UNIT ROOTS

Herman J. Bierens[1]
Pennsylvania State University
(October 30, 2007)

1. Introduction

In this chapter I will explain the two most frequently applied types of unit root tests, namely the Augmented Dickey-Fuller tests [see Fuller (1996), Dickey and Fuller (1979, 1981)], and the Phillips-Perron tests [see Phillips (1987) and Phillips and Perron (1988)]. The statistics and econometrics levels required for understanding the material below are Hogg and Craig (1978) or a similar level for statistics, and Green (1997) or a similar level for econometrics. The functional central limit theorem [see Billingsley (1968)], which plays a key-role in the derivations involved, will be explained in this chapter by showing its analogy with the concept of convergence in distribution of random variables, and by confining the discussion to Gaussian unit root processes.

This chapter is not a review of the vast literature on unit roots. Such a review would entail a long list of descriptions of the many different recipes for unit root testing proposed in the literature, and would leave no space for motivation, let alone proofs. I have chosen for depth rather than breadth, by focusing on the most influential papers on unit root testing, and discussing them in detail, without assuming that the reader has any previous knowledge about this topic.

As an introduction of the concept of a unit root and its consequences, consider the Gaussian AR(1) process $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$, or equivalently $(1 - \beta_1 L)y_t = \beta_0 + u_t$, where L is the lag operator: $Ly_t = y_{t-1}$, and the u_t 's are i.i.d. $N(0,\sigma^2)$. The lag polynomial $1 - \beta_1 L$ has root equal to $1/\beta_1$. If $|\beta_1| < 1$, then by backwards substitution we can write $y_t = \beta_0/(1-\beta_1) + \sum_{j=0}^{\infty}\beta_1^j u_{t-j}$, so that y_t is strictly stationary, i.e., for arbitrary natural numbers $m_1 < m_2 < ... < m_{k-1}$ the joint distribution of $y_t, y_{t-m_1}, y_{t-m_2}, \dots, y_{t-m_{k-1}}$ does not depend on t, but only on the lags or leads $m_1, m_2, .., m_{k-1}$. Moreover, the distribution of y_t, $t > 0$, conditional on $y_0, y_{-1}, y_{-2},...$, then converges to the marginal distribution

[1] This is a slightly revised version of a chapter in Badi Baltagi (Ed.), *Companion in Theoretical Econometrics*, Blackwell Publishers. The useful comments of three referees are gratefully acknowledged.

of y_t if $t \to \infty$. In other words, y_t has a vanishing memory: y_t becomes independent of its past, y_0, y_{-1}, y_{-2},..., if $t \to \infty$.

If $\beta_1 = 1$, so that the lag polynomial $1 - \beta_1 L$ has a unit root, then y_t is called a unit root process. In this case the AR(1) process under review becomes $y_t = y_{t-1} + \beta_0 + u_t$, which by backwards substitution yields for $t > 0$, $y_t = y_0 + \beta_0 t + \sum_{j=1}^{t} u_j$. Thus now the distribution of y_t, $t > 0$, conditional on y_0, y_{-1}, y_{-2},..., is $N(y_0 + \beta_0 t, \sigma^2 t)$, so that y_t has no longer a vanishing memory: a shock in y_0 will have a persistent effect on y_t. The former intercept β_0 now becomes the *drift* parameter of the unit root process involved.

It is important to distinguish stationary processes from unit root processes, for the following reasons:

1. Regressions involving unit root processes may give spurious results. If y_t and x_t are mutually independent unit root processes, i.e. y_t is independent of x_{t-j} for all t and j, then the OLS regression of y_t on x_t for $t = 1,..,n$, with or without an intercept, will yield a significant estimate of the slope parameter if n is large: the absolute value of the t-value of the slope converges in probability to ∞ if $n \to \infty$. We then might conclude that y_t depends on x_t, while in reality the y_t's are independent of the x_t's. This phenomenon is called *spurious regression*.[2] One should therefore be very cautious when conducting standard econometric analysis using time series. If the time series involved are unit root processes, naive application of regression analysis may yield nonsense results.

2. For two or more unit root processes there may exist linear combinations which are stationary, and these linear combinations may be interpreted as long-run relationships. This phenomenon is called *cointegration*[3], and plays a dominant role in modern empirical macroeconomic research.

---

[2] See the chapter on spurious regression in Badi Baltagi (Ed.), *Companion in Theoretical Econometrics*, Blackwell Publishers. This phenomenon can easily be demonstrated by using my free software package *EasyReg*, which is downloadable from website http://econ.la.psu.edu/~hbierens/EASYREG.HTM (Click on "Tools", and then on "Teaching tools").

[3] See the chapter on cointegration in Badi Baltagi (Ed.), *Companion in Theoretical Econometrics*, Blackwell Publishers..

3. Tests of parameter restrictions in (auto)regressions involving unit root processes have in general different null distributions than in the case of stationary processes. In particular, if one would test the null hypothesis $\beta_1 = 1$ in the above AR(1) model using the usual t-test, the null distribution involved is non-normal. Therefore, naive application of classical inference may give incorrectly results. We will demonstrate the latter first, and in the process derive the Dickey-Fuller test [see Fuller (1996), Dickey and Fuller (1979, 1981)], by rewriting the AR(1) model as

$$\Delta y_t = y_t - y_{t-1} = \beta_0 + (\beta_1 - 1)y_{t-1} + u_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \quad \textbf{(1)}$$

say, estimating the parameter α_1 by OLS on the basis of observations $y_0, y_1, ..., y_n$, and then testing the unit root hypothesis $\alpha_1 = 0$ against the stationarity hypothesis $-2 < \alpha_1 < 0$, using the t-value of α_1. In Section 2 we consider the case where $\alpha_0 = 0$ under both the unit root hypothesis and the stationarity hypothesis. In Section 3 we consider the case where $\alpha_0 = 0$ under the unit root hypothesis but not under the stationarity hypothesis.

The assumption that the error process u_t is independent is quite unrealistic for macroeconomic time series. Therefore, in Sections 4 and 5 this assumption will be relaxed, and two types of appropriate unit root tests will be discussed: the Augmented Dickey-Fuller (ADF) tests, and the Phillips-Perron (PP) tests.

In Section 6 we consider the unit root *with* drift case, and we discuss the ADF and PP tests of the unit root with drift hypothesis, against the alternative of trend stationarity.

Finally, Section 7 contains some concluding remarks.

2. The Gaussian AR(1) case without intercept: Part 1

2.1 Introduction

Consider the AR(1) model without intercept, rewritten as[4]

$$\Delta y_t = \alpha_0 y_{t-1} + u_t, \text{ where } u_t \text{ is i.i.d. } N(0, \sigma^2), \quad \textbf{(2)}$$

and y_t is observed for $t = 1,2,..,n$. For convenience I will assume that

[4] The reason for changing the subscript of α from 1 in **(1)** to 0 is to indicate the number of other parameters at the right-hand side of the equation. See also **(39)**.

$$y_t = 0 \; for \; t \le 0. \tag{3}$$

This assumption is, of course, quite unrealistic, but is made for the sake of transparency of the argument, and will appear to be innocent.

The OLS estimator of α_0 is:

$$\hat{\alpha}_0 = \frac{\sum_{t=1}^{n} y_{t-1}\Delta y_t}{\sum_{t=1}^{n} y_{t-1}^2} = \alpha_0 + \frac{\sum_{t=1}^{n} y_{t-1}u_t}{\sum_{t=1}^{n} y_{t-1}^2}. \tag{4}$$

If -2 < α_0 < 0, so that y_t is stationary, then it is a standard exercise to verify that $\sqrt{n}(\hat{\alpha}_0 - \alpha_0) \to N(0, 1-(1+\alpha_0)^2)$ in distribution. On the other hand, if $\alpha_0 = 0$, so that y_t is a unit root process, this result reads: $\sqrt{n}\hat{\alpha}_0 \to N(0,0)$ in distr., hence $\text{plim}_{n\to\infty}\sqrt{n}\hat{\alpha}_0 = 0$. However, we show now that a much stronger result holds, namely that $\hat{\rho}_0 \equiv n\hat{\alpha}_0$ converges in distribution, but the limiting distribution involved is non-normal. Thus, the presence of a unit root is actually advantageous for the efficiency of the OLS estimator $\hat{\alpha}_0$. The main problem is that the t-test of the null hypothesis that $\alpha_0 = 0$ has no longer a standard normal asymptotic null distribution, so that we cannot test for a unit root using standard methods. The same applies to more general unit root processes.

In the unit root case under review we have $y_t = y_{t-1} + u_t = y_0 + \Sigma_{j=1}^t u_j = \Sigma_{j=1}^t u_j$ for t > 0, where the last equality involved is due to assumption **(3)**. Denoting

$$S_t = 0 \; for \; t \le 0, \; S_t = \sum_{j=1}^{t} u_j \quad for \; t \ge 1. \tag{5}$$

and $\hat{\sigma}^2 = (1/n)\Sigma_{t=1}^n u_t^2$, it follows that

$$\frac{1}{n}\sum_{t=1}^{n} u_t y_{t-1} = \frac{1}{2n}\sum_{t=1}^{n}\left((u_t + y_{t-1})^2 - y_{t-1}^2 - u_t^2\right) = \frac{1}{2}\left(\frac{1}{n}\sum_{t=1}^{n} y_t^2 - \frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n}\sum_{t=1}^{n} u_t^2\right)$$
$$= \frac{1}{2}(y_n^2/n - y_0^2/n - \hat{\sigma}^2) = \frac{1}{2}(S_n^2/n - \hat{\sigma}^2), \tag{6}$$

and similarly,

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{n}\sum_{t=1}^{n}(S_{t-1}/\sqrt{n})^2. \tag{7}$$

Next, let

$$W_n(x) = S_{[nx]}/(\sigma\sqrt{n}) \quad \textit{for } x \in [0,1], \tag{8}$$

where $[z]$ means truncation to the nearest integer $\leq z$. Then we have[5]:

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} u_t y_{t-1} &= \frac{1}{2}(\sigma^2 W_n(1)^2 - \hat{\sigma}^2) \\ &= \frac{1}{2}(\sigma^2 W_n(1)^2 - \sigma^2 - O_p(1/\sqrt{n})) = \sigma^2\frac{1}{2}(W_n(1)^2 - 1) + o_p(1),\end{aligned} \tag{9}$$

and

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{n}\sum_{t=1}^{n} \sigma^2 W_n((t-1)/n)^2 = \sigma^2\int W_n(x)^2 dx, \tag{10}$$

where the integral in **(10)** *and below, unless otherwise indicated*, is taken over the unit interval [0,1]. The last equality in **(9)** follows from the law of large numbers, by which $\hat{\sigma}^2 = \sigma^2 + O_p(1/\sqrt{n})$. The last equality in **(10)** follows from the fact that for any power m,

$$\begin{aligned}\int W_n(x)^m dx &= \int_0^1 W_n(x)^m dx = \frac{1}{n}\int_0^n W_n(z/n)^m dz = \frac{1}{n}\sum_{t=1}^{n}\int_{t-1}^{t} W_n(z/n)^m dz \\ &= \frac{1}{n^{1+m/2}}\sum_{t=1}^{n}\int_{t-1}^{t}(S_{[z]}/\sigma)^m dz = \frac{1}{n^{1+m/2}}\sum_{t=1}^{n}(S_{t-1}/\sigma)^m.\end{aligned} \tag{11}$$

