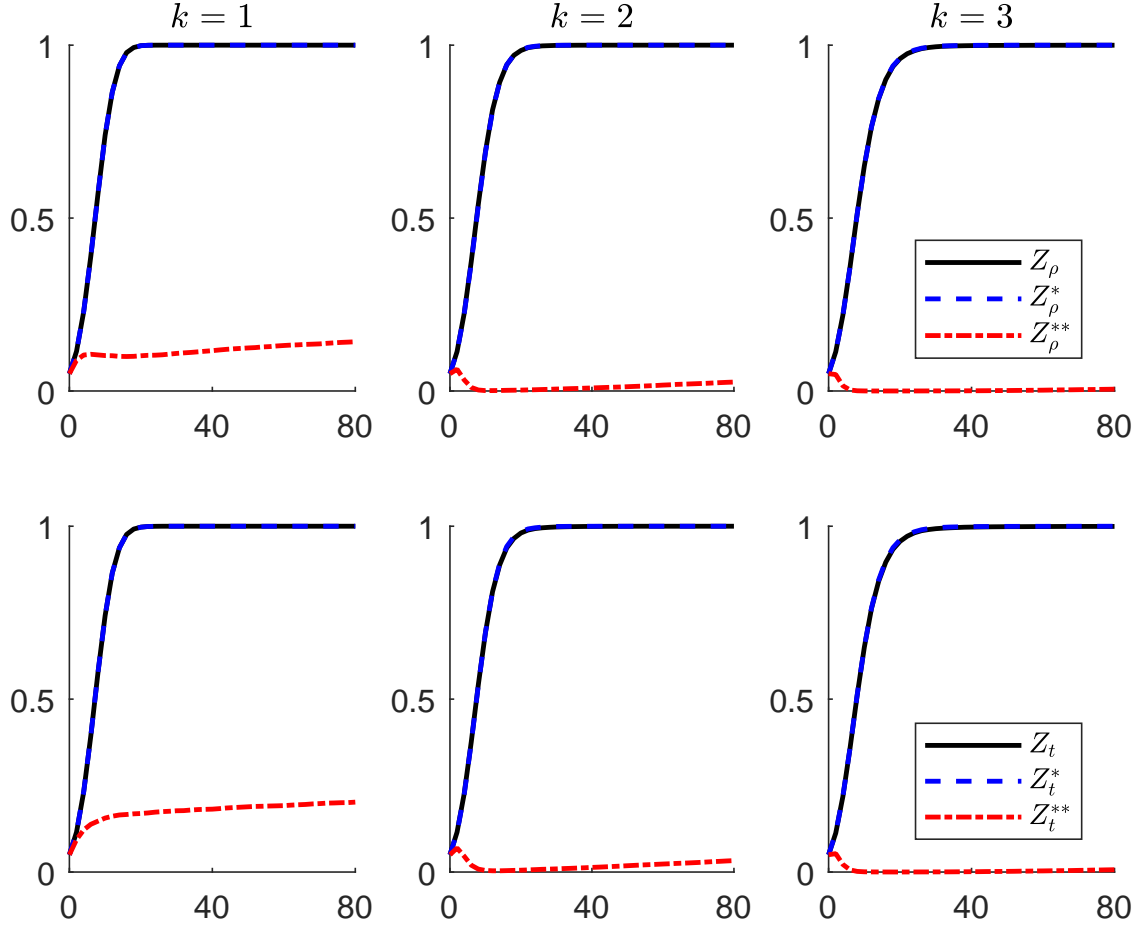
Figure 2: "Empirical" local asymptotic power (LAP) for the six considered test statistics. The upper row displays LAP for the coefficient statistics and the lower row displays LAP for the t-statistics. The horizontal axis displays the value of c.

where ε_t are i.i.d. standard normally distributed and $\gamma \in \{0, 0.3, 0.6, 0.8\}$ governs the amount of serial correlation in u_t. The sample sizes considered are $T \in \{100, 200, 500, 1000\}$ with again 50,000 replications. We consider $x_t = y_t^k$ for $k = 1, 2, 3$ and in Table 2 we display the empirical rejection probabilities at the 5% nominal level.

