LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER OF SCIENCE IN

APPLIED ECONOMETRICS AND FORECASTING

MASTERS FINAL WORK

DISSERTATION

DOUBLE UNIT TESTS IN THE PRESENCE OF STRUCTURAL BREAKS

FRANCISCO ANTÓNIO TEIXEIRA MENDONÇA

ISEG OCTOBER - 2017

MASTER OF SCIENCE IN

APPLIED ECONOMETRICS AND FORECASTING

MASTERS FINAL WORK

DISSERTATION

DOUBLE UNIT TESTS IN THE PRESENCE OF STRUCTURAL BREAKS

FRANCISCO ANTÓNIO TEIXEIRA MENDONÇA

Supervisor:

Professor Nuno Sobreira

ISEG OCTOBER - 2017

DOUBLE UNIT TESTS IN THE PRESENCE OF STRUCTURAL BREAKS

By Francisco Mendonça

We develop new statistical procedures aiming at accessing the presence of exactly two unit roots in a time series, that may have a single shift in its trend function, at a known or unknown date. The test statistics have a non-standard distribution based on functions of Wiener processes.

À minha família

Acknowledgments

I am grateful to Professor Nuno Sobreira for continuously challenging me to go further, and for sharing his knowledge with me. Without him, this work would not be possible.

I also appreciate the valuable inputs from Professor João Nicolau, which helped to improve the quality of this dissertation.

Abstract

The work presented concerns the field of unit root testing, allowing for the possibility of changes in the trend function of a time series. Two tests were designed for the null hypothesis of exactly two-unit roots, one for the case of a known change date, and another for the more likely case of an unknown changepoint. Each test statistic follows a non-standard distribution, which are based on functions of Wiener processes. The percentiles for both distributions were obtained via Monte Carlo simulation. Both tests were applied to several economic variables, and the results suggest that the double unit root hypothesis is a suitable candidate to explain the persistence of the innovations guiding many of those variables.

Keywords: Double unit root test; Structural break; Wiener process

JEL: C12, C22.

Resumo

O trabalho que se apresenta, centra-se no campo dos testes de raiz unitária, permitindo a existência de alterações na função de tendência de uma série temporal. Foram construídos dois testes para a hipótese nula de exatamente duas raízes unitárias, um para o caso de data de quebra conhecida, e outro para o cenário mais provável de uma data de quebra desconhecida. Ambas as estatísticas de teste seguem uma distribuição não convencional, baseada em processos de Wiener. Os percentis destas distribuições foram obtidos via simulação de Monte Carlo. Ambos os testes foram aplicados a várias variáveis económicas e os resultados sugerem que a hipótese de duas raízes unitárias é uma boa candidata para explicar a persistência das inovações em algumas destas séries.

Palavras-Chave: Duas raízes unitárias; Quebras estruturais; Processos de Wiener

JEL: C12, C22.

Contents

List of Tables

List of Figures

1 Introduction

Testing for unit roots has received great attention over the latest decades, and has now become a standard procedure in time series analysis. Some of the most important consequences of the unit root hypothesis are the permanent effect of shocks on the long-run behaviour of a time series, its disruptive effect on conventional statistical inference, forecasting and properties of estimators, namely the Ordinary Least Squares (OLS) estimator.

Perron (1989) made a very important contribution to the unit root literature as he emphasized the empirical relevance of breaks in the trend function of economic and financial data, and its effect on the OLS estimators and unit root tests. Perron proved that the presence of structural breaks dramatically decreased the power of the Dickey Fuller (DF) and Augmented Dickey Fuller (ADF) test statistics. However, Perron's proposed test statistic requires *a priori* knowledge of the eventual date at which the break occurs. This was considered an important limitation and several authors sought to develop unit root tests that do not assume a known break date (Zivot & Andrews, 1992, and Perron, 1997, for example). Others, developed unit root test statistics which allow more than one changepoint in the trend, such as Lumsdaine & Papell (1997).

However, the focus remained on testing for exactly one unit root, and rule out the possibility of more unit roots. In applied work, there have been cases of series which seem to behave according to a double unit root process such as price series, wages, stock variables and population (see, Haldrup, 1998, for example). The conventional tests for two unit roots were developed by Hasza & Fuller (1979), Dickey & Pantula

(1987) and Sen & Dickey (1987). These authors propose tests for the null hypothesis of two-unit roots but without permitting structural breaks in the deterministic components.

The purpose of this thesis is to extend double unit root testing procedures allowing for the presence of breaks in the data. Instead of assuming exactly one unit root under the null hypothesis, it will be assumed two unit roots. Two test statistics are considered. One is valid under the assumption of a known break date and is shown to have a non-standard limiting distribution and its percentiles are obtained through Monte Carlo simulations and provided in this thesis. The second statistic relaxes the known break date assumption and follows the method of Zivot & Andrews (1992) to propose an endogenous change point version of the test. The hypothesis of an unknown break date implies the need for an estimate for the break date, which is taken to be the one that gives the strongest evidence against the null hypothesis. It is shown that the asymptotic distribution of the test statistic is non-standard and again its percentiles are obtained through Monte Carlo simulations and provided in this thesis.

The dissertation is organized as follows. Section 2 reviews the existing literature on the topic of unit root testing with structural changes in the deterministic component. Section 3 briefly addresses double-unit root processes and its characteristics, as well as some important limiting results. Section 4 addresses the trend break model. Section 5 explores the effect of structural breaks on the Hasza & Fuller test and derives the limiting distribution of the OLS estimator under the hypothesis of stationarity and exactly one-unit root, when structural changes are present, but ignored. In Section 6 the test statistics are presented and motivated, and its limiting distributions are derived as well as the simulated percentiles. Section 7 analyses empirical power and size through Monte Carlo simulations. It explores both tests' finite sample properties against several alternatives, including different pairs of autoregressive roots and more complex deterministic functions. In Section 8, the tests are applied to several time series, where we explore and comment on the results. Section 9 is reserved for the concluding remarks and suggestions for future research. The proofs of the theorems can be found in the Mathematical Appendix.

2 Literature Review

This dissertation develops tests for two unit roots in an univariate time series that can be used when there are structural breaks in the trend function.

After the seminal paper of Nelson & Plosser (1982), it was believed that many macroeconomic time series were well described by a unit root process. However, at the time, the existing tests for the null hypothesis of exactly one unit root were based on strict assumptions, namely, parameter constancy of the deterministic components. Yet, we can identify a number of significant events that seem to have altered the path of several economic variables, such as the Great Crash (1929), the First and Second World Wars (1914-1919 and 1939-1945, respectively), the Oil Price Shock (1973), and more recently, the Financial Crisis (2007) that started with the bankruptcy of Lehman Brothers. These events, if not modelled, can potentially invalidate any inference based on unit root tests.

This means that unit roots and structural breaks cannot be treated independently. Perron (1989) made the first breakthrough, showing how structural changes impacted unit root tests. He proposed a test based on an augmented Dickey-Fuller regression that sought to eliminate the effect of the structural change. However, the test was based on the inconvenient assumption of a known break date. Nonetheless, Perron found that many time series, studied by Nelson & Plosser (1982), that were first taught to exhibit a unit root, could instead be described by an I(0) process around a changing trend function.

The paper of Perron (1989) marked the beginning of a new strand in the literature of unit root testing. From that point onwards many authors worked to develop new tests which allowed for an unknown changepoint, whilst other studied the effect of using wrong break dates. For example, Zivot & Andrews (1992) developed the widely used $\inf t$ test, which searches for the break date that maximizes the likelihood of observing the alternative hypothesis. Perron (1997) presents a more complete test that allows for a breakpoint under the null and alternative hypothesis. Other authors opted to explore the behaviour of Perron's (1989) test, such as Montañés (1997), which revealed that an erroneous choice of the changepoint used in Perron's t test caused a significant loss in power in small samples

More recent examples of robust unit root tests include Carrion-i-Silvestre, Kim & Perron (2009) for their GLS-based unit root test that allows for multiple breaks both under the null and alternative hypothesis; Harvey, Leybourne & Taylor (2013) which propose minimum DF statistics in the possible presence of multiple breaks; Cavaliere *et al.* (2011) present a robust unit root test under multiple possible changepoints and non-stationary volatility, using bootstrapped minimum DF test statistics.

Nevertheless, robust tests that allow for more than one unit root under the null hypothesis have not been developed. Since the existing tests allow for exactly one unit root under the null hypothesis, if a time series process exhibits more roots equal to unity, then those tests are theoretically invalid and can lead to wrong conclusions.

The literature for two unit roots is scarcer. Nonetheless, there are some important contributions, namely Hasza & Fuller (1979) for an F-test which allows to test the restrictions imposed by the two unit roots; Dickey & Pantula (1987) for a sequential procedure based on t-statistics which allows for testing multiple unit roots, starting from the highest number of roots and testing down; Sen & Dickey (1987) proposed a t test based on a symmetric OLS estimator. More recently, Haldrup (1994) and Shin & Kim (1999) develop semi-parametric tests for two unit roots. However, double unit root tests carry additional complexity, particularly when concerning the initial conditions. Rodrigues & Taylor (2003) explore the role of the initial conditions on the double unit roots tests, and find that the Hasza-Fuller F test and Sen-Dickey t test, are not invariant to the starting values and that the limit distributions depend on the demeaning method used.

Although tests for two unit roots have been developed, none is robust to structural changes in the deterministic function, thus, this dissertation seeks to address this issue, and proposes two tests that allow for that possibility.

3 The Double-Unit Root Process and its Characteristics

Consider the following dynamic process for $\{y_t\}$, without any type of deterministic component:

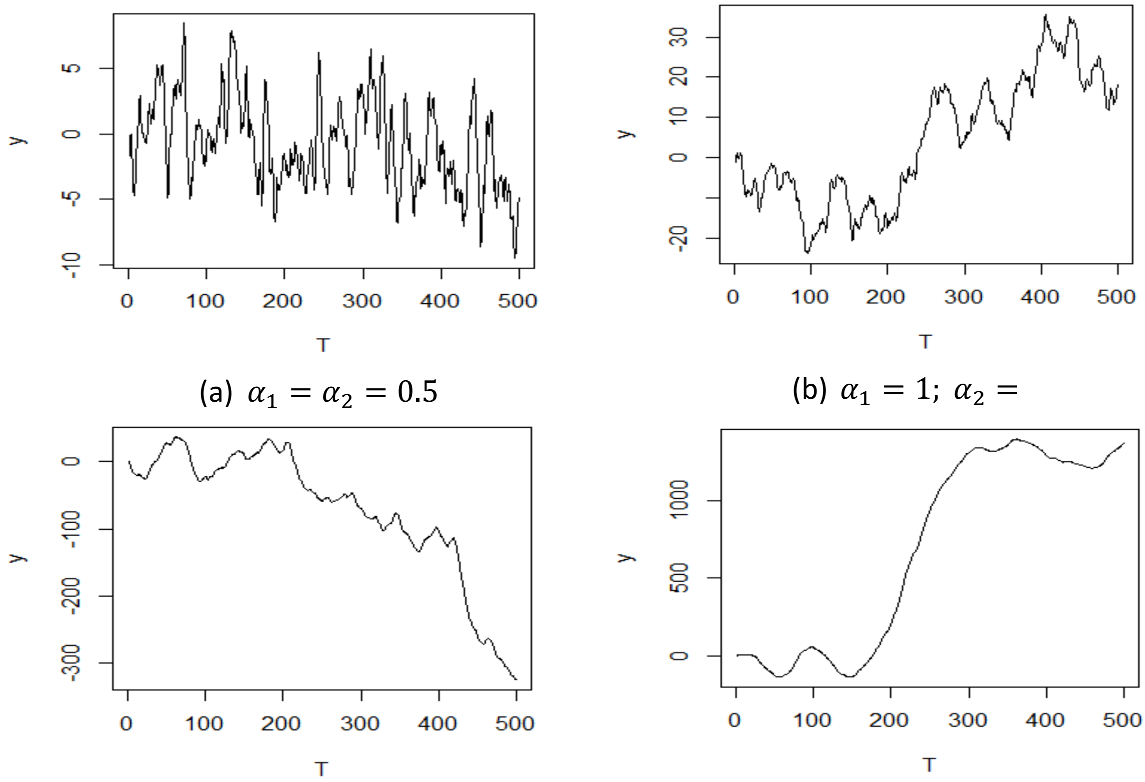
$$(1-\alpha_1 L)(1-\alpha_2 L)y_t = e_t, e_t \sim iid\ (0, \sigma^2) \tag{1}$$

