Pontificia Universidad Católica del Perú

Escuela de Posgrado

Functional Central Limit Theorems and Unit Root Testing

Tesis para optar el grado académico de:

Magister en Matemáticas Aplicadas

Autor:

Juan Carlos Aquino

Asesor:

Gabriel Hender Rodríguez Briones

Miembros del jurado

Loretta Betzabe Rosa Gasco Campos

Luis Hilmar Valdivieso Serrano

Lima, 2011

Functional Central Limit Theorems and Unit Root Testing

Juan Carlos Aquino
*Pontificia Universidad Católica del Perú**

Abstract

This paper analyzes and employs two versions of the Functional Central Limit Theorem within the framework of a unit root with a structural break. Initial attention is focused on the probabilistic structure of the time series to be considered. Later, attention is placed on the asymptotic theory for nonstationary time series proposed by Phillips (1987a), which is applied by Perron (1989) to study the effects of an (assumed) exogenous structural break on the power of the augmented Dickey-Fuller test and by Zivot and Andrews (1992) to criticize the exogeneity assumption and propose a method for estimating an endogenous breakpoint. A systematic method for dealing with efficiency issues is introduced by Perron and Rodríguez (2003), which extends the Generalized Least Squares detrending approach due to Elliott, Rothenberg, and Stock (1996).

Resumen

Este documento analiza y usa dos versiones del Teorema del Límite Central Funcional y su aplicación al contexto de raices unitarias con un quiebre estructural. La atención inicial se enfoca en la estructura probabilística de las series de tiempo a considerarse. Luego, la atención se situa en la teoría asintótica para series de tiempo no estacionarias propuesta por Phillips (1987a), la cual es aplicada por Perron (1989) para estudiar los efectos de un quiebre estructural (asumido) exógeno sobre la potencia de la prueba Dickey-Fuller aumentada y por Zivot y Andrews (1992) para criticar el supuesto de exogeneidad y proponer un método para estimar el punto de quiebre de manera endógena. Un método sistemático para abordar aspectos de eficiencia es introducido por Perron y Rodríguez (2003), quienes extienden el enfoque de extracción de tendencia por Mínimos Cuadrados Generalizados atribuido a Elliott, Rothenberg, y Stock (1996).

JEL Clasification: C12, C22.
Keywords: Hypothesis testing, unit root, structural break, Functional Central Limit Theorem, weak convergence, Wiener process, Ornstein-Uhlenbeck process.

---

*Master's Program in Applied Mathematics.

Contents

List of Tables

List of Figures

1 Introduction

Four decades ago, the empirical study of key macroeconomic variables has been done through the use of ARMA models proposed by Box and Jenkins (1970). In these type of models, first and second moments depend upon time separation but do not depend on the time variable. Hence, these models are covariance stationary[1], whose behaviour reverts to a time invariant unconditional mean and whose methodology is based on the steps of identification, estimation and diagnostic[2].

Consider, for example, the case of a series $\{y_t\}_{t=0}^{T}$ that obeys a first order autoregressive process: $y_t = \mu + \alpha y_{t-1} + u_t$, $t = 1, \ldots, T$, where μ is a constant, $|\alpha| < 1$, y_0 is an initial condition, $u_t \sim N(0, \sigma_u^2)$ and $\sigma_u^2 > 0$. A first conclusion to be extracted is that shocks have an effect on the dependent variable that vanishes as time elapses, an assertion that can be confirmed after manipulating the previous expression: $y_t = \alpha^t y_0 + \mu \sum_{i=1}^{t} \alpha^{t-i} + \sum_{i=1}^{t} \alpha^{t-i} u_i$ or, equivalently, realization of an arbitrary u_i has a decreasing effect on y_t as t increases.

However, assumptions underlying ARMA models are not adequate for modelling macroeconomic series, which exhibit an upward trend along time. Hence, any model that aims to represent macroeconomic data must include such a trend. One of the most popular approachs for this task is the deterministic trend model: $y_t = \mu + \delta t + u_t$, $t = 1, \ldots, T$, μ and δ are constants, $u_t \sim N(0, \sigma_u^2)$ and $\sigma_u^2 > 0$. Since a stationary process is obtained after substracting δt this process is called trend stationary. Notice also that each realization of u_t only has a contemporaneous effect on y_t.

An alternative approach considers the data generating process as an autoregressive one containing a unit root: $y_t = \mu + \alpha y_{t-1} + u_t$, $t = 1, \ldots, T$, μ is a constant, $\alpha = 1$, y_0 is an initial condition, $u_t \sim N(0, \sigma_u^2)$ and $\sigma_u^2 > 0$. In this case, $y_t = y_0 + \mu t + \sum_{i=1}^{t} u_i$ or, equivalently, realization of any u_i has a permanent effect on the level of y_t and the adequate procedure to obtain a stationary series is to work in first differences $\Delta y_t \equiv y_t - y_{t-1}$.

From an economic viewpoint, all of these observations make it necessary to identify the type of process representing macroeconomic data and to understand the long run effects of shocks. Also, based on a predictive perspective this distinction is nontrivial since in the deterministic case the forecasting error has a constant variability whereas in the stochastic case this element has an increasing variability[3].

For these reasons, once provided with macroeconomic data, researchers frequently employ statistical inference procedures to make this distinction. Specifically, one of them consists of stating a (null) hypothesis H_0 about the nature of the data generating process, as opposite to a competing (alternative) hypothesis H_1. The hypothesis attention will be focused on states the autoregressive coefficient as equal to one ($\alpha = 1$) and will be refered as the unit root hyphotesis. Thus, logics behind hypothesis testing is based on the following reasoning: if H_0 were in fact true then any inconsistency with this hypothesis is not likely (although not impossible) to ocurr, so that the probability of incurring in

[1] Hereafter, any reference to a stationary process will be in this sense.
[2] See Enders (2004) for an applied approach to this methodology
[3] See Hamilton (1994) for further details.

	H_0 is true	H_1 is true
H_0 is not rejected	No error	Type II error
H_0 is rejected	Type I error	No error

Table 1: Error types in statistical inference.

Type I error (reject H_0 when it is true) is conventionally set to $0,05$. If an inconsistency is found, this leaves to rejecting H_0 in favor of H_1. Additionally, it is desirable that the probability of incurring in Type II error (avoid rejecting H_0 when H_1 is true) be as lower as possible (see Table 1).

Turning back to the empirical level, the previous framework allows to consider the following autoregressive model (without any deterministic component)

$$y_t = \alpha y_{t-1} + u_t \tag{1}$$

and test

$$H_0 : \alpha = 1 \text{ against } H_1 : |\alpha| < 1. \tag{2}$$

The study of White (1958) is the first one to perform such a procedure: in order to test H_0 with a sample of size T and OLS estimator $\hat{\alpha}$ for parameter α, under the null, it is obtained that

$$\frac{T}{\sqrt{2}}(\hat{\alpha} - 1) \Rightarrow \frac{\int_0^1 W(r) dW(r)}{\int_0^1 W(r)^2 dr} = \frac{1}{2} \frac{W(1)^2 - 1}{\int_0^1 W(r)^2 dr}. \tag{3}$$

In the previous expression $(T/\sqrt{2})(\hat{\alpha} - 1)$ denotes centered and standarized estimator for α, a random variable, and $\Rightarrow$ denotes weak convergence of probability measures. This result is an application of a theorem due to Donsker (1951) and the asymptotic distribution is formulated in terms of functionals of a standard Wiener process W whose details and properties are examined. It is worth to mention that this result is not independent of the correlation between disturbance terms u_t (assumed to be zero) and the fact that there is no specification error when estimating α.

Other study in this line is due to Dickey and Fuller (1979), who assume normal i.i.d. disturbances and develop several one-tailed tests with the following rejection rule: for a given confidence level, if the (properly transformed) centered estimator $\hat{\alpha} - 1$ is low relative to a critical value then the unit root hyphotesis is rejected. In order to understand the previous rule, consider the equation (1) which is equivalent to

$$\Delta y_t = b_0 y_{t-1} + u_t, \tag{4}$$

with $b_0 = \alpha - 1$. Therefore, $\alpha = 1$ holds if and only if $b_0 = 0$. Within this context, the so called Dickey-Fuller (DF) test is simply the t statistic (used when testing for unit roots) for the significance of y_{t-1} in (4). When lagged values of Δy_t are included in (4), the implied t statistic is known as the (lag) augmented Dickey-Fuller test or ADF test.

Analysis is done considering three types of autoregressive models: without intercept nor (deterministic) trend, with intercept but without trend and with both intercept and trend. In this particular study, assumptions let asymptotic distributions be represented through moment generating functions. By using Monte Carlo simulations, the power of

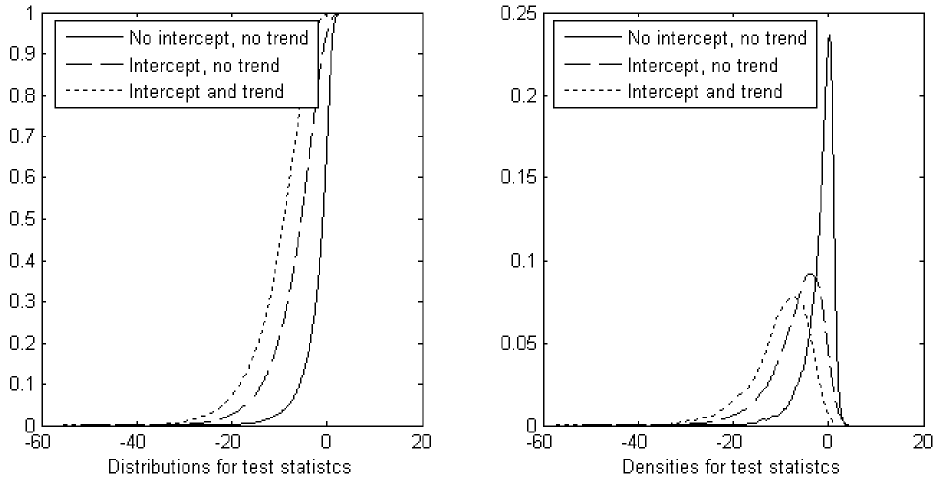

Figure 1: Asymptotic distributions for several specifications.

these tests is compared with those of (autocorrelation based) Q statistics proposed by Box and Pierce (1970). The main results are: first, Q statistics are sistematically less powerful; second, the performance of Dickey-Fuller tests is uniformly superior when there is no mispecificaton error[4]; and third, there is evidence that Dickey-Fuller tests are biased towards not rejecting the null hyphotesis for values of the autoregressive coefficient α arbitrarily close to 1.

A simple way to illustrate the role of specification is given by generating samples from the data generating process $y_t = y_{t-1} + u_t$, $u_t \sim N(0,1)$. The distribution of $T(\hat{\alpha} - 1)$ is plotted under three cases (see Figure 1): when there is no specification error, when intercept is redundant and when both intercept and trend are redundant. It can be appreciated that simulated distributions progressively move to the left and tabulated critical values tend to be higher (in absolute value) as far as redundant regressors are included. This makes the tests biased towards not rejecting the null hyphotesis and, in this sense, their power is reduced.

This brief review shows that, up to the first half of the 1980 decade, unit root econometrics has two well defined limitations: misspecification and local stationary alternatives, and each of them implies an expected loss of power. Additionally, the recurrent use of normal i.i.d. disturbances considerably reduces the applicability of these approaches by applied researchers. Two important advances are produced during the second half of that decade. First, Phillips (1987a) proposes an asymptotic theory under very general conditions for integrated processes, which makes the posterior discussion be done under firmly established foundations and, second, Perron (1989) identifies the presence of a structural break as an element that also reduces the power of the augmented Dickey-Fuller tests.

The reader must also take into account that none of these two advances could have been developed without the notion of weak convergence of probability measures to be discussed. To motivate the need of this concept consider first the Central Limit Theorem which, under conditions that vary along versions, allows for the distribution of the

[4]Intuitively this ocurrs because, for example, it is exploited the knowledge that the intercept is zero.

centered and standarized sample mean to converge to those corresponding to a normal standard distribution. In an analogous fashion, this is a desirable property when dealing with dependent heterogeneously distributed disturbances that do not satisfy the normal i.i.d. assumption in conventional autoregressive models. Indeed, this idea is summarized by several versions of the Functional Central Limit Theorem which, in a wider sense, states that the distribution of standarized partial sums converges to those of a functional of a standard Wiener process W. As described in Brzezniak and Zastawniak (1999), for a fixed value of $r \in [0, 1]$ the density $f_{W(r)}$ of the random variable $W(r)$ is given by the function

$$f_{W(r)}(x) = \frac{1}{\sqrt{2\pi r}} e^{-\frac{x^2}{r}},\ x \in \mathbb{R}.$$

Therefore, in order to deal with advances in this literature two requisites are needed. First, it is required to formally understand both the mathematical and probabilistic structure of data generating processes in order to state the main (weak) convergence results. Second, and most important, it is required to recognize the importance of incorporating econometric problems faced by researchers into the analysis, because their formalization leads to the development of new econometric procedures and testing statistics. This task is frequently made with the help of interesting alternative hypotheses.

With this background at hand, the present paper reviews a selection of theoretical advances in the unit root literature, starting from the second half of the 1980 decade and covering up to several contemporaneous developments. The presentation emphasizes both the relevance of the Functional Central Limit Theorem along the discussion as well as the econometric considerations behind novel approachs. Since time series literature can consider the case of multiple structural breaks, attention is here focused only on a singular structural break. An applied survey that considers multiple breaks can be found in Glynn and Perera (2007).