The table shows that increasing k leads to an increasing tendency of under-rejections, more clearly visible for larger values of γ and smaller values of T. This effect is extremely pronounced for Z_ρ^{**} and Z_t^{**}, where the empirical null rejection probabilities go down to essentially zero even for $\gamma = 0$ and $T = 1000$. These results, however, combine two effects. First, an increasing γ implies larger serial correlation in the first difference of y_t. In case $\gamma = 1$, y_t is actually a process integrated of order two. Now, for a fixed value of $\gamma < 1$ increasing k has a similar effect as letting γ tend

9

γ	100	200	500	1000	100	200	500	1000	100	200	500	1000
		$k=1$				$k=2$				$k=3$		
Panel A: Coefficient Statistic Z_ρ												
0.0	.040	.043	.049	.050	.021	.033	.045	.048	.009	.024	.038	.045
0.3	.047	.051	.055	.055	.012	.026	.042	.048	.005	.014	.031	.040
0.6	.029	.042	.052	.055	.001	.004	.024	.038	.001	.001	.011	.025
0.8	.006	.020	.042	.051	.000	.000	.005	.021	.000	.000	.001	.007
Panel B: t-Statistic Z_t												
0.0	.040	.043	.049	.050	.019	.030	.042	.045	.006	.019	.033	.040
0.3	.047	.051	.055	.056	.010	.023	.039	.045	.004	.011	.025	.035
0.6	.030	.041	.052	.055	.002	.004	.022	.035	.001	.001	.007	.020
0.8	.008	.019	.041	.051	.001	.000	.004	.018	.000	.000	.000	.005
Panel C: Coefficient Statistic Z_ρ^*												
0.0	.040	.043	.049	.050	.022	.034	.044	.048	.010	.024	.038	.046
0.3	.047	.051	.055	.055	.013	.026	.042	.048	.006	.015	.031	.041
0.6	.029	.042	.052	.055	.001	.004	.025	.038	.001	.002	.011	.025
0.8	.006	.020	.042	.051	.000	.000	.005	.021	.001	.000	.001	.008
Panel D: t-Statistic Z_t^*												
0.0	.040	.043	.049	.050	.021	.032	.044	.047	.008	.022	.036	.043
0.3	.047	.051	.055	.056	.011	.025	.041	.047	.005	.013	.029	.039
0.6	.030	.041	.052	.055	.001	.004	.024	.037	.001	.001	.010	.023
0.8	.008	.019	.041	.051	.000	.000	.004	.020	.000	.000	.001	.007
Panel E: Coefficient Statistic Z_ρ^{**}												
0.0	.046	.042	.044	.047	.004	.001	.000	.000	.001	.000	.000	.000
0.3	.052	.042	.041	.043	.006	.002	.000	.000	.002	.000	.000	.000
0.6	.051	.038	.032	.032	.007	.002	.000	.000	.003	.000	.000	.000
0.8	.033	.025	.019	.017	.005	.001	.000	.000	.002	.000	.000	.000
Panel F: t-Statistic Z_t^{**}												
0.0	.044	.039	.044	.047	.004	.001	.000	.000	.001	.000	.000	.000
0.3	.052	.041	.039	.043	.006	.002	.000	.000	.002	.000	.000	.000
0.6	.051	.037	.030	.030	.007	.002	.000	.000	.003	.000	.000	.000
0.8	.033	.025	.018	.015	.005	.001	.000	.000	.002	.000	.000	.000

Table 2: Empirical null rejection probabilities at the 5% nominal level of the null hypothesis $\rho = 1$ for the simulated data using the critical values given in Table 1. The block-columns correspond to the powers $k = 1, 2, 3$ with the four different sample sizes $T = 100, 200, 500, 1000$ within the block-columns. The six panels correspond to the six considered test statistics.

closer to one, i.e., first order serial correlation in x_t is increased. Second, the approximation quality of the empirical null distribution to the asymptotic null distribution is poor at the boundaries of the distributions, leading to size distortions in particular at small significance levels. To gauge this second effect, see Table 3 in the Appendix which displays 50% empirical null rejection probabilities. The table shows that, as expected, the approximation of the (nuisance parameter dependent) finite sample distribution to the asymptotic null distributions is better in the center of the distributions, here exemplified at the median. The combination of the two effects leads to the observed under-rejections. The effects are stronger for Z_ρ^{**} and Z_t^{**} tests, where additionally the "complicated" shape of the distributions renders the simulation of quantiles of the limiting distribution more sensitive than for the other test statistics that have more standard-shaped distributions. This aspect has already been seen when comparing "empirical" and "analytical" LAP of Z_ρ^{**} and Z_t^{**}.