It is well known that when $\alpha_1 = 1$ (or $\alpha_2 = 1$) and $|\alpha_2| < 1$ (or $|\alpha_1| < 1$) we have that $\{y_t\}$ has one unit root. The idea of the DF and ADF type of statistics is to test this individual restriction, under the null hypothesis. When $\alpha_1 = \alpha_2 = 1$ $\{y_t\}$ has two unit roots. However, when a second unit root is allowed, the statistical framework is more complex. In particular, when a test for two unit roots is performed in $\{y_t\}$, the alternative hypothesis is not necessarily a stationary process as in DF and ADF, because under the alternative hypothesis, the process can either be a unit root process ($\alpha_1 = 1$, $|\alpha_2| < 1$ or $|\alpha_1| < 1$, $\alpha_2 = 1$), or I(0) ($|\alpha_1| < 1, |\alpha_2| < 1$) process. Naturally it can also be an explosive process ($|\alpha_2| > 1$ or $|\alpha_1| > 1$ or both) but we do not consider this possibility in this thesis. A second source of additional complexity is the effect of the starting values. To see this, take equation (1) with $\alpha_1 = \alpha_2 = 1$ and solve recursively, denoting the starting values by y_0 and y_{-1}, yielding:

$$y_t = y_0 + (y_0 - y_{-1})t + \sum_{j=1}^{t} \sum_{k=1}^{j} e_k \qquad (2)$$

This equation explicitly shows that, even without adding any deterministic component to the right-hand side of equation (1), the starting values generate a linear trend if $y_{-1} \neq y_0$. Thus, in order for our tests to be invariant to these starting values we must introduce a linear trend to the test equations.

Another characteristic of double integrated processes is its smoothness, when compared to I(0) and one-unit root processes as it can be observed in Figure 1. This is a direct consequence of the double summation of the errors (see equation 2).



(a) $\alpha_1 = \alpha_2 = 0.5$ (b) $\alpha_1 = 1;\ \alpha_2 =$

(c) $\alpha_1 = 1;\ \pi = 0.9$ (d) $\alpha_1 = \alpha_2 = 1$

Figure 1 – Simulated paths for the process given in equation (1) with $e_t \sim iid\ N(0,1)$

4 The Trend Break Model

In this thesis, we assume that the time series under analysis, denoted as y_t, is a realization of the following time series process (DGP):

$$y_t = \boldsymbol{\beta}^{\boldsymbol{i}'} \boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{i}}(\lambda) + x_t,\ t = 1, \ldots, T,\ i = A, B, C \tag{3}$$

where

$$\boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{A}}(\lambda) = \big(1, t, DU_t(\lambda)\big)', \boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{B}}(\lambda) = \big(1, t, DT_t(\lambda)\big)', \boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{C}}(\lambda) = \big(1, t, DU_t(\lambda), DT_t(\lambda)\big)' \tag{4}$$

and

$$\boldsymbol{\beta}^{\boldsymbol{A}'} = (\mu^A, \delta^A, \mu_b^A)',\ \ \boldsymbol{\beta}^{\boldsymbol{B}'} = (\mu^B, \delta^B, \delta_b^B)',\ \ \boldsymbol{\beta}^{\boldsymbol{C}'} = (\mu^C, \delta^C, \mu_b^C, \delta_b^C)'$$

with $DU_t(\lambda) = 1(t > T_b)$ and $DT_t(\lambda) = 1(t > T_b)(t - T_b)$. Here $1(\cdot)$ is the indicator function and $T_b = \lfloor \lambda T \rfloor$ the break date with $\lfloor . \rfloor$ denoting the integer part.

In words, the underlying process is generated as the sum of a deterministic $(\boldsymbol{\beta}^{\boldsymbol{i}'} \boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{i}}(\lambda), i = A, B, C)$ and a stochastic component (x_t). We allow the deterministic component to contain a linear trend and we assume that an exogenous shock may occur at period T_b and cause a structural break in the process. In general, such a break is modelled as a permanent change in the parameters of the trend function after its occurrence. This justifies the functional form of the deterministic part of (3) and the fact that equation (3) is referred in the literature as the "trend break model" (see, for example, Harvey et al., 2009). In particular, we consider three possible formulations of the trend break model: under Model A (i=A) and Model B (i=B) the exogenous shock may cause either a level shift or a slope shift, respectively. Under Model C (i=C), we allow for a simultaneous level and slope shift as a consequence of the break.

The stochastic component is assumed to follow an $AR(2)$ process for simplicity:

$$(1 - \alpha_1 L)(1 - \alpha_2 L)x_t = e_t, \;\; e_t \sim i.i.d\ (0, \sigma^2) \tag{5}$$

The purpose of this thesis is to propose statistical procedures to test the presence of two unit roots in the process generating y_t. Here the relevant null hypothesis is $H_0: \alpha_1 = \alpha_2 = 1$ and the alternatives may be either one-unit root $\left(\alpha_j = 1 \wedge |\alpha_s| < 1, j = 1,2;\ s = 1,2;\ j \neq s\right)$ or no unit roots $(|\alpha_1| < 1 \wedge |\alpha_2| < 1)$. We highlight that, to our knowledge, this is the first statistical testing procedure for the presence of *I(2)ness* in a given time series data allowing for the presence of a possible break in the trend function. This is the main contribution of this thesis.

5 Conventional double unit root tests under a break in trend

As remarked in the previous section, the interest of this thesis lies in testing the null hypothesis of *I(2)ness* in y_t. A number of statistical procedures have been proposed in the literature to test for the presence of two unit roots as mentioned in Section 2. However, all of these tests ignore the problem of structural breaks in the trend. Hence, to motivate this thesis a natural question to ask is: what are the consequences for conventional double unit root tests in terms of size and power when breaks in trend are present in the DGP?

For illustration, we study the behaviour of the Hasza & Fuller test (henceforth HF). Consider the auxiliary regression of the HF test:

$$\Delta^2 y_t = \mu^* + \delta^* t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t \tag{6}$$

Notice that this equation is nested in the statistical framework described by equations (3) and (5) with $\mu_b = \delta_b = 0$, *i.e.*, without a break in trend. In particular, $\mu^* = (1-\alpha_1)(1-\alpha_2)\mu + (\alpha_1+\alpha_2)\delta - 2\alpha_1\alpha_2\delta$, $\delta^* = (1-\alpha_1)(1-\alpha_2)\delta$, $\rho_1 = -(1-\alpha_1)(1-\alpha_2)$, $\rho_2 = \alpha_1\alpha_2 - 1.$ Given that $y_t \sim I(2)$ iff $\alpha_1 = \alpha_2 = 1$, the double unit root null hypothesis is $\rho_1 = \rho_2 = 0$. According to HF, such restrictions are tested with an F-type statistic which follows a non-standard asymptotic distribution under the null. The critical values are well known and can be obtained, for example, from Table 10.2 of Patterson (2011).

To assess the small sample effect of structural breaks on the properties of the HF test, a Monte Carlo experiment is presented. First, 5.000 replications of $\{y_t^i\}$ for $i = A, B$ of length 150 are generated, setting $\lambda = 1/2$, then the empirical size and power of the test are obtained for several values of μ_b and δ_b. For the cases $\alpha_1 = \alpha_2 = 0.8$, and

$\alpha_1 = 1$, $\alpha_2 = 0.8$, empirical power is computed. For the case $\alpha_1 = \alpha_2 = 1$, empirical test size is computed. In all calculations, the test equation is (6), and the null hypothesis tested is $\rho_1 = \rho_2 = 0$.

Table I: *Null Rejection Probabilities. Hasza-Fuller Test*

Model A Simulations: $\mu = 10, \delta = 1$						
$\alpha_1, \alpha_2 \downarrow$	$\mu_b = 0$	$\mu_b = 4$	$\mu_b = 16$	$\mu_b = 30$	$\mu_b = 34$	$\mu_b = 40$
0.8, 0.8	1.00	1.00	1.00	1.00	1.00	1.00
1.0, 0.8	0.616	0.653	0.961	1.00	1.00	1.00
1.0, 1.0	0.063	0.067	0.284	0.759	0.835	0.917
Model B Simulations: $\mu = 10, \delta = 1$						
$\alpha_1, \alpha_2 \downarrow$	$\delta_b = 0$	$\delta_b = 0.4$	$\delta_b = 1.4$	$\delta_b = 2.6$	$\delta_b = 3.4$	$\delta_b = 4$
0.8, 0.8	0.999	0.997	0.928	0.683	0.479	0.334
1.0, 0.8	0.616	0.595	0.454	0.271	0.189	0.135
1.0, 1.0	0.063	0.063	0.065	0.056	0.055	0.057

Table I shows that when the true process has solely a level shift (Model A), the HF test becomes oversized and such an effect becomes more severe with the magnitude of the break. The empirical power of the test also increases apparently as a result of the size distortion. When the true process is Model B (only a break in the slope of the trend), the opposite happens, power is significantly reduced for higher values of δ_b but the size remains almost unchanged and close to the 5% level. Lastly, the simulations for Model C (simultaneous level and slope shift) were not presented because the pattern followed by power and size depend on which effect dominates the other. If μ_b is significantly higher relative to δ_b, then results from Model A apply, whilst the opposite also verifies. To complement our results, we now derive the limiting distribution of the

OLS estimators $\hat{\boldsymbol{\rho}}^i = \left(\hat{\rho}_1^i, \hat{\rho}_2^i\right)' for\ i = A, B, C$ in equation (6), under the alternative hypotheses that $\{x_t\}$ is stationary and I(1).

Theorem 1: Let $\{y_t\}$ be generated by model $i = A, B, C$, according to equation (3).