Moreover, observe from **(11)**, with $m = 1$, that $\int W_n(x)dx$ is a linear combination of i.i.d. standard normal random variables, and therefore normal itself, with zero mean and variance

$$E\left(\int W_n(x)dx\right)^2 = \int\int E\big(W_n(x)W_n(y)\big)dxdy = \int\int\frac{\min\big([nx],[ny]\big)}{n}dxdy \to \int\int\min(x,y)dxdy = \frac{1}{3}. \tag{12}$$

Thus, $\int W_n(x)dx \to N(0,1/3)$ in distribution. Since $\int W_n(x)^2 dx \geq (\int W_n(x)dx)^2$, it follows therefore that $\int W_n(x)^2 dx$ is bounded away from zero:

---

[5] Recall that the notation $o_p(a_n)$, with a_n a deterministic sequence, stands for a sequence of random variables or vectors x_n, say, such that $\text{plim}_{n\to\infty} x_n/a_n = 0$, and that the notation $O_p(a_n)$ stands for a sequence of random variables or vectors x_n such that x_n/a_n is stochastically bounded: $\forall\varepsilon \in (0,1)\ \exists\ M \in (0,\infty)$: $\sup_{n\geq 1} P(|x_n/a_n| > M) < \varepsilon$. Also, recall that convergence in distribution implies stochastic boundedness.

$$\left(\int W_n(x)^2 dx\right)^{-1} = O_p(1). \quad \textbf{(13)}$$

Combining **(9)**, **(10)**, and **(13)**, we now have:

$$\hat{\rho}_0 \equiv n\hat{\alpha}_0 = \frac{(1/n)\Sigma_{t=1}^{n} u_t y_{t-1}}{(1/n^2)\Sigma_{t=1}^{n} y_{t-1}^2} = \frac{(1/2)(W_n(1)^2 - 1) + o_p(1)}{\int W_n(x)^2 dx} = \frac{1}{2}\left(\frac{W_n(1)^2 - 1}{\int W_n(x)^2 dx}\right) + o_p(1). \quad \textbf{(14)}$$

This result does not depend on assumption **(3)**.

2.2 Weak convergence of random functions

In order to establish the limiting distribution of **(14)**, and other asymptotic results, we need to extend the well-known concept of convergence in distribution of random variables to convergence in distribution of a sequence of random functions. Recall that for random variables X_n, X, $X_n \to X$ in distribution if the distribution function $F_n(x)$ of X_n converges pointwise to the distribution function $F(x)$ of X in the continuity points of $F(x)$. Moreover, recall that distribution functions are uniquely associated to probability measures on the Borel sets[6], i.e., there exists one and only one probability measure $\mu_n(B)$ on the Borel sets B such that $F_n(x) = \mu_n((-\infty,x])$, and similarly, $F(x)$ is uniquely associated to a probability measure μ on the Borel sets, such that $F(x) = \mu((-\infty,x])$. The statement $X_n \to X$ in distribution can now be expressed in terms of the probability measures μ_n and μ: $\mu_n(B) \to \mu(B)$ for all Borel sets B with boundary δB satisfying $\mu(\delta B) = 0$.

In order to extend the latter to random functions, we need to define Borel sets of functions. For our purpose it suffices to define Borel sets of continuous functions on [0,1]. Let $C[0,1]$ be the set of all continuous functions on the unit interval [0,1]. Define the distance between two functions f and g in $C[0,1]$ by the sup-norm: $\rho(f,g) = \sup_{0\le x\le 1}|f(x)-g(x)|$. Endowed with this norm, the set

---

[6] The Borel sets in $\mathbb{R}$ are the members of the smallest σ-algebra containing the collection $\mathfrak{C}$, say, of all half-open intervals $(-\infty,x]$, $x \in \mathbb{R}$. Equivalently, we may also define the Borel sets as the members of the smallest σ-algebra containing the collection of open subsets of $\mathbb{R}$. A collection $\mathscr{F}$ of subsets of a set Ω is called a σ-algebra if the following three conditions hold: $\Omega\in\mathscr{F}$; $A\in\mathscr{F}$ implies that its complement also belongs to $\mathscr{F}$: $\Omega\backslash A\in\mathscr{F}$ (hence, the empty set $\emptyset$ belongs to $\mathscr{F}$); $A_n\in\mathscr{F}$, n = 1,2,3,.., implies $\cup_{n=1}^{\infty}A_n\in\mathscr{F}$. The smallest σ-algebra containing a collection $\mathfrak{C}$ of sets is the intersection of all σ-algebras containing the collection $\mathfrak{C}$.

$C[0,1]$ becomes a metric space, for which we can define open subsets, similarly to the concept of an open subset of $\mathbb{R}$: A set B in $C[0,1]$ is open if for each function f in B we can find an $\varepsilon > 0$ such that $\{g \in C[0,1]: \rho(g,f) < \varepsilon\} \subset B$. Now the smallest σ-algebra of subsets of $C[0,1]$ containing the collection of all open subsets of $C[0,1]$ is just the collection of Borel sets of functions in $C[0,1]$.

A random element of $C[0,1]$ is a random function $W(x)$, say, on [0,1], which is continuous with probability 1. For such a random element W, say, we can define a probability measure μ on the Borel sets B in $C[0,1]$ by $\mu(B) = P(W \in B)$. Now a sequence W_n^* of random elements of $C[0,1]$, with corresponding probability measures μ_n, is said to converge weakly to a random element W of $C[0,1]$, with corresponding probability measure μ, if for each Borel set B in $C[0,1]$ with boundary δB satisfying $\mu(\delta B) = 0$, we have $\mu_n(B) \to \mu(B)$. This is usually denoted by: $W_n^* \Rightarrow W$ (on [0,1]). Thus, weak convergence is the extension to random functions of the concept of convergence in distribution.

In order to verify that $W_n^* \Rightarrow W$ on [0,1], we have to verify two conditions. See Billingsley (1968). First, we have to verify that the finite distributions of W_n^* converge to the corresponding finite distributions of W, i.e., for arbitrary points $x_1,..,x_m$ in [0,1], $(W_n^*(x_1),...,W_n^*(x_m)) \Rightarrow (W(x_1),...,W(x_m))$ in distribution. Second, we have to verify that W_n^* is tight. Tightness is the extension of the concept of stochastic boundedness to random functions: for each ε in (0,1) there exists a compact (Borel) set K in $C[0,1]$ such that $\mu_n(K) > 1\text{-}\varepsilon$ for $n = 1,2.,...$ Since convergence in distribution implies stochastic boundedness, we cannot have convergence in distribution without stochastic boundedness, and the same applies to weak convergence: tightness is a necessary condition for weak convergence.

As is well-known, if $X_n \to X$ in distribution, and Φ is a continuous mapping from the support of X into a Euclidean space, then by Slutsky's theorem, $\Phi(X_n) \to \Phi(X)$ in distribution. A similar result holds for weak convergence, which is known as the continuous mapping theorem: If Φ is a continuous mapping from $C[0,1]$ into a Euclidean space, then $W_n^* \Rightarrow W$ implies $\Phi(W_n^*) \to \Phi(W)$ in distribution. For example, the integral $\Phi(f) = \int f(x)^2 dx$ with $f \in C[0,1]$ is a continuous mapping from $C[0,1]$ into the real line, hence $W_n^* \Rightarrow W$ implies that $\int W_n^*(x)^2 dx \to \int W(x)^2 dx$ in distribution.

The random function W_n defined by **(8)** is a step function on [0,1], and therefore not a random element of $C[0,1]$. However, the steps involved can be smoothed by piecewise linear interpolation, yielding a random element W_n^* of $C[0,1]$ such that $\sup_{0 \le x \le 1}|W_n^*(x) - W_n(x)| = o_p(1)$. The finite distributions of W_n^* are therefore asymptotically the same as the finite distributions of W_n. In order to analyze the latter, redefine W_n as

$$W_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nx]} e_t \ for\ x \in [n^{-1},1], \quad W_n(x) = 0\ for\ x \in [0,n^{-1}), \quad e_t\ is\ i.i.d.\ N(0,1). \tag{15}$$

(Thus, $e_t = u_t/\sigma$), and let

$$\begin{aligned} W_n^*(x) &= W_n\left(\frac{t-1}{n}\right) + (nx - (t-1))\left(W_n\left(\frac{t}{n}\right) - W_n\left(\frac{t-1}{n}\right)\right) \\ &= W_n(x) + (nx - (t-1))\frac{e_t}{\sqrt{n}} \quad for\ x \in \left(\frac{t-1}{n},\frac{t}{n}\right],\ t = 1,...,n, \quad W_n^*(0) = 0. \end{aligned} \tag{16}$$

Then

$$\sup_{0\le x\le 1}|W_n^*(x) - W_n(x)| \le \frac{\max_{1\le t\le n}|e_t|}{\sqrt{n}} = o_p(1). \tag{17}$$

The latter conclusion is not too hard an exercise.[7]

It is easy to verify that for *fixed* $0 \le x < y \le 1$ we have

$$\begin{aligned} \begin{pmatrix} W_n(x) \\ W_n(y) - W_n(x) \end{pmatrix} &= \frac{1}{\sqrt{n}}\begin{pmatrix} \sum_{t=1}^{[nx]}e_t \\ \sum_{t=[nx]+1}^{[ny]}e_t \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{[nx]}{n} & 0 \\ 0 & \frac{[ny]-[nx]}{n} \end{pmatrix}\right) \\ &\to \begin{pmatrix} W(x) \\ W(y)-W(x) \end{pmatrix}\ in\ distr., \end{aligned} \tag{18}$$

where $W(x)$ is a random function on [0,1] such that for $0 \le x < y \le 1$,

$$\begin{pmatrix} W(x) \\ W(y) - W(x) \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x & 0 \\ 0 & y-x \end{pmatrix}\right). \tag{19}$$

---

[7] Under the assumption that e_t is i.i.d. $N(0,1)$,

$$P\left(\max_{1\le t\le n}|e_t| \le \varepsilon\sqrt{n}\right) = \left(1 - 2\int_{\varepsilon\sqrt{n}}^{\infty}\frac{\exp(-x^2/2)}{\sqrt{2\pi}}dx\right)^n \to 1$$

for arbitrary $\varepsilon > 0$.