4   Summary and Conclusions

We have shown that the coefficient and t-statistic Phillips (1987) unit root tests have nuisance parameter free limiting distributions when applied to polynomial of integrated processes driven by linear process errors. This result, based on kernel weighted sum limit results, closes the – from today's perspective – incomplete analysis in Wagner (2012). In addition to the two standard tests we have considered four additional tests based on different modifications to remove additive bias terms. It turns out that the behavior of the modified Z_ρ^* and Z_t^* statistics is very similar to that of the original Z_ρ and Z_t statistics with respect to both LAP and empirical null rejections. The modifications using Itô's Lemma, Z_ρ^{**} and Z_t^{**}, perform significantly worse.

Acknowledgements

References

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Darvas, Z. (2008). Estimation Bias and Inference in Overlapping Autoregressions: Implications for the Target-Zone Literature. *Oxford Bulletin of Economics and Statistics* **70**, 1–22.

Grossman, G.M. and A.B. Krueger (1993). Environmental Impacts of a North American Free Trade Agreement. In Garber, P. (Ed.) *The Mexico-US Free Trade Agreement*, 13–56, MIT Press, Cambridge.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer.

Labson, B.S. and P.L. Crompton (1993). Common Trends in Economic Activity and Metals Demand: Cointegration and the Intensity of Use Debate. *Journal of Environmental Economics and Management* **25**, 147–161.

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.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stypka, O., M. Wagner, P. Grabarczyk and R. Kawka (2017). The Asymptotic Validity of "Standard" Fully Modified OLS Estimation of Cointegrating Polynomial Regressions. *IHS Economics Series, Working Paper 333*.

Vogelsang, T.J. and M. Wagner (2013). A Fixed-*b* Perspective on the Phillips-Perron Unit Root Tests. *Econometric Theory* **29**, 609–628.

Svensson, L.E.O. (1992). An Interpretation of Recent Research on Exchange Rate Target Zones. *Journal of Economic Perspectives* **6**, 119–144.

Wagner, M. (2012). The Phillips Unit Root Tests for Polynomials of Integrated Processes *Economics Letters* **114**, 299–303.

Wagner, M. (2015). The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* **30**, 948–967.

Wagner, M. and S.H. Hong (2016). Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Econometric Theory* **32**, 1289–1315.

Appendix: Proofs

Proof of Proposition 1:

The arguments in the proofs are unchanged if instead of x_t only the leading term y_t^k is considered in the regression (3):

$$y_t^k \;=\; \rho y_{t-1}^k + v_t, \tag{21}$$

where we keep for brevity unchanged notation for ρ and v_t. That it suffices to consider (21) stems from the fact that asymptotically only the highest order terms matter, as discussed briefly already in Wagner (2012) and as can be seen in more detail also in Stypka *et al.* (2017). Thus, we consider

$$\hat{\rho} \;:=\; \frac{\sum_{t=2}^{T} y_t^k y_{t-1}^k}{\sum_{t=2}^{T} y_{t-1}^{2k}}. \tag{22}$$

Under Assumption 1 and – used later – Assumption 2, the asymptotic behavior of the OLS estimator and its t-statistic are derived in Wagner (2012, Proposition 1, p. 300), i.e.

$$T(\hat{\rho}-1) \;\Rightarrow\; \frac{k\int_0^1 B_u(r)^{2k-1}dB_u(r) + \left(k(2k-1)\lambda_u + \binom{k}{2}\sigma_u^2\right)\int_0^1 B_u(r)^{2(k-1)}dr}{\int_0^1 B_u(r)^{2k}dr},$$

$$t_\rho \;\Rightarrow\; \frac{k\int_0^1 B_u(r)^{2k-1}dB_u(r) + \left(k(2k-1)\lambda_u + \binom{k}{2}\sigma_u^2\right)\int_0^1 B_u(r)^{2(k-1)}dr}{k\sigma_u\sqrt{\int_0^1 B_u(r)^{2(k-1)}dr \int_0^1 B_u(r)^{2k}dr}},$$

where $B_u(r) = \omega_u^{1/2}W(r)$ is Brownian motion with variance ω_u.

It thus remains to establish the asymptotic behavior of $\hat{\sigma}_v^2$, $\hat{\lambda}_v$ and $\hat{\omega}_v$, which can be derived using similar arguments as in Stypka *et al.* (2017, Theorem 1 and Corollary 1). More precisely it can be shown that

$$\frac{1}{T^{k-1}}\begin{pmatrix} \hat{\sigma}_v^2 \\ \hat{\lambda}_v \\ \hat{\omega}_v \end{pmatrix} \;\Rightarrow\; \begin{pmatrix} \sigma_u^2 \\ \lambda_u \\ \omega_u \end{pmatrix} k^2 \int_0^1 B_u(r)^{2(k-1)}dr, \tag{23}$$

which in conjunction with $\frac{1}{T^{k+1}}\sum_{t=2}^{T} y_{t-1}^{2k} \Rightarrow \int_0^1 B_u(r)^{2k}dr$ leads upon combining and simplifying the terms to the results for Z_ρ and Z_t.