This paper is organized as follows. Section 2 describes the probabilistic structure of disturbance sequences involved, a building block for this literature. Section 3 details a general version of the Functional Central Limit Thorem that covers a wide class of disturbance processes. Section 4 presents the asymptotic theory for integrated time series proposed by Phillips (1987a). Section 5 generalizes the former framework in order to consider the so called near-integrated processes, as made by Phillips (1987b). Section 6 studies linear processes and the class of modified or M tests proposed by Stock (1999), which are meant to be employed in later developments. Section 7 details the warning made by Perron (1989) about the effects of structural breaks on the power of Dickey-Fuller statistics and the methodology proposed for dealing with an (assumed) exogenous break. Section 8 covers the critique made by Zivot and Andrews (1992) to this exogeneity assumption and the new test proposed, which involves estimating an endogenous structural break. Since none of the two previous studies deals with the power loss due to local-to-unity alternatives, section 9 illustrates the results of Perron and Rodríguez (2003), who develop efficient (power increasing) unit root tests under structural break and extend the results obtained by Elliott, Rothenberg, and Stock (1996) for linear processes. Section 10 concludes with a retrospective view about the developments in statistical inference with integrated series and the role played by the theory of difussion processes.

Part I

Asymptotic Theory

2 The structure of weakly dependent heterogeneously distributed disturbances

2.1 Some motivation

Most of the econometric theory to be covered is related with extensions of the following autoregressive model: $y_t = \alpha y_{t-1} + u_t$, $t = 1, 2, \ldots$. The main objetive here is to contrast the null hypothesis $H_0 : \alpha = 1$ when a sample of T observations $\{y_t\}_{t=1}^{T}$ is available, and the previous section introduced this task in some detail. However, a major limitation is given by the assumption that the unobservable disturbance sequence $\{u_t\}_{t=1}^{\infty}$ is composed by i.i.d. normal random variables. Thus, the empirical applicablity of several procedures is heavily restricted and it becomes desirable to cover a case intended to be as general as possible. This case is formalized by considering a sequence of disturbance terms $\{u_t\}_{t=1}^{\infty}$ that are dependent and heterogeneously distributed. A way to control the extent in which dependence occurs, such that permits to derive convergence results, is to define a measure of dependence among random variables contained in a sequence. For this measure to be well defined it needs to be refered to a specific probabilistic structure. Conditions that bound the extent of dependence are called mixing conditions. Results exposed here follow both White (1984) and Herrndorf (1984).

2.2 Mixing conditions

Consider a probabilistic space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space containing all of the possible results for an experiment, $\mathcal{F}$ is a set of events of Ω (σ-field) and $P : \mathcal{F} \to [0, 1]$ is a probability measure ($P(\Omega) = 1$) over events contained in $\mathcal{F}$. Next, consider a sequence of random variables $\{u_t\}_{t=1}^{\infty}$ (that is, $u_t : \Omega \to \mathbb{R}$ is a Borel-measurable real function for all t) on $(\Omega, \mathcal{F}, P)$. Let m and n denote two positive integers and consider a track of disturbances $\{u_t : n \leq t \leq n + m\}$. Since it will be needed to assign probabilities to events involving random variables contained in such a track, and since such events need to be included into a family with a σ-field structure, it becomes necessary to define the σ-field generated by random variables contained in the track as the smallest σ-field that contains events for which each u_t, $t = n, \ldots, n + m$, is mesurable.

Definition 1 *Let $\mathcal{B}$ denote the Borel σ-field on $\mathbb{R}$. The Borel σ-field generated by the random variables included in the track $\{u_t : n \leq t \leq n + m\}$, $\mathcal{B}_n^{n+m} = \sigma(u_t : n \leq t \leq n+m)$, is the smallest σ-field that contains events $[u_{t_i} \in B_i, n \leq t_i \leq n+m]$ with $B_i \in \mathcal{B}$.*

Intuitively, $\mathcal{B}_n^{n+1}$ is the smallest collection of events that allows to assign probabilities to events, for example, of the form $\{\omega \in \Omega : u_n(\omega) < a_1 \text{ and } u_{n+1}(\omega) < a_2\} \in \mathcal{F}$, where $a_1, a_2 \in \mathbb{R}$.

The notion of mixing is needed to explicit the fact that, although two arbitrary sets of random variables can exhibit dependence, this vanishes as time separation increases[5].

In order to illustrate the former idea, consider the track composed by the first n elements of $\{u_t\}_{t=1}^{\infty}$ and denote it by $\{u_t\}_{t=1}^{n}$. Within this track, two non-overlapping

[5]Notice that the idea of progressive lack of dependence includes that of ergodicity and asymptotic independence.

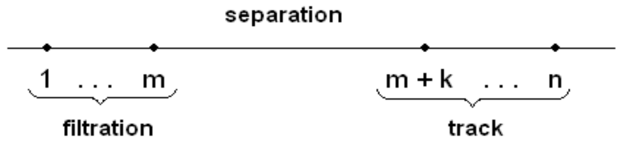

Figure 2: Dependence and mixing coefficients.

subtracks can be identified: a first one starting at u_1 and a second one ending at u_n. Let $k \geq 1$ denote the difference between time indexes corresponding to the last element of the first subtrack (denoted by $m \geq 1$) and the first element of the second subtrack (see Figure 2). Of course, the previous characterization does not completely determine both subtracks but allows for several cases. Indeed, the following definition of mixing coefficients employs the previous observations in order to quantify, given the first n elements of a sequence, the dependence between random variables separated by k periods at least.

Definition 2 *The mixing coefficients of the sequence* $\{u_t\}_{t=1}^{\infty}$ *are*

$$\alpha_n(k) = \begin{cases} \sup\limits_{\substack{A \in \sigma(u_t : 1 \leq t \leq m) \\ B \in \sigma(u_t : m+k \leq t \leq n) \\ 1 \leq m \leq n-k}} |P(A \cap B) - P(A)P(B)| & \text{for } k \leq n-1 \\ 0 & \text{for } k \geq n \end{cases}$$

Intuitively, for $n \geq 1$ given, $\alpha_n(k)$ measures how far dependence among events contained in the σ-fields $\mathcal{H} = \sigma(u_t : 1 \leq t \leq m)$ and $\mathcal{G} = \sigma(u_t : m + k \leq t \leq n)$ is situated from the independence case. $k \geq 1$ denotes time separation between these two sets of random variables (see Figure 2). If $\mathcal{H}$ and $\mathcal{G}$ were independent then for any $h \in \mathcal{H}$ and $g \in \mathcal{G}$ condition $P(g \cap h) = P(g) P(h)$ must hold or, equivalently $\alpha_n(k) = 0$.

Since mixing coefficients only take into account a finite number of disturbances (i.e. the first n random variables), this notion is extended to consider the highest dependence among random variables separated by at least k periods.

Definition 3 *The strong mixing coefficient of the sequence* $\{u_t\}_{t=1}^{\infty}$ *is*

$$\alpha(k) = \sup_{n \in \mathbb{N}} \alpha_n(k), \text{ for } k \in \mathbb{N}.$$

Thus, $\alpha(k)$ provides a measure of dependence. If $\alpha(k) = 0$ for some k, events separated by k periods are independent. Also, if $\alpha(k) \to 0$ as $k \to \infty$, sequence $\{u_t\}_{t=1}^{\infty}$ is said to be strong mixing, so that the notion of asymptotic independence is considered too. For future reference, it is useful to emphasize for a strong mixing sequence the velocity at which $\alpha(k)$ tends to zero or, equivalently, the rate of decay of $\alpha(k)$. This will be denoted by $\alpha(k) = O(k^{-\nu})$ for some $\nu > 0$[6].

[6] Let $\{a_t\}$ and $\{b_t\}$ denote two sequences of positive real variables. Then $a_t = O(b_t)$ if there exists $M > 0$ such that $|a_t/b_t| \leq M$ for all t.

3 The Functional Central Limit Theorem

3.1 The Skorohod topology

The logics behind the Functional Central Limit Theorem relies on the convergence of a sequence of standarized partial sums of disturbances u_t. The limit for this new sequence is W a standard Wiener process. Correspondingly, elements of these sequence of partial sums are contained on $D = D[0,1]$ the space of right-continous functions whose left limits exists everywhere on the unit interval, also refered as càdlàg[7] functions.

Convergence above mentioned must be understood as weak convergence of a sequence of random functions. As will be shown, in order to guarantee convergence results it is sufficient to endow D with a metric d such that (D, d) is a complete separable space, so that the limit of any convergent sequence of elements contained in D is also contained in D. Concepts and results here discussed are strongly based on Billingsley (1968), although this presentation follows Davidson (1994). The following definition characterizes the properties of the functions hereafter to be considered.

Definition 4 *$D[0,1]$ is the space of functions $x : [0,1] \to \mathbb{R}$ satisfying the following conditions:*

1. $\lim_{t \to r^+} x(t) = x(r)$ *for* $r \in [0,1)$,
2. $\lim_{t \to r^-} x(t)$ *exists for* $r \in (0,1]$,
3. $x(1) = \lim_{t \to 1^-} x(t)$.

Thus, only first class discontinuities are admited. A first metric to be considered for D is the uniform metric d_U, defined as

$$d_U(x,y) = \sup_r |x(r) - y(r)|\,,\ x, y \in D.$$

This metric states that two functions are arbitrairly close if the maximum difference between ordinates corresponding to the same abcissa is small. Metric space (C, d_U) is complete but, since $C \subset D$, completeness does not necessarily generalize to (D, d_U). In fact, it is not difficult to show that the limit of sequences of càdlàg functions in fact does not necessarily lie on D under d_U. Thus, (D, d_U) is not a complete space and the strategy adopted by Billingsley (1968) consists in metrizing D as a separable complete space by introducing the Skorohod metric.

Definition 5 (Skorohod metric) *Let Λ be the collection of all homeomorphisms[8] $\lambda : [0,1] \to [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$. The Skorohod metric is defined as*

$$d_S(x,y) = \inf_{\lambda \in \Lambda} \{\varepsilon > 0 : \sup_r |\lambda(r) - r| \leq \varepsilon \ \textit{and} \ \sup_r |x(r) - y(\lambda(r))| \leq \varepsilon\}.$$

[7] In Frech: "continue à droite, limitée à gauche".

[8] A homeomorphism (or bicontinuous function) is a continuous function that has a continuous inverse function.

This metric is defined in order to overcome the following key limitation in the (D, d_U) space: given two càdlàg function $x, y \in D$, under the uniform metric x and y are arbitrairly near to each other only if the distance between the functions is uniformly small, whereas the Skorohod metric also takes into account the fact that the distance between the arguments of these functions is small.

Metric space (D, d_S) induces a topological space. As usual, an open ball of radius $r > 0$ around $x \in D$ is defined as $B(x, r) = \{y \in D : d_S(x, y) < r\}$. Open balls like the previous one generate a topology on (D, d_S) referred to as the Skorohod topology and denoted by T_S. In this sense the topological space (D, T_S) is a metrizable topological space.

However, D is not complete under d_S yet. For this purpose, a new equivalent metric (the Billingsley metric) to d_S is introduced such that these two metrics induce the same topology in D, the Skorohod topology. The only difference now lies in the fact that the new metric space is complete.

Definition 6 (Billingsley metric) *Let Λ be the collection of all homeomorphisms $\lambda : [0,1] \to [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$, satisfying*

$$\|\lambda\| = \sup_{t \neq s} \left| \log \left(\frac{\lambda(t) - \lambda(s)}{t - s} \right) \right| < \infty.$$

The Billingsley metric is

$$d_B(x, y) = \inf_{\lambda \in \Lambda} \{\varepsilon > 0 : \|\lambda\| \leq \varepsilon, \ \sup |x(t) - y(\lambda(t))| \leq \varepsilon\}.$$

The next two results formalize the fact commented above.

Theorem 1 *In D, metrics d_B and d_S are equivalent.*

Proof. See Davidson (1994), Theorem 28.7, p. 464. ■

Theorem 2 *The space (D, d_B) is complete.*

Proof. See Davidson (1994), Theorem 28.8, p. 464. ■

3.2 The main theorem (Herrndorf, 1984)

The main result to be considered in this section is a generalization of the Central Limit Theorem for the case of functional spaces such as D, known as the Functional Central Limit Theorem. In order to understand the theorem, concepts previously defined are complemented with additional conditions for the disturbance sequence $\{u_t\}_{t=1}^{\infty}$ and, specifically, for the sequence of partial sums $S_T = \sum_{t=1}^{T} u_t$. First, disturbances are required to have zero mean and finite variance

$$E(u_t) = 0, \ E(u_t^2) < \infty \text{ for } t = 1, 2, \ldots. \tag{5}$$

Second, variance of partial sums must converge

$$\lim_{T \to \infty} E(T^{-1} S_T^2) = \sigma^2 > 0 \text{ for some } \sigma > 0. \tag{6}$$

Consider now the space D endowed with the Skorohod topology with Borel σ-field $\mathcal{B}$ and define random functions $W_T : \Omega \to \mathbb{R}$ by

$$W_T(r) = \frac{1}{\sqrt{T}\sigma} S_{\lfloor rT \rfloor},\ r \in [0,1],\ T = 1, 2, \ldots$$

where $\lfloor \cdot \rfloor$ denotes the integer part of its argument. Each W_T is a measurable map from $(\Omega, \mathcal{F})$ into $(D, \mathcal{B})$. Sequence $\{W_T\}_{T=1}^{\infty}$ is said to satisfy the invariance principle if it is weakly convergent to a standard Wiener process W on D. For the development of this result, let $\|u\|_\beta$ be defined as

$$\begin{aligned} \|u\|_\beta &= (E\,|u|^\beta)^{1/\beta} \text{ for } \beta \in [1, \infty) \\ \|u\|_\beta &= \operatorname{ess\,sup} |u| \quad \text{for } \beta = \infty. \end{aligned}$$

As will be shown in the next section, the following version of the Functional Central Limit Theorem is the starting point for all of the recent literature on unit roots. This result is due to Herrndorf (1984).

