Understanding the Functional Central Limit Theorems with Some Applications to Unit Root Testing with Structural Change[*]

Juan Carlos Aquino[**]

Gabriel Rodríguez[***]

ABSTRACT

The application of different unit root statistics is by now a standard practice in empirical work. Even when it is a practical issue, these statistics have complex nonstandard distributions depending on functionals of certain stochastic processes, and their derivations represent a barrier even for many theoretical econometricians. These derivations are based on rigorous and fundamental statistical tools which are not (very) well known by standard econometricians. This paper aims to fill this gap by explaining in a simple way one of these fundamental tools: namely, the Functional Central Limit Theorem. To this end, this paper analyzes the foundations and applicability of 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. Thereafter, attention is focused 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 Rodriguez (2003), which extends the Generalized Least Squares detrending approach due to Elliot *et al.* (1996). An empirical application is provided.

Keywords: Unit Root Testing, Structural Break, Functional Central Limit Theorem, Ornstein-Uhlenbeck Process.

JEL Codes: C12, C22

** Banco Central de Reserva del Perú.

*** Mailing address: Gabriel Rodríguez, Department of Economics, Pontificia Universidad Católica del Perú, Av. Universitaria 1801, Lima 32, Lima, Peru; Telephone: +511-626-2000 (4998); Fax: +511-626-2874; E-mail address: gabriel.rodriguez@pucp.edu.pe.

El Teorema del Límite Central Funcional con algunas aplicaciones a raíces unitarias con cambios estructurales

RESUMEN

Hoy en día es una práctica estándar de trabajo empírico la aplicación de diferentes estadísticos de contraste de raíz unitaria. A pesar de ser un aspecto práctico, estos estadísticos poseen distribuciones complejas y no estándar que dependen de funcionales de ciertos procesos estocásticos y sus derivaciones representan una barrera incluso para varios econometristas teóricos. Estas derivaciones están basadas en herramientas estadísticas fundamentales y rigurosas que no son (muy) bien conocidas por econometristas estándar. El presente artículo completa esta brecha al explicar en una forma simple una de estas herramientas fundamentales la cual es el Teorema del Límite Central Funcional. Por lo tanto, este documento analiza los fundamentos y la aplicabilidad de dos versiones del Teorema del Límite Central Funcional dentro del marco de una raíz unitaria con un quiebre estructural. La atención inicial se centra en la estructura probabilística de las series de tiempo 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 las pruebas Dickey-Fuller aumentadas y por Zivot y Andrews (1992) para criticar el supuesto de exogeneidad y proponer un método para estimar un punto de quiebre endógeno. Un método sistemático para tratar con aspectos de eficiencia es introducido por Perron y Rodríguez (2003), el cual extiende el enfoque de Mínimos Cuadrados Generalizados para eliminar los componentes determinísticos de Elliot *et al.* (1996). Se presenta además una aplicación empírica.

Palabras clave: Prueba de Raíz Unitaria, Quiebre Estructural, Teorema del Límite Central Funcional, Proceso Ornstein-Uhlenbeck.

JEL Codes: C12, C22

1. INTRODUCTION

The application of different unit root statistics is by now a standard practice in empirical work. In spite of being a practical issue, these statistics have complex nonstandard distributions that depend on the functionals[1] of some stochastic processes, and their derivations represent a challenge even for many theoretical econometricians. These derivations are based on rigorous and fundamental statistical tools which are not (very) well known by standard econometricians. This paper aims to plug this gap by explaining in a simple way one of these fundamentals tools: the Functional Central Limit Theorem. To this end, this paper analyzes the foundations and applicability of two related versions of the Functional Central Limit Theorem within the framework of a unit root with a structural break.

---

[1]   A functional can be understood as a map that takes a real-valued function as the input argument and returns a real number. Naturally, this idea extends to the case of random real-valued functions.

Four decades ago, the empirical study of key macroeconomic variables was done through the use of the ARMA models proposed by Box and Jenkins (1970). In this type of models, first and second moments depend upon time separation but do not depend on the time variable itself. Hence, these models are covariance stationary[2], whose behavior reverts to a time invariant unconditional mean and where the associated methodology is based on the steps of identification, estimation and diagnostic[3].

However, the assumptions underlying ARMA models are not suitable for modeling macroeconomic series, which usually exhibit an upward trend over time. Hence, any model that aims at representing macroeconomic data must include such a trend. One of the most popular approach to this task is the deterministic trend model: $y_t = \mu + \delta t + u_t$, $t = 1,\ldots, T$, where 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 subtracting 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 autoregressive, containing a unit root: $y_t = \mu + \alpha y_{t-1} + u_t$, where $t = 1,\ldots,T$, 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, the realization of any u_i has a permanent effect on the level of y_t and the appropriate procedure to obtain a stationary series is to work on first differences $\Delta y_t = y_t - y_{t-1}$.

From an economic viewpoint, these two approaches require the identification of the type of processes representing macroeconomic data and to understand the long term effects of shocks. Also, based on a predictive perspective, this distinction is nontrivial since in the deterministic trend model the forecasting error has a constant variability whereas in the stochastic case this element has an increasing variability[4].

Turning back to empirical concerns, the unit root framework allows us to consider 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 and α are constants, y_0 is an initial condition, $u_t \sim N(0, \sigma_u^2)$ and $\sigma_u^2 > 0$. A first conclusion to be arrived at is that the effect of shocks on the dependent variable is linked to the unrestricted value of α, an assertion that can be confirmed after manipulating the previous expression: $y_t = \alpha^t y_0 + \mu \sum_{i=1}^{t} \alpha^{t-i} + \sum_{i=1}^{t} \alpha^{t-i} u_i$. For $\mu = 0$, the process reduces to

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

and allows for testing

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

---

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

The study by White (1958) was the first to perform such a procedure: in order to test H_0 against H_1 with a sample of size T and the OLS estimator $\hat{\alpha}$ for parameter α, under the null hypothesis, he calculated 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 a centered and standardized estimator for α, a random variable, and $\Rightarrow$ denotes weak convergence of probability measures. This result was an application of a theorem due to Donsker (1951) and the asymptotic distribution was formulated in terms of functionals of a standard Wiener process W whose details and properties are to be examined. It is worth mentioning that this result is not independent of the correlation between the disturbance terms u_t (assumed to be zero in this case for sake of simplicity) and the fact that there is no specification error when estimating α.

Another study in this line was that of Dickey and Fuller (1979), who assumed normal i.i.d. disturbances and developed several one-tailed tests with the following rejection rule: for a given confidence level, if the (properly transformed) centered estimator $\hat{\alpha}-1$ yields a value that is low relative to a critical value, then the unit root hypothesis is rejected. In order to understand the previous rule, consider 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 true if and only if $b_0 = 0$. In this context, the 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.

The analysis by Dickey and Fuller (1979) is done by considering three types of autoregressive models: without an intercept or (deterministic) trend, with an intercept but without a trend, and with both an intercept and a trend. In this particular study, assumptions allow the asymptotic distributions to be represented through moment generating functions. Monte Carlo simulations allow the authors to compare the power of these tests with those of (autocorrelation-based) Q statistics proposed by Box and Pierce (1970). Their main results are: firstly, Q statistics are systematically less powerful; secondly, the performance of Dickey-Fuller test is uniformly superior when there is no misspecification error[5]; and thirdly, there is evidence that Dickey-Fuller tests are biased towards not rejecting the null hypothesis for values of the autoregressive coefficient α arbitrarily close to 1.

---

[5]   For example, this occurs because the knowledge of the true value for the intercept (equal to zero) is being exploited.

For our purposes, a simple way to illustrate the role of specification is provided 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 the intercept is redundant and when both intercept and trend are redundant. It can be appreciated that the 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 test biased towards not rejecting the null hypothesis and, in this sense, their power is reduced.