Proof of Corollary 1:

The result for Z_ρ^* and Z_t^* is straightforward since the terms additionally subtracted in (10) and (11) converge exactly to the second terms in the numerators of the Z_ρ and Z_t statistics (divided by the denominator) given in (8) and (9).

Let us therefore consider the two remaining statistics Z_ρ^{**} and Z_t^{**} in some more detail. It suffices to consider, e.g., Z_ρ^{**}, since the arguments are entirely analogous for both:

$$
\begin{aligned}
Z_\rho^{**} \Rightarrow \quad & \frac{k \int_0^1 B_u(r)^{2k-1} dB_u(r) + \left(k(2k-1)\lambda_u + \binom{k}{2}\sigma_u^2\right) \int_0^1 B_u(r)^{2(k-1)} dr}{\int_0^1 B_u(r)^{2k} dr} \\
& + \frac{k^2 \sigma_u^2 \int_0^1 B_u(r)^{2(k-1)} dr}{2 \int_0^1 B_u(r)^{2k} dr} \\
= \quad & \frac{2k \int_0^1 B_u(r)^{2k-1} dB_u(r) + \left(2k(2k-1)\lambda_u + (k(k-1)+k^2)\sigma_u^2\right) \int_0^1 B_u(r)^{2(k-1)} dr}{2 \int_0^1 B_u(r)^{2k} dr} \\
= \quad & \frac{2k \int_0^1 B_u(r)^{2k-1} dB_u(r) + 2k(2k-1)\omega_u \frac{1}{2} \int_0^1 B_u(r)^{2(k-1)} dr}{2 \int_0^1 B_u(r)^{2k} dr} \\
= \quad & \frac{B(1)^{2k}}{2 \int_0^1 B_u(r)^{2k} dr} = \frac{W(1)^{2k}}{2 \int_0^1 W_u(r)^{2k} dr}, \qquad\qquad (24)
\end{aligned}
$$

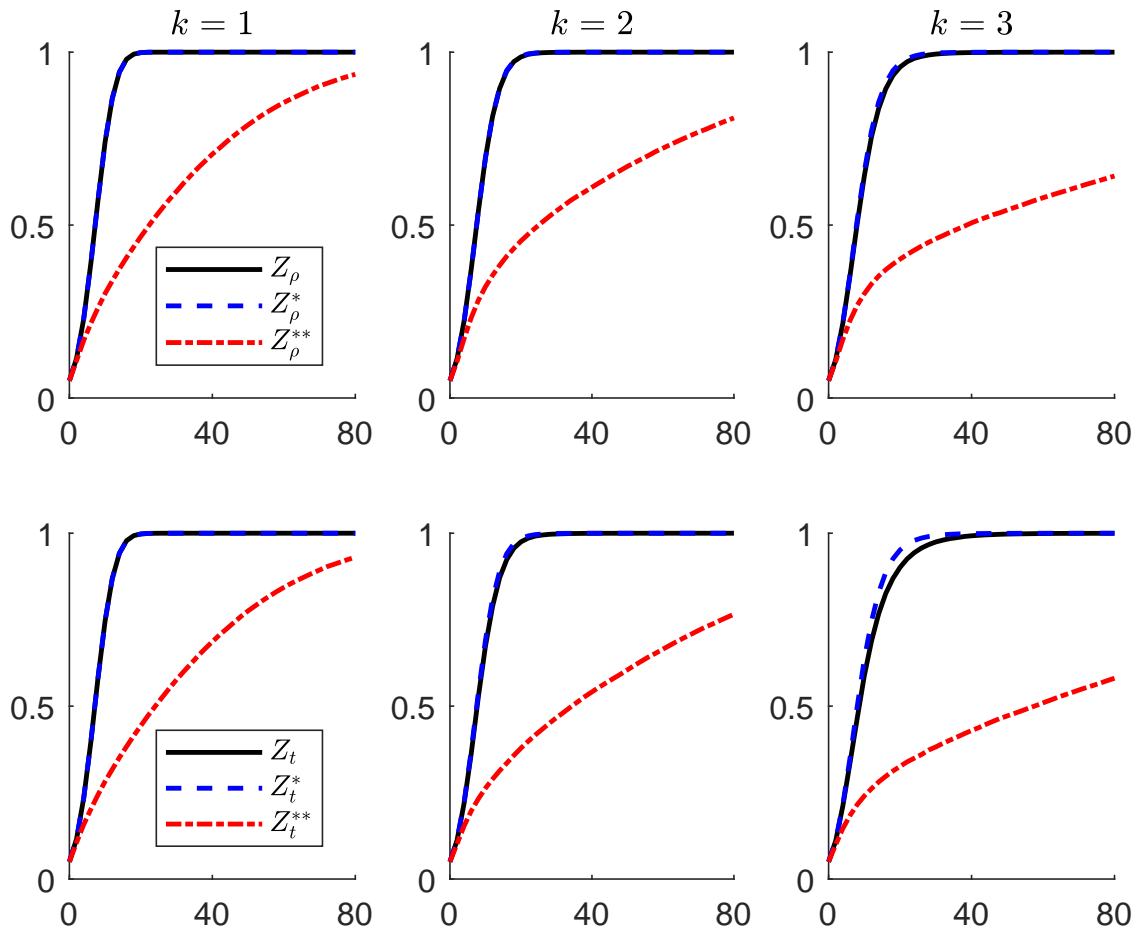using $B_u(r) = \omega_u^{1/2} W(r)$, $\omega_u = 2\lambda_u + \sigma_u^2$ and Itô's Lemma.