Theorem 3 (Herrndorf, 1984 Corollary 1 p. 142) *Let $\beta \in (2, \infty]$ and $\gamma = 2/\beta$. If $\{u_t\}_{t=1}^{\infty}$ satisfies* (5), (6) *and*

$$\textstyle\sum_{k=1}^{\infty} \alpha(k)^{1-\gamma} < \infty \text{ and } \limsup_{t \in \mathbb{N}} \|u_t\|_\beta < \infty,$$

then $W_T \Rightarrow W$ as $T \to \infty$.

Proof. See Herrndorf (1984), Corollary 1, p. 148. ■

4 Asymptotics for integrated processes (Phillips, 1987a)

The two previous sections stated the probabilistic foundations for econometric developments to be considered in the following lines. The first of these works is due to Phillips (1987a), who develops a quite general asymptotic theory for processes that contain a unit root.

4.1 Probabilistic structure of time series with a unit root

The first study to develop a quite general framework for testing unit roots is due to Phillips (1987a). This study establishes weak dependence conditions, among others, for the disturbance sequence in order to propose a new asymptotic theory and develop new testing statistics. Exposition here is focused on the first task because of their application in subsequent studies. The approach starts by considering a data generating process for a sequence $\{y_t\}_{t=1}^{\infty}$ that satisfies

$$y_t = \alpha y_{t-1} + u_t,\ t = 1, 2, \ldots \tag{7}$$

with

$$\alpha = 1. \tag{8}$$

Under such a representation $y_t = S_t + y_0$, where $S_t = \sum_{i=1}^{t} u_i$ and y_0 is a random initial state whose distribution is assumed to be known. Interest is placed here on the limiting distribution of standarized partial sums defined by

$$W_T(r) = \begin{cases} \frac{1}{\sqrt{T}\sigma} S_{\lfloor Tr \rfloor} & \text{if } \frac{j-1}{T} \leq r < \frac{j}{T},\ j = 1, \ldots, T, \\ \frac{1}{\sqrt{T}\sigma} S_T & \text{if } r = 1, \end{cases} \tag{9}$$

where σ is a positive constant. Notice that the sample paths $W_T(r)$ lie in D. It is worth to emphasize that Phillips (1987a) endows D with the uniform metric d_U and this is done in order to show that each random function $W_T(r)$ lies on D. Also the adoption of assumptions about disturbances $\{u_t\}_{t=1}^{\infty}$ less restrictive than i.i.d. allows to demonstrate that $W_T(r)$ weakly converges to a standard Wiener process $W(r)$ through a direct application of the Functional Central Limit Theorem developed by Herrndorf (1984). Assumptions regarding $\{u_t\}_{t=1}^{\infty}$ are grouped in the following statement and are intended to be as general as possible.

Assumption 1 (Phillips, 1987 p. 280) *Disturbance sequence* $\{u_t\}_{t=1}^{\infty}$ *satisfies*

1. $E(u_t) = 0$ *for* $t = 1, 2, \ldots,$
2. $\sup_t E|u_t|^{\beta} < \infty$ *for some* $\beta > 2$,
3. $\sigma^2 = \lim_{T\to\infty} T^{-1} E(S_T^2)$ *exists and* $\sigma^2 > 0$, *with* $S_T = \sum_{t=1}^{T} u_t$,
4. $\{u_t\}_{t=1}^{\infty}$ *is strong mixing, with strong mixing coefficients* $\alpha(k)$ *that satisfy*

$$\textstyle\sum_{k=1}^{\infty} \alpha(k)^{1-2/\beta} < \infty. \tag{10}$$

As usual, condition 1 imposes a zero mean disturbance for every t. Condition 2 bounds the probability of outliers: the higher β the lower the probability of outliers. As long as such $\beta > 2$ exists, all of the lower absolute moments of each u_t (including the second one) are finite. Condition 3 is conventional along central limit theory, concerning the convergence of the average variance of partial sums S_T. Condition 4 bounds the temporal dependence among disturbances contained in $\{u_t\}_{t=1}^{\infty}$, and elements covered in previous sections allows to assert that although dependence can exists between any pair of disturbances, this vanishes as time separation increases. Hence, two random disturbances sufficiently distant along time are almost independent. Finally, sumability condition (10) is satisfied as long as the mixing decay rate is $\alpha(k) = O(k^{-\nu})$ for some $\nu > 0$ such that $-\nu(1 - 2/\beta) < 1$ or, equivalently $\nu > \beta/(\beta - 2)$.

It is interesting to notice that as T increases the constant sections conforming $W_T(r) \in D$ reduce their size and discontinuities become less perceptible (see Figure 3), reflecting how this sequence of random functions in D converge to a random function in C, the standard Wiener process. This property is exploited by Phillips (1987a) through two lemmas. The first lemma is the Functional Central Limit Theorem shown in Theorem 3 and the second result is widely known as the Continous Mapping Theorem and states conditions under which convergence to a Wiener process can be preserved along (almost everywhere) continous transformations.

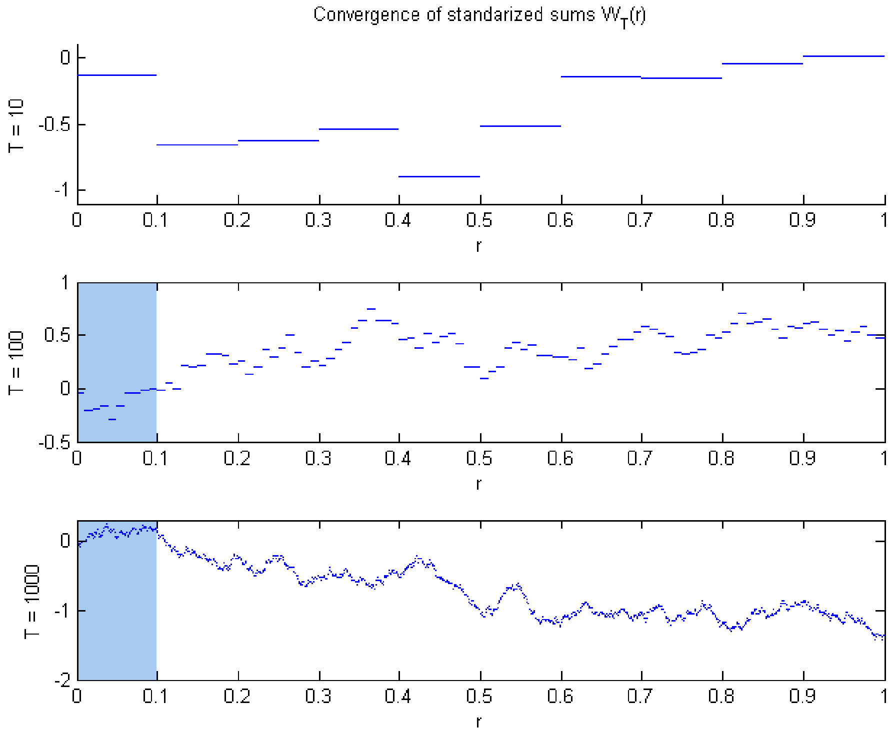

Figure 3: Convergence of standarized sums.

Lemma 4 (Phillips, 1987 p. 281) *If* $\{u_t\}_{t=1}^{\infty}$ *satisfies Assumption 1 then, as* $T \to \infty$, $W_T \Rightarrow W$ *a standard Wiener process on* C.

Proof. See Herrndorf (1984), Corollary 1, p. 142. ■

Lemma 5 (Phillips, 1987 p. 281) *If* $W_T \Rightarrow W(r)$ *as* $T \to \infty$ *and* h *is a continous functional on* D *a.e. then* $h(W_T) \Rightarrow h(W)$ *as* $T \to \infty$.

Proof. See Billingsley (1968), Corollary 1 p. 31. ■

4.2 Some asymptotic theory for econometricians

The importance of the two previous lemmas relies on the fact that they allow the derivation of convergence rules often employed by theoretical econometricians. These rules are summarized in the next theorem.

Theorem 6 (Phillips, 1987 p. 282) *If* $\{u_t\}_{t=1}^{\infty}$ *satisfies Assumption 1 and if*

$$\sup_t |u_t|^{\beta+\varepsilon} < \infty \text{ for some } \varepsilon > 0$$

(where $\beta > 2$ *is the same as that in Assumption 1), then as* $T \to \infty$*:*

1. $T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 \, dr$,
2. $T^{-1} \sum_{t=1}^{T} y_{t-1}(y_t - y_{t-1}) \Rightarrow (\sigma^2/2)(W(1)^2 - \sigma_u^2/\sigma^2)$,

3. $T(\hat{\alpha}-1) \Rightarrow \frac{1}{2}(W(1)^2 - \sigma_u^2/\sigma^2)/\int_0^1 W^2(r)\,dr$,

4. $\hat{\alpha} \xrightarrow{p} 1$,

5. $t_{\hat{\alpha}} \Rightarrow (\sigma/2\sigma_u)(W(1)^2 - \sigma_u^2/\sigma^2)/\{\int_0^1 W(r)^2\,dr\}^{1/2}$,

where $\sigma_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(u_t^2)$, $\sigma^2 = \lim_{T\to\infty} E(T^{-1}S_T^2)$ *and* W *is a standard Wiener process on* C*.*

Proof. See Phillips (1987a), Theorem 3.1 p. 296. ■

In the previous theorem, results 1 and 2 constitute derivation rules for limiting distributions. Result 3 is focused on the limiting distribution of the statistic $T(\hat{\alpha}-1)$, which corrects the results of White (1958)[9], among others. Result 4 states the consistency of the OLS estimator $\hat{\alpha}$ in the presence of a unit root and under the general case of dependent and heterogeneously distributed disturbances. Finally, result 5 shows the asymptotic distribution of the t statistic used when testing for unit roots. It is worth to mention that under (7) and (8) the t statistic does not follow a Student's t distribution. Since $W(1)$ follows a normal standard distribution, $W(1)^2$ follows a chi-squared distribution with one degree of freedom. However, the functional $\int_0^1 W(r)^2\,dr$ is a random variable with a rather complex distribution, so that usual distributions (normal, chi-squared, t and F) employed in the stationary case are not relevant for the subsequent analysis.

In this way, results let Phillips (1987a) to propose (after developing consistent estimators for parameters σ_u^2 and σ^2) two new test statistics for the unit root hypothesis often refered as the Z tests. Although it is important to remember that both (7) and (8) correspond only to the case of a unit without drift nor deterministic trend, the importance of this study relies on providing a general theory on test statistics for the unit root hypothesis. Distributions considerd here differ from those involved in the stationary case ($|\alpha| < 1$). Obviously, this methodology is well suited for extensions that include both drift and deterministic trend, derived by Phillips and Perron (1988), and constitute the starting point for the study of the unit root test under structural break in the following sections.

5 Asymptotics for near-integrated processes (Phillips, 1987b)

For later discussion on the asymptotic power of unit root tests against alternative hypotheses that consider autoregressive coefficients near to one, it will be useful to consider generalizations of integrated processes often referred as near-integrated processes and studied in detail by Phillips (1987b). For this case, time series $\{y_t\}_{t=1}^{\infty}$ is assumed to be generated according to the following model

$$y_t = \alpha y_{t-1} + u_t,\ t = 1, 2, \ldots \quad (11)$$
$$\alpha = e^{c/T},\ -\infty < c < \infty. \quad (12)$$

[9] See equation (3).

In the above model, initial condition y_0 is allowed to be any random variable whose distribution is fixed and independent of T. The constant c is interpreted as a non-centrality parameter that quantifies deviations from the unit root null hypothesis that holds when $c = 0$

$$H_0 : \alpha = 1. \tag{13}$$

Under (13), $\{y_t\}_{t=0}^{\infty}$ is an integrated process of order 1 or $I(1)$ process. Additionally, any $c \neq 0$ in (12) represents a local alternative to H_0. For future reference, the next definition formally establishes this distinction.

Definition 7 *A time series* $\{y_t\}_{t=1}^{\infty}$ *that is generated by* (11) *and* (12) *with* $c \neq 0$ *is called near-integrated. When* $c = 0$*, in* (12)*,* $\{y_t\}_{t=1}^{\infty}$ *is also called integrated.*

The main objetive of the present section is to present an asymptotic theory for this type of processes. Naturally, results and properties are indexed by the parameter c.

5.1 Probabilistic structure of time series with a near-to-unit root

For a wide applicability of this asymptotic theory, general assumptions concerning the disturbance sequence $\{u_t\}_{t=0}^{\infty}$ are necessary. For this reason, the following mixing conditions about the behaviour of disturbances $\{u_t\}_{t=0}^{\infty}$ (hereby now familiar) are adopted and summarized in the next statement.

Assumption 2 (Phillips, 1987b p. 537) *Disturbance sequence* $\{u_t\}_{t=1}^{\infty}$ *satisfies*

1. $E(u_t) = 0$ *for* $t = 1, 2, \ldots,$
2. $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$ *for some* $\beta > 2$ *and* $\varepsilon > 0$*,*
3. $\sigma^2 = \lim_{T\to\infty} T^{-1}E(S_T^2)$ *exists and* $\sigma^2 > 0$*, with* $S_T = \sum_{t=1}^{T} u_t$*,*
4. $\{u_t\}_{t=1}^{\infty}$ *is strong mixing, with strong mixing coefficients* $\alpha(k)$ *that satisfy*

$$\textstyle\sum_{k=1}^{\infty} \alpha(k)^{1-2/\beta} < \infty. \tag{14}$$

Notice that Assumptions 1 and 2 are quite similar and the only difference relies on the existence of $\varepsilon > 0$ such that the existence of $\sup_t E|u_t|^{\beta+\varepsilon}$ holds. By the other hand, it will be convenient to represent stochastic limit theory by means of extensive use of certain difussion process. This process can be interpreted as the continous time version of an AR(1) process.