Figure 1. Asymptotic distributions for several specifications
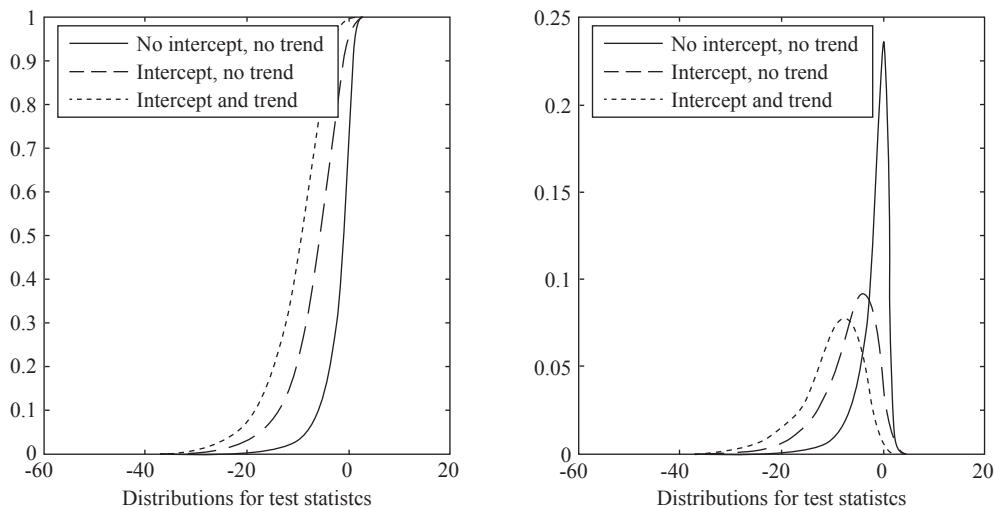
Source: authors' calculations.

So far, this brief review shows that, for the first half of the 1980s, unit root econometrics exhibited two well-defined limitations: vulnerability to misspecification and to local stationary alternatives, and that 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 were produced during the second half of that decade. Firstly, Phillips (1987a) proposed an asymptotic theory under very general conditions for integrated processes, which meant that the subsequent discussion was to be conducted under firmly established foundations, and secondly, Perron (1989) identified the presence of a structural break as an element that also reduced the power of the augmented Dickey-Fuller tests.

The reader must also take into account that none of these two advances could have been devised without discussing the notion of weak convergence of probability measures. To understand the need for this concept, it is necessary to consider first the classical Central Limit Theorem which, under conditions that vary along different versions,

allows for the distribution of the centered and standardized 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 any of the normal i.i.d. assumptions 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 standardized 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

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

Therefore, in order to fully understand the advances in this literature, two requisites are needed. Firstly, one must formally understand both the mathematical and probabilistic structure of the data-generating processes in order to state the main (weak) convergence results. Secondly, and most importantly, one needs to recognize the importance of incorporating particular problems faced by researchers into the analysis, because their formalization leads to the development of new specific procedures and testing statistics. This task is frequently undertaken by employing creative alternative hypotheses, which help to identify current limitations.

This paper reviews a selection of theoretical advances in the unit root literature, starting from the second half of the 1980s and finishing with several contemporary developments. The presentation emphasizes both the relevance of the Functional Central Limit Theorem to the discussion as well as the econometric considerations behind novel approaches. Since the time series literature can consider the case of multiple structural breaks, attention is focused here 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 the disturbance sequences involved, a building block for this literature. Section 3 details a general version of the Functional Central Limit Theorem that covers a wide range 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 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 is intended to be employed in later developments. Section 7 presents three econometric applications of the above theories in the context of unit root testing when structural change is present. Section 7.1 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 7.2 covers the critique made by Zivot and Andrews (1992) to this exogeneity assumption and the new test

proposed by them, 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 7.3 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 Elliot *et al.* (1996) for linear processes. Section 8 shows some empirical applications. Section 9 concludes with a retrospective overview of the developments in statistical inference with integrated series and the role played by the theory of diffusion processes.

2. ASYMPTOTIC THEORY: THE STRUCTURE OF WEAKLY DEPENDENT AND HETEROGENEOUSLY DISTRIBUTED DISTURBANCES

Most of the econometric theory to be reviewed by us in this paper is related with extensions of the following autoregressive model: $y_t = \alpha y_{t-1} + u_t$, $t = 1, 2, \ldots$. The main objective 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 imposed 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 applicability of several procedures would be 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[6]. A way to control the extent to which this dependence occurs, that permits convergence results to be derived, is to define a measure of dependence among the random variables contained in a sequence. For this measure to be well-defined it needs to be associated with a specific probabilistic structure. The conditions that bind the extent of dependence are called mixing conditions. Results expressed below follow both White (1984) and Herrndorf (1984).

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 $P : \mathcal{F} \rightarrow [0,1]$ 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, $u_t: \Omega \rightarrow \mathbb{R}$ 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 we will need to assign probabilities to events involving the 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 the 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 measurable.

---

[6]   Typically, textbook treatment of time series analysis assumes a sequence of independent and identically distributed (i.i.d.) disturbances. Within our exposition, the independence assumption is relaxed by dealing with potentially dependent disturbances. Additionally, homogeneity of distributions corresponding to disturbances is relaxed by considering a wider family of distributions.

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*

1. *all the sets of the form $\times_{i=1}^{n-1} \mathbb{R} \times_{i=n}^{n+m} B_i \times_{i=n+m+1}^{\infty} \mathbb{R}$ with $B_i \in \mathcal{B}$,*
2. *the complement A^c of each set A in $\mathcal{B}_n^{n+m}$, and*
3. *the union $\cup_{i=1}^{\infty} A_i$ of each sequence $\{A_i\}$ contained in $\mathcal{B}_n^{n+m}$.*

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 make explicit the fact that, although two arbitrary sets of random variables can exhibit dependence; this vanishes as time separation increases[7].

Figure 2. Dependence and mixing coefficients
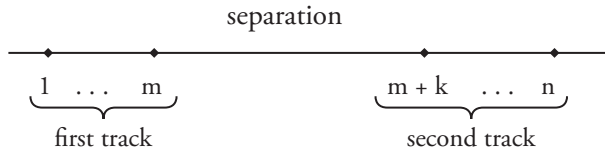
In order to illustrate the former idea, consider the track composed by the first n elements of $\{u_t\}_{t=1}^{\infty}$ and denote it by $\{u_t\}_{t=1}^{n}$. Within this track, two non-overlapping subtracks can be identified: a first one starting at u_1 and a second one ending at u_n. Let $k \geq 1$ 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 it allows for several cases. Indeed, the following definition of mixing coefficients employs the previous observations in order to quantify the dependence between random variables separated by k periods at least, given the first n elements of a sequence.

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

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

---

7   Note that the idea of progressive lack of dependence includes ergodicity and asymptotic independence.

Intuitively, given the first $n \geq 1$ elements of $\{u_t\}_{t=1}^{\infty}$, $\alpha_n(k)$ 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 independent 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 $g \in \mathcal{G}$, the condition $P(g \cap h) = P(g)P(h)$ must hold true or, equivalently, it must be true that $\alpha_n(k) = 0$.

Since mixing coefficients only takes into account a finite number of disturbances (i.e. the first n random variables), this notion is extended to consider the highest magnitude of 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*

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

Therefore, $\alpha(k)$ provides a measure of dependence. If $\alpha(k) = 0$ for some k, events separated by k periods are independent. Also, if $\alpha(k) \to \infty$ as $k \to \infty$, the sequence $\{u_t\}_{t=1}^{\infty}$ is said to be strong mixing, and the notion of asymptotic independence must be 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 $\alpha(k) = O(k^{-v})$ for some $v > 0$.

3. THE FUNCTIONAL CENTRAL LIMIT THEOREM

3.1. The Skorohod topology

The logic behind the Functional Central Limit Theorem relies on the convergence of a sequence of standardized partial sums of disturbances u_t. The limit for this new sequence is W, a standard Wiener process. Correspondingly, the elements of this sequence of partial sums are contained on $D = D[0,1]$ which is the space of right-continuous functions whose left limit exists everywhere on the unit interval, also referred to as càdlàg[8] functions.