Suppose we estimate equation (6), neglecting the level/slope shift. Then, as $T \rightarrow \infty$:

(1) Under the alternative hypothesis that $|\alpha_s| < 1$, $s = 1,2$:

(a) Under Model A

$(a.1)$ $\sqrt{T}(\hat{\rho}_1 - \rho_1) \rightarrow^d N[0, 2\sigma^2\tau]$

$(a.2)$ $\sqrt{T}(\hat{\rho}_2 - \rho_2) \rightarrow^d N[0, \sigma^2\tau]$

Where

$$\tau = \frac{(-3\lambda^4 + 6\lambda^3 - 4\lambda^2 + \lambda)\mu_b^2 + \gamma_0}{(\gamma_0^2 - \gamma_1^2) + \left((12\lambda^3 - 6\lambda^4 - 8\lambda^2 + 2\lambda)\gamma_0 + (6\lambda^4 - 12\lambda^3 + 8\lambda^2 - 2\lambda)\gamma_1\right)\mu_b^2}$$

(b) *Under Models B and C*

$(b.1)$ $T^{\frac{3}{2}}(\hat{\rho}_1 - \rho_1) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_1]$

$(b.2)$ $\sqrt{T}(\hat{\rho}_2 - \rho_2) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_2]$

Where

$\theta_1 = [\boldsymbol{V}]^{-1}{}_{3,3}$

$\theta_2 = [\boldsymbol{V}]^{-1}{}_{4,4}$

Where $[\boldsymbol{V}]^{-1}$ is the inverse of the matrix $\boldsymbol{V}$ given by:

$$\mathbf{V} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2}\delta_b(1-\lambda)^2 & \delta_b(1-\lambda) \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{2}\delta_b(1-\lambda)^2 \\ \frac{1}{2}\delta_b(1-\lambda)^2 & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{2}\delta_b^2(1-\lambda)^2 \\ \delta_b(1-\lambda) & \frac{1}{2}\delta_b(1-\lambda)^2 & \frac{1}{2}\delta_b^2(1-\lambda)^2 & \delta_b^2(1-\lambda) + 2(\gamma_0 - \gamma_1) \end{bmatrix}$$

(2)

(3) Under the hypothesis of exactly one-unit root in $\{y_t\}$

(c) Under Model A

Since the stochastic trend dominates the level shift, the asymptotic distributions are invariant to the break.

(d) Under Models B and C

$$(d.1) \quad T^{\frac{3}{2}}\left(\hat{\rho}_1^i - \rho_1^i\right) \rightarrow^p \theta^{-1}(G_1 - G_2)$$

$$(d.2) \quad \sqrt{T}\left(\hat{\rho}_2^i - \rho_2^i\right) \rightarrow^p \theta^{-1}(G_3 - G_4)$$

For $i = B$ *or* C

Where:

$$(1)\ \theta = \frac{1}{12}\delta_b^2(\lambda-1)^3\left(4\gamma_0 - \lambda\delta_b^2(12\lambda^6 - 24\lambda^5 + 15\lambda^4 - 15\lambda^3 + 24\lambda^2 - 16\lambda + 4)\right)$$

$$(2)\ \mathrm{G}_1 = \left(\gamma_0 - \lambda\delta_b^2(3\lambda^3 - 6\lambda^2 + 4\lambda - 1)\right) \times \left(\sigma\delta_b(\lambda^3 - \lambda^2 - 2\lambda + 2)W(1) - \sigma\delta_b(1-\lambda)W(\lambda) - \sigma\delta_b\left(\int_\lambda^1 W(r)dr + (\lambda-1)^2(1+2\lambda)\int_0^1 W(r)dr\right)\right)$$

$$(3)\ \mathrm{G}_2 = \frac{1}{2}\lambda^2\delta_b^2(\lambda-1)^2(2\lambda-1) \times \left(\sigma\left(\delta_b(2+2\lambda-3\lambda^2) - \sqrt{\gamma_0}\right)W(1) - \delta_b\sigma W(\lambda) - 6\lambda(1-\lambda)\delta_b\sigma\int_0^1 W(r)dr\right)$$

$$(4)\ \mathrm{G}_3 = \frac{1}{6}\delta_b^2(\lambda-1)^2(3\lambda^2 - 6\lambda^3) \times \left(\sigma\delta_b(\lambda^3 - \lambda^2 - 2\lambda + 2)W(1) - \sigma\delta_b(1-\lambda)W(\lambda) - \sigma\delta_b\left(\int_\lambda^1 W(r)dr + (\lambda-1)^2(1+2\lambda)\int_0^1 W(r)dr\right)\right)$$

$$(5)\ \mathrm{G}_4 = \frac{1}{6}\delta_b^2(\lambda-1)^2(3\lambda^2 - 6\lambda^3) \times \left(\sigma\left(\delta_b(2+2\lambda-3\lambda^2) - \sqrt{\gamma_0}\right)W(1) - \delta_b\sigma W(\lambda) - 6\lambda(1-\lambda)\delta_b\sigma\int_0^1 W(r)dr\right)$$

Where $W(r)$ *is a standard Brownian motion.*

Part (1) of Theorem 1 establishes that, under Model A, both estimators are normally distributed, with asymptotic variance depending on the parameters μ_b and λ. However, in small samples, the estimators are increasingly biased as μ_b grows, as can be inferred from Table I. For Models B and C, the limiting distribution is also Normal, with asymptotic variance as a function of δ_b. From Table I, we find that in small samples both estimators are biased towards zero zero, when δ_b grows. For models B and C, the rate of converge of $\hat{\rho}_1$ $(T^{\frac{3}{2}})$ is higher than the common $\sqrt{T}$, due to the fact that the slope change (DT_t) is present in the DGP but neglected in the test regression.

For the second part of the theorem, for Models B and C, the limiting distribution of the OLS estimator depend on the slope shift parameter and the break fraction. According to the simulations in Table I, for a given break fraction, both distributions concentrate around zero when the slope shift parameter increases. The result for Model A is not presented because the limiting distribution is not affected by the break parameters, since the stochastic trend asymptotically dominates a level shift. However, the small sample distribution clearly depends on these parameters, as can be seen from Table I.

To sum up, changes in the deterministic component affect the outcome of the HF test. The need for a consistent testing strategy follows from the results so far. For Models B and C, the problem is the low power against stationary, and single unit root alternatives, whilst for Model A it is a size problem. In fact, considering Model A, the test will always reject the null hypothesis for a sufficiently large shift, thus leading to the conclusion that the underlying process has less than two unit roots, when in fact it can be a double unit root process, a single unit root process or an I(0) process with a shift in its level. Therefore, inference based on the simple Hasza-Fuller test cannot lead to any valid conclusion. A similar argument holds for Models B and C.

6 Double unit root tests in the presence of a trend break

In this section, we present alternative testing procedures aimed at solving the problems considered in the previous section. First, the Hasza-Fuller test is extended to the case of a known break date. The methodology is similar to Perron (1989). Then a test allowing for a break at an unknown date is devised. This test is closely related to the $inf - t$ test of Zivot & Andrews (1992).

6.1 Known Break Date

In this section we closely follow Perron (1989).

Consider the following modified test regression for Model C, where the dummy variables are included:

$$y_t = \mu + \delta t + \mu_b DU_t + \delta_b DT_t + x_t$$

$$(1 - \alpha_1 L)(1 - \alpha_2 L)x_t = e_t \sim i.i.d\ (0, \sigma^2)$$

These equations can be combined, and after suitable transformations we arrive to:

$$\Delta^2 y_t = \mu^* + \delta^* t + \theta_1 DT_t + \theta_2 DU_t + \theta_3 \Delta DU_t + \theta_4 \Delta^2 DU_t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t \tag{7}$$

Where, $\mu^* = (1 - \alpha_1)(1 - \alpha_2)\mu + (\alpha_1 + \alpha_2)\delta - 2\alpha_1\alpha_2\delta$, $\delta^* = (1 - \alpha_1)(1 - \alpha_2)\delta$, $\theta_1 = (1 - \alpha_1)(1 - \alpha_2)\delta_b$, $\theta_2 = (1 - \alpha_1)(1 - \alpha_2)\mu_b + (2 - \alpha_1 - \alpha_2)\delta_b$, $\theta_3 = (2 - \alpha_1 - \alpha_2)\mu_b + \alpha_1\alpha_2\delta_b$, $\theta_4 = \alpha_1\alpha_2\mu_b$, $\rho_1 = -(1 - \alpha_1)(1 - \alpha_2)$, $\rho_2 = \alpha_1\alpha_2 - 1$.

Model A follows by fixing $\delta_b = 0$ and Model B by setting $\mu_b = 0$.

The OLS estimator of equation (7) attains exact invariance to the break parameters, thus, that equation should be used to test the null hypothesis $\rho_1 = \rho_2 = 0$. However, the following regression is asymptotically equivalent to (7):

$$\Delta^2 y_t = \mu^* + \delta^* t + \theta_1 DT_t + \theta_2 DU_t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$$

This is because ΔDU_t and $\Delta^2 DU_t$ are pulse dummies, and thus, $o_p(1)$.

Hence, we can derive the limiting distribution of the test statistic with the reduced equation.

Additionally, in empirical applications it might be needed to include lagged second differences to equation (7) to account for the presence of serial correlation in the error term e_t.

In the following theorem, we study the limit distributions of the test statistic $F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$ from equation (7):

Theorem 2: Let $\{y_t\}$ be generated according to equation (3). Additionally, let $F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$ denote the statistic used for testing the nullity of both ρ_1^i and ρ_2^i in equations (7) for $i = A, B, C$. Then, under the null hypothesis that $\rho_1 = \rho_2 = 0$, as $T \to \infty$:

$(a)\quad F_{\hat{\rho}}^{i}(\lambda) \to^{d} (2S_i)^{-1} A_i$

Where

$$A_i = \left(\eta_2^i(\lambda)\left[\xi_1^i(\lambda)\right]^2 - 2\eta_3^i(\lambda)\xi_1^i(\lambda)\xi_2^i(\lambda) + \eta_1^i(\lambda)\left[\xi_2^i(\lambda)\right]^2\right)$$

$$S_i = \left(\eta_1^i(\lambda)\eta_2^i(\lambda) - \left[\eta_3^i(\lambda)\right]^2\right)$$

$i = A, B, C$

Where:

$$\xi_1^i(\lambda) = \left\{\int_0^1 \Upsilon_1 dW(r) - \int_0^1 \boldsymbol{Z^i}(\lambda,r)'dW(r)\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_1 dr\right\}$$

$$\xi_2^i(\lambda) = \left\{\int_0^1 \Upsilon_2 dW(r) - \int_0^1 \boldsymbol{Z^i}(\lambda,r)'dW(r)\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\}$$

$$\eta_1^i(\lambda) = \int_0^1 \left\{\Upsilon_1 - \boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_1 dr\right\}^2 dr$$

$$\eta_2^i(\lambda) = \int_0^1 \left\{\Upsilon_2 - \boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\}^2 dr$$

$$\eta_3^i(\lambda) = \int_0^1 \left\{\Upsilon_1 - \boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_1 dr\right\} \times \left\{\Upsilon_2 - \boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\} dr$$

$$\Upsilon_1 = [rV(r) + V_r(r) + W_r(r)]$$

$$\Upsilon_2 = [V(r) + W(r)]$$

Where $V(r)$ and $W(r)$ are two independent standard Brownian motions and $V_r(r) = \int_0^r V(s)ds$ and $W_r(r) = \int_0^r W(u)du$.