Definition 8 (Ornstein-Uhlenbeck process) *A Ornstein-Uhlenbeck process is a functional of the form* $W_c(r) = \int_0^r e^{(r-s)c}dW(s)$ *that satisfies the stochastic differential equation*

$$dW_c(r) = cW_c(r)dr + dW(r),\ W_c(0) = 0. \tag{15}$$

Equation (15) is called the *Ornstein-Uhlenbeck* or *Langevin equation.* It is a particular case of the equation

$$dX(t) = b(t, X(t))dt + \sigma(t, X(t))dW(t), \tag{16}$$

where $b(t, X(t))$, $\sigma(t, X(t)) \in \mathbb{R}$ and $W(t)$ is a Wiener process with $t \in [0, \infty)$ (Oksendal 2000). Equation (15) can also be written as

$$W_c(r) = W(r) + c\int_0^r e^{(r-s)c}W(s)ds$$

and the effect of the non-centrality parameter c becomes even more evident.

5.2 More asymptotic theory for econometricians

If parameter c were fixed it would be natural to expect, based on (12), $\alpha \to 1$ as $T \to \infty$. However, in this framework the speed of convergence of α towards 1 is controlled at $O(T^{-1})$. Equivalently, such a speed is not too fast so that the effect of c on the main results does not vanish[10]. This observation leads to the following derivation rules and properties for regression-based statistics.

Lemma 7 (Phillips, 1987b p. 539) *If $\{y_t\}$ is a near-integrated time series generated by* (11) *and* (12) *then, as $T \to \infty$:*

1. $T^{-1/2} y_{[Tr]} \Rightarrow \sigma W_c(r)$,
2. $T^{-3/2} \sum_{t=1}^{T} y_t \Rightarrow \sigma \int_0^1 W_c(r)dr$,
3. $T^{-2} \sum_{t=1}^{T} y_t^2 \Rightarrow \sigma^2 \int_0^1 W_c(r)^2 dr$,
4. $T^{-1} \sum y_{t-1} u_t \Rightarrow \sigma^2 \int_0^1 W_c(r)dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2)$, *with* $\sigma_u = \lim_{T\to\infty} T^{-1} \sum E(u_t^2)$.

Proof. See Phillips (1987b), Lemma 1, p. 539. ■

Theorem 8 (Phillips, 1987b p. 540) *If $\{y_t\}$ is a near-integrated time series generated by* (11) *and* (12) *then, as $T \to \infty$:*

1. $T(\hat{\alpha} - \alpha) \Rightarrow \{\int_0^1 W_c(r)dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2)\}/ \int_0^1 \{W_c(r)\}^2 dr$,
2. $\hat{\alpha} \xrightarrow{p} 1$, $s^2 \xrightarrow{p} \sigma_u^2$,
3. $t_\alpha \Rightarrow (\sigma/\sigma_u)\{\int_0^1 W_c(r)dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2)\}/[\int_0^1 \{W_c(r)\}^2 dr]^{1/2}$.

Proof. See Phillips (1987b), Theorem 1, p. 540. ■

Up to this point, the theory presented can be used in the analysis of the power of unit root tests under local alternatives. For a non-centrality parameter c arbitrarily close to 0 it is easy to show that $e^{c/T} \approx 1 + c/T$ and this is the approach usually employed in unit root testing. A brief illustration of this procedure can be found, for example, in Phillips (1988).

6 Linear processes and modified tests

6.1 Motivation

Although the reader must have noticed that the so called mixing conditions are intended to be a powerful tool that allows the derivation of weak convergence results for a wide class of processes, Phillips and Solo (1992) pointed out that, since much of the time series analysis is concerned with parametric models that fall in the class of linear processes, mixing conditions exhibit a major drawback. The reason is quite simple since not all linear processes are strong mixing. In spite of this, they propose a turnback to linear processes as the main focus for developing time series asymptotics. Under the linear

[10] Since $c = T \ln \alpha$, c depends on T. To simplify notation, however, this dependence is avoided.

model class, Phillips and Solo (1992) make extensive use of the algebraic Beveridge-Nelson decomposition (see Appendix A) to demonstrate the Functional Central Limit Theorem once provided with a disturbance sequence $\{\varepsilon_t\}_{t=0}^{\infty}$ that is a martingale difference sequence (see Appendix B), strongly uniformly integrable (see Appendix C) with dominating random variables $\{Z_t\}_{t=0}^{\infty}$ such that $E(Z_t^{2+\eta}) < \infty$ for some $\eta > 0$, further $T^{-1}\sum_{t=1}^{T} E(\varepsilon_t^2|\mathcal{F}_{t-1}) \overset{a.s.}{\to} \sigma_\varepsilon^2 > 0$, where $\mathcal{F}_t$ is the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$. Given the latter notation it is now possible to establish the following

Theorem 9 (Phillips and Solo, 1992) *Suposse that* $\{u_t\}_{t=0}^{\infty}$ *is the linear process*

$$u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} = C(L)\varepsilon_t,\ C(L) = \sum_{j=0}^{\infty} c_j L^j$$

with $0 < C(1) \equiv \sum_{j=0}^{\infty} c_j < \infty$ *and* $\sum_{j=0}^{\infty} c_j^2 < \infty$. *If* $\sum_{j=1}^{\infty} j|c_j| < \infty$, *then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t \Rightarrow \sigma_\varepsilon C(1) W(r).$$

Proof. See Phillips and Solo (1992) Theorem 3.4 p. 983. ■

Although the latter theorem is less general than versions previously presented, it will be frequently used in posterior work, specially along the developments due to Stock (1999).

6.2 The M class of integration tests (Stock, 1999)

Stock (1999) proposed a new class of statistics that directly test the implication that an integrated process has a growing variance having an order of probability[11] of $T^{-1/2}$ ($O_p(T^{-1/2})$). Since the remaining of this paper deals with this class of tests under several frameworks, the general class is examined in some detail. First, suppose the following data ganerating process for $\{y_t\}_{t=1}^{\infty}$

$$y_t = \delta_t(\beta) + \sum_{i=1}^{t} u_i,$$

$t = 1, \ldots, T$. That is, under the null hypothesis the series $\{y_t\}_{t=1}^{\infty}$ can be written as the sum of a purely deterministic component $\delta_t(\beta)$ (with finite dimensional vector β estimated by $\hat{\beta}$) and an integrated or I(1) component that is the partial sum of weakly stationary or I(0) terms. Let the long run variance of u_t be denoted by $\sigma^2 = 2\pi s_u(0)$, where $s_u(0)$ is the spectral density of u_t at frequency zero and, for $r \in [0, 1]$, let

$$\begin{aligned} S_T(r) &= \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor rT \rfloor} u_i, \text{ and} \\ D_T(r, \beta) &= \delta_{\lfloor rT \rfloor}(\beta) \end{aligned}$$

be càdlàg versions of the components of the discrete time process. As expected, this is done in order to apply the Functional Central Limit Theorem. Such functionals are assumed to satisfy the following

Assumption 3 (Stock, 1999 p. 137) *The following conditions hold:*

[11] Let $\{y_t\}$ denote a sequence of random variables and let $\{a_t\}$ denote a sequence of positive non-stochastic real numbers. Then $y_t = O_p(a_t)$ if for each $\varepsilon > 0$ there exists $M > 0$ such that $P(|y_t|/a_t > M) < \varepsilon$.

1. $S_T \Rightarrow \sigma W$, *where* $0 < \sigma^2 < \infty$, *and*

2. $\sqrt{T}\{D_T(\cdot, \hat{\beta}) - D_T(\cdot, \beta)\} \Rightarrow \sigma D$, *where* $D \in D[0,1]$ *has a distribution that does not depend on* β *or on the nuisance parameters describing the distribution of* $\{u_t\}$.

In line with the proposal of Phillips and Solo (1992), Stock (1999) focuses on linear processes

$$\begin{aligned} u_t &= C(L)\varepsilon_t, \\ \textstyle\sum_{j=1}^{\infty} j|c_j| &< \infty, \\ C(1) &\neq 0 \end{aligned}$$

where ε_t is a martingale difference sequence (m.d.s.) with

$$E[\varepsilon_t|\mathcal{F}_{t-1}] = 0, \text{ and} \tag{17}$$

$$\sup_t E[\varepsilon_t^{2+\kappa}|\mathcal{F}_{t-1}] < \infty \text{ for some } \kappa > 0. \tag{18}$$

As usual, condition (17) imposes zero-mean disturbances whereas condition (18) bounds the probability of outliers in a similar fashion to condition 2 presented in Assumption 1. Also, although the deterministic component $\delta_t(\beta)$ is designed to potentially contain polynomial and further general trends, the following cases are here considered for obvious reasons.

1. No deterministic trend: $\delta_t(\beta) = 0$. In this case there is no need of detrending. For completeness, let the "detrended" series be $y_t^0 \equiv y_t$.

2. Constant: $\delta_t(\beta) = \beta_0$. In this case β_0 is estimated by $\hat{\beta}_0 = \bar{y} = T^{-1}\sum_{t=1}^{T} y_t$ and the demeaned series is $y_t^{\mu} \equiv y_t - \bar{y}$.

3. Linear trend: $\delta_t(\beta) = \beta_0 + \beta_1(t/T)$. If (β_0, β_1) is estimated by the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1)$ then the detrended series is $y_t^{\tau} \equiv y_t - \hat{\beta}_0 - \hat{\beta}_1(t/T)$. Normalization of the known part of the deterministic component is done for its continous time analogous to lie in the interval $[0,1]$.

The three former cases are enough for subsequent analysis. Since limiting representation in Assumption 3 depends on the nuisance parameter σ^2, it is assumed that there exists a consistent estimator $\hat{\sigma}^2$ for σ^2.

Assumption 4 (Stock, 1999 p. 137) *Under the null hypothesis* $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.

Elements for the development of the new class of tests are based on both the Functional Central Limit Theorem and the Continous Mapping Theorem. For each case considered, define S_T^d as the scaled stochastic process formed using the respective detrended series

$$S_T^d(r) = \frac{1}{\sqrt{T\hat{\sigma}^2}} y_{\lfloor rT \rfloor}^d,\ d = 0, \mu, \tau,$$

for $r \in [0,1]$. If Assumptions 3 and 4 hold, then

$$S_T^d \Rightarrow S^d = W - \tilde{D}, \text{ for certain } \tilde{D} \in D[0,1]. \tag{19}$$

For the three functional forms of the deterministic component, the following theorem shows the specific form that $\tilde{D}$ adopts.

Theorem 10 (Stock, 1999 p. 137) *Assume that Assumptions 3 and 4 hold.*

1. *If* $\delta_t(\beta) = 0$*, then*
$$S_T^0(r) = \frac{1}{\sqrt{T\hat{\sigma}^2}} y_{\lfloor rT \rfloor}^0 \Rightarrow W(r).$$

2. *If* $\delta_t(\beta) = \beta_0$*, then*
$$S_T^\mu(r) = \frac{1}{\sqrt{T\hat{\sigma}^2}} y_{\lfloor rT \rfloor}^\mu \Rightarrow S^\mu(r) = W(r) - \int_0^1 W(s)ds.$$

3. *If* $\delta_t(\beta) = \beta_0 + \beta_1(t/T)$*, then*
$$S_T^\tau(r) = \frac{1}{\sqrt{T\hat{\sigma}^2}} y_{\lfloor rT \rfloor}^\tau$$
$$\Rightarrow S^\tau(r) = W(r) - (4 - 6r)\int_0^1 W(s)ds - (-6 + 12r)\int_0^1 sW(s)ds,$$

Proof. See Stock (1999), Theorem 1 p. 139. ■

This latter result is one of the cornerstones for the proposed class of tests. Also, it follows from the Continous Mapping Theorem that if (19) holds and g is a continous function $g : D[0,1] \to \mathbb{R}$, then
$$g(S_T^d) \Rightarrow g(S^d). \tag{20}$$
Let $M^d = \{m : D[0,1] \to \mathbb{R}\}$ be the collection of functionals that satisfy the following conditions:

1. m is continous,

2. there exists c_υ, $|c_\upsilon| < \infty$, such that $P[m(S^d) \leq c_\upsilon] = \upsilon$ for all $\upsilon \in (0,1)$, and

3. $g(0) < c_\upsilon$ for all $\upsilon \in (0,1)$.

The class M^d, refered only to continous functionals of S_T^d, groups test statistics for the null hypothesis that y_t is I(1) against the alternative that it is I(0). Since S_T^d represents any of the three detrended series mentioned, under the null hypothesis $m(S_T^d)$ has an asymptotic distribution with critical values that depend on the functional m, whereas under a fixed alternative y_t is I(0), which suggests the construction of one tailed tests of level υ of the form:
$$\text{reject } H_0 : y_t \sim I(1) \text{ if } m(S_T^d) \leq c_\upsilon.$$
This approach, as Stock (1999) asserts, suggests working backwards from the desired asymptotic representation to the actual test statistic. The fact that the form of the function $m(\cdot)$ does not depend on the type of detrending emphasizes that the steps of eliminating the deterministic components and testing for unit root are distinct: detrending a series when it is not required does not affect the size of the tests since $\tilde{D}$ does not depend on β. In contrast, failing to detrend a series that contains a trend typically leads to a loss of consistency and an incorrect asymptotic size.

In summary, always detrending a series before hypothesis testing does not affect the size and this is a desirable property. Once size is guaranteed to be fixed, power increasing procedures can be perfomed. The next two subsections illustrate the main idea behind: if certain test statistic V has a limiting distribution characterized as the functional m of certain diffusion process S^d

$$V \Rightarrow m(S^d),$$

this asymptotic distribution can also be written as the limiting one of a respective modified test statistic for detrended data $m(S_T^d)$:

$$m(S_T^d) \Rightarrow m(S^d),$$

such that V and its modified version $m(S_T^d)$ are asymptotically equivalent.