The concept of convergence mentioned above must be understood as the weak convergence of a sequence of random functions. As will be shown, in order to guarantee the convergence results, it is sufficient to endow D with a metric d in such a way 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. The concepts and results discussed here are strongly based upon Billingsley (1968), although our presentation follows Davidson (1994). The following definition characterizes the properties of the functions to be considered hereafter.

---

[8]   In French: "continue à droite, limitée à gauche".

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

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

Therefore, only first class discontinuities are admitted. A first candidate for a suitable metric for D is the uniform metric d_U, defined as

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

The above metric states that two functions are arbitrarily close if the maximum difference between ordinates corresponding to the same abscissa is small. In that case, the metric space (C, d_U)[9] is complete but, since $C \subset D$, completeness does not necessarily generalize to (C, d_U). In fact, it is not difficult to show that the limit of sequences of càdlàg functions does not necessarily lie on D under d_U. Thus, (C, 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*[10] *λ: $[0,1] \rightarrow [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$. The Skorohod metric is defined as*

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

This metric is defined in order to overcome the following key limitation in the (C, d_U) space: given two càdlàg functions $x, y \in D$, under the uniform metric x and y are arbitrarily 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.

The 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) that is 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[11] metric (the Billingsley metric) to d_S is introduced in such a way 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.

---

[9] $C = C([0,1])$ is the space of all continuous functions on the unit interval.
[10]   A homeomorphism (or bicontinuous function) is a continuous function that has a continuous inverse function.
[11]   Consider a set X and two metrics g_1 and g_2 defined on X. g_1 and g_2 are said to be *topologically equivalent* or *equivalent* if they generate the same metric topology on X.

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

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

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 result (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, the 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 $\{S_T\}_{t=1}^{\infty}$ where $S_T = \sum_{t=1}^{T} u_t$. First, the disturbances are required to have zero mean and finite variance

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

Second, the 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 the random functions $W_T \colon \Omega \to \mathbb{R}$ by

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

where $\lfloor \cdot \rfloor$ denotes the integer part of its argument. Each W_T is a measurable map from $(\Omega, \mathcal{F})$ into $(D, \mathcal{B})$. The 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

$$\|u\|_\beta = (E|u|^\beta)^{1/\beta} \text{ for } \beta \in [1, \infty)$$

$$\|u\|_\beta = \text{ess sup} |u| \text{ for } \beta \in \infty.$$

As will be shown in the next section, the following version of the Functional Central Limit Theorem is the starting point for 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),*

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

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 the econometric developments to be considered in the following lines. The first of these works is due to Phillips (1987a), who developed a rather 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 general framework for testing unit roots was due to Phillips (1987a). This study established weak dependence conditions 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 its application in subsequent studies. One 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 standardized partial sums defined by

$$W_T(r) = \begin{cases} \dfrac{1}{\sqrt{T}\sigma} S_{\lfloor Tr \rfloor} & \text{if } \dfrac{j-1}{T} \leq r < \dfrac{j}{T}, j = 1,\ldots,T, \\[3mm] \dfrac{1}{\sqrt{T}\sigma} S_T & \text{if } r = 1, \end{cases} \tag{9}$$

where σ is a positive constant. Note that the sample paths $W_T(r)$ lie in D. It is worth emphasizing 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. In addition, by adopting assumptions about the disturbances that are less restrictive than i.i.d., Phillips (1987a) showed 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, 1987a, p. 280) *The disturbance sequence $\{u_t\}_{t=1}^{\infty}$ satisfies the following conditions:*

1. $E(u_t) = 0$ *for $t = 1,2,\ldots$,*
2. $\sup_t E|u_t|^{\beta} < \infty$ *for $\beta > 2$,*
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*

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

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 the disturbances contained in $\{u_t\}_{t=1}^{\infty}$, and the elements covered in previous sections allow it to be asserted that although dependence can exist between any pair of disturbances, it vanishes as time separation increases. Hence, any two random disturbances sufficiently distant along time are almost independent. Finally, the summability condition (10) is satisfied as long as the mixing decay rate is $\alpha(k) = O(k^{-v})$ for some $v > 0$ in such a way that $-v(1 - 2/\beta) < 1$ or, equivalently $v > \beta/(\beta - 2)$.

It is interesting to notice that as T increases the constant sections conforming $W_T(r) \in D$ shrink, and their discontinuities become less perceptible (see Figure 3), reflecting how this sequence of random functions in D converges 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. The second result is widely known as the Continuous Mapping Theorem and states conditions under which convergence to a Wiener process can be preserved (almost everywhere) along continuous transformations.

Figure 3. Convergence of standardized sums

Convergence of $W_T(r)$
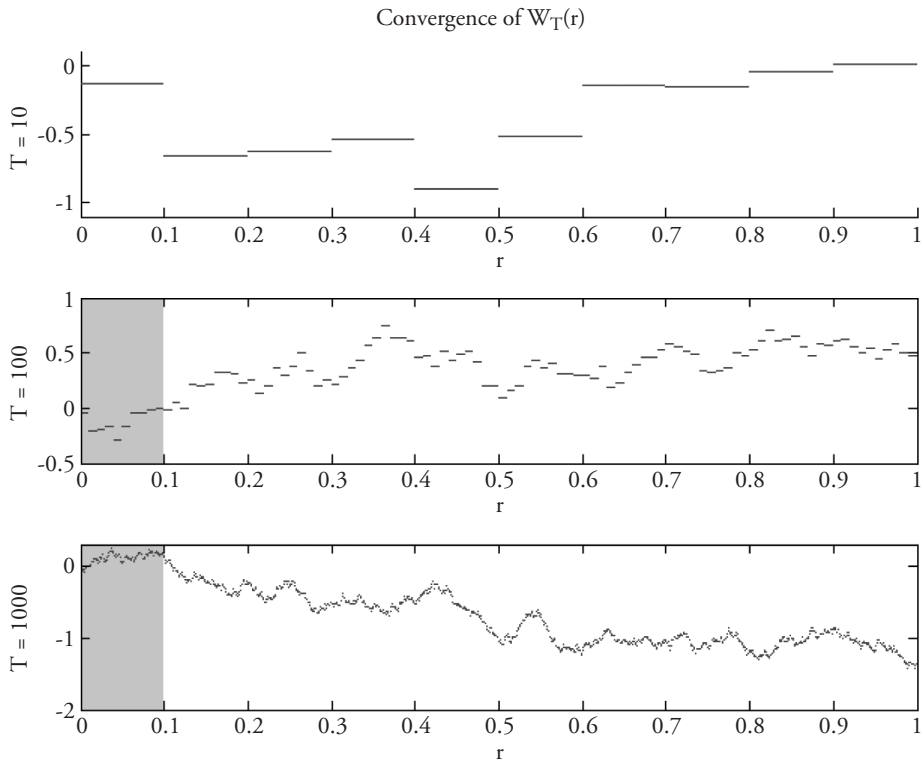
Source: authors' calculations.

Lemma 4 (Phillips, 1987a, 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, 1987a, p. 281) *If $W_T \Rightarrow W(r)$ as $T \to \infty$ and h is a continuous functional on D almost everywhere (a.e.) then $h(W_T) \Rightarrow h(W)$ as $T \to \infty$.*

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

4.2. An 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, 1987a, 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. $T(\hat{\alpha} - 1) \Rightarrow \dfrac{1}{2}(W(1)^2 - \sigma_u^2 / \sigma^2) \Big/ \int_0^1 W^2(r) dr,$

4. $\hat{\alpha} \overset{p}{\to} 1,$

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

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)[12], 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 mentioning 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, which implies that usual distributions (normal, chi-squared, *t* and *F*) employed in the stationary case are not relevant for the subsequent analysis.

Following the previous considerations, Phillips (1987a) proposed (after developing consistent estimators for the parameters σ_u^2 and σ^2) two new test statistics for the unit root hypothesis often referred to as the *Z* tests. Although it is important to remember that both (7) and (8) correspond only to the case of a unit root without drift or deterministic trend, the importance of this study lies on its providing us with a general theory on test statistics for the unit root hypothesis. Distributions considered 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 that we develop in the following sections.