The representation of the limit distribution of $F_{\hat{\rho}}^i(\lambda)$ is free of nuisance parameters and it is only a function of λ. In particular, it does not depend on the magnitude of the break, thus allowing for hypothesis testing under the assumption of a known break date. Table II presents the selected simulated percentiles of the asymptotic distribution of $F_{\hat{\rho}}^i(\lambda)$ for $i = A, B, C$. These critical values were obtained through Monte Carlo simulations. First we simulate $T = 1000$ random $N(0,1)$ variates, and generate x_t according to equation (5), fixing $\alpha_1 = \alpha_2 = 1$. Then, for each model $i =$

A, B, C, the value of λ is fixed, as shown in each column entry of Table II, y_t is generated as in (3) and the test regressions (7) are estimated. Finally, we calculate the value of the test statistic. We repeat this process 5000 times for each λ from 0.1 to 0.9, and obtain the desired percentiles with the 5000 obtained test statistics.

Some key features are worth mention regarding these critical values. First, for all values of λ, the critical values are greater than those of the conventional HF test. Moreover, the critical values are clearly influenced by λ, exhibiting a symmetric behaviour around $\lambda = 0.5$, and achieving its maximum around that same value. Secondly, as $\lambda \rightarrow 0$ or 1, the critical values get closer to those tabulated by Hasza & Fuller (1989), which is also to be expected from the asymptotic derivations.

Table II: *Selected Percentiles of the Asymptotic Distribution of* $F_{\hat{\rho}}^{i}(\lambda)$

$Model\ A$									
$\lambda \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90%	8.212	9.952	9.073	9.103	9.019	9.186	9.094	8.917	8.188
95%	9.288	10.269	10.379	10.367	10.205	10.397	10.412	10.196	9.327
99%	11.537	12.648	12.922	12.830	12.559	12.509	13.060	12.501	11.731
$Model\ B$									
$\lambda \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90%	8.201	9.422	10.761	11.602	11.841	11.523	10.765	9.427	8.194
95%	9.295	10.784	12.054	12.915	13.136	13.014	12.110	10.819	9.297
99%	11.509	13.145	14.599	15.672	15.816	15.789	15.167	13.592	11.705
$Model\ C$									

$\lambda \rightarrow$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90%	8.189	9.465	10.797	11.573	11.873	11.531	10.744	9.435	8.175
95%	9.275	10.724	12.019	12.949	13.163	12.937	12.108	10.841	9.321
99%	11.466	13.158	14.818	15.688	15.669	15.747	15.145	13.553	11.705

6.2 Unknown Break Date

The first part of the text is concerned with testing for two unit roots allowing for the possibility of a single break in trend at a known date. However, in practice, we seldom know the true break date, whether that's caused by lagged decisions made by the agents, which not always coincide with the announcement of economic events or simply because the investigator does not have any *a priori* information about a possible shift in the trend function. Therefore, if parameters change at an unknown date, the empirical researcher either ignores it or chooses a date which will likely be wrong. In those two cases, statistical inference will be misleading. A third choice is to search for the break date. Whenever a systematic search is made, the test presented in section 6.1 is no longer valid, because it treats the break date as exogenous, and thus, the test is likely to indicate the rejection of the null hypothesis, when in fact it is true.

Therefore, we propose an alternative test procedure, which is an extension of Zivot & Andrews (1992) to the case of two unit roots. The first step is to reformulate our null hypothesis. In what follows, we will no longer consider that the null hypothesis is a double unit root process with a break in the deterministic trend. Under the null hypothesis, we will assume a pure double unit root process.

$$(1-L)^2 y_t = e_t \tag{8}$$

Because the null hypothesis implicit in Equation (8) is different, the new test equations are defined by (9), without the need for additional dummy variables.

$$\Delta^2 y_t = \mu^* + \delta^* t + \theta_1 DT_t + \theta_2 DU_t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t \tag{9}$$

The key difference is that we now treat λ as unknown, thus $DU_t(\lambda)$ and $DT_t(\lambda)$ are also unknown.

To proceed with the test, we need an estimate of λ, denoted by $\hat{\lambda}$. Under our maintained hypothesis, $\hat{\lambda}$ is such that the evidence for a process with a trend shift is greatest. Thus, if we construct an algorithm that searches across every possible λ, for the greatest evidence in favour of the alternative hypothesis, and compare the value of the test statistic computed with $\hat{\lambda}$ with a threshold mark, and then we will be able to test our null hypothesis.

Since we are dealing with an F test, the scheme consistent with the argument above is to choose the greatest value of the sequence of test statistics, computed across every $\lambda \in (0,1)$. The next theorem provides the limiting distribution for the test statistic of interest.

Theorem 3: Let $\{y_t\}$ be generated according to Equation (8), with the error sequence $\{e_t\}$ i.i.d, with $E(e_t) = 0$ and $E(e_t^2) = \sigma^2 > 0$. Additionally, let $F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$ be the test statistic computed for equations (9) *for $i = A, B, C$, for a given $\lambda \in (0,1)$. Then, as $T \to \infty$:*

(a) $\sup_{\lambda\in\Lambda} F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda) \to^{d} \sup_{\lambda\in\Lambda}((2S_i)^{-1}A_i)$ $\qquad (i = A, B, C)$

Where S_i and A_i have the same expressions as in Theorem 2.

The proof is given in Appendix B.

Theorem 3 provides the limiting distribution of the test statistics used to test the null hypothesis of two unit roots. These are functions of Wiener processes and λ.

Additionally, these distributions are valid only when the error term $\{e_t\}$ in equation (9) is $i.i.d.$ The percentiles for these distributions were obtained through Monte Carlo simulation and are presented in Table III. For each test statistic, we construct a sequence of $N(0,1)$ random variables $\{e_t\}_{t=1}^{T}$ with T given in each row of Table III. Then, we construct the process $\{y_t\}_{t=1}^{T}$ based on equation (8). Next, we run $T-2$ regressions (9), one for each λ, from $\lambda = 2/T$ to $\lambda = (T-1)/T$, and calculate the test statistic $F_{\hat{\rho}}^{i}(\lambda)$ for each. Finally, we take the supremum of the $T-2$ test statistics. We repeat this process 5000 times to obtain the percentiles of the distribution of $\sup_{\lambda\in\Lambda} F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$.

The results in Theorem 3 also show that the test statistic is asymptotically invariant to the magnitude of the break. However, in small samples, the distribution will depend on these parameters. This is a direct consequence of the asymmetry imposed under the null and alternative hypothesis, that is, under the null hypothesis, changes in the level or slope of the series can only be explained by exogenous shocks coming from the error distribution. This problem does not arise in the test for a known break date because a level (slope) shift is allowed both under the null and the alternative hypotheses. This is clearly one of the limitations of the proposed test for an unknown changepoint. However, the effect of structural breaks on test size and power vanish asymptotically.

Table III: Selected percentiles of the distribution of $sup_{\lambda\in\Lambda} F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$

Model A				
Sample Size ↓	90%	95%	97.5%	99%
50	17.972	20.102	22.185	25.306
100	15.747	17.476	19.059	21.191
200	15.049	16.563	17.968	19.853

>500	14.712	15.943	17.122	18.408
Model B				
Sample Size ↓	90%	95%	97.5%	99%
50	19.052	21.764	24.073	26.377
100	16.032	17.719	19.554	21.596
200	14.953	16.645	17.817	19.117
>500	14.019	15.367	16.622	18.183
Model C				
Sample Size ↓	90%	95%	97.5%	99%
50	21.950	24.388	26.556	30.175
100	18.712	20.585	22.253	24.619
200	17.601	18.691	20.339	22.383
>500	16.761	18.049	19.299	20.861

Testing for the null hypothesis goes as follows: Choose the Model (A, B or C) that best describe the data from (9), and estimate it by OLS for break fractions, λ, going from $j = 2/T$ until $j = (T-1)/T$. For each possible break fraction, it might be needed to augment the test regression with lagged second differences of y_t to remove the effect of autocorrelated errors on the properties of the test statistics. The number k of lags can be determined with the GTS (General-to-Specific) methodology with the usual p-value of 10% starting from $k = k_{max}$. Then compute $F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda)$ for each λ, and choose the greatest entry of the sequence of test statistics, and compare it with the respective critical value presented in Table III. Reject the null hypothesis if the value of the test statistic is greater than the chosen critical value.

6.3 Rejecting the Null Hypothesis

One key issue when testing for two unit roots is how to proceed when the test leads to the rejection of the null hypothesis. In this case, the series can either be I(0) or have a unit root. Several tests have been devised to test the null of exactly one unit root when there is a break in the deterministic trend, such as Perron (1989), Zivot & Andrews (1992) and more recently Perron (1997). Therefore, when rejecting the null hypothesis of two unit roots, one natural suggestion is to proceed sequentially and in a second step apply one of these tests for one unit root and conclude whether the series is I(0) or I(1). Caution should be taken when a second test is performed sequentially, because the overall test size increases with the application of a new individual test. The difference between the significance level defined for each individual test and the probability of Type I error of the sequential procedure is lwft for future research.

7 Size and Power Simulations

We now assess the finite sample power and size of the proposed tests. The data generating processes for the Monte Carlo simulations are given by equations $(3)\ i = A, B, C$ and (5) with $e_t \sim i.i.d\ N(0,1)$. The number of replications is set to 2500, and the sample size is 50, 100, 200, 300, 400 and 500. The nominal size was fixed at 5% and the asymptotic critical values from Table II were used for the test with a known break date. The reported results are for Model A only, but Monte Carlo simulations for Models B and C undertaken by the author of this thesis[1] show identical outcomes. For the test with unknown break date the critical values from the Model A entry in Table III were used. Because the null hypothesis assumes that no change occurs in the deterministic function, we also investigate the robustness of the test to structural

[1] The simulation results for models B and C are available upon request.

changes under the null. In those simulations, the sample size is fixed in T=150 and $\lambda = \frac{1}{2}$, while varying the break magnitude. Lastly, we inspect the behaviour of the test for an unknown changepoint when multiple breaks occur, both under the null, and under different alternatives, for Models A and B. In this case the sample size is also T=150.

Table IV: *Null Rejection Probabilities – Known Changepoint Test*

$\alpha_1, \alpha_2 \downarrow$	50	100	200	300	400	500
1.0, 1.0	0.046	0.018	0.021	0.034	0.033	0.041
1.0, 0.9	0.112	0.067	0.128	0.305	0.577	0.818
1.0, 0.7	0.287	0.249	0.434	0.782	0.960	0.998
0.9, 0.9	0.188	0.199	0.567	0.922	0.995	1.000
0.9, 0.8	0.315	0.427	0.861	0.966	0.999	1.000
0.99, 0.99	0.066	0.033	0.040	0.055	0.082	0.108

Table V: *Null Rejection Probabilities – Unknown Changepoint Test*

$\alpha_1, \alpha_2 \downarrow$	50	100	200	300	400	500
1.0, 1.0	0.213	0.100	0.068	0.055	0.057	0.053
1.0, 0.9	0.338	0.348	0.562	0.842	0.974	1.000
1.0, 0.7	0.732	0.928	1.000	1.000	1.000	1.000
0.9, 0.9	0.465	0.648	0.987	1.000	1.000	1.000
0.9, 0.8	0.658	0.914	1.000	1.000	1.000	1.000
0.99, 0.99	0.228	0.132	0.134	0.138	0.138	0.168

Starting with the test for a known breakpoint (Table IV), the empirical size is very close to nominal 5%, even when the sample includes only 50 observations. When the sample grows larger, the simulated size draws a U-shaped trajectory, with a minimum close to 2% when the sample size is around 100 observations. However, it tends to the nominal value of 5% with successive increases in T.