6.3 The modified Sargan-Bhargava test

One of the test statistics to be covered along the section 9 is due to Sargan and Bhargava (1983) for the model

$$y_t = \beta_0 + \textstyle\sum_{s=1}^{t} \alpha^{t-s} \varepsilon_s, \tag{21}$$

where $\varepsilon_t \sim N(0, \sigma^2)$, $t = 1, \ldots, T$ and $(\alpha, \beta_0, \sigma^2)$ is a vector of unknown parameters. The authors propose the following Durbin-Watson statistic for a regression of y_t against a constant

$$SB_\mu = \frac{\sum_{t=2}^{T} (\Delta y_t)^2}{\sum_{t=1}^{T} (y_t^\mu)^2},$$

where $y_t^\mu \equiv y_t - \bar{y}$. For the case where there exists a linear deterministic trend Bhargava (1986) considers the extension

$$y_t = \beta_0 + \beta_1 t + \textstyle\sum_{s=1}^{t} \alpha^{t-s} \varepsilon_s, \tag{22}$$

where $\varepsilon_t \sim N(0, \sigma^2)$, $t = 1, \ldots, T$ and $(\alpha, \beta_0, \beta_1, \sigma^2)$ is a vector of unknown parameters. A similar test is proposed

$$SB_\tau = \frac{\sum_{t=2}^{T} (\Delta y_t)^2}{\sum_{t=1}^{T} (y_t^\tau)^2},$$

where $y_t^\mu \equiv y_t - \tilde{\beta}_0 - \tilde{\beta}_1 (t/T)$,

$$\begin{aligned} \tilde{\beta}_0 &= \bar{y} - \frac{1}{2} \frac{T+1}{(T-1)} (y_T - y_1), \\ \tilde{\beta}_1 &= \frac{T}{T-1} (y_T - y_1). \end{aligned}$$

For both tests, Stock (1999) derives the limiting distribution

$$T^{-1} SB_d \Rightarrow \frac{\sigma^2}{var(\Delta y_t)} \textstyle\int_0^1 S^d(r)^2 dr, \text{ for } d = \mu, \tau. \tag{23}$$

After noticing in (21) and (22) that $\sigma^2 = var(\Delta y_t)$, (23) can be written as

$$T^{-1} SB_d \Rightarrow \textstyle\int_0^1 S^d(r)^2 dr, \text{ for } d = \mu, \tau.$$

Now, notice that the functional

$$m_{SB}(f) = \textstyle\int_0^1 f(r)^2 dr$$

is also involved in the limiting distribution of the following functional in $D[0, 1]$

$$\frac{1}{\sqrt{T}} \textstyle\sum_{t=1}^{T} (y_t^d)^2.$$

This latter statistic will be refered as the modified Sargan-Bhargava or MSB test.

6.4 A modified Z test

For a model that contains a constant deterministic component, Phillips (1987a) and Phillips and Perron (1988) propose the test statistic

$$Z_\alpha = T(\hat{\alpha} - 1) - \frac{1}{2}\frac{\hat{\sigma}^2 - \hat{\sigma}_u^2}{T^{-2}\sum_{t=1}^{T} y_{t-1}^2}, \tag{24}$$

where

$$\begin{aligned}
\hat{\alpha} &= \textstyle\sum_{t=2}^{T} y_t^\mu y_{t-1}^\mu / \sum_{t=2}^{T} (y_{t-1}^\mu)^2, & (25)\\
\hat{u}_t &= y_t^\mu - \hat{\alpha} y_{t-1}^\mu, & (26)\\
\hat{\sigma}^2 &= T^{-1}\textstyle\sum_{t=2}^{T} \hat{u}_t^2 + 2\sum_{j=1}^{l} T^{-1} \sum_{t=j+2}^{T} \hat{u}_t \hat{u}_{t-j} \text{ and} & (27)\\
\hat{\sigma}_u^2 &= T^{-1}\textstyle\sum_{t=1}^{T} (y_t - \hat{\alpha} y_{t-1})^2. & (28)
\end{aligned}$$

Since $\sum_{t=1}^{T} y_{t-1}\Delta y_t = (1/2)(y_T^2 - \Delta y_t^2)$, Z_α test in (24) can be written as

$$Z_\alpha = \frac{1}{2}\frac{S_T(1)^2 - 1}{T^{-1}\sum_{t=1}^{T-1} S_T(t/T)^2} - \frac{1}{2}T^{-2}(\hat{\alpha} - 1),$$

and, provided that $\hat{\alpha} - 1 \xrightarrow{p} 0$, asymptotic distribution is

$$Z_\alpha \Rightarrow \frac{1}{2}\frac{W(1)^2 - 1}{\int_0^1 W(r)^2 dr}.$$

This latter expression suggests the use of the following functional

$$m_{Z_\alpha}(f) = \frac{1}{2}\frac{f(1)^2 - 1}{\int_0^1 f(s)^2 ds},$$

as shown by Stock (1999). For the study of Perron and Rodríguez (2003) to be covered in section 9, the modified Z_α test will be refered to as the MZ_α test.

Part II

Econometric Applications

7 Exogenous structural break (Perron, 1989)

In the previous sections, the foundations for the study of inference with nonstationary time series have been established. Now, subsequent sections extend the analysis to the case in which a structural break is present. This literature starts with the identification of key limitations concerning ADF tests.

After the work of Dickey and Fuller (1979), several empirical studies were done in order to test for the existence of unit roots along macroeconomic variables. Most of these empirical results favored such an hypothesis and the perception that macroeconomic variables were characterized by stochastic trends became popular. One of the most influent studies in this empirical literature is done by Nelson and Plosser (1982). In this study, 14 macroeconomic variables for the US economy were employed. Under the stochastic trend perspective, a series that exhibits an upward sloping behaviour and an abrupt reduction (see Figure 4a) is interpreted as a consequence of an atipic realization of u_t (situated in the left tail of its distribution) for the process $y_t = \mu + y_{t-1} + u_t$. However, the same behaviour can be interpreted as a trend stationary process $y_t = \mu_t + \delta t + u_t$ whose intercept changes its value from, say, μ_1 to $\mu_2 < \mu_1$ (see Figure 4b).

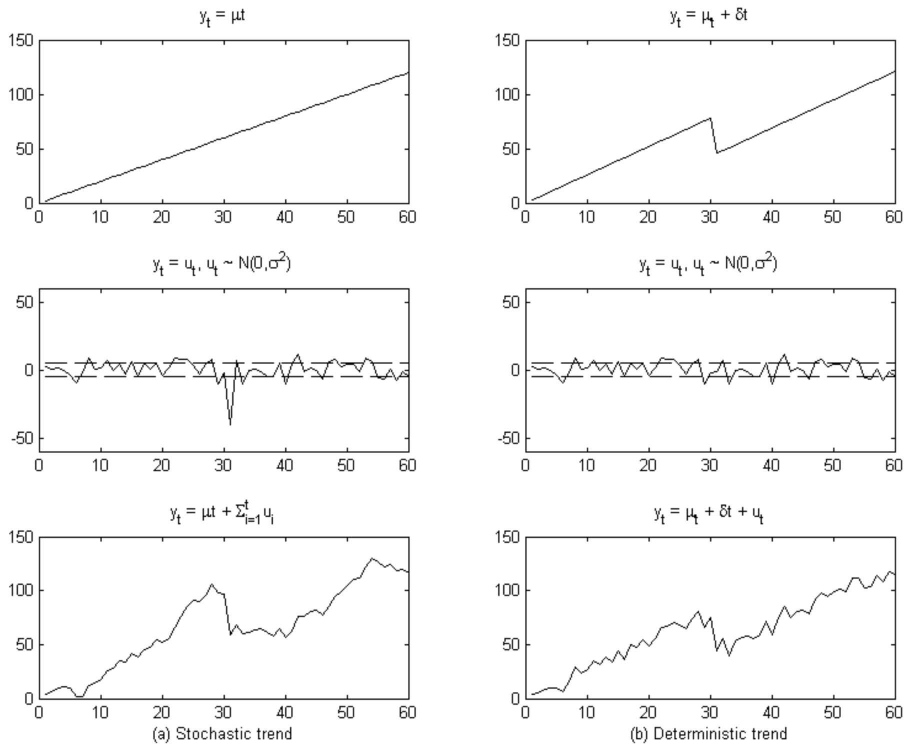

Figure 4: Shifts under stochastic and deterministic trend frameworks.

Indeed, Perron (1989) emphasizes this latter interpretation and asserts that

> "... most macroeconomic variables are trend stationary if one allows a single change in the intercept of the trend function after 1929 and a single change in the slope of the trend function after 1973".

Null hyphotesis	Alternative hyphotesis
Model A $y_t = \mu + y_{t-1} + \theta D(T_B)_t + u_t$	**Model A** $y_t = \mu_1 + (\mu_2 - \mu_1) DU_t + \delta t + u_t$
Model B $y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1) DU_t + u_t$	**Model B** $y_t = \mu + \delta_1 t + (\delta_2 - \delta_1) DT_t^* + u_t$
Model C $y_t = \mu_1 + y_{t-1} + \theta D(T_B)_t + (\mu_2 - \mu_1) DU_t + u_t$	**Model C** $y_t = \mu_1 + \delta_1 t + (\mu_2 - \mu_1) DU_t + (\delta_2 - \delta_1) DT_t + u_t$
where $D(T_B)_t = 1$ if $t = T_B + 1$, 0 otherwise; $DU_t = 1$ if $t > T_B$, 0 otherwise	where $DT_t^* = t - T_B$ if $t > T_B$, 0 otherwise; $DT_t = t$ if $t > T_B$, 0 otherwise

Table 2: Null and alternative hypotheses considered by Perron (1989).

7.1 The key motivation

Perron (1989) considers atipic events as interventions to the deterministic component of the model, and this allows to distinguish between what can be explained or not by the disturbance term. Additionally, the date for this intervention is assumed to be known by the researcher. Because there exist two competing interpretations (above mentioned) for time series with an abrupt shift, models considered by Perron (1989) are summarized in Table 2.

In Table 2, θ, μ, μ_1, μ_2, δ, δ_1 and δ_2 are parameters, $A(L) u_t = B(L) e_t$ and $e_t \sim i.i.d.(0, \sigma_e^2)$. $A(L)$ and $B(L)$ are pth and qth order polinomials. That is, $\{u_t\}$ is an $ARMA(p, q)$ process with p and q possibly unknown. This assumption allows $\{y_t\}$ to represent general processes. In this sense, different specifications allow for different models:

1. Under the null hypothesis, model A contains a dummy variable that equals 1 only inmediatly after T_B (a one time change of the intercept), whereas under the alternative hypothesis the series is trend stationary with a permanent shift in the intercept of the trend function after T_B (see Figure 5).

2. For model B, under the null hypothesis a permanent change in the intercept is allowed after T_B; whereas under the alternative hypothesis only a permanent shift is allowed in the slope of the deterministic component.

3. Finally, model C allows both the two shifting types simultaneously: a shift in level acompained by a shift in slope.

In this way, Perron (1989) introduces a third interpretation to the discussion (see Figure 5) in order to identify limitations present in already known testing statistics.

A first attempt to discriminate between the two approaches included in Figure 4 could be through the use of DF tests. However, by using numerical experiments, Perron

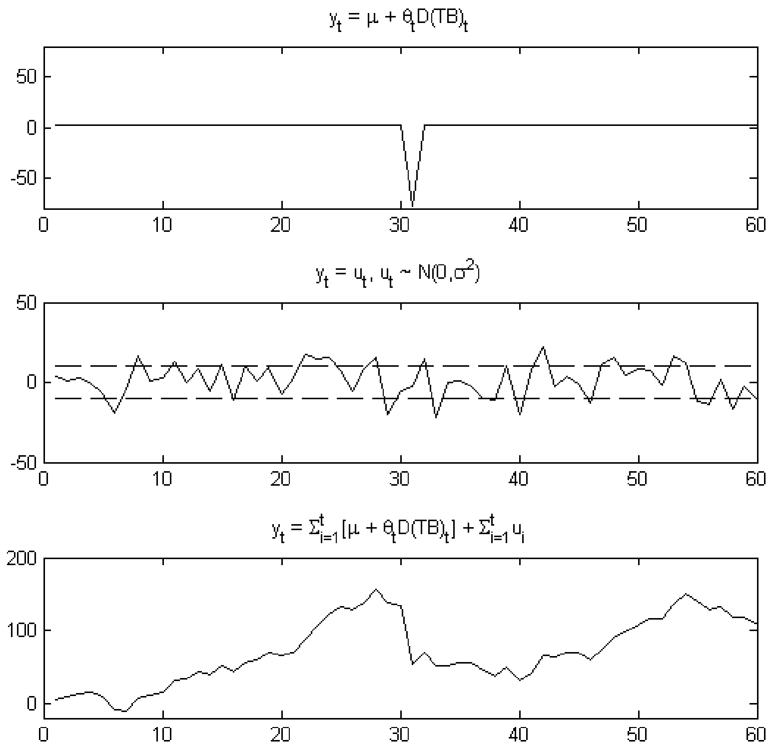

Figure 5: The "Crash" model.

(1989) examines the performance of these class of tests under the alternative hypothesis. Specifically, Monte Carlo simulations reveal that when the data generating process is described as by model A under the alternative, DF tests tend to detect a spurious unit root that does not vanish, even asymptotically. Therefore, a power loss is expected. This property is also derived at the theoretical level (Perron 1989, Theorem 1) and for this result to be as general as possible, assumptions identical to those adopted by Phillips (1987a) concerning the innovations sequence $\{u_t\}$ are adopted and summarized as follows.