---

[12] See equation (3).

5. ASYMPTOTICS FOR NEAR-INTEGRATED PROCESSES (PHILLIPS, 1987B)

For later discussion of the asymptotic power of unit root tests against alternative hypotheses that consider autoregressive coefficients close but not equal to one, it will be useful to consider generalizations of integrated processes often referred to as near-integrated processes and studied in detail by Phillips (1987b). The focus is on the time series $\{y_t\}_{t=1}^{\infty}$ which is assumed to be generated according to the following model

$$y_t = \alpha y_{t-1} + u_t, \, t = 1,2,\ldots \tag{11}$$

$$\alpha = e^{c/T}, \, -\infty < c < \infty. \tag{12}$$

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 true when $c = 0$:

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

Under (13), $\{y_t\}_{t=0}^{\infty}$ is an integrated process of order 1 or an $I(1)$ 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 $c = 0$ (i.e. $\alpha = 1$) in (12), $\{y_t\}_{t=1}^{\infty}$ is also called integrated.*

The main objective of the present section is to present an asymptotic theory for these types 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, some general assumptions concerning the disturbance sequence $\{u_t\}_{t=0}^{\infty}$ are necessary. For this reason, the following mixing conditions about the behaviour of the disturbances $\{u_t\}_{t=0}^{\infty}$ (hereby now familiar) are adopted and summarized in the next statement.

Assumption 2 (Phillips, 1987b, p. 537) *The 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*

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

Notice that Assumptions 1 and 2 are quite similar and the only difference relies on the existence of $\varepsilon > 0$ in such a way that the existence of $\sup_t E|u_t|^{\beta+\varepsilon}$ holds true.

On the other hand, it will be useful to represent the stochastic limit theory by means of an extensive use of certain diffusion process, which can be interpreted as the continuous time version of an AR(1) process.

Definition 8 (Ornstein-Uhlenbeck process) *An Ornstein-Uhlenbeck process is a functional W_c 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 following differential equation in term of a continuous-time stochastic process $X(t)$

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

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 THEORIES FOR ECONOMETRICIANS

If the value of the parameter c was fixed, it would be natural to expect, based on (12), that $\alpha \to 1$ as $T \to \infty$. However, within this framework the speed of convergence of α towards 1 is controlled at $O(T^{-1})$. Equivalently, such a speed is not so fast that the effect of c on the main results in Section 4 does not vanish[13]. This observation leads to the following derivation of 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. $T^{-1} \sum_{t=1}^{T} y_{t-1}u_t \Rightarrow \sigma^2 \int_0^1 W_c(r)dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2),$ *with σ_u is defined by* $\sigma_u = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2).$

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

---

[13] Since $c = T \ln(\alpha)$.

Theorem 8 (Phillips, 1987b, p. 540) *If $\{y_t\}_{t=0}^{\infty}$ is a near-integrated time series generated by (11) and (12) then, as $T \to \infty$:*

1. $T(\hat{\alpha} - \alpha) \Rightarrow \left\{ \int_0^1 W_c(r)dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2) \right\} \Big/ \int_0^1 \{W_c(r)\}^2 \, dr,$

2. $\hat{\alpha} \overset{p}{\to} 1$, $s^2 \overset{p}{\to} \sigma_u^2$, with $s^2 = T^{-1} \sum_{t=1}^{T}(y_t - \hat{\alpha}y_{t-1})^2,$

3. $t_\alpha \Rightarrow (\sigma/\sigma_u) \left\{ \int_0^1 W_c(r)dW(r) + \frac{1}{2}(1 - \sigma_u^2/\sigma^2) \right\} \Big/ \left[\int_0^1 \{W_c(r)\}^2 \, dr \right]^{1/2}.$

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

Up to this point, the theory presented can be employed 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 adopted in unit root testing. A brief illustration of this procedure can be found, for example, in Phillips (1988).

6. LINEAR PROCESSES AND MODIFIED UNIT ROOT TESTS

6.1. MOTIVATION

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

Under the class of linear models, 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 A (see Appendix B), strongly uniformly integrable (see Appendix C) with dominating random variables $\{Z_t\}_{t=0}^{\infty}$ in such a way that $E(Z_t^{2+\eta}) < \infty$ for some $\eta > 0$, as well as $T^{-1} \sum_{t=1}^{T} E(\varepsilon_t^2 | \mathcal{F}_{t-1}) \overset{a.s.}{\to} \sigma_\varepsilon^0 > 0$, where $\mathcal{F}_t$ denotes the σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$. Given the latter notation it is now possible to establish the following:

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

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

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 Functional Central Limit Theorem is less general than the versions previously presented, it has been used frequently in subsequent work, especially along the developments by 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[14] of $T^{-1/2}(O_p(T^{-1/2}))$. Since the remainder 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 generating process for $\{y_t\}_{t=1}^\infty$:

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

for $t = 1, \ldots, T$. 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 the 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 term 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; then, for $r \in [0,1]$ we define the following functionals:

$$S_T(r) = \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor rT \rfloor} u_i, \ \ and$$

$$D_T(r, \beta) = \delta_{\lfloor rT \rfloor}(\beta),$$

which are both càdlàg versions of the components of the discrete time process. Our goal is to apply the Functional Central Limit Theorem. Such functionals are assumed to satisfy the following:

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

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

---

[14] 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 $P(|y_t|/a_t > M) < \varepsilon$.

2. $\sqrt{T}\{D_t(.,\hat{\beta}) - D_t(.,\beta)\} \Rightarrow \sigma D$, where $D \in D[0,1]$ *has a distribution that does not depend on β or on σ, 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

$$u_t = C(L)\varepsilon_t,$$

$$\sum_{j=1}^{\infty} j|c_j| < \infty,$$

$$C(1) \neq 0$$

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, we consider the following three cases:

1. No deterministic trend: $\delta_t(\beta) = 0$. In this case there is no need for detrending. For the sake of 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 continuous time analogous to lie in the interval $[0,1]$.

The three former cases are enough for subsequent analysis. Since the 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 $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$.*

The elements for the development of the new class of tests are based on both the Functional Central Limit Theorem and the Continuous Mapping Theorem. For each of the cases considered, we 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) *Suppose 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*

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

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 class of tests proposed. Also, it follows from the Continuous Mapping Theorem that if (19) holds and g is a continuous function $g: D[0,1] \rightarrow \mathbb{R}$, then

$$g(S_T^d) \Rightarrow g(S^d). \tag{20}$$

Let $M^d = \{m: D[0,1] \rightarrow \mathbb{R}\}$ be the collection of functionals that satisfy the following conditions:

1. m is continuous,
2. there exists c_v, $|c_v| < +\infty$, in such a way that $P[m(S^d) \leq c_v] = v$ for all $v \in (0,1)$ and
3. $m(0) < c_v$ for all $v \in (0,1)$.