This random function $W(x)$ is called a standard Wiener process, or Brownian motion. Similarly, for arbitrary fixed x,y in [0,1],

$$\begin{pmatrix} W_n(x) \\ W_n(y) \end{pmatrix} \rightarrow \begin{pmatrix} W(x) \\ W(y) \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x & \min(x,y) \\ \min(x,y) & y \end{pmatrix} \right) \ in\ distr. \tag{20}$$

and it follows from **(17)** that the same applies to W_n^*. Therefore, the finite distributions of W_n^* converge to the corresponding finite distributions of W. Also, it can be shown that W_n^* is tight [see Billingsley (1968)]. Hence, $W_n^* \Rightarrow W$, and by the continuous mapping theorem,

$$(W_n^*(1), \int W_n^*(x)dx, \int W_n^*(x)^2dx, \int xW_n^*(x)dx)^T \rightarrow (W(1), \int W(x)dx, \int W(x)^2dx, \int xW(x)dx)^T \tag{21}$$

in distr. This result, together with **(17)**, implies that:

LEMMA 1. *For W_n defined by* **(15)**, $(W_n(1), \int W_n(x)dx, \int W_n(x)^2dx, \int xW_n(x)dx)^T$ *converges jointly in distribution to* $(W(1), \int W(x)dx, \int W(x)^2dx, \int xW(x)dx)^T$.

2.3 Asymptotic null distributions

Using Lemma 1, it follows now straightforwardly from **(14)** that:

$$\hat{\rho}_0 \equiv n\hat{\alpha}_0 \rightarrow \rho_0 \equiv \frac{1}{2}\left(\frac{W(1)^2 - 1}{\int W(x)^2dx} \right) \ in\ distr. \tag{22}$$

The density[8] of the distribution of ρ_0 is displayed in Figure 1, which clearly shows that the distribution involved is non-normal and asymmetric, with a fat left tail.

---

[8] This density is actually a kernel estimate of the density of $\hat{\rho}_0$ on the basis of 10,000 replications of a Gaussian random walk $y_t = y_{t-1} + e_t$, $t = 0,1,...,1000$, $y_t = 0$ for $t < 0$. The kernel involved is the standard normal density, and the bandwidth $h = c.s10000^{-1/5}$, where s is the sample standard error, and $c = 1$. The scale factor c has been chosen by experimenting with various values. The value $c = 1$ is about the smallest one for which the kernel estimate remains a smooth curve; for smaller values of c the kernel estimate becomes wobbly. The densities of ρ_1, τ_1, ρ_2, and τ_2 in Figures 2-6 have been constructed in the same way, with $c = 1$.

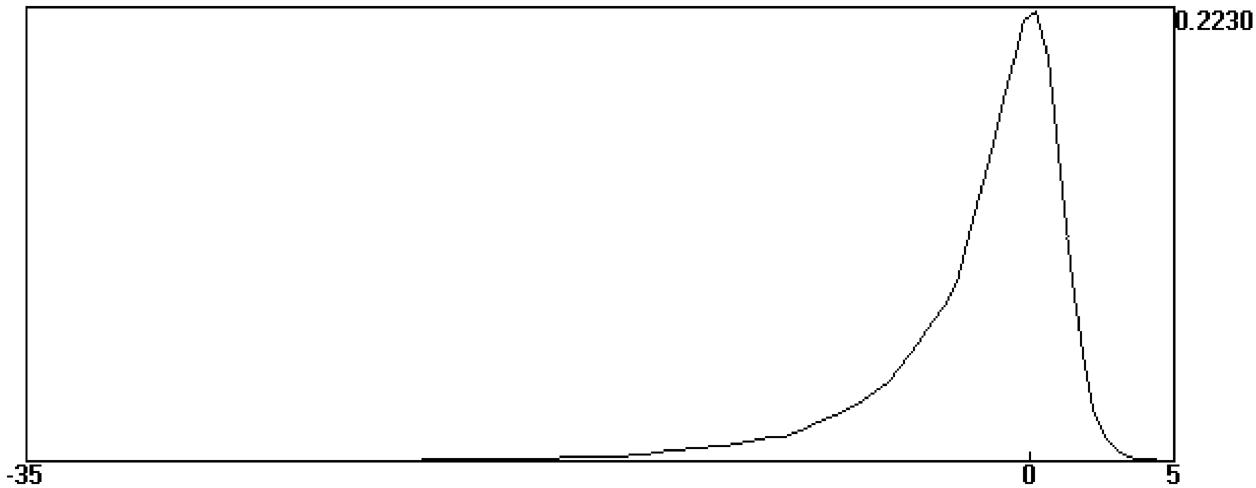

Figure 1: Density of ρ_0

Also the limiting distribution of the usual t-test statistic of the null hypothesis $\alpha_0 = 0$ is non-normal. First, observe that due to **(10)**, **(22)**, and Lemma 1, the residual sum of squares (*RSS*) of the regression **(2)** under the unit root hypothesis is:

$$RSS = \sum_{t=1}^{n} (\Delta y_t - \hat{\alpha}_0 y_{t-1})^2 = \sum_{t=1}^{n} u_t^2 - (n\hat{\alpha}_0)^2(1/n^2)\sum_{t=1}^{n} y_{t-1}^2 = \sum_{t=1}^{n} u_t^2 + O_p(1). \quad \textbf{(23)}$$

Hence $RSS/(n-1) = \sigma^2 + O_p(1/n)$. Therefore, similarly to **(14)** and **(22)**, the Dickey-Fuller t-statistic $\hat{\tau}_0$ involved satisfies:

$$\hat{\tau}_0 \equiv n\hat{\alpha}_0 \frac{\sqrt{(1/n^2)\sum_{t=1}^{n} y_{t-1}^2}}{\sqrt{RSS/(n-1)}} = \frac{(W_n(1)^2 - 1)/2}{\sqrt{\int W_n(x)^2 dx}} + o_p(1) \to \tau_0 \equiv \frac{(W(1)^2 - 1)/2}{\sqrt{\int W(x)^2 dx}} \text{ in distr.} \quad \textbf{(24)}$$

Note that the unit root tests based on the statistics $\hat{\rho}_0 \equiv n\hat{\alpha}_0$ and $\hat{\tau}_0$ are left-sided: under the alternative of stationarity, $-2 < \alpha_0 < 0$, we have $\text{plim}_{n\to\infty}\hat{\alpha}_0 = \alpha_0 < 0$, hence $\hat{\rho}_0 \to -\infty$ in probability at rate n, and $\hat{\tau}_0 \to -\infty$ in probability at rate $\sqrt{n}$.

The non-normality of the limiting distributions ρ_0 and τ_0 is no problem, though, as long one is aware of it. The distributions involved are free of nuisance parameters, and asymptotic critical values of the unit root tests $\hat{\rho}_0$ and $\hat{\tau}_0$ can easily be tabulated, using Monte Carlo simulation. In particular,

$$P(\tau_0 \leq -1.95) = 0.05, \quad P(\tau_0 \leq -1.62) = 0.10, \tag{25}$$

(see Fuller 1996, p. 642), whereas for a standard normal random variable *e*,

$$P(e \leq -1.64) = 0.05, \quad P(e \leq -1.28) = 0.10 \tag{26}$$

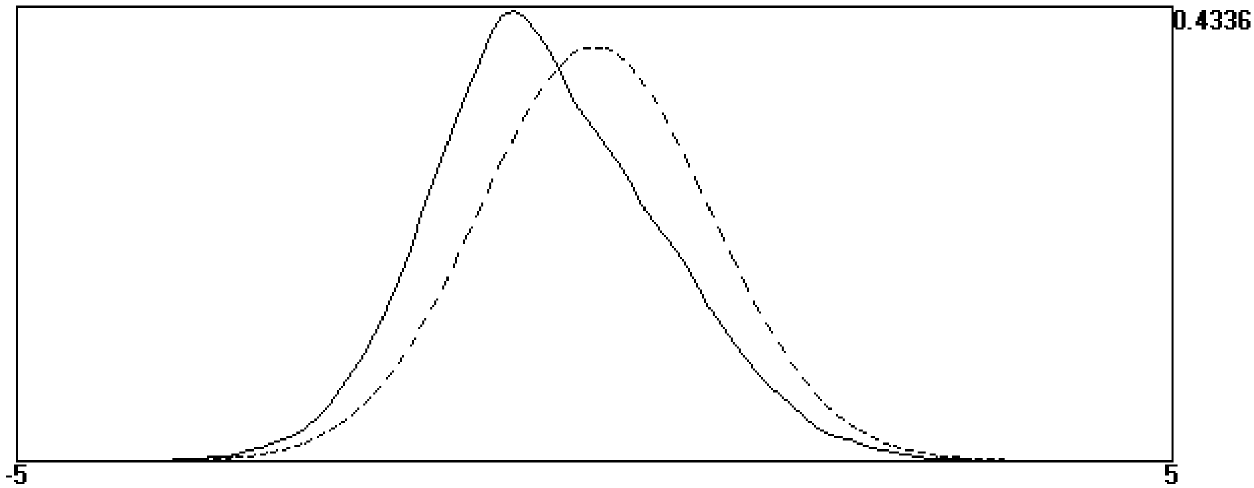

Figure 2: *Density of* τ_0 *compared with the standard normal density* (*dashed curve*)

In Figure 2 the density of τ_0 is compared with the standard normal density. We see that the density of τ_0 is shifted to left of the standard normal density, which causes the difference between **(25)** and **(26)**. Using the left-sided standard normal test would result in a type 1 error of about twice the size: compare **(26)** with

$$P(\tau_0 \leq -1.64) \approx 0.09, \quad P(\tau_0 \leq -1.28) \approx 0.18 \tag{27}$$

3. The Gaussian AR(1) case with intercept under the alternative of stationarity

If under the stationarity hypothesis the AR(1) process has an intercept, but not under the unit root hypothesis, the AR(1) model that covers both the null and the alternative is:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \quad where \ \alpha_0 = -c\alpha_1. \tag{28}$$

If -2 < α_1 < 0, then the process y_t is stationary around the constant *c*:

$$y_t = -c\alpha_1 + (1+\alpha_1)y_{t-1} + u_t = \sum_{j=0}^{\infty}(1+\alpha_1)^j(-c\alpha_1 + u_{t-j}) = c + \sum_{j=0}^{\infty}(1+\alpha_1)^j u_{t-j}, \qquad \textbf{(29)}$$

hence $E(y_t^2) = c^2 + (1-(1+\alpha_1)^2)^{-1}\sigma^2$, $E(y_t y_{t-1}) = c^2 + (1+\alpha_1)(1-(1+\alpha_1)^2)^{-1}\sigma^2$, and the OLS estimator **(4)** of α_0 in model **(2)** satisfies

$$\operatorname*{plim}_{n\to\infty}\hat{\alpha}_0 = \frac{E(y_t y_{t-1})}{E(y_{t-1}^2)} - 1 = \frac{\alpha_1}{1 + (c/\sigma)^2(1-(1+\alpha_1)^2)}, \qquad \textbf{(30)}$$

which approaches zero if $c^2/\sigma^2 \to \infty$. Therefore, the power of the test $\hat{\rho}_0$ will be low if the variance of u_t is small relative to $[E(y_t)]^2$. The same applies to the t-test $\hat{\tau}_0$. We should therefore use the OLS estimator of α_1 and the corresponding t-value in the regression of Δy_t on y_{t-1} *with* intercept.