Assumption 5 (Perron, 1989, p. 1371) *Disturbance sequence* $\{u_t\}_{t=1}^{\infty}$ *satisfies*

1. $E(u_t) = 0$ *for all* t*;*

2. $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$ *for some* $\beta > 2$ *and* $\varepsilon > 0$*;*

3. $\sigma^2 = \lim_{T\to\infty} T^{-1} E(S_T^2)$ *exists and* $\sigma^2 > 0$*, where* $S_T = \sum_{t=1}^{T} u_t$*;*

4. $\{u_t\}_{t=1}^{\infty}$ *is strong mixing with strong mixing coefficients* $\alpha(k)$ *that satisfy*

$$\sum_{k=1}^{\infty} \alpha(k)^{1-2\beta} < \infty.$$

As expected, the Functional Central Limit Theorem due to Herrndorf (1984) can still be employed in this case. Specifically, Assumption 2 allows for the generalization of the asymptotic theory included in Theorem 6 (Perron 1989, Lemma A.3), now under the presence of a breakfraction $\lambda \in (0,1)$. The next subsection presents the strategy adopted and the main results.

7.2 Structure of the model and main findings

Because of the caveats when using DF tests, the strategy adopted by Perron (1989) consists on developing a unit root test under structural break. That is, the null hypothesis specifies the model as an autoregressive model that simultaneously contains both a unit root and a sudden shift (either on slope, intercept or both).

The two statistics of interest are generalizations of the Z-tests proposed by Phillips (1987a). The intuition behind is simple: since the researcher is assumed to know the breakfraction λ, this effect must be removed from data. Thus, let $\{\tilde{y}_t^i\}$ denote detrended data under model i ($i = A$, B, C). Furthermore, let $\tilde{\alpha}^i$ be the least squares estimator of $\tilde{\alpha}^i$ in the following regression

$$\tilde{y}_t^i = \tilde{\alpha}^i \tilde{y}_{t-1}^i + \tilde{e}_t, \tag{29}$$

where $i = A$, B, C; $t = 1, 2, \ldots, T$. If the null hypothesis were in fact true, the value of $\tilde{\alpha}^i$ must be near to one or, equivalently, bias $\tilde{\alpha}^i - 1$ must be near to zero. Formally, the next theorem presents the asymptotic distribution of both standarized bias $T\left(\tilde{\alpha}^i - 1\right)$ and t statistic $t_{\tilde{\alpha}^i}$ along several specifications.

Theorem 11 (Perron, 1989, p. 1373) *Let the process $\{y_t\}$ be generated under the null hypothesis of model i ($i = A$, B, C) with the innovation sequence $\{u_t\}$ satisfying Assumption 5. Let $\Rightarrow$ denote weak convergence in distribution and $\lambda = T_B/T$ for all T. Then, as $T \to \infty$:*

$$a)\ T\left(\tilde{\alpha}^i - 1\right) \Rightarrow H_i/K_i; \qquad b)\ t_{\tilde{\alpha}^i} \Rightarrow (\sigma/\sigma_u)\, H_i/\left(g_i K_i\right)^{1/2};$$

where

$$H_A = g_A D_1 - D_5\psi_1 - D_6\psi_2; \qquad K_A = g_A D_2 - D_4\psi_2 - D_3\psi_1;$$

$$H_B = g_B D_1 + D_5\psi_3 + D_8\psi_4; \qquad K_B = g_B D_2 + D_7\psi_4 + D_3\psi_3;$$

$$H_C = g_C D_9 + D_{13}\psi_5 - D_{14}\psi_6; \qquad K_C = g_C D_{10} - D_{12}\psi_6 + D_{11}\psi_5;$$

with

$$\psi_1 = 6D_4 + 12D_3; \qquad \psi_2 = 6D_3 + (1-\lambda)^{-1}\lambda^{-1} D_4;$$

$$\psi_3 = (1+2\lambda)(1-\lambda)^{-1} D_7 - (1+3\lambda) D_3;$$

$$\psi_4 = (1+2\lambda)(1-\lambda)^{-1} D_3 - (1-\lambda)^{-3} D_7;$$

$$\psi_5 = D_{12} - D_{11}; \qquad \psi_6 = \psi_5 + (1-\lambda)^2 D_{12}/\lambda^3;$$

and

$$D_1 = (1/2)\left(W(1)^2 - \sigma_u^2/\sigma^2\right) - W(1)\int_0^1 W(r)\,dr;$$

$$D_2 = \int_0^1 W(r)^2\,dr - [\int_0^1 W(r)\,dr]^2;$$

$$D_3 = \int_0^1 rW(r)\,dr - (1/2)\int_0^1 W(r)\,dr;$$

$$D_4 = \int_0^\lambda W(r)\,dr - \lambda\int_0^1 W(r)\,dr;$$

$$D_5 = W(1)/2 - \int_0^1 W(r)\,dr;\ D_6 = W(\lambda) - \lambda W(1);$$

$$D_7 = \int_\lambda^1 rW(r)\,dr - \lambda\int_\lambda^1 W(r)\,dr - \left((1-\lambda)^2/2\right)\int_0^1 W(r)\,dr;$$

$$D_8 = \left(\left(1-\lambda^2\right)/2\right)W(1) - \int_\lambda^1 W(r)\,dr;$$

$$D_9 = \int_0^1 W(r)^2\,dr - \lambda^{-1}[\int_0^\lambda W(r)\,dr]^2 - (1-\lambda)^{-1}[\int_\lambda^1 W(r)\,dr]^2;$$

$$D_{10} = \left(W(1)^2 - \sigma_u^2/\sigma^2\right)/2 - \lambda^{-1}W(\lambda)\int_0^\lambda W(r)\,dr - (W(1) - W(\lambda))(1-\lambda)^{-1}\int_\lambda^1 W(r)\,dr;$$

$$D_{11} = \int_0^1 rW(r)\,dr - (1/2)(1+\lambda)\int_0^1 W(r)\,dr + (1/2)\int_0^\lambda W(r)\,dr;$$

$$D_{12} = \int_0^\lambda rW(r)\,dr - (\lambda/2)\int_0^\lambda W(r)\,dr;$$

$$D_{13} = (1-\lambda)W(1)/2 + W(\lambda)/2 - \int_0^1 W(r)\,dr;$$

$$D_{14} = \lambda W(\lambda)/2 - \int_0^\lambda W(r)\,dr;$$

$$g_A = 1 - 3(1-\lambda)\lambda; \qquad g_B = 3\lambda^3; \qquad g_C = 12(1-\lambda)^2;$$

where $\sigma^2 = \lim_{T\to\infty} E\left[T^{-1}S_T^2\right]$, $S_T = \sum_{t=1}^T u_t$, $\sigma_u^2 = \lim_{T\to\infty} E[T^{-1}\sum_{t=1}^T u_t^2]$ *and* W *is a standard Wiener process on* C.

Proof. See Perron (1989), Theorem 2 p. 1393. ■

The reader must take into account that the previous limiting distributions depend, besides λ, on nuisance parameters σ^2 and σ_u^2. The finding of consistent estimators for the variance of innovations σ_u^2 and the long run variance of partial sums σ^2 constitues an empirical issue. In the case of weakly stationary innovations, $\sigma^2 = 2\pi f(0)$ where $f(0)$ is the spectral density of $\{u_t\}$ evaluated at the zero frequency. Even more, Perron (1989) mentions that when the sequence $\{u_t\}$ is independent and identically distributed, $\sigma^2 = \sigma_u^2$ and in that case the limiting distributions are invariant with respect to nuisance parameters, except λ.

With this theoretical results and the tabulation of critical values through Monte Carlo simulation, evidence is found against the unit root hypothesis for the series studied by Nelson and Plosser (1982). Thus, the relevance of the results of Perron (1989) lies in the

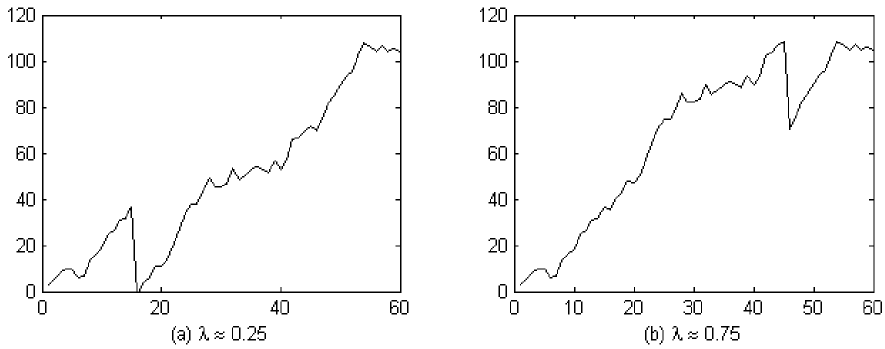

Figure 6: Sample paths under different breakfractions.

analysis of the performance of ADF tests when misspecification is present. As will be shown below, misspecification becomes crucial for the identification of desirable properties of new tests to be proposed. By the other hand, results generalize the tests due to Phillips (1987a) and the inference procedure assumes knowledge of both the existence of structural break and the breakfraction value. Subsequent studies progressively avoid this two assumptions and include desireable properties.

8 Endogenous structural break (Zivot and Andrews, 1992)

8.1 A simple reason for relaxing exogeneity

Before the formal analysis corresponding to this section, it is important to illustrate the main argument held by Zivot and Andrews (1992) against Perron (1989) through the following example. First, consider two sample paths as described in Figure 6. Under Perron's perspective, applied researchers are going to choose a breakfraction near to 0.25 for the first sample path, whereas they are more likely to choose a breakfraction near to 0.75 for the second one. Thus, breakfraction is not longer exogenous since the previous selections are based on *a priori* inspection of data, which incorporates an implicit selection rule behind. This fact is going to be exploited formally and will lead to the use of the Functional Central Limit Theorem under somewhat different conditions.

8.2 The approach

The first one of the two assumptions above mentioned is avoided by Zivot and Andrews (1992). They consider not an exogenous breakfraction but an endogenous one that has to be estimated. As they assert:

> "If one takes the view that these events are endogenous, then the correct unit root testing procedure would have to account for the fact that that the breakpoints in Perron's regressions are data dependent. The null hypothesis of interest in these cases is a unit root process with drift that excludes any structural change. The relevant alternative hypothesis is still a trend stationary process that allows for a one time break in the trend function. Under

Null hypothesis	Alternative hyphotesis
Model A	**Model A**
$y_t = \mu + y_{t-1} + u_t$	$y_t = \mu_1 + (\mu_2 - \mu_1)\, DU_t + \delta t + u_t$
Model B	**Model B**
$y_t = \mu + y_{t-1} + u_t$	$y_t = \mu + \delta_1 t + (\delta_2 - \delta_1)\, DT_t^* + u_t$
Model C	**Model C**
$y_t = \mu + y_{t-1} + u_t$	$y_t = \mu_1 + \delta_1 t + (\mu_2 - \mu_1)\, DU_t$ $+ (\delta_2 - \delta_1)\, DT_t + u_t$
where	
$DU_t = 1$ if $t > T_B$, 0 otherwise	$DT_t^* = t - T_B$ if $t > T_B$, 0 otherwise
$DT_t = t$ if $t > T_B$, 0 otherwise	

Table 3: Null and alternative hypotheses considered by Zivot and Andrews (1992).

> the alternative, however, we assume that we do not know exactly when the breakpoint ocurrs".

As noticed, attention is turned back to competing approaches shown in Figure 4 and formalized in Table 3. Additionally, while the tests developed by Perron (1989) are conditional on a given breakfraction $\lambda \in (0,1)$, Zivot and Andrews (1992) attemp to transform these tests into unconditional ones by designing an estimation method for λ.

It is important to mention that conventional wisdom in applied econometrics considers the so called Zivot-Andrews tests as unit root tests under structural break. By definition, this is not true since the null hypothesis considers only a unit root and no other deterministic component. By the other hand, in line with the structural change literature under unknown changepoint Zivot and Andrews (1992) suggest to choose the breakfraction λ that gives the least favorable result for the null hypothesis $H_0 : \alpha^i = 1$ $(i = A,\, B,\, C)$ using the one sided t statistic $t_{\hat{\alpha}}(\lambda)$ when small values of the statistic lead to the rejection of the null. Let $\hat{\lambda}^i_{\text{inf}}$ denote such a value for model i, then $t_{\hat{\alpha}}[\lambda^i_{\text{inf}}] \equiv \inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda)$ where Λ is a specified closed subset of $(0,1)$. For models A, B and C, t statistics are obtained from the following regression equations:

$$y_t = \hat{\mu}^A + \hat{\theta}^A DU_t(\hat{\lambda}) + \hat{\beta}^A t + \hat{\alpha}^A y_{t-1} + \textstyle\sum_{j=1}^k \hat{c}_j^A \Delta y_{t-1} + \hat{e}_t, \tag{30}$$

$$y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT_t^*(\hat{\lambda}) + \hat{\alpha}^B y_{t-1} + \textstyle\sum_{j=1}^k \hat{c}_j^B \Delta y_{t-1} + \hat{e}_t, \text{ and} \tag{31}$$

$$y_t = \hat{\mu}^C + \hat{\theta}^C DU_t(\hat{\lambda}) + \hat{\beta}^C t + \hat{\gamma}^C DT_t^*(\hat{\lambda}) + \hat{\alpha}^C y_{t-1} + \textstyle\sum_{j=1}^k \hat{c}_j^C \Delta y_{t-1} + \hat{e}_t \tag{32}$$

respectively, where parameter estimates are denoted with a hat and $\hat{e}_t$ is the residual term. In (30)-(32) $DU_t(\lambda) = 1$ if $t > T\lambda$ and 0 otherwise and $DT_t^*(\lambda) = t - T\lambda$ if $t > T\lambda$ and 0 otherwise. The number of extra lags k is here included to potentially take into account correlation between disturbances and $\hat{\lambda}$ denotes the estimated value of λ. In order to make the results as simple as possible, the author consider first the case $k = 0$

(no correlation among disturbances). In contrast to the work of Perron (1989), when correlation between disturbances is present, it is restricted to be of the ARMA structure. It is worth to mention that this structure is a particular case of mixing processes and this implies that the Functional Central Limit Theorem still can be applied.