The class of tests M^d referred only to continuous functionals of S_T^d and it grouped 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 test of level v of the form:

$$\text{reject } H_0: y_t \sim I(1) \text{ if } m(S_T^d) \leq c_v.$$

This approach, as Stock (1999) asserted, suggests working backwards from the desired asymptotic representation to the actual test statistic. The fact that the form of the functional m does not depend on the type of detrending emphasizes that the steps of eliminating the deterministic components and testing for a unit root are distinct;

detrending a series when it is not required does not affect the size of the tests (although it can affect power) 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 thus is a desirable property. Once the size is guaranteed to be fixed, power increasing procedures can be performed. The next two subsections illustrate the main idea behind the following: if a given V_T test statistic has a limiting distribution characterized as the functional m of certain diffusion process S^d, that is

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

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

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

In such a way that V_T 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 section 7.3 is that by Sargan and Bhargava (1983) for the model

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

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 proposed 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) considered the following extension:

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

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 was proposed:

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

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

$$\tilde{\beta}_0 = \bar{y} - \frac{1}{2} \frac{T+1}{(T-1)} (y_T - y_1),$$

$$\tilde{\beta}_1 = \frac{T}{(T-1)} (y_T - y_1).$$

For both tests SB_{μ} and SB_{τ}, Stock (1999) derived their limiting distribution

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

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

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

Now, note that the functional

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

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

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

This latter statistic will be referred to 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) proposed the following test statistic:

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

where

$$\hat{\alpha} = \sum_{t=2}^T y_t^{\mu} y_{t-1}^{\mu} \Big/ \sum_{t=2}^T (y_{t-1}^{\mu})^2, \qquad (25)$$

$$\hat{u}_t = y_t^{\mu} - \hat{\alpha} y_{t-1}^{\mu}, \qquad (26)$$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=2}^T \hat{u}_t^2 + 2 \sum_{j=1}^l T^{-1} \sum_{t=j+2}^l \hat{u}_t \hat{u}_{t-j} \text{ and} \qquad (27)$$

$$\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\alpha} y_{t-1})^2. \qquad (28)$$

Since $\sum_{t=1}^{T} y_{t-1}\Delta y_t = (1/2)(y_T^2 - \Delta y_t^2)$, the test Z_α in (24) can be rewritten 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 $\hat\alpha - 1 \overset{p}{\to} 0$, its 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 Rodriguez (2003) to be covered in section 7.3, the modified Z_α test will be referred to as the MZ_α test.

7. ECONOMETRIC APPLICATIONS

7.1. Exogenous Structural Break (Perron, 1989)

The previous sections have established the foundations for studying the inference in the case of nonstationary time series. Henceforth we 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 a hypothesis and the perception that macroeconomic variables were characterized by stochastic trends became popular. One of the most influential studies in this empirical literature was conducted by Nelson and Plosser (1982), who studied fourteen macroeconomic variables for the US economy. Under the stochastic trend perspective, a series that exhibits an upward sloping behavior and an abrupt reduction (see Figure 4a) is interpreted as the consequence of an atypical realization of u_t (situated on 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 the 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

(a) Stochastic trend   (b) Deterministic trend
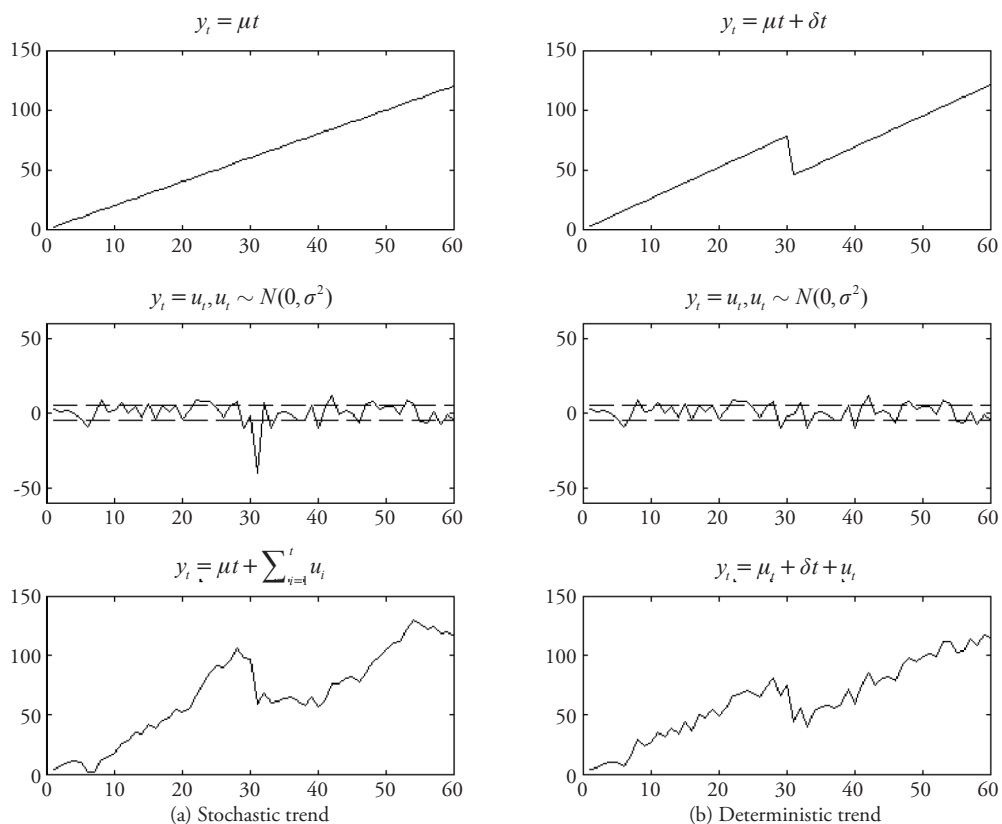
Source: authors' calculations.

Indeed, Perron (1989) switched emphasis to this latter interpretation and asserted 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" (Perron, 1989, p. 1962-63).

Perron (1989) considered atypical events as interventions on the deterministic component of the model, and this allowed him to distinguish what can be explained by the disturbance term from what cannot. Additionally, the date of this intervention is assumed to be known by the researcher. Because there are two competing interpretations mentioned above for time series with an abrupt shift, the models considered by Perron (1989) are summarized in Table 1.

Table 1. Null and alternative hypotheses considered by Perron (1989)

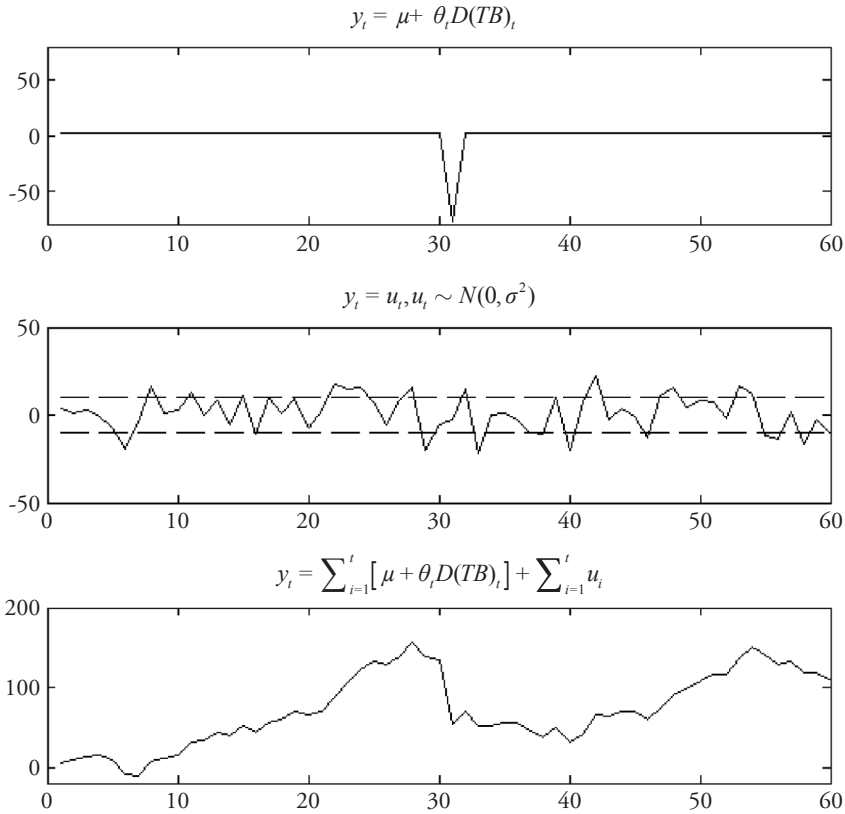| Null hypothesis	Alternative hypothesis
Model A	**Model A**
$y_t = \mu + y_{t-1} + \theta D(T_B) + u_t$	$y_t = \mu_1 + (\mu_2 - \mu_1)DU_T + \delta t + u_t$
Model B	**Model B**
$y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + u_t$	$y_t = \mu + \delta_1 t + (\delta_2 - \delta_1)DT_t^* + u_t$
Model C	**Model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

In Table 1, θ, μ, μ_1, μ_2, δ, δ_1 and δ_2 are parameters, and $A(L)u_t = B(L)e_t$ where $e_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. $A(L)$ and $B(L)$ are pth and qth order polynomials. That is, $\{u_t\}$ is an $ARMA(p,q)$ process with p and q possibly unknown. This assumption allows $\{y_t\}$ to represent the 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 immediately 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 accompanied by a shift in slope.