The test also displays decent power to reject the null in finite samples against several configurations of the alternative hypothesis. In the somewhat extreme case of both roots equal to 0.99, the test presents very low power, close to the nominal size. However, even in this case, we notice that power tends to increase with the sample size, albeit at a very slow rate.

For the test with an unknown changepoint (Table V), there is a modest size distortion in small samples. The test has finite sample power to reject the null, and lacks power only in the extreme cases of both inverse roots equal to 0.99, although in the latter case, convergence to maximum power is much faster.

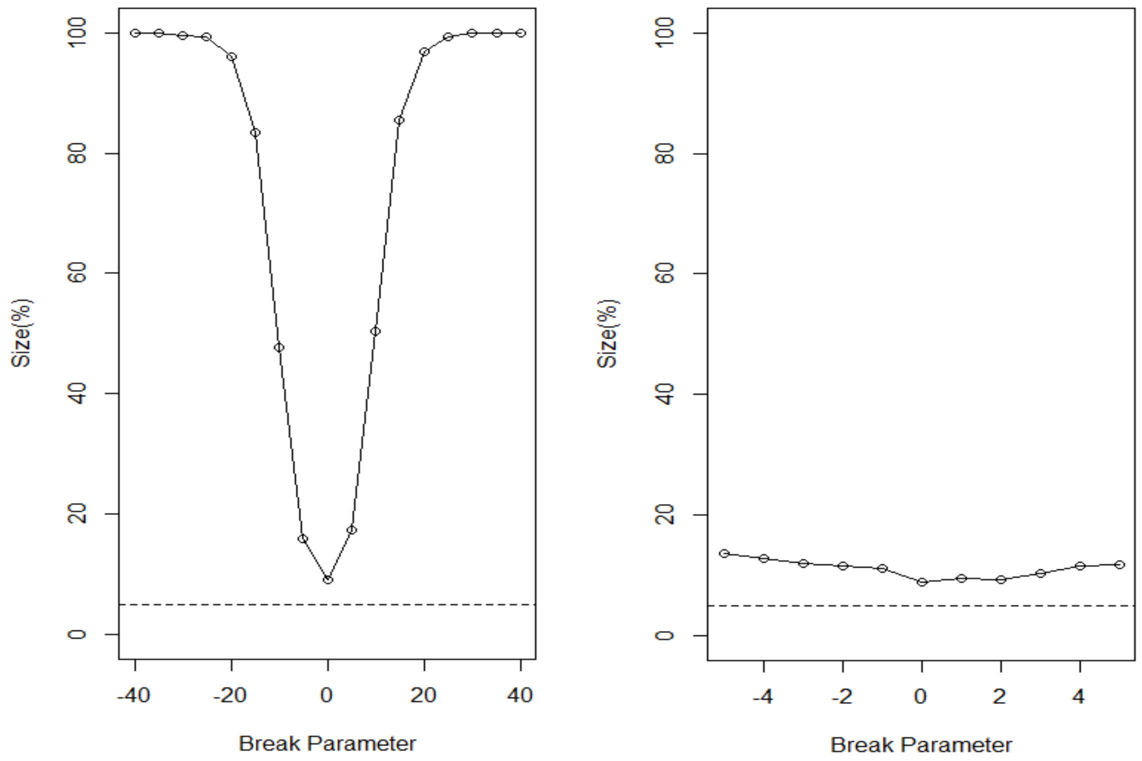


Figure 2 - Size as a function of the break parameters. Model A (left) and Model B (right)

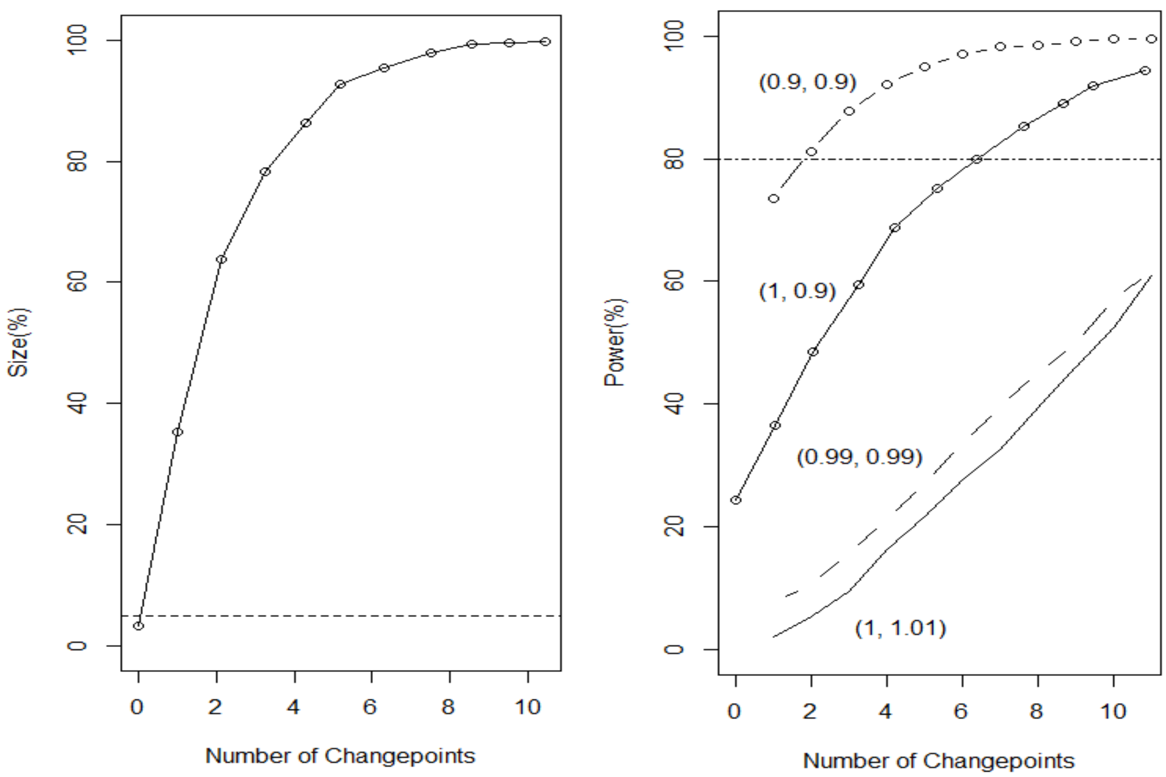


Figure 3 Size (left) and Power (right) as a function of the number of changepoints for Model A

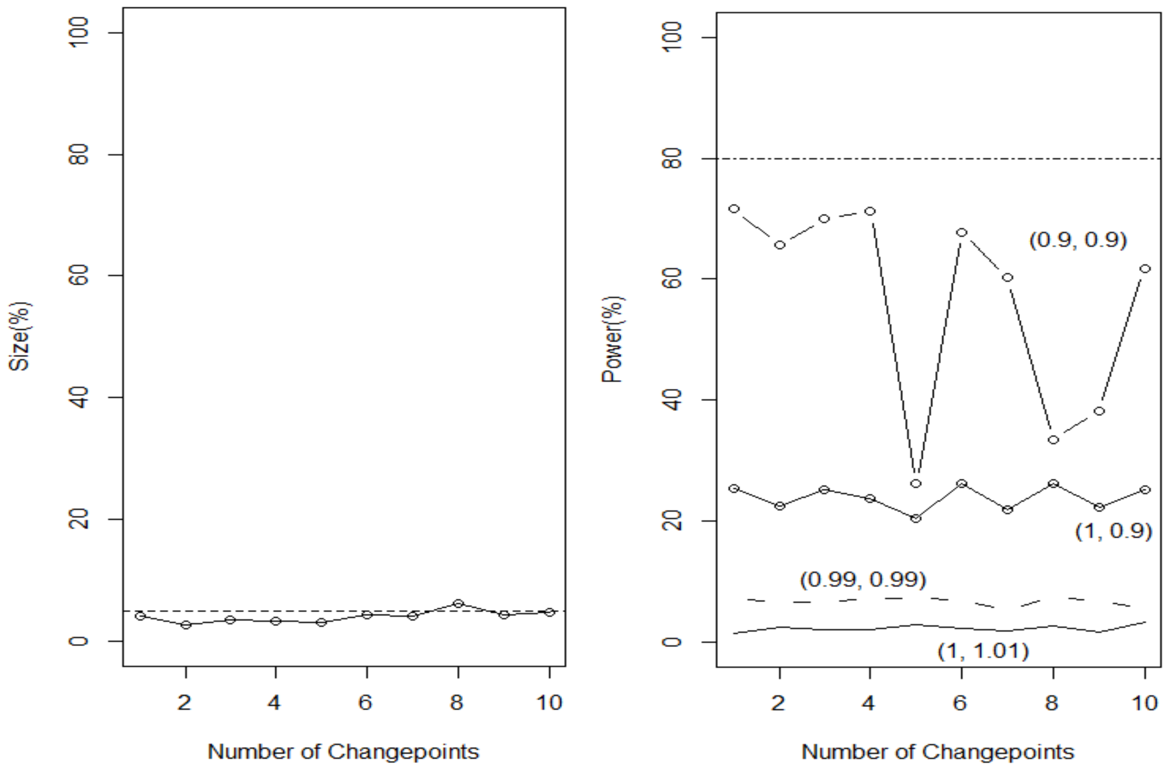


Figure 4 - Size (left) and Power (right) as a function of the number of changepoints for Model B

Figures 2-4 show the behaviour of the test for an unknown break date, under different configurations. Figure 2 shows that a level shift (Model A) under the null hypothesis leads to an oversized test, which is aggravated when the break parameter grows larger. However, from Theorem 3, this problem disappears when the sample size grows large. If instead of a level shift, we have a slope change (Model B) the effect is almost negligible. The test becomes oversized, but the size distortion increases very slowly with the magnitude of the break. Again, Theorem 3 ensures that this problem does not exist in large samples.

Figure 3 shows a similar exercise; however, this time it is the number of breakpoints that changes. Here, both size and power increase with the number of breaks.

Figure 4 shows the effect of multiple changepoints, but for Model B. Here, the test size and power remain almost unchanged.

8 Empirical Applications

We now apply the previous tests to a dataset composed of monthly, quarterly and annual series. Information about the complete dataset can be found in the *Data Appendix*. One relevant aspect to the application of the proposed tests is the choice of k, *i.e.*, the number of lagged second-differences, $\Delta^2 y_{t-i}, i = 1, \ldots, k$, to be added in equation $(7.i)$. We opted to use a conventional GTS strategy with p-value=0.1, starting from $k = k_{max}$. The k_{max} chosen depended on the frequency of each time series. For the annual series we used $k_{max} = 4$ or 6, for the quarterly series we used $k_{max} = 8$ and for the monthly series we used $k_{max} = 24$. For each model, the sequence of test statistics was calculated. Once the sequence of test statistics is obtained, we use the estimate of the break date of each model, which is the one that corresponds to the $\sup_{\lambda \in \Lambda} F_{\hat{\boldsymbol{\rho}}}^{i}(\lambda_0)$, for $i = A, B, C$, and rerun the test regression with those estimated break dates and calculate the AIC and BIC information criteria. The chosen model is that which minimizes these criteria. In the case of different models chosen, the Schwartz criteria is favoured because it tends to choose a more parsimonious model. The results are given in Table VI.