Denoting $\bar{y}_{-1} = (1/n)\sum_{t=1}^{n} y_{t-1}$, $\bar{u} = (1/n)\sum_{t=1}^{n} u_t$, the OLS estimator of α_1 is:

$$\hat{\alpha}_1 = \alpha_1 + \frac{\sum_{t=1}^{n} u_t y_{t-1} - n\bar{u}\,\bar{y}_{-1}}{\sum_{t=1}^{n} y_{t-1}^2 - n\bar{y}_{-1}^2}. \qquad \textbf{(31)}$$

Since by **(8)**, $\sqrt{n}\bar{u} = \sigma W_n(1)$, and under the null hypothesis $\alpha_1 = 0$ and the maintained hypothesis **(3)**,

$$\bar{y}_{-1}/\sqrt{n} = \frac{1}{n\sqrt{n}}\sum_{t=1}^{n} S_{t-1} = \sigma\int W_n(x)dx, \qquad \textbf{(32)}$$

where the last equality follows from **(11)** with $m = 1$, it follows from Lemma 1, similarly to **(14)** and **(22)** that

$$\hat{\rho}_1 \equiv n\hat{\alpha}_1 = \frac{(1/2)(W_n(1)^2 - 1) - W_n(1)\int W_n(x)dx}{\int W_n(x)^2dx - (\int W_n(x)dx)^2} + o_p(1)$$
$$\to \rho_1 \equiv \frac{(1/2)(W(1)^2 - 1) - W(1)\int W(x)dx}{\int W(x)^2dx - (\int W(x)dx)^2} \text{ in distr.} \qquad \textbf{(33)}$$

The density of ρ_1 is displayed in Figure 3. Comparing Figures 1 and 3, we see that the density of ρ_1 is farther left of zero than the density of ρ_0, and has a fatter left tail.

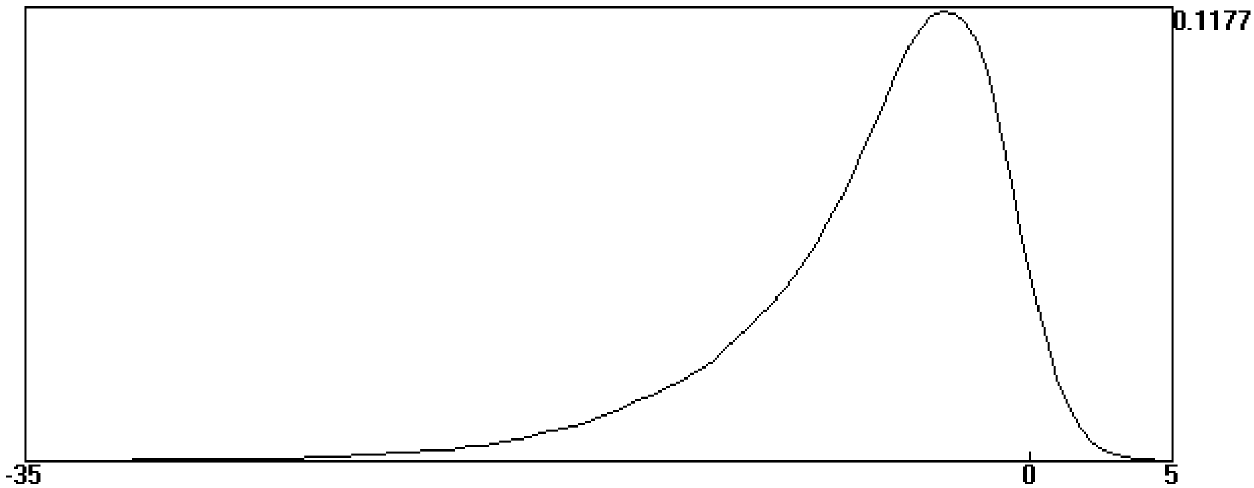

Figure 3: *Density of* ρ_1

As to the t-value $\hat{\tau}_1$ of α_1 in this case, it follows similarly to **(24)** and **(33)** that under the unit root hypothesis,

$$\hat{\tau}_1 \to \tau_1 \equiv \frac{(1/2)(W(1)^2 - 1) - W(1)\int W(x)dx}{\sqrt{\int W(x)^2 dx - (\int W(x)dx)^2}} \ in\ distr. \quad \textbf{(34)}$$

Again, the results **(33)** and **(34)** do not hinge on assumption **(3)**.

The distribution of τ_1 is even farther away from the normal distribution than the distribution of τ_0, as follows from comparison of **(26)** with

$$P(\tau_1 \leq -2.86) = 0.05, \quad P(\tau_1 \leq -2.57) = 0.1 \quad \textbf{(35)}$$

See again Fuller (1996, p. 642). This is corroborated by Figure 4, where the density of τ_1 is compared with the standard normal density.

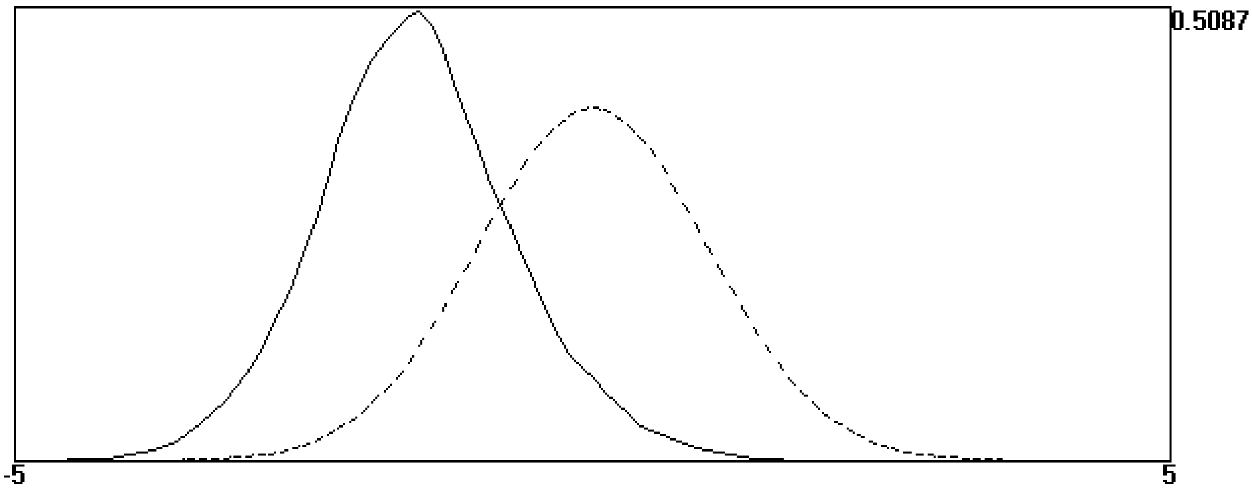

Figure 4: *Density of* τ_1 *compared with the standard normal density* (*dashed curve*)

We see that the density of τ_1 is shifted even more to the left of the standard normal density than in Figure 2, hence the left-sided standard normal test would result in a dramatically higher type 1 error than in the case without an intercept: compare

$$P(\tau_1 \le -1.64) \approx 0.46, \quad P(\tau_1 \le -1.28) \approx 0.64 \qquad \textbf{(36)}$$

with **(26)** and **(27)**.

4. General AR processes with a unit root, and the Augmented Dickey-Fuller test

The assumption made in Sections 2 and 3 that the data-generating process is an AR(1) process, is not very realistic for macroeconomic time series, because even after differencing most of these time series will still display a fair amount of dependence. Therefore we now consider an AR(p) process:

$$y_t = \beta_0 + \Sigma_{j=1}^{p}\beta_j y_{t-j} + u_t, \quad u_t \sim i.i.d.\ N(0,\sigma^2) \qquad \textbf{(37)}$$

By recursively replacing y_{t-j} by $\Delta y_{t-j} + y_{t-1-j}$ for $j = 0,1,..,p$-1, this model can be written as

$$\Delta y_t = \alpha_0 + \Sigma_{j=1}^{p-1}\alpha_j \Delta y_{t-j} + \alpha_p y_{t-p} + u_t, \quad u_t \sim i.i.d.\ N(0,\sigma^2), \qquad \textbf{(38)}$$

where $\alpha_0 = \beta_0$, $\alpha_j = \Sigma_{i=1}^{j}\beta_i - 1$, $j = 1,..,p$.. Alternatively and equivalently, by recursively replacing y_{t-p+j} by $y_{t-p+j+1} - \Delta y_{t-p+j+1}$ for $j = 0,1,..,p$-1, model **(37)** can also be written as

$$\Delta y_t = \alpha_0 + \Sigma_{j=1}^{p-1}\alpha_j \Delta y_{t-j} + \alpha_p y_{t-1} + u_t, \quad u_t \sim i.i.d.\ N(0,\sigma^2), \qquad \textbf{(39)}$$

where now $\alpha_j = -\Sigma_{i=1}^{j}\beta_i, \quad j = 1,..,p-1, \ \alpha_p = \Sigma_{i=1}^{p}\beta_i - 1$.

If the AP(p) process **(37)** has a unit root, then clearly $\alpha_p = 0$ in **(38)** and **(39)**. If the process **(37)** is stationary, i.e., all the roots of the lag polynomial $1 - \Sigma_{j=1}^{p}\beta_j L^j$ lie outside the complex unit circle, then $\alpha_p = \Sigma_{j=1}^{p}\beta_j - 1 < 0$ in **(38)** and **(39)**.[9] The unit root hypothesis can therefore be tested by testing the null hypothesis $\alpha_p = 0$ against the alternative hypothesis $\alpha_p < 0$, using the t-value $\hat{t}_p$ of α_p in model **(38)** or model **(39)**. This test is known as the Augmented Dickey-Fuller (ADF) tests.