In this way, Perron (1989) introduced 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 (1989) examined the performance of this class of tests under the alternative hypothesis. Specifically, Monte Carlo simulations revealed that when the data generating process was described as being by model A under the alternative, DF tests tended to detect a spurious unit root that did not vanish, even asymptotically. Therefore, a power loss was expected. Perron (1989, Theorem 1) derived this property also at the theoretical level, and he adopted Phillips's (1987a) assumptions concerning the innovation sequence $\{u_t\}$ in order for his results to be as general as possible.

Figure 5. The "Crash" model

$$y_t = \mu + \theta_t D(TB)_t$$

$$y_t = u_t, u_t \sim N(0, \sigma^2)$$

$$y_t = \sum_{i=1}^{t} \left[\mu + \theta_t D(TB)_t \right] + \sum_{i=1}^{t} u_i$$

Source: authors' calculations.

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 by 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 $\lambda \in (0,1)$ breakfraction. The next subsection presents the strategy adopted and the main results.

7.1.1. Structure of the Model and Main Findings

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

The two test statistics of interest were generalizations of the Z-tests proposed by Phillips (1987a). The intuition behind it was simple: since the researcher was assumed to know the λ breakfraction, this effect must be removed from the data. Thus, let $\{\tilde{y}_t^i\}$ denote the 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 $\tilde{\alpha}^i$ should lie sufficiently close to one or, equivalently, the bias $\tilde{\alpha}^i - 1$ must lie close to zero. Formally, the next theorem presents the asymptotic distribution of both the standardized bias $T(\tilde{\alpha}^i - 1)$ and the 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) $T(\tilde{\alpha}^i - 1) \Rightarrow H_i/K_i;$ a) $t_{\tilde{\alpha}^i} (\sigma/\sigma_u)H_i/(g_iK_i)^{1/2};$

where

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

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

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

with

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

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

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

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

Proof. See Perron (1989), Theorem 2 p. 1393, and Appendix D for an extended definition of coefficients. ∎

The reader must take into account that the previous limiting distributions depend, besides λ, on the 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 constitutes 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) mentioned that when the sequence $\{u_t\}$ was independent and identically distributed, $\sigma^2 = \sigma_u^2$, and in that case the limiting distributions were invariant with respect to the nuisance parameters, except λ.

    With these theoretical results and the tabulation of critical values through Monte Carlo simulations, evidence was found against the unit root hypothesis for the series studied by Nelson and Plosser (1982). Thus, the relevance of the results by Perron (1989) lies in the analysis of the performance of ADF tests when misspecification was present. As will be shown below, misspecification becomes crucial for the identification of desirable properties of new tests to be proposed. On the other hand, these results generalized the tests by Phillips (1987a) and the inference procedure assumed knowledge of both the existence of structural break and the breakfraction value. Subsequent studies progressively avoided these two assumptions and included desirable properties.

7.2. Endogenous Structural Break (Zivot and Andrews, 1992)

7.2.1. 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. From Perron's perspective, applied researchers will 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, the breakfraction is no longer exogenous since the previous selections are based on an *a priori* inspection of data, which incorporates an implicit selection rule behind it. This fact is going to be exploited formally and will lead to the use of the Functional Central Limit Theorem under somewhat different conditions.

Figure 6. Sample paths under different breakfractions

(a) $\lambda \approx 0.25$ (b) $\lambda \approx 0.75$
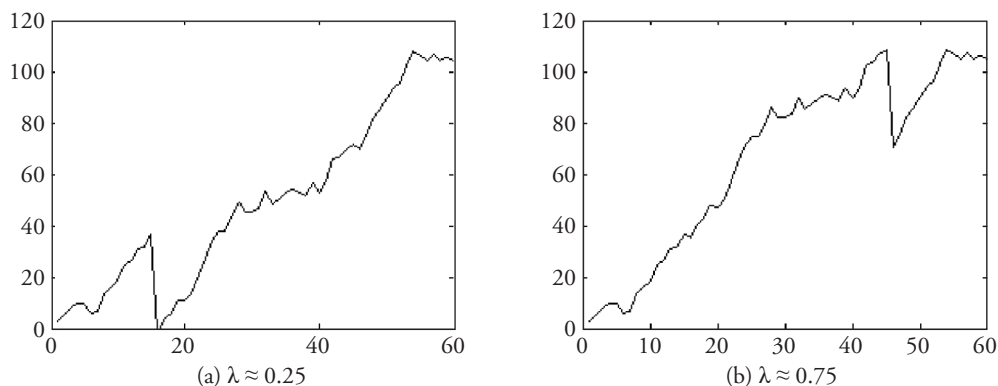
Source: authors' calculations.

7.2.2. The Approach

The first of the abovementioned two assumptions 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 the alternative, however, we assume that we do not know exactly when the breakpoint occurs" (Zivot and Andrews, 1992, p. 252).

As noticed, attention is turned back to the competing approaches shown in Figure 4 and formalized in Table 2. Additionally, while the tests developed by Perron (1989) are conditional on a given breakfraction $\lambda \in (0,1)$, Zivot and Andrews (1992) attempted to transform these tests into unconditional ones by designing an estimation method for λ.

Table 2. Null and alternative hypotheses considered by Zivot and Andrews (1992).

Null hypothesis	Alternative hypothesis
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	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	

It is important to mention that conventional wisdom in applied econometrics considers the 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. On the other hand, in line with the structural change literature under an unknown changepoint, Zivot and Andrews (1992) suggested to choose the λ breakfraction that gave 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 $\hat{\lambda}_{\text{inf}}^i$ denote such a value for model i, then $t_{\hat{\alpha}}[\lambda_{\text{inf}}^i] \equiv \inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda)$ 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:

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

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

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

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, while $DT_t^* = t - T\lambda$ if $t > T\lambda$ and 0 otherwise. The number of extra lags k is included here to potentially take into account the correlation between disturbances and $\hat{\lambda}$ denotes the estimated value of λ. In order to make the results as simple as possible, the authors considered first the case where $k = 0$ (no correlation among disturbances). In contrast with the work of Perron (1989), when correlation between disturbances is present, it is restricted to being of the ARMA structure. It is worth mentioning that this structure is a particular case of mixing processes and that this implies that the Functional Central Limit Theorem can still be applied.

For unit root 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, the critical values in Perron (1989) cannot be employed for unit root testing. Consider an estimated λ with minimum t statistic. Then, the decision rule can be summarized as

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

where $\kappa_{\inf, v}^i$ denotes the asymptotic critical value of $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i} (\lambda)$ for a size equal to v. By definition, critical values are larger than (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 $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}$ ($i = A, B, C$) are needed.

7.2.3. Asymptotic Distribution Theory

In order to obtain the limiting distribution for their proposed statistic, Zivot and Andrews (1992) made use of the framework suggested by Ouliaris, Park and Phillips (1989), which allowed for a compact form for their results. It is worth mentioning that this framework was also used by Perron (1989) when the objective was to develop a generalization for his main theorem on the case of disturbances that express autocorrelation. Attention is here focused on i.i.d. disturbances. The following two definitions are necessary for the understanding 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 presented in the next theorem[15].

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, and 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,*

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

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 mentioning that when a correlation of the ARMA type is allowed, the previous result can be extended in order to obtain an autoregressive estimate of 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 in Phillips (1987a). That is, the probability of outliers is controlled and such an assumption was also adopted in subsequent work.

7.3. Efficient Unit Root Testing under Structural Break
   (Perron and Rodríguez, 2003)

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

1. Deterministic trend and size. Most of the earlier unit root tests under less restrictive assumptions were extensions of augmented Dickey-Fuller tests and therefore their asymptotic distributions depended on whether or not a deterministic component had been added to the regression equation. According to Stock (1999) this problem could be solved by first detrending the series and next performing (robust) modified unit root tests in such a way that size is not affected.