The test under the unknown changepoint framework was applied for all series, except the Greek government debt. For the Greek Government Debt series, the test for a known break date was used, since there is a clear level shift in the last quarter of 2013. For the monthly series, the null hypothesis was rejected for all but three series when using the $\sup_{\lambda \in \Lambda} F_{\hat{\boldsymbol{\rho}}}(\lambda)$. From Figures 1 and 2 (left panels) it is known that the proposed test has excessive size-distortions when a level shift truly exists, even if the null hypothesis is true. Structural changes are a common feature of long span data, thus, it is possible that the null hypothesis is rejected due to the presence of a level shift. This problem, and an eventual solution, will be discussed shortly.

For the quarterly series, we highlight the fact that for the Portuguese and Greek General Government debts, the null hypothesis was not rejected meaning that innovations have a very persistent behaviour. In fact, if we assume that the data can be described by an $ARIMA(p, 2, q)$ process, then it is easy to understand and justify the marked increase in debt observed in the latest years (several positive shocks), and why it has been difficult to lower said debt (past shocks have its effects increased over time).

For the annual time series, there are some differences in both tests, namely for some population series, which is an interesting outcome. For six population series, the HF test rejects the null hypothesis, whilst the supremum test does not. Non-rejection of the null hypothesis for the population series suggests that, with the right policies (positive shocks), it might be possible to reverse the downward trend verified in many developed countries. For Portugal, however, the null hypothesis was rejected when using $\sup_{\lambda\in\Lambda} F^{A}_{\hat{\boldsymbol{\rho}}}(\lambda)$, but not rejected when using the common HF test, which seems to go against the findings in Table I. However, note that the $\sup_{\lambda\in\Lambda} F^{A}_{\hat{\boldsymbol{\rho}}}(\lambda)$ might be rejecting because there really is a level shift in 1974, which coincides with a major event in Portugal.

Table VI: *Tests for the Double Unit Root Null Hypothesis*

Variable	T	$\widehat{T_b}$	HF test	Model	$\sup_{\lambda\in\Lambda} F^{i}_{\hat{\rho}}(\lambda)$	Conclusion
PIPCE	699	2/1973	14.72	A	18.65	H0 (1%)
PIPCECORE	699	2/1973	14.12	A	18.30	H0 (1%)
PIPCEDG	699	3/1980	4.65	B	28.27	H1
PIPCEF	699	1/1973	11.58	A	30.85	H1
CPIALL	686	10/1981	5.75	C	26.99	H1
CPICORE	700	2/1973	4.34	A	15.28	H0
TCCOUT	891	2/2006	29.62	A	35.95	H1
EMPLOYNF	940	11/1997	39.82	B	53.05	H1
M1	700	12/1993	11.55	A	21.37	H1
M2	700	1/1987	10.70	A	19.95	H1
PIGDP	281	1/1973	4.37	A	14.39	H0
PIPCEHOUSE	281	10/1973	6.03	A	13.94	H0
PIGPDI	281	Q1/1973	14.43	A	28.76	H1
TFD	204	Q1/1990	5.61	B	12.09	H0
PTGGOVDEBT	68	Q1/2012	4.94	B	11.19	H0
ESGGOVDEBT	68	Q1/2003	13.54	B	335.29	H1
GRGGOVDEBT*	68	Q4/2013	12.24	A	-	H0 (1%)
ITGGOVDEBT	68	Q3/2003	58.90	A	65.71	H1
USNRULR	68	Q1/1990	9.03	A	13.99	H0
USNRUSR	273	Q3/1975	12.60	B	17.73	H0 (2.5%)
ROUTM	273	Q1/2008	23.94	A	38.71	H1
OUTNFB	121	Q1/2006	19.04	C	93.72	H1
COMPNFB	281	Q1/1982	6.48	A	28.64	H1
PTCPBNAS	281	Q2/2008	11.38	A	38.19	H1
PTCP	88	Q2/2008	24.34	A	37.08	H1
PTCPUB	87	Q2/2010	11.10	A	19.11	H0
USMFI	88	1980	16.02	B	30.59	H1
USRMFI	60	1969	26.38	B	30.55	H1
USFDTOGDP	60	1964	10.94	B	42.13	H1
USFDHP	76	1975	16.65	A	21.93	H0 (2.5%)
USFDDSL	76	2000	1.68	B	9.20	H0
USFGSL	71	1995	9.36	C	18.74	H0

BEL	71	1998	7.11	B	16.19	H0
CAN	56	1998	14.09	A	18.25	H0
CHN	56	1996	10.22	B	60.99	H1
EMU	56	1974	10.44	B	17.42	H0
EUU	56	1976	11.01	B	16.88	H0
FIN	56	1999	17.46	A	37.23	H1
FRA	56	1973	11.43	B	17.15	H0
ITA	56	1989	3.47	B	21.12	H0
PRT	56	1974	9.14	A	39.17	H1
ESP	56	1993	11.54	A	15.64	H0
DE	56	1989	13.60	A	25.40	H0 (1%)
US	86	1947	3.72	C	26.44	H0 (2.5%)

One of the limitations of the proposed test for an unknown changepoint is that it tends to reject the null hypothesis more often than it should, when a change in the deterministic function exists. If the data series studied is not too long, then it is plausible that, should it be present, a very reduced number of changepoints exist. With this in mind, for those series in Table VI for which the null was rejected, the sample was split in two using the estimated break date. Then, the test for two-unit roots was applied to the longest segment. Table VII shows the results. The green entries indicate those series for which the conclusion of the test changed, yellow indicate those for which the conclusion did not change, and the grey series were not re-tested since the null had been not rejected before. The letters in brackets, (L) and (R), indicate whether the left or right subsample was used.

Having repeated the test, it was possible to reverse the conclusion in seven series. However, splitting the samples in two led to small subsamples in five of those series, thus, it is necessary to be cautious before not rejecting the null hypothesis.

Nonetheless, there are some interesting non-rejections of the null, such as the Spanish debt, the US Debt to GDP ratio, and the Portuguese population suggesting that, for the subsample considered, innovations have a very persistent impact, which can be well described by the two-unit root model.

Table VII: *Tests for the Double Unit Root Null Hypothesis - Revised*

Variable	T	$\widehat{T_b}$	HF test	Model	$\sup_{\lambda \in \Lambda} F_{\hat{\rho}}^{i}(\lambda)$	Conclude
PIPCE	419	2/1973	14.72	A	18.65	H0 (1%)
PIPCECORE	419	2/1973	14.12	A	18.30	H0 (1%)
PIPCEDG	363 (R)	8/1980	14.74	B	20.24	H1
PIPCEF	372 (R)	9/1980	16.38	B	23.84	H1
CPIALL	298 (R)	8/2008	24.51	A	30.79	H1
CPICORE	420	2/1973	4.34	A	15.28	H0
TCCOUT	410 (L)	5/1950	23.36	B	29.23	H1
EMPLOYNF	490 (R)	11/1964	41.65	A	46.11	H1
M1	197 (R)	9/2008	8.14	A	16.83	H0
M2	255 (R)	5/1995	9.25	A	39.17	H1
PIGDP	169	1/1973	4.37	A	14.39	H0
PIPCEHOUSE	169	10/1973	6.03	A	13.94	H0
PIGPDI	123 (R)	Q1/1981	15.81	B	25.24	H1
TFD	122	Q1/1990	5.61	B	12.09	H0
PTGGOVDEBT	41	Q1/2012	4.94	B	11.19	H0
ESGGOVDEBT	68	Q1/2003	13.54	B	335.29	H1
GRGGOVDEBT*	68	Q4/2013	12.24	A	-	H0 (1%)
ITGGOVDEBT	32 (R)	Q1/2007	18.66	A	54.28	H1
USNRULR	164	Q1/1990	9.03	A	13.99	H0
USNRUSR	164	Q3/1975	12.60	B	17.73	H0 (2.5%)
ROUTM	60 (R)	Q2/2000	14.05	A	25.13	H0 (1%)
OUTNFB	166 (L)	Q1/1961	23.54	C	87.27	H1
COMPNFB	99 (L)	Q2/1965	12.08	B	24.39	H1
PTCPBNAS	38 (R)	Q3/1999	38.18	C	79.33	H1
PTCP	38 (R)	Q1/2000	45.66	B	62.55	H1
PTCPUB	53	Q2/2010	11.10	A	19.11	H0

USMFI	25 (R)	2000	12.66	B	21.55	H0
USRMFI	32 (R)	2000	20.23	B	24.25	H0 (1%)
USFDTOGDP	35 (R)	2008	11.75	A	14.83	H0
USFDHP	46	1975	16.65	A	21.93	H0 (2.5%)
USFDDSL	43	2000	1.68	B	9.20	H0
USFGSL	43	1995	9.36	C	18.74	H0
BEL	56	1998	7.11	B	16.19	H0
CAN	56	1998	14.09	A	18.25	H0
CHN	34 (R)	1974	13.99	B	61.26	H1
EMU	56	1974	10.44	B	17.42	H0
EUU	56	1976	11.01	B	16.88	H0
FIN	36 (R)	1969	14.01	A	61.20	H1
FRA	56	1973	11.43	B	17.15	H0
ITA	56	1989	3.47	B	21.12	H0
PRT	36 (R)	2006	6.29	C	20.60	H0
ESP	56	1993	11.54	A	15.64	H0
DE	56	1989	13.60	A	25.40	H0 (1%)
US	88	1947	3.72	C	26.44	H0 (2.5%)

9 Conclusions

Structural changes affect the limiting distribution of the OLS estimator, and can distort the inference based on non-robust methods. In this dissertation, we extended the test proposed in Hasza & Fuller (1979) for two-unit roots, to cover the case of a possible change in the deterministic function.

Two different approaches were developed, one that exploits information available about the break date, and other that assumes no prior knowledge about it. Under the known break date hypothesis, the test regression from the common Hasza-Fuller test was adapted to include the information available. The test statistic has a non-standard limiting distribution under the null hypothesis, which does not depend on nuisance parameters, and its critical values are obtained via Monte Carlo simulation. For the case of an unknown break date, a *supremum* approach was taken in order to pinpoint the date that gives the highest evidence for the alternative hypothesis. The test statistic has a non-standard limiting distribution, and is also free of nuisance parameters under the null hypothesis. However, due to the way the null hypothesis is formulated, the small sample distribution is perturbed by the break parameters. Nonetheless, this is only a problem under model A with integrated errors, and it quickly dissipates as the sample size grows large.

The tests were subject to a series of simulations in order to verify its robustness against several alternatives. Both tests are asymptotically consistent and have the correct size. One of the limitations of the test for an unknown break date is the excessive size distortions, in small samples, when facing a level shift under the null hypothesis. However, this problem vanishes asymptotically. This issue does not arise when there is a slope change under the null. Another issue emerges when there is more than one changepoint but, again only for model A and it vanishes with the increase of the sample size.

In section 8, both tests were applied to several time series which are suitable candidates to be explained by the two-unit root model. The null hypothesis was not rejected for 23 of those time series, with special attention to the Portuguese and Greek General Government debts. Other interesting results concern the natural rates of unemployment for the United States and for many of the population series. It seems that the double-unit root model is a suitable candidate to explain some major economic variables.