We will show now for the case $p = 2$, with intercept under the alternative, i.e.,

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 y_{t-1} + u_t, \quad u_t \sim i.i.d.\ N(0,\sigma^2), \ t = 1,...,n. \qquad \textbf{(40)}$$

that under the unit root (without drift[10]) hypothesis the limiting distribution of $n\hat{\alpha}_p$ is proportional to the limiting distribution in **(33)**, and the limiting distribution of $\hat{t}_p$ is the same as in **(34)**.

Under the unit root hypothesis, i.e., $\alpha_0 = \alpha_2 = 0, \ |\alpha_1| < 1$, we have

$$\begin{aligned}\Delta y_t &= \alpha_1 \Delta y_{t-1} + u_t = (1-\alpha_1 L)^{-1}u_t = (1-\alpha_1)^{-1}u_t + [(1-\alpha_1 L)^{-1} - (1-\alpha_1)^{-1}]u_t \\ &= (1-\alpha_1)^{-1}u_t - \alpha_1(1-\alpha_1)^{-1}(1-\alpha_1 L)^{-1}(1-L)u_t = (1-\alpha_1)^{-1}u_t + v_t - v_{t-1},\end{aligned} \qquad \textbf{(41)}$$

say, where $v_t = -\alpha_1(1-\alpha_1)^{-1}(1-\alpha_1 L)^{-1}u_t = -\alpha_1(1-\alpha_1)^{-1}\Sigma_{j=0}^{\infty}\alpha_1^j u_{t-j}$ is a stationary process. Hence:

$$\begin{aligned}y_t/\sqrt{n} &= y_0/\sqrt{n} + v_t/\sqrt{n} - v_0/\sqrt{n} + (1-\alpha_1)^{-1}(1/\sqrt{n})\Sigma_{j=1}^{t}u_j \\ &= y_0/\sqrt{n} + v_t/\sqrt{n} - v_0/\sqrt{n} + \sigma(1-\alpha_1)^{-1}W_n(t/n)\end{aligned} \qquad \textbf{(42)}$$

and therefore, similarly to **(6)**, **(7)**, and **(32)**, it follows that

[9] To see this, write $1 - \Sigma_{j=1}^{p}\beta_j L^j = \Pi_{j=1}^{p}(1-\rho_j L)$, so that $1 - \Sigma_{j=1}^{p}\beta_j = \Pi_{j=1}^{p}(1-\rho_j)$, where the $1/\rho_j$'s are the roots of the lag polynomial involved. If root $1/\rho_j$ is real valued, then the stationarity condition implies $-1 < \rho_j < 1$, so that $1 - \rho_j > 0$. If some roots are complex-valued, then these roots come in complex-conjugate pairs, say $1/\rho_1 = a+i.b$ and $1/\rho_2 = a-i.b$, hence $(1-\rho_1)(1-\rho_2) = (1/\rho_1-1)(1/\rho_2-1)\rho_1\rho_2 = ((a-1)^2+b^2)/(a^2+b_2) > 0$.

[10] In the sequel we shall suppress the statement "without drift". A unit root process is from now on by default a unit root without drift process, except if otherwise indicated.

$$(1/n)\sum_{t=1}^{n} y_{t-1}/\sqrt{n} \; = \; \sigma(1-\alpha_1)^{-1}\int W_n(x)dx \; + \; o_p(1), \tag{43}$$

$$(1/n^2)\sum_{t=1}^{n} y_{t-1}^2 \; = \; \sigma^2(1-\alpha_1)^{-2}\int W_n(x)^2 dx \; + \; o_p(1), \tag{44}$$

$$\begin{aligned}(1/n)\sum_{t=1}^{n} u_t y_{t-1} \; &= \; (1/n)\sum_{t=1}^{n} u_t\left((1-\alpha_1)^{-1}\sum_{j=1}^{t-1} u_j \; + \; y_0 \; + \; v_{t-1} \; - \; v_0\right) \\ &= \; (1-\alpha_1)^{-1}(1/n)\sum_{t=1}^{n} u_t\sum_{j=1}^{t-1} u_j \; + \; (y_0 - v_0)(1/n)\sum_{t=1}^{n} u_t \; + \; (1/n)\sum_{t=1}^{n} u_t v_{t-1} \\ &= \; \frac{(1-\alpha_1)^{-1}\sigma^2}{2}\left(W_n(1)^2 \; - \; 1\right) \; + \; o_p(1)\end{aligned} \tag{45}$$

Moreover,

$$\operatorname*{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} \Delta y_{t-1} \; = \; E(\Delta y_t) \; = \; 0, \quad \operatorname*{plim}_{n\to\infty}(1/n)\sum_{t=1}^{n} (\Delta y_{t-1})^2 \; = \; E(\Delta y_t)^2 \; = \; \sigma^2/(1-\alpha_1^2) \tag{46}$$

and

$$\begin{aligned}(1/n)\sum_{t=1}^{n} y_{t-1}\Delta y_{t-1} \; &= \; (1/n)\sum_{t=1}^{n} (\Delta y_{t-1})^2 \; + \; (1/n)\sum_{t=1}^{n} y_{t-2}\Delta y_{t-1} \\ &= \; (1/n)\sum_{t=1}^{n} (\Delta y_{t-1})^2 \; + \; \frac{1}{2}\left((1/n)\sum_{t=1}^{n} y_{t-1}^2 \; - \; (1/n)\sum_{t=1}^{n} y_{t-2}^2 \; - \; (1/n)\sum_{t=1}^{n} (\Delta y_{t-1})^2\right) \\ &= \; \frac{1}{2}\left((1/n)\sum_{t=1}^{n} (\Delta y_{t-1})^2 \; + \; y_{n-1}^2/n \; - \; y_{-1}^2/n\right) \; = \; \frac{1}{2}\left(\sigma^2/(1-\alpha_1^2) \; + \; \sigma^2(1-\alpha_1)^{-2}W_n(1)^2\right) \; + \; o_p(1)\end{aligned} \tag{47}$$

hence

$$(1/n)\sum_{t=1}^{n} y_{t-1}\Delta y_{t-1}/\sqrt{n} \; = \; O_p(1/\sqrt{n}). \tag{48}$$

Next, let $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)^T$ be the OLS estimator of $\alpha = (\alpha_0, \alpha_1, \alpha_2)^T$. Under the unit root hypothesis we have

$$\begin{pmatrix} \sqrt{n}\hat{\alpha}_0 \\ \sqrt{n}(\hat{\alpha}_1 - \alpha_1) \\ n\hat{\alpha}_2 \end{pmatrix} \; = \; \sqrt{n}D_n\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xu} \; = \; \left(D_n^{-1}\hat{\Sigma}_{xx}D_n^{-1}\right)^{-1}\sqrt{n}D_n^{-1}\hat{\Sigma}_{xu}, \tag{49}$$

where

$$D_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sqrt{n} \end{pmatrix}, \tag{50}$$

$$\hat{\Sigma}_{xx} = \begin{pmatrix} 1 & (1/n)\Sigma_{t=1}^{n}\Delta y_{t-1} & (1/n)\Sigma_{t=1}^{n} y_{t-1} \\ (1/n)\Sigma_{t=1}^{n}\Delta y_{t-1} & (1/n)\Sigma_{t=1}^{n}(\Delta y_{t-1})^2 & (1/n)\Sigma_{t=1}^{n} y_{t-1}\Delta y_{t-1} \\ (1/n)\Sigma_{t=1}^{n} y_{t-1} & (1/n)\Sigma_{t=1}^{n} y_{t-1}\Delta y_{t-1} & (1/n)\Sigma_{t=1}^{n} y_{t-1}^2 \end{pmatrix}, \tag{51}$$

and

$$\hat{\Sigma}_{xu} = \begin{pmatrix} (1/n)\Sigma_{t=1}^{n} u_t \\ (1/n)\Sigma_{t-1}^{n} u_t\Delta y_{t-1} \\ (1/n)\Sigma_{t=1}^{n} u_t y_{t-1} \end{pmatrix}. \tag{52}$$

It follows from **(43)** through **(48)** that

$$D_n^{-1}\hat{\Sigma}_{xx}D_n^{-1} = \begin{pmatrix} 1 & 0 & \sigma(1-\alpha_1)^{-1}\int W_n(x)dx \\ 0 & \sigma^2/(1-\alpha_1^2) & 0 \\ \sigma(1-\alpha_1)^{-1}\int W_n(x)dx & 0 & \sigma^2(1-\alpha_1)^{-2}\int W_n(x)^2dx \end{pmatrix} + o_p(1), \tag{53}$$

hence, using the easy equality

$$\begin{pmatrix} 1 & 0 & a \\ 0 & b & 0 \\ a & 0 & c \end{pmatrix}^{-1} = \frac{1}{c-a^2}\begin{pmatrix} c & 0 & -a \\ 0 & b^{-1}(c-a^2) & 0 \\ -a & 0 & 1 \end{pmatrix},$$

it follows that

$$\left(D_n^{-1}\hat{\Sigma}_{xx}D_n^{-1}\right)^{-1} = \frac{\sigma^{-2}(1-\alpha_1)^2}{\int W_n(x)^2dx - (\int W_n(x)dx)^2}$$

$$\times \begin{pmatrix} \sigma^2(1-\alpha_1)^{-2}\int W_n(x)^2dx & 0 & -\sigma(1-\alpha_1)^{-1}\int W_n(x)dx \\ 0 & \dfrac{\int W_n(x)^2dx-(\int W_n(x)dx)}{(1-\alpha_1^2)(1-\alpha_1)^2} & 0 \\ -\sigma(1-\alpha_1)^{-1}\int W_n(x)dx & 0 & 1 \end{pmatrix} + o_p(1). \quad \textbf{(54)}$$

Moreover, it follows from **(8)** and **(45)** that

$$\sqrt{n}D_n^{-1}\hat{\Sigma}_{xu} = \begin{pmatrix} \sigma W_n(1) \\ (1/\sqrt{n})\Sigma_{t=1}^n u_t\Delta y_{t-1} \\ \sigma^2(1-\alpha_1)^{-2}(W_n(1)^2-1)/2 \end{pmatrix} + o_p(1). \quad \textbf{(55)}$$

Combining **(49)**, **(54)** and **(55)**, and using Lemma 1, it follows now easily that

$$\frac{n\hat{\alpha}_2}{1-\alpha_1} = \frac{\frac{1}{2}(W_n(1)^2 - 1) - W_n(1)\int W_n(x)dx}{\int W_n(x)^2dx - (\int W_n(x)dx)^2} + o_p(1) \rightarrow \rho_1 \ \ in\ distr., \quad \textbf{(56)}$$

where ρ_1 is defined in **(33)**. Along the same lines it can be shown:

THEOREM 1. *Let y_t be generated by* **(39)**, *and let* $\hat{\alpha}_p$ *be the OLS estimator of* α_p. *Under the unit root hypothesis, i.e.,* $\alpha_p = 0$ *and* $\alpha_0 = 0$, *the following hold*: *If model* **(39)** *is estimated without intercept, then* $n\hat{\alpha}_p \rightarrow (1-\Sigma_{j=1}^{p-1}\alpha_j)\rho_0$ *in distr., where* ρ_0 *is defined in* **(22)**. *If model* **(39)** *is estimated with intercept, then* $n\hat{\alpha}_p \rightarrow (1-\Sigma_{j=1}^{p-1}\alpha_j)\rho_1$ *in distr., where* ρ_1 *is defined in* **(33)**. *Moreover, under the stationarity hypothesis,* $\text{plim}_{n\rightarrow\infty}\hat{\alpha}_p = \alpha_p < 0$, *hence* $\text{plim}_{n\rightarrow\infty}n\hat{\alpha}_p = -\infty$, *provided that in the case where the model is estimated without intercept this intercept,* α_0, *is indeed zero.*

Due to the factor $1 - \Sigma_{j=1}^{p-1}\alpha_j$ in the limiting distribution of $n\hat{\alpha}_p$ under the unit root hypothesis, we cannot use $n\hat{\alpha}_p$ directly as a unit root test. However, it can be shown that under the unit root hypothesis this factor can be consistently estimated by $1 - \sum_{j=1}^{p-1}\hat{\alpha}_j$, hence we can use $n\hat{\alpha}_p/|1 - \sum_{j=1}^{p-1}\hat{\alpha}_j|$ as a unit root test statistic, with limiting distribution given by **(22)** or **(33)**. The reason for the absolute value is that under the alternative of stationarity the probability limit of $1 - \sum_{j=1}^{p-1}\hat{\alpha}_j$ may be negative[11].

The actual ADF test is based on the t-value of α_p, because the factor $1 - \sum_{j=1}^{p-1}\alpha_j$ will cancel out in the limiting distribution involved. We will show this for the AR(2) case.

First, it is not too hard to verify from **(43)** through **(48)**, and **(54)**, that the residual sum of squares RSS of the regression **(40)** satisfies:

$$RSS = \sum_{t=1}^{n} u_t^2 + O_p(1). \qquad \textbf{(57)}$$

This result carries over to the general AR(p) case, and also holds under the stationarity hypothesis. Moreover, under the unit root hypothesis it follows easily from **(54)** and **(57)** that the OLS standard error, s_2, say, of $\hat{\alpha}_2$ in model **(40)** satisfies:

$$ns_2 = \sqrt{\frac{\big(RSS/(n-3)\big)\sigma^{-2}(1-\alpha_1)^2}{\int W_n(x)^2dx - (\int W_n(x)dx)^2}} + o_p(1) = \frac{1-\alpha_1}{\sqrt{\int W_n(x)^2dx - (\int W_n(x)dx)^2}} + o_p(1), \qquad \textbf{(58)}$$

hence it follows from **(56)** that the t-value $\hat{t}_2$ of $\hat{\alpha}_2$ in model **(40)** satisfies **(34)**. Again, this result carries over to the general AR(p) case:

THEOREM 2. *Let y_t be generated by* **(39)**, *and let $\hat{t}_p$ be t-value of the OLS estimator of α_p. Under the unit root hypothesis, i.e., $\alpha_p = 0$ and $\alpha_0 = 0$, the following hold: If model* **(39)** *is estimated without intercept, then $\hat{t}_p \to \tau_0$ in distr., where τ_0 is defined in* **(24)**. *If model* **(39)** *is estimated with intercept, then $\hat{t}_p \to \tau_1$ in distr., where τ_1 is defined in* **(34)**. *Moreover, under the stationarity*

---

[11] For example, let $p = 2$ in **(37)** and **(39)**. Then $\alpha_1 = -\beta_1$, hence if $\beta_1 < -1$ then $1 - \alpha_1 < 0$. In order to show that $\beta_1 < -1$ can be compatible with stationarity, assume that $\beta_1^2 = 4\beta_2$, so that the lag polynomial $1 - \beta_1 L - \beta_2 L^2$ has two common roots $-2/|\beta_1|$. Then the AR(2) process involved is stationary for $-2 < \beta_1 < -1$.

hypothesis, $\text{plim}_{n\to\infty}\hat{t}_p/\sqrt{n} < 0$, *hence* $\text{plim}_{n\to\infty}\hat{t}_p = -\infty$, *provided that in the case where the model is estimated without intercept this intercept,* α_0, *is indeed zero.*

5. ARIMA processes, and the Phillips-Perron test

The ADF test requires that the order p of the AR model involved is finite, and correctly specified, i.e., the specified order should not be smaller than the actual order. In order to analyze what happens if p is misspecified, suppose that the actual data-generating process is given by **(39)** with $\alpha_0 = \alpha_2 = 0$ and $p > 1$, and that the unit root hypothesis is tested on the basis of the assumption that $p = 1$. Denoting $e_t = u_t/\sigma$, model **(39)** with $\alpha_0 = \alpha_2 = 0$ can be rewritten as

$$\Delta y_t = (\textstyle\sum_{j=0}^{\infty}\gamma_j L^j)e_t = \gamma(L)e_t, \quad e_t \sim i.i.d.\ N(0,1), \tag{59}$$

where $\gamma(L) = \alpha(L)^{-1}$, with $\alpha(L) = 1 - \sum_{j=1}^{p-1}\alpha_j L^j$. This data-generating process can be nested in the auxiliary model

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \quad u_t = \gamma(L)e_t,\ e_t \sim i.i.d.\ N(0,1). \tag{60}$$

We will now determine the limiting distribution of the OLS estimate $\hat{\alpha}_1$ and corresponding t-value $\hat{t}_1$ of the parameter α_1 in the regression **(60)**, derived under the assumption that the u_t's are independent, while in reality **(59)** holds.

Similarly to **(41)** we can write $\Delta y_t = \gamma(1)e_t + v_t - v_{t-1}$, where $v_t = [(\gamma(L)-\gamma(1))/(1-L)]e_t$ is a stationary process. The latter follows from the fact that by construction the lag polynomial $\gamma(L)-\gamma(1)$ has a unit root, and therefore contains a factor 1-L. Next, redefining $W_n(x)$ as

$$W_n(x) = (1/\sqrt{n})\sum_{t=1}^{[nx]} e_t \quad if\ x \in [n^{-1},1], \quad W_n(x) = 0 \quad if\ x \in [0,n^{-1}), \tag{61}$$

it follows similarly to **(42)** that

$$y_t/\sqrt{n} = y_0/\sqrt{n} + v_t/\sqrt{n} - v_0/\sqrt{n} + \gamma(1)W_n(t/n), \tag{62}$$

hence

$$y_n/\sqrt{n} = \gamma(1)W_n(1) + O_p(1/\sqrt{n}), \tag{63}$$

and similarly to **(43)** and **(44)** that

$$\bar{y}_{-1}/\sqrt{n} = \frac{1}{n}\sum_{t=1}^{n} y_{t-1}/\sqrt{n} = \gamma(1)\int W_n(x)dx + o_p(1), \tag{64}$$

and

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 = \gamma(1)^2\int W_n(x)^2 dx + o_p(1). \tag{65}$$

Moreover, similarly to **(6)** we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}(\Delta y_t)y_{t-1} &= \frac{1}{2}\left(y_n^2/n - y_0^2/n - \frac{1}{n}\sum_{t=1}^{n}(\Delta y_t)^2\right)\\ = \frac{1}{2}\left(\gamma(1)^2 W_n(1)^2 - \frac{1}{n}\sum_{t=1}^{n}(\gamma(L)e_t)^2\right) &+ o_p(1) = \gamma(1)^2\frac{1}{2}(W_n(1) - \lambda) + o_p(1),\end{aligned} \tag{66}$$

where

$$\lambda = \frac{E(\gamma(L)e_t)^2}{\gamma(1)^2} = \frac{\Sigma_{j=0}^{\infty}\gamma_j^2}{(\Sigma_{j=0}^{\infty}\gamma_j)^2}. \tag{67}$$

Therefore, **(33)** now becomes:

$$n\hat{\alpha}_1 = \frac{(1/2)(W_n(1)^2 - \lambda) - W_n(1)\int W_n(x)dx}{\int W_n(x)^2 dx - (\int W_n(x)dx)^2} + o_p(1) \to \rho_1 + \frac{0.5(1-\lambda)}{\int W(x)^2 dx - (\int W(x)dx)^2} \tag{68}$$

in distr., and **(34)** becomes:

$$\hat{t}_1 = \frac{(1/2)(W_n(1)^2 - \lambda) - W_n(1)\int W_n(x)dx}{\sqrt{\int W_n(x)^2 dx - (\int W_n(x)dx)^2}} + o_p(1) \to \tau_1 + \frac{0.5(1-\lambda)}{\sqrt{\int W(x)^2 dx - (\int W(x)dx\,t)^2}} \tag{69}$$

in distr. These results carry straightforwardly over to the case where the actual data-generating process is an ARIMA process $\alpha(L)\Delta y_t = \beta(L)e_t$, simply by redefining $\gamma(L) = \beta(L)/\alpha(L)$.

The parameter $\gamma(1)^2$ is known as the long-run variance of $u_t = \gamma(L)e_t$:

$$\sigma_L^2 = \lim_{n\to\infty} var[(1/\sqrt{n})\Sigma_{t=1}^{n}u_t] = \gamma(1)^2 \tag{70}$$

which in general is different from the variance of u_t itself:

$$\sigma_u^2 = var(u_t) = E(u_t^2) = E(\Sigma_{j=0}^{\infty}\gamma_j e_{t-j})^2 = \Sigma_{j=0}^{\infty}\gamma_j^2. \tag{71}$$

If we would know σ_L^2 and σ_u^2, and thus $\lambda = \sigma_u^2/\sigma_L^2$, then it follows from **(64)**, **(65)**, and Lemma 1, that

$$\frac{\sigma_L^2 - \sigma_u^2}{(1/n^2)\Sigma_{t=1}^n (y_{t-1} - \bar{y}_{-1})^2} \rightarrow \frac{1 - \lambda}{\int W(x)^2 dx - (\int W(x)dx)^2} \text{ in distr.} \qquad \textbf{(72)}$$

It is an easy exercise to verify that this result also holds if we replace $y_{t\text{-}1}$ by y_t and $\bar{y}_{-1}$ by $\bar{y} = (1/n)\Sigma_{t=1}^n y_t$. Therefore it follows from **(68)** and **(72)** that,

THEOREM 3. (*Phillips-Perron test 1*) *Under the unit root hypothesis, and given consistent estimators* $\hat{\sigma}_L^2$ *and* $\hat{\sigma}_u^2$ *of* σ_L^2 *and* σ_u^2, *respectively, we have*

$$\hat{Z}_1 = n\left(\hat{\alpha}_1 - \frac{(\hat{\sigma}_L^2 - \hat{\sigma}_u^2)/2}{(1/n)\Sigma_{t=1}^n (y_t - \bar{y})^2}\right) \rightarrow \rho_1 \text{ in distr.} \qquad \textbf{(73)}$$

This correction of **(68)** has been proposed by Phillips and Perron (1988) for particular estimators $\hat{\sigma}_L^2$ and $\hat{\sigma}_u^2$, following the approach of Phillips (1987) for the case where the intercept α_0 in **(60)** is assumed to be zero.