---

[15]   Although independent, the derivation presented here occurs in a similar fashion to those reported by Banerjee *et al.* (1992).

2. Structural break and power. Perron (1989) illustrated how deterministic trends that contained a break could induce spurious unit roots in Dickey-Fuller tests. Following Stock (1999), a trend with structural break could 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 by following the near-integrated time series approach proposed by Phillips (1987b) and developed by Elliot *et al.* (1996) for the case of no structural break. Therefore, an extension is called for.

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

7.3.1. Data Generating Process

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

$$y_t = \psi'z_t + u_t, \tag{33}$$

$$u_t = \alpha u_{t-1} + v_t, \text{ and} \tag{34}$$

for $t = 1, \ldots, T$. Perron and Rodríguez (2003) considered two models for testing the presence of structural change, summarized in Table 3. 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), adopted the following specification:

Table 3. Deterministic components considered by Perron and Rodríguez (2003)

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

Assumption 6 (Perron and Rodríguez, 2003 p. 3) *The following conditions must 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 an 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,* $T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \sigma W(r)$, *where* $\Rightarrow$ *denotes weak convergence in distribution and* $W(r)$ *is the standard Wiener process defined on* $C[0,1]$ *the space of continuous functions on the interval* $[0,1]$.

7.3.2. *GLS* Detrending and *M* tests

First, denote the transformed data by

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

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

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

The data is transformed in order to make the results dependent on the parameter $\bar{\alpha}$. The goal here is to derive an optimal unit root test 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 be explicitly obtained. The *M* tests, studied in section 6, are defined by

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

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

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

with local detrended data defined by $\tilde{y}_t = y_t - \psi'z_t$ 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

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

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

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

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

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

7.3.3. Asymptotic Distributions

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

Theorem 13 (Perron and Rodriguez, 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 $\overline{\alpha} = 1 + \overline{c}/T$, 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*

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

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

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

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

where

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

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

and $V_{c\overline{c}}^{(1)}(r,\lambda)=W_c(r)-rb_3$, $V_{c\overline{c}}^{(2)}(r,\lambda)=b_4(r-\lambda)[W_c(r)-rb_3-(1/2)(r-\lambda)b_4]$ *with* $w_c(r)$ *the Ornstein-Uhlenbeck process as 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

$$b_1 \;=\; (1-\overline{c})W_c(1)+c^2\int_0^1 rW_c(r)dr,$$

$$b_1 \;=\; (1-\overline{c}+\lambda\overline{c})W_c(1)+\overline{c}^2\int_\lambda^1 W_c(r)(r-\lambda)dr-W_c(\lambda),$$

$$b_3 \;=\; \lambda_1 b_1 + \lambda_2 b_2,$$

$$b_4 \;=\; \lambda_2 b_1 + \lambda_3 b_2,$$

$$\lambda_1 \;=\; d/\Theta,$$

$$\lambda_2 \;=\; -m/\Theta,$$

$$\lambda_3 \;=\; a/\Theta,$$

$$b \;=\; 1-\lambda-c+2c\lambda-c\lambda^2-c^2\lambda+c^2\lambda^2+(c^2/3)(1-\lambda^3),$$

$$m \;=\; 1-\lambda-c+c\lambda+(c^2/2)\lambda^3+(c^2/3)(1-\lambda^3),$$

$$a \;=\; 1-c+c^2/3 \text{ and}$$

$$\Theta \;=\; ad-m^2.$$

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

7.3.4. A Feasible Point Optimal Test with Known Breakdate

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 the 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 Elliot *et al.* (1996), if the data distribution were known then the Neyman-Pearson Lemma would suggest the optimal point alternative against any other point alternative hypothesis, and in such circumstances a power envelope could be derived[16].

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

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

where $S(\overline{\alpha}, \lambda)$ and $S(1, \lambda)$ are the sums of squares from GLS, detrending under both $\alpha = \overline{\alpha}$ and $\alpha = 1$, respectively. Under the assumption of a known λ breakfraction, different values for $\overline{\alpha}$ 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, Elliot *et al.* (1996) proposed a feasible optimal point test P_T^{GLS} defined by

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

where its distribution is derived in the following theorem:

Theorem 14 (Perron and Rodriguez, 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 $\overline{\alpha} = 1 + \overline{c}/T$. 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*

$$P_T^{GLS}(c, \overline{c}, \lambda) \Rightarrow M(c, 0, \lambda) - M(c, \overline{c}, \lambda)$$

$$- 2\overline{c} \int_0^1 W_c(r) dW(r) + (\overline{c}^2 - 2\overline{c}c) \int_0^1 W_c(r)^2 dr - \overline{c} \equiv H^{P_T^{GLS}}(c, \overline{c}, \lambda)$$

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

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

---

[16]   As Elliot *et al.* (1996) mentioned, the Gaussian power envelope is upper bound to the asymptotic power function for tests of the unit root hypothesis when the data are generated by $y_t = d_t + u_t$ and $u_t = \alpha u_{t-1} + v_t$ but under ideal conditions. Namely, the process $\{v_t\}$ 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.

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

$$\begin{bmatrix} \overline{c}^2/3 - \overline{c} + 1 & (1-\lambda)(1-\overline{c}) + \overline{c}^2(2+\lambda^3 - 3)/6 \\ & \overline{c}^2(1-\lambda^3)/3 - \overline{c}(1-\lambda^2)(1+\lambda\overline{c}) + (1-\lambda)(1+\lambda\overline{c})^2 \end{bmatrix}.$$

Proof. See Perron and Rodríguez (2003), Theorem 2, p. 34. ■

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 a certain critical value. Since the distribution for the tests was derived both under the null and the alternative hypothesis, the (asymptotic) power function can be explicit by means of the probability of rejecting the null under the alternative. Such a function is given by

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

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

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

and v is the size of the test. Therefore, different values of λ generate different power functions.

7.3.5. A Feasible Point Optimal Test with Unknown Breakdate and the Power Envelope

The previous subsections refer 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

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

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 bind critical values, a situation that arises in the context of tests for structural change. Using Theorem 14, the following result is obtained:

$$P_{T,*}^{GLS}(c,\overline{c}) \Rightarrow \sup_{\lambda \in [\varepsilon, 1-\varepsilon]} M(c,0,\lambda) - \sup_{\lambda \in [\varepsilon, 1-\varepsilon]} M(c,\overline{c},\lambda)$$

$$-2\overline{c} \int_0^1 W_c(r) dW(r) + (\overline{c}^2 - 2\overline{c}c) \int_0^1 W_c(r)^2 dr - \overline{c} \equiv H^{P_{T,*}^{GLS}}(c,\overline{c}).$$

Accordingly, the asymptotic Gaussian power envelope is given by

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

where the critical value is $b^{P_{T,*}^{GLS}}(c)$ so that $P\left[H^{P_{T,*}^{GLS}}(0,c) < b^{P_{T,*}^{GLS}}(c)\right] = v$. It must be pointed out that Elliot *et al* (1996) recommended a value for $\bar{c}$ so that $\pi^*(c) = 0.5$. Using Monte Carlo simulations, Perron and Rodriguez (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 to which power functions lie close 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, however, several sets of simulations are performed (one for each value of λ in $[\varepsilon, 1 - \varepsilon]$).