The tests proposed are F type of statistics, and thus, bilateral in nature. This means that additional power gains can be achieved by exploring unilateral t-tests, by following for example the Dickey & Pantula approach. However, their method is more complex since it requires changing the test regression in each step. Another issue that should be solved in the unknown changepoint test starts by admitting a trend break model even under the null hypothesis, to prevent excessive size distortions when there is a level shift and the error is integrated of order two. Finally, it would be interesting to develop tests that can accommodate multiple breaks.

References

Carrion-i-Silvestre, J. L., Kim, D., & Perron, P. (2009). GLS-based unit root tests with multiple structural breaks under both the null and the alternative hypotheses. *Econometric Theory*, 25(6), 1754-1792.

Cavaliere, G., Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2011). Testing for unit roots in the presence of a possible break in trend and nonstationary volatility. *Econometric Theory*, 27(5), 957-991.

Dickey, D. A., & Pantula, S. G. (1987). Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, 5(4), 455-461.

Haldrup, N. (1994). Semi-parametric Tests for Double Unit Roots. *Journal of Business & Economic Statistics*, 12, 109–122

Haldrup, N. (1998). A Review of the Econometric Analysis of I(2) Variables. *Journal of Economic Surveys*, 12, 595– 650.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton university press.

Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory*, 25(4), 995-1029.

Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2013). Testing for unit roots in the possible presence of multiple trend breaks using minimum Dickey–Fuller statistics. *Journal of Econometrics*, 177(2), 265-284.

Hasza, D. P., & Fuller, W. A. (1979). Estimation for autoregressive processes with unit roots. *The Annals of Statistics*, 1106-1120.

Kim, D., & Perron, P. (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics*, 148(1), 1-13.

Lumsdaine, R. L., & Papell, D. H. (1997). Multiple trend breaks and the unit-root hypothesis. *The review of economics and statistics*, *79*(2), 212-218.

Montañés, A. (1997). Level shifts, unit roots and misspecification of the breaking date. *Economics Letters,* 54(1), 7-13.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: some evidence and implications. *Journal of Monetary Economics*, 10(2), 139-162.

Patterson, K. (2011). *Unit root tests in time series volume 1: key concepts and problems*. Springer.

Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, 1361-1401.

Perron, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of econometrics*, *80*(2), 355-385.

Rodrigues, P. M., & Taylor, A. R. (2003). On Tests for Double Differencing: Some Extensions and the Role of Initial Values. *Documento de trabajo,* 23.

Sen, D. L., & Dickey, D. A. (1987). Symmetric test for second differencing in univariate time series. *Journal of Business & Economic Statistics,* 5(4), 463-473.

Shin, D.W. and Kim, H.J. (1999). Semiparametric Tests for Double Unit Roots Based on Symmetric Estimators. *Journal of Business & Economic Statistics*, 17, 67–73.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics,* 20(1), 25-44.

A Data Appendix

Table VIII: *Dataset Description*

Variable	Description	Freq.	Nº Obs.
PIPCE	USA - Personal Consumption Expenditure Price Index (2009=100)	M	699
PIPCECORE	USA - Personal Consumption Expenditure Price Index, less Food and Energy (2009=100)	M	699
PIPCEDG	USA - Personal Consumption Expenditure Price Index, Durable Goods (2009=100)	M	699
PIPCEF	USA - Personal Consumption Expenditure, Food (2009=100)	M	699
CPIALL	USA - Consumer Price Index, All Items (2010=100)	M	686
CPICORE	USA - Consumer Price Index, less Food and Energy (2010=100)	M	686
TCCOUT	USA - Total Consumer Credit Owned and Securitized, Outstanding Bn. $	M	891
EMPLOYNF	USA - Total Employment, Non-Farm	M	940
M1	USA - M1 Money Stock	M	700
M2	USA - M2 Money Stock	M	700
PIGDP	USA - Gross Domestic Product Price Index (2009=100)	Q	281
PIPCEHOUSE	USA - Personal Consumption Expenditure Price Index – Housing (2009=100)	Q	281
PIGPDI	USA - Gross Private Domestic Investment Price Index (2009=100)	Q	281
TFD	USA - Total Federal Debt, Million $ (no sa)	Q	204
PTGGOVDEBT	Portuguese General Government Debt, Million €	Q	68
ESGGOVDEBT	Spanish General Government Debt, Million €	Q	68
GRGGOVDEBT	Greek General Government Debt, Million €	Q	68
ITGGOVDEBT	Italian General Government Debt, Million €	Q	68
USNRULR	USA – Natural Rate of Unemployment, Long Run	Q	68
USNRUSR	USA – Natural Rate of Unemployment, Short Run	Q	68
ROUTM	USA - Real Output - Manufacturing	Q	121
OUTNFB	USA – Non Farm Business Sector – Real Output	Q	281
COMPNFB	USA – Non Farm Business Sector – Compensation per Hour	Q	281
PTCPBNAS	PT – Private Consumption. Services and non Food Goods	Q	88
PTCP	PT – Private Consumption	Q	88
PTCPUB	PT – Public Consumption	Q	88
USMFI	Mean Family Income – USA	A	63
USRMFI	Real Mean Family Income – USA	A	63

USFDTOGDP	Federal Debt % of GDP - USA	A	77
USFDHP	Federal Debt Held by the Public - USA	A	77
USFBDSL	Financial Businesses – Debt Securities and Liabilities - USA	A	72
USFGSL	Federal Government – Debt Securities and Liabilities – USA	A	72
BEL	Total Population, Belgium	A	56
CAN	Total Population, Canada	A	56
CHN	Total Population, China	A	56
EMU	Total Population, Euro Area (19)	A	56
EUU	Total Population, European Union (28)	A	56
FIN	Total Population, Finland	A	56
FRA	Total Population, France	A	56
ITA	Total Population, Italy	A	56
PRT	Total Population, Portugal	A	56
ESP	Total Population, Spain	A	56
DE	Total Population, Germany	A	56
US	Total Population, USA	A	56

B Mathematical Appendix

Proof of Theorem 1

For convenience, we write here the data generation process.

$$(A.1)\quad y_t = \boldsymbol{\beta}^{\boldsymbol{i}'}\boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{i}}(\boldsymbol{\lambda}) + x_t$$

$$(A.2)\quad (1-\alpha_1 L)(1-\alpha_2 L)x_t = e_t$$

$$(A.3)\quad e_t \sim^{iid} (0, \sigma^2)$$

We also write here the test equation:

$$(A.4)\quad \Delta^2 y_t = \mu^* + \delta^* t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$$

$$\Leftrightarrow y_t = \mu^* + \delta^* t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

, where $\phi_1 = \rho_1 + \rho_2 + 2$ and $\phi_2 = -\rho_2 - 1$.

Note that, we are estimating the test equation (A.4), as if the data generating process was a pure double unit root process, instead of the trend break model in (A.1). This can be understood as an omitted variables problem.

Part 1 – model A

Under the hypothesis $|\alpha_j| < 1$ for $j = 1,2$ for model A, we have $\boldsymbol{\beta}^{\boldsymbol{A}'} = (\mu, \delta, \mu_b)$ and $\boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{A}}(\boldsymbol{\lambda}) = (1, t, DU_t)'$.

Since a linear trend is present in the DGP, the regressors y_{t-k} $(k = 1, 2)$ and t will be asymptotically perfectly collinear, since the trend asymptotically dominates the remaining terms, thus we need to modify the test regression, by adding and subtracting $\phi_k[\mu^* + \delta^*(t-k)]$ for $k = 1, 2$. This transformation was suggested in Sims, Stock & Watson (1990) and is presented in detail in Hamilton, J. (1994).

Following with that transformation, we get:

$$(A.5)\quad y_t = \mu^{**} + \delta^{**} t + \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + e_t$$

Where:

$$\mu^{**} = \mu^*(1 + \phi_1 + \phi_2) - \delta^*(\phi_1 + 2\phi_2)$$

$$\delta^{**} = (1 + \phi_1 + \phi_2)$$

$$\tilde{y}_{t-k} = y_{t-k} - \mu^* - \delta^*(t-k)$$

$$k = 1,2$$

Note that, whilst we are transforming the test equation, $\tilde{y}_{t-k}$ is obtained from the DGP (A.1).

The transformation from (A.4) to (A.5) can be summarized in matrix form. Write (A.5) as $y_t = \boldsymbol{x}'_{\boldsymbol{t}}\boldsymbol{\beta} + e_t = \boldsymbol{x}'_{\boldsymbol{t}}\boldsymbol{H}'[\boldsymbol{H}']^{-\mathbf{1}}\boldsymbol{\beta} + e_t = \widetilde{\boldsymbol{x}}'_{\boldsymbol{t}}\widetilde{\boldsymbol{\beta}} + e_t$, where the "~" indicates the transformed vector of variables and parameters.

The matrix $\boldsymbol{H}'$ can be found by solving the system of equations:

$[\boldsymbol{H}']^{-\mathbf{1}}\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}$

Which yields $[\boldsymbol{H}']^{-\mathbf{1}} = \begin{bmatrix} 1 & 0 & \mu^* - \delta^* & \mu^* - 2\delta^* \\ 0 & 1 & \delta^* & \delta^* \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \boldsymbol{H}' =$
$\begin{bmatrix} 1 & 0 & \delta^* - \mu^* & 2\delta^* - \mu^* \\ 0 & 1 & -\delta^* & -\delta^* \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Now, the OLS estimation error of the transformed regression is given by:

$$(A.6)\quad \left(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}\right) = \left[\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \widetilde{\boldsymbol{x}}_{\boldsymbol{t}}[\widetilde{\boldsymbol{x}}_{\boldsymbol{t}}]'\right]^{-\mathbf{1}} \left[\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \widetilde{\boldsymbol{x}}_{\boldsymbol{t}} e_t\right] = \left[\boldsymbol{H}\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{t}}\boldsymbol{x}'_{\boldsymbol{t}}\right)\boldsymbol{H}'\right]^{-\mathbf{1}} \boldsymbol{H}\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{t}}\, e_t\right)$$

Consider the diagonal matrix $\boldsymbol{\Gamma} = \mathrm{diag}\left(\sqrt{T}, T^{\frac{3}{2}}, \sqrt{T}, \sqrt{T}\right)$.

Premultiply (A.6) by $\boldsymbol{\Gamma}$:

$$\boldsymbol{\Gamma}\left(\widehat{\widetilde{\boldsymbol{\beta}}} - \widetilde{\boldsymbol{\beta}}\right) = \left[\boldsymbol{\Gamma}^{-\mathbf{1}}\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \widetilde{\boldsymbol{x}}_{\boldsymbol{t}}[\widetilde{\boldsymbol{x}}_{\boldsymbol{t}}]'\right)\boldsymbol{\Gamma}^{-\mathbf{1}}\right]^{-\mathbf{1}} \left[\boldsymbol{\Gamma}^{-\mathbf{1}}\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \widetilde{\boldsymbol{x}}_{\boldsymbol{t}} e_t\right)\right]$$

The first term is equal to:

$$\boldsymbol{\Gamma}^{-\mathbf{1}}\left(\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \widetilde{\boldsymbol{x}}_{\boldsymbol{t}}[\widetilde{\boldsymbol{x}}_{\boldsymbol{t}}]'\right)\boldsymbol{\Gamma}^{-\mathbf{1}} =$$

$$\begin{bmatrix} 1 & T^{-2}\sum t & T^{-1}\sum \tilde{y}_{t-1} & T^{-1}\sum \tilde{y}_{t-2} \\ T^{-2}\sum t & T^{-3}\sum t^2 & T^{-2}\sum t\tilde{y}_{t-1} & T^{-2}\sum t\tilde{y}_{t-2} \\ T^{-1}\sum \tilde{y}_{t-1} & T^{-2}\sum t\tilde{y}_{t-1} & T^{-1}\sum \tilde{y}_{t-1}^2 & T^{-1}\sum \tilde{y}_{t-1}\tilde{y}_{t-2} \\ T^{-1}\sum \tilde{y}_{t-2} & T^{-2}\sum t\tilde{y}_{t-2} & T^{-1}\sum \tilde{y}_{t-1}\tilde{y}_{t-2} & T^{-1}\sum \tilde{y}_{t-2}^2 \end{bmatrix}$$

Where all summations go from 1 to T

The terms $T^{-2}\sum t$ and $T^{-3}\sum t^2$ converge to $1/2$ and $1/3$ respectively.