It is desirable to choose the estimators $\hat{\sigma}_L^2$ and $\hat{\sigma}_u^2$ such that under the stationarity alternative, $\text{plim}_{n\rightarrow\infty}\hat{Z}_1 = -\infty$. We show now that this is the case if we choose

$$\hat{\sigma}_u^2 = \frac{1}{n}\sum_{t=1}^n \hat{u}_t^2, \quad \text{where } \hat{u}_t = \Delta y_t - \hat{\alpha}_0 - \hat{\alpha}_1 y_{t-1}, \qquad \textbf{(74)}$$

and $\hat{\sigma}_L^2$ such that $\bar{\sigma}_L^2 = \text{plim}_{n\rightarrow\infty}\hat{\sigma}_L^2 \geq 0$ under the alternative of stationarity.

First, it is easy to verify that $\hat{\sigma}_u^2$ is consistent under the null hypothesis, by verifying that **(57)** still holds. Under stationarity we have $\text{plim}_{n\rightarrow\infty}\hat{\alpha}_1 = cov(y_t, y_{t-1})/var(y_t) - 1 = \alpha_1^*$, say, $\text{plim}_{n\rightarrow\infty}\hat{\alpha}_0 = -\alpha_1^* E(y_t) = \alpha_0^*$, say, and $\text{plim}_{n\rightarrow\infty}\hat{\sigma}_u^2 = (1 - (\alpha_1^*+1)^2)var(y_t) = \sigma_*^2$, say. Therefore,

$$\text{plim}_{n\rightarrow\infty}\hat{Z}_1/n = -0.5(\alpha_1^{*\,2} + \bar{\sigma}_L^2/var(y_t)) < 0. \qquad \textbf{(75)}$$

Phillips and Perron (1988) propose to estimate the long-run variance by the Newey-West (1987) estimator

$$\hat{\sigma}_L^2 \ = \ \hat{\sigma}_u^2 \ + \ 2\sum_{i=1}^{m} [1 \ - \ i/(m+1)](1/n) \sum_{t=i+1}^{n} \hat{u}_t \hat{u}_{t-i}, \tag{76}$$

where $\hat{u}_t$ is defined in **(74)**, and m converges to infinity with n at rate $o(n^{1/4})$. Andrews (1991) has shown (and we will show it again along the lines in Bierens (1994)) that the rate $o(n^{1/4})$ can be relaxed to $o(n^{1/2})$. The weights $1 - j/(m+1)$ guarantee that this estimator is always positive. The reason for the latter is the following. Let $u_t^* = u_t$ for $t = 1,..,n$, and $u_t^* = 0$ for $t < 1$ and $t > n$. Then,

$$\hat{\sigma}_L^{*2} \ \equiv \ \frac{1}{n}\sum_{t=1}^{n+m}\left(\frac{1}{\sqrt{m+1}}\sum_{j=0}^{m} u_{t-j}^*\right)^2 \ = \ \frac{1}{m+1}\sum_{j=0}^{m}\frac{1}{n}\sum_{t=1}^{n+m} u_{t-j}^{*2} \ + \ 2\frac{1}{m+1}\sum_{j=0}^{m-1}\sum_{i=1}^{m-j}\frac{1}{n}\sum_{t=1}^{n+m} u_{t-j}^* u_{t-j-i}^*$$

$$= \ \frac{1}{m+1}\sum_{j=0}^{m}\frac{1}{n}\sum_{t=1-j}^{n+m-j} u_t^{*2} \ + \ 2\frac{1}{m+1}\sum_{j=0}^{m-1}\sum_{i=1}^{m-j}\frac{1}{n}\sum_{t=1-j}^{n+m-j} u_t^* u_{t-i}^* \tag{77}$$

$$= \ \frac{1}{n}\sum_{t=1}^{n} u_t^2 \ + \ 2\frac{1}{m+1}\sum_{j=0}^{m-1}\sum_{i=1}^{m-j}\frac{1}{n}\sum_{t=i+1}^{n} u_t u_{t-i} \ = \ \frac{1}{n}\sum_{t=1}^{n} u_t^2 \ + \ 2\frac{1}{m+1}\sum_{i=1}^{m}(m+1-i)\frac{1}{n}\sum_{t=i+1}^{n} u_t u_{t-i}$$

is positive, and so is $\hat{\sigma}_L^2$. Next, observe from **(62)** and **(74)** that

$$\hat{u}_t \ = \ u_t \ - \ \sqrt{n}\hat{\alpha}_1\gamma(1)W_n(t/n) \ - \ \hat{\alpha}_1 v_t \ + \ \hat{\alpha}_1(v_0 - y_0) \ - \ \hat{\alpha}_0. \tag{78}$$

Since

$$E|(1/n)\Sigma_{t=1+i}^{n} u_t W_n((t-i)/n)| \ \leq \ \sqrt{(1/n)\Sigma_{t=1+i}^{n} E(u_t^2)}\sqrt{(1/n)\Sigma_{t=1+i}^{n} E(W_n((t-i)/n)^2)} \ = \ O(1),$$

it follows that $(1/n)\Sigma_{t=1+i}^{n} u_t W_n((t-i)/n) = O_p(1)$. Similarly, $(1/n)\Sigma_{t=1+i}^{n} u_{t-i} W_n(t/n) = O_p(1)$. Moreover, $\hat{\alpha}_1 = O_p(1/n)$, and similarly, it can be shown that $\hat{\alpha}_0 \ = \ O_p(1/\sqrt{n})$. Therefore, it follows from **(77)** and **(78)** that

$$\hat{\sigma}_L^2 \ - \ \hat{\sigma}_L^{*2} \ = \ O_p(1/n) \ + \ O_p\left(\sum_{i=1}^{m}[1 - i/(m+1)]/\sqrt{n}\right) \ = \ O_p(1/n) \ + \ O_p(m/\sqrt{n}). \tag{79}$$

A similar result holds under the stationarity hypothesis. Moreover, substituting $u_t = \sigma_L^2 e_t + v_t - v_{t-1}$, and denoting $e_t^* = e_t$, $v_t^* = v_t$ for $t = 1,..,n$, $v_t^* = e_t^* = 0$ for $t < 1$ and $t > n$, it is easy to verify that under the unit root hypothesis,

$$\hat{\sigma}_L^{*2} = \frac{1}{n}\sum_{t=1}^{n+m}\left(\sigma_L\frac{1}{\sqrt{m+1}}\sum_{j=0}^{m} e_{t-j}^{*} + \frac{v_t^{*}-v_{t-m}^{*}}{\sqrt{m+1}}\right)^2$$

$$= \sigma_L^2\frac{1}{n}\sum_{t=1}^{n+m}\left(\frac{1}{\sqrt{m+1}}\sum_{j=0}^{m} e_{t-j}^{*}\right)^2 + 2\sigma_L\frac{1}{n}\sum_{t=1}^{n+m}\left(\frac{1}{\sqrt{m+1}}\sum_{j=0}^{m} e_{t-j}^{*}\right)\left(\frac{v_t^{*}-v_{t-m}^{*}}{\sqrt{m+1}}\right) \qquad \textbf{(80)}$$

$$+ \frac{1}{n}\sum_{t=1}^{n+m}\left(\frac{v_t^{*}-v_{t-m}^{*}}{\sqrt{m+1}}\right)^2 = \sigma_L^2 + O_p(\sqrt{m/n}) + O_p(1/\sqrt{m}) + O_p(1/m).$$

A similar result holds under the stationarity hypothesis. Thus:

THEOREM 4. *Let m increase with n to infinity at rate* $o(n^{1/2})$. *Then under both the unit root and stationarity hypothesis,* $\text{plim}_{n\to\infty}(\hat{\sigma}_L^2 - \hat{\sigma}_L^{*2}) = 0$. *Moreover, under the unit root hypothesis,* $\text{plim}_{n\to\infty}\hat{\sigma}_L^{*2} = \sigma_L^2$, *and under the stationarity hypothesis,* $\text{plim}_{n\to\infty}\hat{\sigma}_L^{*2} > 0$. *Consequently, under stationarity, the Phillips-Perron test satisfies* $\text{plim}_{n\to\infty}\hat{Z}_1/n < 0$.

Finally, note that the advantage of the PP test is that there is no need to specify the ARIMA process under the null hypothesis. It is in essence a nonparametric test. Of course, we still have to specify the Newey-West truncation lag m as a function of n, but as long as $m = o(\sqrt{n})$, this specification is asymptotically not critical.

6. Unit root with drift versus trend stationarity

Most macroeconomic time series in (log) levels have an upwards sloping pattern. Therefore, if they are (covariance) stationary, then they are stationary around a deterministic trend. If we would conduct the ADF and PP tests in Sections 4 and 5 to a linear trend stationary process, we will likely accept the unit root hypothesis, due to the following. Suppose we conduct the ADF test under the hypothesis $p = 1$ to the trend stationary process $y_t = \beta_0 + \beta_1 t + u_t$, where the u_t's are i.i.d. $N(0,\sigma^2)$. It is a standard exercise to verify that then $\text{plim}_{n\to\infty} n\hat{\alpha}_1 = 0$, hence the ADF and PP tests in sections 4 and 5 have no power against linear trend stationarity!

Therefore, if one wishes to test the unit root hypothesis against linear trend stationarity, then a trend term should be included in the auxiliary regressions **(39)** in the ADF case, and in **(60)** in the

PP case: Thus the ADF regression **(39)** now becomes

$$\Delta y_t = \alpha_0 + \Sigma_{j=1}^{p-1}\alpha_j\Delta y_{t-1} + \alpha_p y_{t-1} + \alpha_{p+1}t + u_t, \quad u_t \sim i.i.d.\ N(0,\sigma_2) \tag{81}$$

where the null hypothesis of a unit root with drift corresponds to the hypothesis $\alpha_p = \alpha_{p+1} = 0$, and the PP regression becomes:

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + u_t, \quad u_t = \gamma(L)e_t,\ e_t \sim i.i.d.\ N(0,1). \tag{82}$$

The asymptotic null distributions of the ADF and PP tests for the case with drift are quite similar to the ADF test without an intercept. The difference is that the Wiener process $W(x)$ is replaced by the detrended Wiener process:

$$W^{**}(x) = W(x) - 4\int W(z)dz + 6\int zW(z)dz + 6x\left(\int W(z)dz - 2\int zW(z)dz\right)x$$

After some tedious but not too difficult calculations it can be shown that effectively the statistics $n\hat{\alpha}_p/(1-\Sigma_{j=1}^{p}\alpha_j)$ and $\hat{t}_p$ are asymptotically equivalent to the Dickey-Fuller tests statistics $\hat{\rho}_0$ and $\hat{\tau}_0$, respectively, applied to detrended time series.