For testing, intuition relies on the following reasoning: if H_0 were in fact true then the minimum t statistic should not significantly differ from zero, whereas if H_1 were true then H_0 should be rejected and an estimated value for λ would be provided for the alternative trend stationary specification. When λ is estimated, critical values in Perron (1989) cannot be employed for unit root testing. Consider an estimated λ with minimum t statistic. Then, decission rule can be summarized as

$$\text{reject } H_0 \text{ if } \inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda) < \kappa^i_{\inf,\upsilon},\ i = A, B, C,$$

where $\kappa^i_{\inf,\upsilon}$ denotes the asymptotic critical value of $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda)$ for a size equal to υ. By definition, critical values are as bigger (in absolute value) to those calculated on the basis of an arbitrary λ. Thus, the tests built by Perron (1989) are biased towards rejecting the null. In order to formally establish this distinction, distributions for the statistics $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda)$ ($i = A$, B, C) are needed.

8.3 Asymptotic distribution theory

In order to obtain the limiting distribution for their proposed statistic, Zivot and Andrews (1992) make use of the framework suggested by Ouliaris, Park, and Phillips (1989), which allows for a compact form for their results. It is worth to mention that this framework is also used by Perron (1989) when the objetive is to develop a generalization for his main theorem to the case of disturbances that exhibit autocorrelation. Attention is here focused on i.i.d. disturbances. The following two definitions are necessary for the undestanding of the main theorem.

Definition 9 $L_2[0,1]$ *is the Hilbert space of square integrable functions on* $[0,1]$ *with inner product* $\langle f,g\rangle \equiv \int_0^1 fg$ *for* f, $g \in L_2[0,1]$.

Definition 10 $W^i(\lambda,r)$ *is the stochastic process on* $[0,1]$ *that is the projection residual in* $L_2[0,1]$ *of a Wiener process projected onto the subspace generated by the following:*

1. *for* $i = A$: 1, r, $du(\lambda,r)$;
2. *for* $i = B$: 1, r, $dt^*(\lambda,r)$; *and*
3. *for* $i = C$: 1, r, $du(\lambda,r)$, $dt^*(\lambda,r)$

where $du(\lambda,r) = 1$ *if* $r > \lambda$, 0 *otherwise and* $dt^*(\lambda,r) = r - \lambda$ *if* $r > \lambda$ *and* 0 *otherwise.*

Asymptotic distribution is now given in the next theorem[12].

[12] Although independent, the derivation here presented is done in a similar fashion to those reported by Banerjee, Lumsdaine, and Stock (1992).

Theorem 12 (Zivot and Andrews, 1992 Theorem 1 p. 256) *Let $\{y_t\}$ be generated under the null hypothesis and let the disturbances $\{u_t\}$ be i.i.d., mean 0, variance σ^2 random variables with $0 < \sigma^2 < \infty$. Let $t_{\hat{\alpha}^i}(\lambda)$ denote the t statistic for testing $\alpha^i = 1$ computed from either (30), (31), or (32) with $k = 0$ for Models $i = A$, B, and C, respectively. Let Λ be a closed subset of $(0, 1)$. Then,*

$$\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda) \Rightarrow \inf_{\lambda\in\Lambda}[\int_0^1 W^i(\lambda, r)^2\, dr]^{-1/2}[\int_0^1 W^i(\lambda, r)\, dW(r)]$$

for $i = A$, B, and C, where $\Rightarrow$ denotes convergence in distribution.

Proof. See Zivot and Andrews (1992), Appendix A, p. 266. ■

It is worth to mention that when correlation of the ARMA type is allowed, the previous result can be extended in order to obtain an autoregressive estimate the spectral density of e_t at the zero frequency. This empirical issue is addressed by authors with the help of an assumption similar to assumption 2 of Phillips (1987a). That is, the probability of outliers is controlled and such an assumption will also be adopted in posterior work.

9 Efficient unit root testing under structural break (Perron and Rodríguez, 2003)

9.1 Motivation

Based on elements contained in the previous sections, two features can be identified along the unit root literature:

1. Deterministic trend and size. Most of the earlier unit root tests under less restrictive assumptions are extensions to augmented Dickey-Fuller tests and thus the asymptotic distributions depend on whether a deterministic component has been added or not into the regression equation. According to Stock (1999) this problem can be solved by first detrending the series and performing (robust) modified unit root tests such that size is not affected.

2. Structural break and power. Perron (1989) illustrated how deterministic trends that contain a break can induce spurious unit roots in Dickey-Fuller tests. Following Stock (1999), a trend with structural break can be incorporated in the detrending process. Since it is guaranteed that size will not be affected, it becomes desirable to increase the power of the tests against local alternatives. Such a procedure can be done following the near-integrated time series approach proposed by Phillips (1987b) and developed by Elliott, Rothenberg, and Stock (1996) in the case of no structural break. Thus, an extension is called to.

Within this framework, Perron and Rodríguez (2003) extend the modified or M tests, analyzed in detail by Ng and Perron (2001), to the case in which there exists a structural break in the trend function.

Structural change in slope	Structural change in trend and slope
Model A	**Model B**
$\psi' z_t = \beta_1 t + \beta_2 DT_t^*$	$\psi' z_t = \mu_1 + \mu_2 DU_t + \beta_1 t + \beta_2 DT_t^*$
where	where
$DT_t^* = t - T_B$ if $t > T_B$, 0 otherwise	$DU_t = 1$ if $t > T_B$, 0 otherwise

Table 4: Deterministic components considered by Perron and Rodríguez (2003).

9.2 Data generating process

Observed series $\{y_t\}_{t=0}^{T}$ is assumed to be generated according to

$$y_t = \psi' z_t + u_t, \text{ and} \tag{33}$$
$$u_t = \alpha u_{t-1} + v_t \tag{34}$$

for $t = 1, \ldots, T$. Following the framework in Perron (1989), the test to be proposed considers a structural break under the null. Perron and Rodríguez (2003) consider two models for structural change, summarized in Table 4. A model with structural change in the intercept is not considered since its limiting distribution is the same as those corresponding to both intercept and slope. For disturbances the authors, following Phillips and Solo (1992), adopt the following

Assumption 6 (Perron and Rodríguez, 2003 p. 3) *The following conditions hold:*

1. $u_0 = 0$, *and*
2. *the noise function is* $v_t = \sum_{i=0}^{\infty} \gamma_i \varepsilon_{t-i}$ *where* $\sum_{i=0}^{\infty} i|\gamma_i| < \infty$ *and where* $\{\varepsilon_t\}$ *is a m.d.s. The process* $\{v_t\}$ *has a non-normalized spectral density at frequency zero given by* $\sigma^2 = \sigma_\varepsilon^2 \gamma(1)^2$, *where* $\sigma_\varepsilon^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{\infty} E(\varepsilon_t^2)$. *Furthermore,* $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} v_t \Rightarrow \sigma W(r)$, *where* $\Rightarrow$ *denotes weak convergence in distribution and* $W(r)$ *is the standard Wiener process defined on* $C[0,1]$ *the space of continous functions on the interval* $[0,1]$.

9.3 *GLS* detrending and *M* tests

First, define transformed data by

$$\tilde{y}_t^{\bar{\alpha}} = (y_0, (1 - \bar{\alpha}L)\, y_t), \quad \tilde{z}_t^{\bar{\alpha}} = (z_0, (1 - \bar{\alpha}L)\, z_t), \quad t = 0, \ldots, T,$$

and let $\hat{\psi}$ be the estimator that minimizes (35)

$$S^*(\psi, \bar{\alpha}, \lambda) = \textstyle\sum_{t=0}^{T} (y_t^{\bar{\alpha}} - \psi' \tilde{z}_t^{\bar{\alpha}})^2. \tag{35}$$

Data is transformed in order to make results dependent on parameter $\bar{\alpha}$. The goal here is to derive an optimal unit root tests against a local alternative hypothesis. In this sense, later a computed value for $\bar{\alpha}$ will be necessary. Based on Phillips (1987b), both null and alternative hypotheses can be summarized by means of a near-integrated process. In (34), the autoregressive coefficient can be written as

$$\alpha = 1 + \frac{c}{T}.$$

Then, under the null $c = 0$ whereas under the alternative $c < 0$ and the power function can thus be explicity obtained. The M tests, studied in section 6, are defined by

$$MZ_{\alpha}^{GLS}(\lambda) = \frac{1}{2}\frac{T^{-1}\tilde{y}_T^2 - \hat{\sigma}^2}{T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2}, \quad (36)$$

$$MSB^{GLS}(\lambda) = (T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2/\hat{\sigma}^2)^{1/2}, \quad (37)$$

$$MZ_{t}^{GLS}(\lambda) = \frac{1}{2}\frac{T^{-1}\tilde{y}_T^2 - \hat{\sigma}^2}{(\hat{\sigma}^2 T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2)^{1/2}}, \quad (38)$$

with local detrended data defined by $\tilde{y}_t = y_t - \hat{\psi}' z_t$ where $\hat{\psi}$ minimizes (35). The term $\hat{\sigma}^2$ is an autoregressive estimate of the spectral density at frequency zero of v_t, defined as

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_{vk}^2}{[1 - \hat{b}(1)]^2},$$

$$\hat{\sigma}_{vk}^2 = \frac{\sum_{t=k+1}^{T}\hat{v}_{tk}^2}{T - k},$$

$$\hat{b}(1) = \sum_{j=1}^{k}\hat{b}_j,$$

where $\hat{b}_j$ and $\{\hat{v}_{tk}\}$ are obtained from the auxiliar ADF regression

$$\Delta\tilde{y}_t = b_0\tilde{y}_{t-1} + \sum_{j=1}^{k} b_j\Delta\tilde{y}_{t-j} + v_{tk}. \quad (39)$$

9.4 Asymptotic distributions

The next theorem presents the limiting distribution of the testing statistics for fixed values of c, $\bar{c}$ and λ.

Theorem 13 (Perron and Rodríguez, 2003 p. 7) *Let $\{y_t\}_{t=0}^T$ be generated by model (33) with $\alpha = 1 + c/T$, MZ_{α}^{GLS}, MSB^{GLS} and MZ_t^{GLS} be defined by (36)-(38) with data obtained from local GLS detrending ($\tilde{y}_t$) at $\bar{\alpha} = 1 + \bar{c}/T$, and ADF^{GLS} be the t statistic for testing $b_0 = 0$ in the regression (39). Also, $\hat{\sigma}^2$ is a consistent estimate of σ^2. For models A and B*

$$MZ_{\alpha}^{GLS}(\lambda) \Rightarrow \frac{1}{2}\frac{K_1(c,\bar{c},\lambda)}{K_2(c,\bar{c},\lambda)} \equiv H^{MZ_{\alpha}^{GLS}}(c,\bar{c},\lambda),$$

$$MSB^{GLS}(\lambda) \Rightarrow [K_2(c,\bar{c},\lambda)]^{1/2} \equiv H^{MSB^{GLS}}(c,\bar{c},\lambda),$$

$$MZ_{t}^{GLS}(\lambda) \Rightarrow \frac{1}{2}\frac{K_1(c,\bar{c},\lambda)}{[K_2(c,\bar{c},\lambda)]^{1/2}} \equiv H^{MZ_{t}^{GLS}}(c,\bar{c},\lambda),$$

$$ADF^{GLS}(\lambda) \Rightarrow \frac{1}{2}\frac{K_1(c,\bar{c},\lambda)}{[K_2(c,\bar{c},\lambda)]^{1/2}} \equiv H^{ADF^{GLS}}(c,\bar{c},\lambda),$$

where

$$K_1(c,\bar{c},\lambda) = V_{c\bar{c}}^{(1)}(1,\lambda)^2 - 2V_{c\bar{c}}^{(2)}(1,\lambda) - 1,$$

$$K_2(c,\bar{c},\lambda) = \int_0^1 V_{c\bar{c}}^{(1)}(r,\lambda)^2 dr - 2\int_{\lambda}^1 V_{c\bar{c}}^{(2)}(r,\lambda)dr,$$

and $V^{(1)}_{c\bar{c}}(r,\lambda) = W_c(r) - rb_3$, $V^{(2)}_{c\bar{c}}(r,\lambda) = b_4(r-\lambda)[W_c(r) - rb_3 - (1/2)(r-\lambda)b_4]$ *with* $W_c(r)$ *the Ornstein-Uhlenbeck process that is the solution to the stochastic differential equation*

$$dW_c(r) = cW_c(r)dr + dW(r) \text{ with } W_c(r) = 0.$$

Also, b_3 *and* b_4 *are defined by*

$$\begin{aligned}
b_3 &= \lambda_1 b_1 + \lambda_2 b_2, \\
b_4 &= \lambda_2 b_1 + \lambda_3 b_2, \\
b_1 &= (1-\bar{c})W_c(1) + \bar{c}^2 \textstyle\int_0^1 rW_c(r)dr, \\
b_2 &= (1-\bar{c}+\lambda\bar{c})W_c(1) + \bar{c}^2 \textstyle\int_\lambda^1 W_c(r)(r-\lambda)dr - W_c(\lambda), \\
\lambda_1 &= d/\Theta, \\
\lambda_2 &= -m/\Theta, \\
d &= 1-\lambda-c+2c\lambda-c\lambda^2-c^2\lambda+c^2\lambda^2+(c^2/3)(1-\lambda^3), \\
m &= 1-\lambda-c+c\lambda+(c^2/2)\lambda^3+(c^2/3)(1-\lambda^3), \\
a &= 1-c+c^2/3, \\
\Theta &= ad-m^2 \text{ and} \\
\lambda_3 &= a/\Theta.
\end{aligned}$$

Proof. See Perron and Rodríguez (2003), Theorem 1, p. 22. ■

9.5 Asymptotic power function

As Phillips (1988) pointed out, the discriminatory power of unit root tests is low against local alternatives near but not equal to unity because under both hypotheses distributions are quite similar. The main idea behind efficiency relies on the increase of power or, equivalently, the probability of rejecting a false alternative hypothesis. As mentioned by Elliott, Rothenberg, and Stock (1996), if data distribution were known then the Neyman-Pearson Lemma would suggest the optimal point alternative against any other point alternative hypothesis and in such circunstances a power envelope can be derived[13].