Table 4. Empirical Application

Series	(1) ADF statistic	(2) Perron statistic		(3) Zivot-Andrews statistic		(4) Perron-Rodríguez statistic	
	C + T	C	BT	C	BT	BT	BT
1. GDP	-1.638	-0.119	-2.077	-4.447	-4.707**	-5.405 ***	-3.226
(1980:I-2011:II)		[1990:III]	[1990:III]	[1988:III]	[1988:III]	[1988:III]	[1990:II]
2. Absorption	-1.471	-0.502	-2.129	-4.048	-3.992	-3.164	-3.184
(1980:I-2011:II)		[1990:III]	[1990:III]	[1988:III]	[1988:III]	[1990:I]	[1990:I]
3. Consumption	-1.521	-1.638	-2.326	-5.368**	-5.347***	-3.258	-3.340
(1980:I-2011:II)		[1990:III]	[1990:III]	[1988:III]	[1988:III]	[2004:II]	[2004:II]
4. CPI	-0.955	-4.273**	-1.712	-4.386	-3.921	-2.910	-3.149
(1994:I-2011:II)		[2000:IV]	[2000:IV]	[2001:II]	[2007:IV]	[1994:I]	[1994:I]
5. Exchange Rate	-0.756	-4.039**	-4.738***	-5.687***	-4.967***	-2.060	-2.163
(1991:I-2011:II)		[2005:I]	[2005:I]	[1997:III]	[2006:II]	[2002:IV]	[2002:IV]
6. Unemployment	-9.615***	-10.089***	-10.225***	-10.684***	-11.245***	-3.163	-3.936 *
(Jan2001-Jun2011)		[Dec2000]	[Dec2005]	[Nov2005]	[Sep2005]	[Sep2005]	[Set2005]
7. Real Exchange Rate	-3.397	-3.805***	-3.391	-4.672*	-4.659**	-3.300	-3.274
(Jan1994-Jun2011)		[Sep1998]	[Sep1998]	[Oct2005]	[Oct2005]	[Feb2006]	[Feb2006]
8. Exports Price Index	-1.221	-2.043	-1.764	-3.013	-3.379	-2.979	-2.909
(Jan1991-Jun2011)		[Dec2008]	[Dec2008]	[Jun1997]	[Aug2008]	[Jan2004]	[Ene2004]
9. Imports Price Index	-1.194	-1.715	-1.713	-2.707	-4.390*	-3.851 **	-3.772 *
(Jan1991-Jun2011)		[Dec2008]	[Dec2008]	[Nov2006]	[Nov2000]	[May2003]	[May2003]

Notes: Variables that exhibited a seasonal pattern were adjusted with the programs TRAMO and SEATS in automatic mode. Excepting unemployment, all variables are expressed in logs; for augmented Dickey-Fuller statistics in column (1), (*), (**) and (***) indicate that the null hypothesis is rejected at the 10%,5% and 1% level of significance, respectively; for column (2), exogenous (fixed) break date is reported with in square brackets. C stands for "Crash" model where as BT stands for "Breaking Trend" model; for column(3), estimated break dates are reported with in square brackets. C stands for "Crash" model where as BT stands for "Breaking Trend" model; for column (4), estimated break dates are reported with in square brackets. The modified AIC is employed as in Perron and Rodriguez (2003).
Source: authors' calculations.

8. EMPIRICAL APPLICATIONS

A set of nine macroeconomic variables have been selected for the Peruvian economy: gross domestic product (GDP), absorption, consumption, consumer price index (CPI), nominal exchange rate (domestic currency by Dollar), unemployment rate, real exchange rate, export price index and import price index. Frequencies of data vary according to availability and the covered sample period appears below the name of each variable. Table 4 presents results of application of four different unit root statistics. The first is the standard augmented Dickey-Fuller (ADF) statistic without structural change. The second statistic is the ADF proposed by Perron (1989) with an exogenous breakpoint. The third one is the ADF proposed by Zivot and Andrews (1992) based on the infimum method of selecting the break point. In the two columns, two models are estimated: the crash and the breaking trend models denoted by C and BT, respectively. The remaining column is based on Perron and Rodríguez (2003) using the breaking trend model. Statistics in Column (4) are based on the infimum method to select the break point using the MZ_t^{GLS} and ADF^{GLS} statistics, respectively. The lag length has been selected using the MAIC proposed by Perron and Ng (2001).

The results indicate the different results that may be obtained from the application of the different unit root statistics with or without structural change. Therefore, the results also show the difficulties of obtaining a unique answer regarding the stationarity or nonstationarity of the variables. Overall, we obtain mixed evidence for or against the stationarity of the analyzed variables. Mostly, no rejection of the null hypothesis of a unit root is found for the consumption, the CPI or the export price index. On the other hand, mostly rejection of the null hypothesis is found for the unemployment rate, the import price index, and the nominal exchange rate. Mixed evidence is found for the GDP and the real exchange rate.

9. CONCLUSIONS

The application of different unit root statistics is by now a standard practice in empirical work. Even when it is a practical issue, these statistics have complex non-standard distributions depending on the functionals of certain stochastic processes, and their derivations represent a barrier even for many theoretical econometricians. These derivations are based on rigorous and fundamental statistical tools which are not (very) well known for standard econometricians. This paper aims to fix this gap by explaining in a simple way one of these fundamental tools: namely, the Functional Central Limit Theorem. Therefore, this paper analyzes the foundations and applicability of two versions of the Functional Central Limit Theorem within the framework of a unit root with a structural break. 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 separation of unit root test sizes from the specific form of the deterministic component, a problem not solved in earlier works. Tools developed also allow the analytical tractability of several problems within this literature: the presence of structural breaks and the low power against local alternatives. For the issue of structural breaks (Perron 1989), first detrending in the series has shown itself 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), while a power function can be derived and maximized. When the two issues are combined, the result is an efficient test (Perron & Rodriguez 2003).

APPENDIX

A. 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 polynomial. Then

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

where

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

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

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

B. 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, or m.d.s.

C. 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

$$P(|Z_t| \geq x) \; \leq \; c P(|Z| \geq x)$$

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

D. SOME THEOREMS OF UNIT ROOT WITH STRUCTURAL CHANGE

Functionals in Perron (1989), Theorem 2 p. 1393 are defined as follows:

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

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

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

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

$$D_5 = W(1)/2 - \int_0^1 (r)dr$$

$$D_6 = W(\lambda) - \lambda W(1);$$

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

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

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

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

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

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

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

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

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

where

$$\sigma^2 = \lim_{T\to\infty} E[T^{-1}S_T^2]; \; S_T = \sum_{t=1}^T u_t; \; \sigma_u^2 = \lim_{T\to\infty} E\left[T^{-1}\sum_{t=1}^T u_t^2\right]$$

and W is a standard Wiener process on C.

REFERENCES

Aquino, J. C. (2011). Functional Central Limit Theorems and Unit Root Testing. Unpublished M.Sc. Dissertation, Department of Sciences, Section of Mathematics, Pontificia Universidad Católica del Perú.

Banerjee, A., R. L. Lumsdaine & 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), 271-287.

Bhargava, A. (1986). On the Theory of Testing for Unit Roots in Observed Time Series. *Review of Economic Studies*, 53(3), 369-384.

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

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

Box, G. E. P. & 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. & T. Zastawniak (1999). *Basic Stochastic Processes*. Springer.

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

Dickey, D. A. & W. A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427-431.

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 & J. H. Stock. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, 64(4), 813-836.

Enders, W. (2004). *Applied Econometric Time Series*. John Wiley & Sons, Inc.

Glynn, J. & 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. & C. I. Plosser. (1982). Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *Journal of Monetary Economics*, 10(2), 139-162.

Ng, S. & 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 & P. C. B. Phillips. (1989). Testing for a Unit Root in the Presence of a Maintained Trend. In: B. Raj (ed.). *Advances in Econometrics and Modeling*, Kluwern Academic.

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

Perron, P. & 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.

Phillips, P. C. B. (1987b). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika*, 74(3), 535-547.

Phillips, P. C. B. (1988). Regression Theory for Near-Integrated Time Series. *Econometrica*, 56(5), 1021-1043.

Phillips, P. C. B. & P. Perron. (1988). Testing for a Unit Root in Time Series Regression. *Biometrika*, 75(2), 335-346.

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

Sargan, J. D. & A. Bhargava. (1983). Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk. *Econometrica*, 51(1), 153-174.

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

White, J. S. (1958). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *The Annals of Mathematical Statistics*, 29(4), 1188-1197.

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

Zivot, E. & 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), 251-270.