THEOREM 5. *Let y_t be generated by* **(81)**, *and let $\hat{\alpha}_p$ and $\hat{t}_p$ be the OLS estimator and corresponding t-value of α_p. Under the unit root with drift hypothesis, i.e.,* $\alpha_p = \alpha_{p+1} = 0$, *we have* $n\hat{\alpha}_p \to (1-\Sigma_{j=1}^{p-1}\alpha_j)\rho_2$ *and $\hat{t}_p \to \tau_2$ in distr., where*

$$\rho_2 = \frac{1}{2}\left(\frac{W^{**}(1) - 1}{\int W^{**}(x)^2dx}\right), \quad \tau_2 = \frac{1}{2}\left(\frac{W^{**}(1) - 1}{\sqrt{\int W^{**}(x)^2dx}}\right).$$

Under the trend stationarity hypothesis, $\text{plim}_{n\to\infty}\hat{\alpha}_p = \alpha_p < 0$, *hence* $\text{plim}_{n\to\infty}\hat{t}_p/\sqrt{n} < 0$.

The densities of ρ_2 and τ_2 (the latter compared with the standard normal density), are displayed in Figures 5 and 6, respectively. Again, these densities are farther to the left, and heavier left-tailed, than the corresponding densities displayed in Figures 1-4. The asymptotic 5% and 10% critical values of the Dickey-Fuller t-test are:

$$P(\tau_2 < -3.41) = 0.05, \quad P(\tau_2 < -3.13) = 0.10$$

Moreover, comparing **(26)** with

$$P(\tau_2 \leq -1.64) \approx 0.77, \quad P(\tau_2 \leq -1.28) \approx 0.89,$$

we see that the standard normal tests at the 5% and 10% significance level would reject the correct unit root with drift hypothesis with probabilities of about 0.77 and 0.89, respectively!

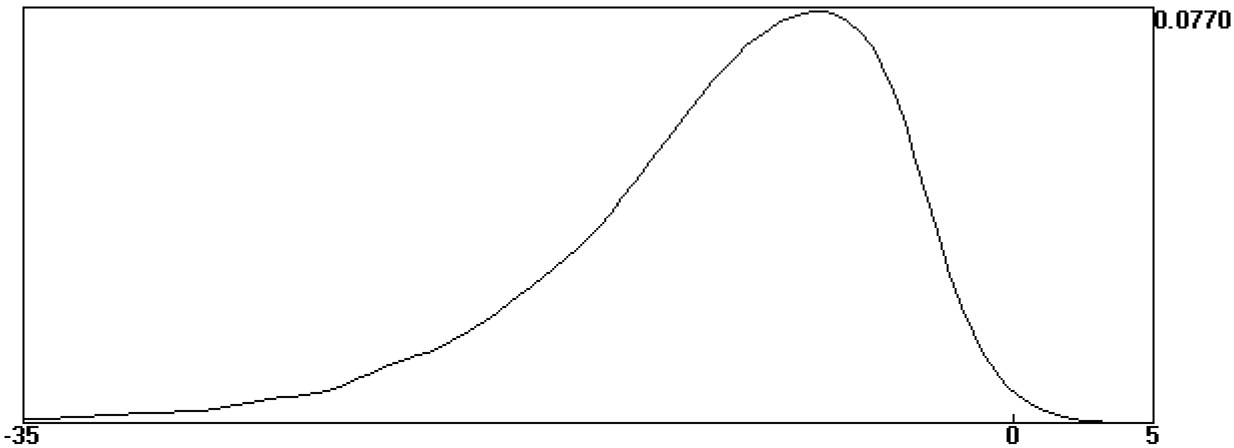

Figure 5: *Density of* ρ_2

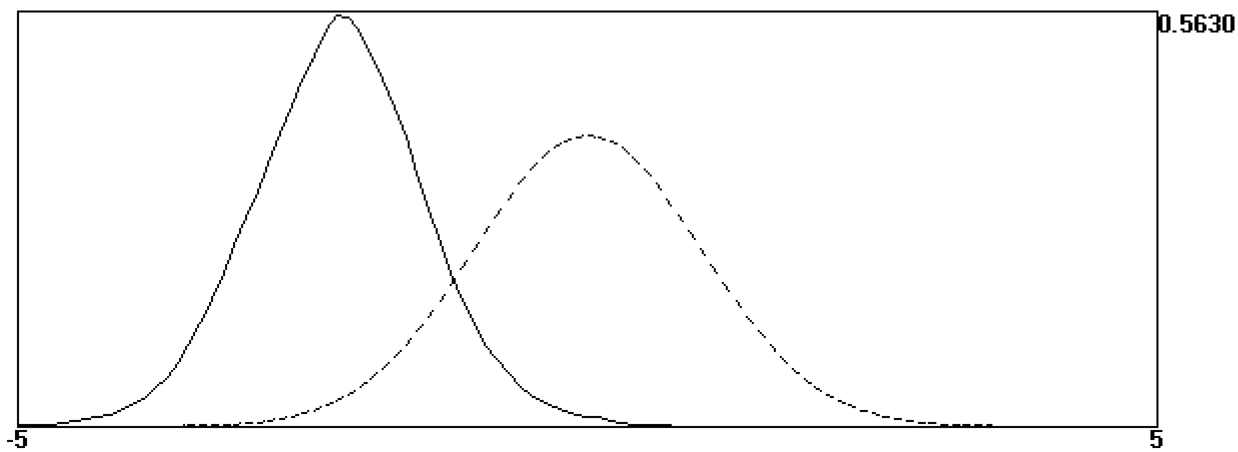

Figure 6: *Density of* τ_2 *compared with the standard normal density* (*dashed curve*)

A similar result as in Theorem 5 can be derived for the PP test, on the basis of the OLS estimator of α_1 in, and the residuals $\hat{u}_t$ of, the auxiliary regression **(82)**:

THEOREM 6. (*Phillips-Perron test 2*) *Let* $\hat{r}_t$ *be the residuals of the OLS regression of* y_t *on t and a constant, and let* $\hat{\sigma}_u^2$ *and* $\hat{\sigma}_L^2$ *be as before, with the* $\hat{u}_t$*'s the OLS residuals of the auxiliary* *regression* **(82)**. *Under the unit root with drift hypothesis,*

$$\hat{Z}_2 = n\left(\hat{\alpha}_1 - \frac{(\hat{\sigma}_L^2 - \hat{\sigma}_u^2)/2}{(1/n)\Sigma_{t=1}^n \hat{r}_t^2}\right) \to \rho_2 \ \textit{in distr.}, \qquad \textbf{(83)}$$

whereas under trend stationarity $\text{plim}_{n\to\infty}\hat{Z}_2/n < 0$.

7. Concluding remarks

In the discussion of the ADF test we have assumed that the lag length p of the auxiliary regression **(81)** is fixed. It should be noted that we may choose p as a function of the length n of the time series involved, similarly to the truncation width of the Newey-West estimator of the long-run variance in the Phillips-Perron test. See Said and Dickey (1984).

We have seen that the ADF and Phillips-Perron tests for a unit root against stationarity around a constant have almost no power if the correct alternative is linear trend stationarity. However, the same may apply to the tests discussed in section 6 if the alternative is trend stationarity with a broken trend. See Perron (1988,1989,1990), Perron and Vogelsang (1992), and Zivot and Andrews (1992), among others.

All the tests discussed so far have the unit root as the null hypothesis, and (trend) stationarity as the alternative. However, it is also possible to test the other way around. See Bierens and Guo (1993), and Kwiatkowski et.al. (1992). The latter test is known as the KPSS test.

Finally, note that the ADF and Phillips-Perron tests can easily be conducted by various econometric software packages, for example the commercial software packages TSP, EViews, RATS, and my freeware *EasyReg International*.[12]

[12] *EasyReg International* can be downloaded from URL
http://econ.la.psu.edu/~hbierens/EASYREG.HTM.
EasyReg International also contains my own unit root tests, Bierens (1993,1997), Bierens and Guo (1993), and the KPSS test.

References

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimators, *Econometrica* 59, 817-858.

Bierens, H.J., 1993, Higher order sample autocorrelations and the unit root hypothesis, *Journal of Econometrics* 57, 137-160.

Bierens, H.J., 1997, Testing the unit root hypothesis against nonlinear trend stationarity, with an application to the price level and interest rate in the U.S, *Journal of Econometrics* 81, 29-64.

Bierens, H.J., 1994, *Topics in advanced econometrics: estimation, testing and specification of cross-section and time series models* (Cambridge University Press, Cambridge, U.K.).

Bierens, H.J. and S. Guo, 1993, Testing stationarity and trend stationarity against the unit root hypothesis, *Econometric Reviews* 12, 1-32.

Billingsley (1968), P., 1968, *Convergence of probability measures* (John Wiley, New York).

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive times series with a unit root, *Journal of the American Statistical Association* 74, 427-431.

Dickey, D.A. and W.A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057-1072.

Fuller, W.A., 1996, *Introduction to statistical time series* (John Wiley, New York).

Green, W., 1997, *Econometric analysis* (Prentice Hall, Upper Saddle River, NJ).

Hogg, R.V. and A.T. Craig, 1978, *Introduction to mathematical statistics* (Macmillan, London).

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, 1992, Testing the null of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159-178.

Newey, W.K. and K.D. West, 1987, A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Perron, P., 1988, Trends and random walks in macroeconomic time series: further evidence from a new approach, *Journal of Economic Dynamics and Control* 12, 297-332.

Perron, P., 1989, The great crash, the oil price shock and the unit root hypothesis, *Econometrica* 57, 1361-1402.

Perron, P., 1990, Testing the unit root in a time series with a changing mean, *Journal of Business and Economic Statistics* 8, 153-162.

Perron, P. and T.J. Vogelsang, 1992, Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics* 10, 301-320.

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 of unknown order, *Biometrika* 71, 599-607.

Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the great crash, the oil price shock, and the unit root hypothesis, *Journal of Business and Economic Statistics* 10, 251-270.