Figure 3: "Analytical" local asymptotic power (LAP) for the six considered test statistics. The upper row displays LAP for the coefficient statistics and the lower row displays LAP for the t-statistics. The horizontal axis displays the value of c.

γ	100	200	500	1000	100	200	500	1000	100	200	500	1000
		$k=1$				$k=2$				$k=3$		
Panel A: Coefficient Statistic Z_ρ												
0.0	.490	.496	.493	.500	.478	.495	.496	.501	.451	.482	.493	.500
0.3	.526	.526	.518	.519	.502	.517	.517	.519	.464	.499	.509	.515
0.6	.528	.531	.524	.524	.472	.503	.514	.519	.391	.463	.495	.507
0.8	.505	.523	.526	.527	.373	.465	.504	.516	.234	.377	.465	.494
Panel B: t-Statistic Z_t												
0.0	.492	.496	.494	.501	.478	.495	.497	.502	.457	.490	.500	.507
0.3	.523	.525	.516	.518	.496	.513	.515	.517	.463	.501	.513	.520
0.6	.523	.527	.522	.522	.459	.497	.510	.516	.379	.463	.497	.513
0.8	.492	.516	.522	.524	.337	.450	.495	.511	.201	.364	.466	.496
Panel C: Coefficient Statistic Z_ρ^*												
0.0	.490	.496	.493	.500	.479	.493	.495	.500	.451	.478	.490	.497
0.3	.526	.526	.518	.519	.506	.521	.518	.519	.466	.497	.507	.513
0.6	.528	.531	.524	.524	.481	.509	.516	.520	.398	.464	.493	.506
0.8	.505	.523	.526	.527	.391	.470	.507	.518	.256	.383	.465	.493
Panel D: t-Statistic Z_t^*												
0.0	.492	.496	.494	.501	.477	.494	.496	.501	.450	.482	.492	.499
0.3	.523	.525	.516	.518	.500	.516	.516	.518	.462	.498	.508	.514
0.6	.523	.527	.522	.522	.468	.502	.514	.519	.386	.461	.494	.507
0.8	.492	.516	.522	.524	.360	.460	.501	.514	.226	.371	.462	.493
Panel E: Coefficient Statistic Z_ρ^{**}												
0.0	.496	.501	.502	.501	.477	.491	.497	.498	.434	.478	.493	.496
0.3	.498	.503	.502	.500	.460	.484	.494	.496	.349	.446	.487	.494
0.6	.481	.495	.500	.499	.392	.457	.486	.492	.191	.332	.457	.484
0.8	.409	.459	.485	.493	.251	.366	.459	.482	.066	.164	.352	.446
Panel F: t-Statistic Z_t^{**}												
0.0	.494	.501	.498	.499	.478	.490	.497	.497	.457	.482	.494	.497
0.3	.513	.515	.509	.509	.475	.490	.499	.501	.397	.471	.493	.496
0.6	.504	.512	.511	.511	.426	.473	.493	.498	.190	.379	.477	.489
0.8	.453	.491	.504	.509	.261	.407	.474	.490	.043	.148	.399	.471

Table 3: Empirical null rejection probabilities at the 50% nominal level of the null hypothesis $\rho = 1$ for the simulated data using the critical values given in Table 1. For further explanations see caption of Table 2.