However, although within this framework a uniformly most powerful (UMP) tests is not attainable, it is possible to define an optimal test for $\alpha = 1$ against the alternative $\alpha = \bar{\alpha}$. Even more, if v_t were i.i.d. then such a test is given by the likelihood ratio statistic which, under the normality assumption, equals the following difference

$$L(\lambda) \equiv S(\bar{\alpha},\lambda) - S(1,\lambda),$$

where $S(\bar{\alpha},\lambda)$ and $S(1,\lambda)$ are the sums of squares from GLS detrending both under $\alpha = \bar{\alpha}$ and $\alpha = 1$, respectively. Under the assumption of a known breakfraction λ,

[13] As Elliott, Rothenberg, and Stock (1996) mention, the Gaussian power envelope is an upper bound to the asymptotic power function for tests of the unit root hypothesis when the data are generated by

$$y_t = d_t + u_t \text{ and } u_t = \alpha u_{t-1} + \nu_t$$

but under "ideal" conditions. Namely, the process $\{\nu_t\}$ has a moving average representation involving independent standard normal random variables, the initial condition u_0 is 0 and the deterministic component d_t is known. Such unrealistic assumptions are made in order to employ the Neyman-Pearson theory.

different values for $\bar{\alpha}$ lead to a family of point optimal tests and a gaussian envelope for testing $\alpha = 1$. Furthermore, in order to allow for correlation between errors v_t, Elliott, Rothenberg, and Stock (1996) propose a feasible optimal point test P_T^{GLS} defined by

$$P_T^{GLS}(c, \bar{c}, \lambda) = \frac{S(\bar{\alpha}, \lambda) - \bar{\alpha} S(1, \lambda)}{\hat{\sigma}^2}, \tag{40}$$

and its distribution is derived in the following

Theorem 14 (Perron and Rodríguez, 2003 p. 7) *Let $\{y_t\}$ be generated by (33) with $\alpha = 1 + c/T$. Let P_T^{GLS} be defined by (40) with data obtained from local GLS detrending ($\tilde{y}_t$) at $\bar{\alpha} = 1 + \bar{c}/T$. Also, let $\hat{\sigma}^2$ be a consistent estimate of σ^2. The limit distribution of the P_T^{GLS} under Models A and B is given by*

$$\begin{aligned} P_T^{GLS}(c, \bar{c}, \lambda) &\Rightarrow M(c, 0, \lambda) - M(c, \bar{c}, \lambda) \\ &\quad - 2\bar{c} \textstyle\int_0^1 W_c(r) dW(r) + (\bar{c}^2 - 2\bar{c}c) \int_0^1 W_c(r)^2 dr - \bar{c} \equiv H^{P_T^{GLS}}(c, \bar{c}, \lambda) \end{aligned}$$

where $M(c, \bar{c}, \lambda) = A(c, \bar{c}, \lambda) B(\bar{c}, \lambda)^{-1} A(c, \bar{c}, \lambda)$ with $A(c, \bar{c}, \lambda)$ a 2×1 vector defined by

$$\left[\begin{array}{c} W(1) + (c - \bar{c}) \int_0^1 W_c(r) dr - \bar{c} \int_0^1 r dW(r) - (c - \bar{c}) \bar{c} \int_0^1 r W_c(r) dr \\ (1 + \lambda \bar{c})([W(1) - W(\lambda)] + (c - \bar{c}) \int_\lambda^1 W_c(r) dr) - \bar{c} \int_\lambda^1 r dW(r) - (c - \bar{c}) \bar{c} \int_\lambda^1 r W_c(r) dr \end{array} \right]$$

and $B(\bar{c}, \lambda)$ is a symmetric matrix with entries

$$\left[\begin{array}{cc} \bar{c}^2/3 - \bar{c} + 1 & (1 - \lambda)(1 - \bar{c}) + \bar{c}^2(2 + \lambda^3 - 3)/6 \\ & \bar{c}^2(1 - \lambda^3)/3 - \bar{c}(1 - \lambda^2)(1 + \lambda \bar{c}) + (1 - \lambda)(1 + \lambda \bar{c})^2 \end{array} \right].$$

Proof. See Perron and Rodríguez (2003), Theorem 2, p. 24. ■

The reader must remember that any test statistic is also a random variable and rejecting the unit root hypothesis is an event in which the test statistic lies below some critical value. Since distribution for the tests was derived both under the null and the alternative hypothesis, the (asymptotic) power function can be explicited by means of the probability of rejecting the null under the alternative. Such a function is given by

$$\pi(c, \lambda) \equiv P[H^{P_T^{GLS}}(c, c, \lambda) < b^{P_T^{GLS}}(c, \lambda)],$$

where the critical value $b^{P_T^{GLS}}(c, \lambda)$ is determined by the probability of Type I error

$$P[H^{P_T^{GLS}}(0, c, \lambda) < b^{P_T^{GLS}}(c, \lambda)] = \upsilon,$$

and υ is the size of the test. Thus, different values of λ generate different power functions.

9.6 A feasible point optimal test

The previous subsections are referred to the case in which the breakfraction λ is known. In practice, however, this parameter is required to be estimated by applied researchers. For this reason the feasible version of the statistic in (40) is given by

$$P_{T,*}^{GLS}(c, \bar{c}) = \left\{ \inf_{\lambda \in [\varepsilon, 1-\varepsilon]} S(\bar{\alpha}, \lambda) - \inf_{\lambda \in [\varepsilon, 1-\varepsilon]} \bar{\alpha} S(1, \lambda) \right\} \Big/ \hat{\sigma}^2. \tag{41}$$

The principle behind (41) is the same as in (40). The main difference relies on the trimming parameter ε introduced. This latter parameter is usually set to 0.15 in order to bound critical values, a situation that arises in the context of tests for structural change. Using Theorem 14, the following result is obtained

$$P_{T,*}^{GLS}(c,\bar{c}) \Rightarrow \sup_{\lambda\in[\varepsilon,1-\varepsilon]} M(c,0,\lambda) - \sup_{\lambda\in[\varepsilon,1-\varepsilon]} M(c,\bar{c},\lambda) \\ - 2\bar{c}\int_0^1 W_c(r)dW(r) + (\bar{c}^2 - 2\bar{c}c)\int_0^1 W_c(r)^2 dr - \bar{c} \equiv H^{P_{T,*}^{GLS}}(c,\bar{c}) \,.$$

Accordingly, the asymptotic Gaussian power envelope is given by

$$\pi^*(c) \equiv P[H^{P_{T,*}^{GLS}}(c,c) < b^{P_{T,*}^{GLS}}(c)],$$

where the critical value $b^{P_{T,*}^{GLS}}(c)$ is such that $P[H^{P_{T,*}^{GLS}}(0,c) < b^{P_{T,*}^{GLS}}(c)] = \upsilon$. It must be pointed out that Elliott, Rothenberg, and Stock (1996) recommended a value for $\bar{c}$ such that $\pi^*(\bar{c}) = 0.5$. Using Monte Carlo simulation, Perron and Rodríguez (2003) found that $\bar{c} = -22.5$.

It must be emphasized that, within this literature, the idea behind power increasing unit root tests is related to the extent in which power functions are near to the Gaussian power envelope (the benchmark case). When λ is known, only one set of simulations is performed in order to obtain the power function corresponding to that value. When λ is unknown, on the contrary, several sets of simulations are performed (one for each value of λ in $[\varepsilon, 1-\varepsilon]$).

10 Conclusions

The present paper has analyzed the foundations and the applicability of the Functional Central Limit Theorem to the task of developing unit root tests. As shown, unit root tests can be described as functionals of stochastic processes such as the standard Wiener process and the Ornstein-Uhlenbeck process.

Therefore, a general framework involving mixing conditions (Phillips 1987a) generalizes the results obtained under the assumption of normal i.i.d. disturbances (Dickey and Fuller 1979). Also, the analysis of modified tests (Stock 1999) allows to separate the size of unit root tests from the specific form of the deterministic component, a problem not solved in earlier works. Tools developed also allow the analitical tractability of several problem within this literature: the presence of strucutral breaks and the low power against local alternatives. For the issue of structural breaks (Perron 1989), first detrending the series has shown to be a robust procedure, so that asymptotic size is not affected (Stock 1999). For the issue of increasing power, asymptotic distribution can be derived by means of Ornstein-Uhlenbeck processes both under the null and local alternatives (Phillips 1987b) and a power function can be derived and maximized. When the two issues are combined, the result is an efficient test when a structural break is present under the null (Perron and Rodríguez 2003).

References

BANERJEE, A., R. L. LUMSDAINE, AND J. H. STOCK (1992): "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10(3), 271–87.

BHARGAVA, A. (1986): "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53(3), 369–84.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. John Wiley & Sons, Inc.

BOX, G. E. P., AND G. M. JENKINS (1970): *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Inc.

BOX, G. E. P., AND D. A. PIERCE (1970): "Distribution of the Autocorrelations in Autoregressive Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–26.

BRZEZNIAK, Z., AND T. ZASTAWNIAK (1999): *Basic Stochastic Processes*. Springer.

DAVIDSON, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autorregresive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–31.

DONSKER, M. (1951): "An Invariance Principle for Certain Probability Limit Theorems," *Memoirs of the American Mathematical Society*, 6, 1–12.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–36.

ENDERS, W. (2004): *Applied Econometric Time Series, Second Edition*. John Wiley & Sons, Inc.

GLYNN, J., AND N. PERERA (2007): "Unit Root Tests and Structural Breaks: A Survey with Applications," *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 3(1), 63–79.

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.

HERRNDORF, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12(1), 141–53.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and random walks in macroeconmic time series : Some evidence and implications," *Journal of Monetary Economics*, 10(2), 139–62.

NG, S., AND P. PERRON (2001): "LAG Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69(6), 1519–54.

OKSENDAL, B. (2000): *Stochastic Differential Equations*. Springer.

Ouliaris, S., J. Y. Park, and P. C. B. Phillips (1989): "Testing for a Unir Root in the Presence of a Maintained Trend," in *Advances in Econometrics Modelling*, ed. by B. Raj. Kluwern Academic.

Perron, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57(6), 1361–1401.

Perron, P., and G. Rodríguez (2003): "GLS detrending, efficient unit root tests and structural change," *Journal of Econometrics*, 115(1), 1–27.

Phillips, P. C. B. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55(2), 277–301.

——— (1987b): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74(3), 535–47.

Phillips, P. C. B. (1988): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56(5), 1021–43.

Phillips, P. C. B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75(2), 335–46.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for Linear Processes," *The Annals of Statistics*, 20(2), 971–1001.

Sargan, J. D., and A. Bhargava (1983): "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51(1), 153–74.

Stock, J. (1999): "A Class of Tests for Integration and Cointegration," in *Cointegration: Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, ed. by R. F. Engle, and H. White, pp. 135–167. Oxford University Press.

White, H. (1984): *Asymptotic Theory for Econometricians*. Academic Press.

White, J. S. (1958): "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *The Annals of Mathematical Statistics*, 29(4), 1188–97.

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 & Economic Statistics*, 10(3), 251–70.

A Appendix. Beveridge-Nelson decomposition

Based on Phillips and Solo (1992), let the operator $C(L) = \sum_{j=0}^{\infty} c_j L^j$ be a lag polinomial. Then

$$C(L) = C(1) - (1 - L)\tilde{C}(L)$$

where

$$\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j, \; \tilde{c}_j = \sum_{k=j+1}^{\infty} c_k.$$

If $p \geq 1$, then

$$\sum_{j=1}^{\infty} j^p |c_j|^p < \infty \text{ implies } \sum_{j=0}^{\infty} |\tilde{c}_j|^p < \infty \text{ and } |C(1)| < \infty.$$

If $p < 1$, then

$$\sum_{j=1}^{\infty} j |c_j|^p < \infty \text{ implies } \sum_{j=0}^{\infty} |\tilde{c}_j|^p < \infty.$$

B Appendix. Martingale difference sequence

Let $\{x_t\}$ and $\{y_t\}$ denote two stochastic processes. Then $\{y_t\}$ is a martingale difference sequence with respect to $\{x_t\}$ if its expectation, conditional to past values of $\{x_t\}$, is zero. Formally

$$E[y_t | x_{t-1}, x_{t-2}, \ldots] = 0, \text{ for all } t.$$

When the expectation of $\{y_t\}$, conditional to its own past values, is zero then $\{y_t\}$ is said to be a martingale difference sequence (m.d.s.).

C Appendix. Strongly uniform integrability

Let $\{Z_t\}_{t=1}^{\infty}$ be a sequence of random variables adapted to the filtration $\{\mathcal{F}_t\}_{t=1}^{\infty}$. For Phillips and Solo (1992), $\{Z_t\}$ is said to be strongly uniformly integrable (s.u.i.) if there exists a dominating random variable Z for which $E(|Z|) < \infty$ and

$$P(|Z_t| \geq x) \leq cP(|Z| \geq x)$$

for each $x \geq 0$, $t \geq 1$ and for some constant c.