The terms evolving sums of $\tilde{y}_{t-k}$ can be shown to have the following limiting results:

$$T^{-1}\sum \tilde{y}_{t-i}\tilde{y}_{t-j} \to^p \gamma_{|i-j|} + \mu_b^2(1-\lambda),\ \ i,j = 1,2$$

$$T^{-1}\sum \tilde{y}_{t-k} \to^p \mu_b(1-\lambda),\ \ k = 1,2$$

$$T^{-1}\sum \tilde{y}_{t-k}^2 \to^p \gamma_0 + \mu_b^2(1-\lambda),\ \ k = 1,2$$

Finally, $T^{-2}\sum t\tilde{y}_{t-k} \to^p \frac{1}{2}\mu_b(1-\lambda)^2,\ \ k = 1, 2.$

Therefore, $\boldsymbol{\Gamma}^{-1}\left(\sum_{t=1}^{T}\tilde{\boldsymbol{x}}_t[\tilde{\boldsymbol{x}}_t]'\right)\boldsymbol{\Gamma}^{-1} \to^{p} \mathbf{V}$

Next, consider the vector $\left[\boldsymbol{\Gamma}^{-1}\left(\sum_{t=1}^{T}\tilde{\boldsymbol{x}}_t e_t\right)\right] = \boldsymbol{T}^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{\xi}_t$ where

$\boldsymbol{\xi}_t = \begin{bmatrix} e_t \\ \left({}^{t}/_{T}\right)e_t \\ \tilde{y}_{t-1}e_t \\ \tilde{y}_{t-2}e_t \end{bmatrix}$ which is a martingale difference sequence with covariance matrix given by $E[\boldsymbol{\xi}_t\boldsymbol{\xi}_t'] = \sigma^2\mathbf{V_t}$, where $T^{-1}\sum_{t=1}^{T}\mathbf{V_t} \to \mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} 1 & \frac{1}{2} & \mu_b(1-\lambda) & \mu_b(1-\lambda) \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{2}\mu_b(1-\lambda)^2 & \frac{1}{2}\mu_b(1-\lambda)^2 \\ \mu_b(1-\lambda) & \frac{1}{2}\mu_b(1-\lambda)^2 & \mu_b^2(1-\lambda)+\gamma_0 & \gamma_1+\mu_b^2(1-\lambda) \\ \mu_b(1-\lambda) & \frac{1}{2}\mu_b(1-\lambda)^2 & \gamma_1+\mu_b^2(1-\lambda) & \mu_b^2(1-\lambda)+\gamma_0 \end{bmatrix}$$

Thus, appealing to the Martingale Difference Sequence Central Limit Theorem, we have that $\boldsymbol{\Gamma}^{-1}\left(\sum_{t=1}^{T}\tilde{\boldsymbol{x}}_t e_t\right) \to^{d} \boldsymbol{N}(\mathbf{0}, \sigma^2\mathbf{V})$ and, applying Slutsky's theorem, it follows:

$$\boldsymbol{\Gamma}\left(\hat{\tilde{\boldsymbol{\beta}}} - \tilde{\boldsymbol{\beta}}\right) \to^{d} \boldsymbol{N}(\mathbf{0}, \boldsymbol{V}^{-1}\sigma^2\mathbf{V}\mathbf{V}^{-1}) = \boldsymbol{N}(\mathbf{0}, \sigma^2\mathbf{V}^{-1}).$$

Finally, recalling the transformation matrix $\boldsymbol{H}$ we find that the OLS estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ are identical in the transformed and untransformed regressions, hence their distribution is also the same. Therefore,

$$\sqrt{T}\left(\hat{\phi}_1 - \phi_1\right) \to^{d} N[0, \sigma^2\tau]$$

$$\sqrt{T}\left(\hat{\phi}_2 - \phi_2\right) \to^{d} N[0, \sigma^2\tau]$$

Where:

$$\tau = \frac{(-3\lambda^4 + 6\lambda^3 - 4\lambda^2 + \lambda)\mu_b^2 + \gamma_0}{(\gamma_0^2 - \gamma_1^2) + \left((12\lambda^3 - 6\lambda^4 - 8\lambda^2 + 2\lambda)\gamma_0 + (6\lambda^4 - 12\lambda^3 + 8\lambda^2 - 2\lambda)\gamma_1\right)\mu_b^2}$$

The final step links the estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ to $\hat{\rho}_1$ and $\hat{\rho}_2$, respectively, through the linear relation

$\begin{bmatrix}\rho_1 \\ \rho_2\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & -1\end{bmatrix}\begin{bmatrix}\phi_1 \\ \phi_2\end{bmatrix} + \begin{bmatrix}-1 \\ -1\end{bmatrix}$ which can be estimated by $\begin{bmatrix}\hat{\rho}_1 \\ \hat{\rho}_2\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & -1\end{bmatrix}\begin{bmatrix}\hat{\phi}_1 \\ \hat{\phi}_2\end{bmatrix} + \begin{bmatrix}-1 \\ -1\end{bmatrix}$.

Doing $\begin{bmatrix}\hat{\rho}_1 - \rho_1 \\ \hat{\rho}_2 - \rho_2\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & -1\end{bmatrix}\begin{bmatrix}\hat{\phi}_1 - \phi_1 \\ \hat{\phi}_2 - \phi_2\end{bmatrix}$ and multiplying both sides by $\sqrt{T}$, we obtain:

$\begin{bmatrix}\sqrt{T}(\hat{\rho}_1 - \rho_1) \\ \sqrt{T}(\hat{\rho}_2 - \rho_2)\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & -1\end{bmatrix}\begin{bmatrix}\sqrt{T}\left(\hat{\phi}_1 - \phi_1\right) \\ \sqrt{T}\left(\hat{\phi}_2 - \phi_2\right)\end{bmatrix}$, which in turn implies:

$\sqrt{T}(\hat{\rho}_1 - \rho_1) \to^d N[0, 2\sigma^2\tau]$

$\sqrt{T}(\hat{\rho}_2 - \rho_2) \to^d N[0, \sigma^2\tau]$

As intended.

Part 1 – models B and C

The proof for model A was based on a transformed regression, which helped to deal with the correlation between t, y_{t-1} and y_{t-2} in large samples. To prove the results for models B and C, that transformation is not enough, because even if we deal with the linear trend in the DGP, we are still left with the term $\delta_b DT_t$ which will still cause the perfect collinearity problem in large samples referred for model A, but this time between y_{t-1} and y_{t-2}. Thus, we start by transforming the regression (A.4) in the same way as before, but then proceed to transform the equation an additional time to remove the problem of the neglected slope shift.

Start with equation (A.5). Add and subtract $\phi_2 \tilde{y}_{t-1}$, which yields the following equation

$(A.7)\ \ y_t = \mu^* + \delta^* t + \phi_1^* \breve{y}_{t-1} + \phi_2^* \Delta\breve{y}_{t-1} + e_t = \breve{\boldsymbol{x}}_t' \breve{\boldsymbol{\beta}} + e_t$

Where $\phi_1^* = \phi_1 + \phi_2$ and $\phi_2^* = \phi_2$

Where $\breve{y}_{t-1} = \mu_b DU_{t-1} + \delta_b DT_{t-1} + x_{t-1}$ and $\Delta\breve{y}_{t-1} = \mu_b \Delta DU_{t-1} + \delta_b DU_{t-1} + \Delta x_{t-1}$.

This time, the matrix $\boldsymbol{H}'$ that solves $[\boldsymbol{H}']^{-1}\boldsymbol{\beta} = \breve{\boldsymbol{\beta}}$ is given by:

$$[\boldsymbol{H}']^{-1} = \begin{bmatrix} 1 & 0 & \mu^* - \delta^* & \mu^* - 2\delta^* \\ 0 & 1 & \delta^* & \delta^* \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \boldsymbol{H}' = \begin{bmatrix} 1 & 0 & \delta^* - \mu^* & \delta^* \\ 0 & 1 & -\delta^* & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

And, like before, the OLS estimation error is given by

$$(A.8)\ \ \left(\hat{\breve{\boldsymbol{\beta}}} - \breve{\boldsymbol{\beta}}\right) = \left[\sum_{\boldsymbol{t}=\boldsymbol{1}}^{\boldsymbol{T}} \breve{\boldsymbol{x}}_{\boldsymbol{t}} [\breve{\boldsymbol{x}}_{\boldsymbol{t}}]'\right]^{-\boldsymbol{1}} \left[\sum_{\boldsymbol{t}=\boldsymbol{1}}^{\boldsymbol{T}} \breve{\boldsymbol{x}}_{\boldsymbol{t}} e_t\right] = \left[\boldsymbol{H}\left(\sum_{\boldsymbol{t}=\boldsymbol{1}}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{t}} \boldsymbol{x'}_{\boldsymbol{t}}\right)\boldsymbol{H}'\right]^{-\boldsymbol{1}} \boldsymbol{H}\left(\sum_{\boldsymbol{t}=\boldsymbol{1}}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{t}}\, e_t\right)$$

$$= \boldsymbol{H}^{-\boldsymbol{1}}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$$

Before proceeding, we need to check the order of convergence of the following moments:

Lemma A.1

$$\sum_{t=1}^{T}\breve{y}_t=\sum_{t=1}^{T}\mu_b DU_{t-1}+\delta_b DT_{t-1}+x_{t-1}$$
$$=\mu_b T(1-\lambda)+\frac{1}{2}\delta_b T(1-\lambda)(T-\lambda T+1)+\sqrt{T}\left[T^{-\frac{1}{2}}\sum_{t=1}^{T}x_{t-1}\right]$$

$$\sum_{t=1}^{T}\breve{y}_t^2=\frac{1}{6}T(1-\lambda)[T^2(2\lambda^2\delta_b-4\lambda\delta+2\lambda)+T(3\delta_b^2+6\delta_b\mu_b-3\lambda\delta_b-6\lambda\mu_b\delta_b)$$
$$+(6\mu_b^2-6\mu_b\delta_b+\delta_b^2)]+2\mu_b\sqrt{T}\left[T^{-\frac{1}{2}}\sum_{t=\lambda T+1}^{T}x_{t-1}\right]$$
$$+2\delta_b T^{\frac{3}{2}}\left[T^{-\frac{3}{2}}\sum_{t=\lambda T+1}^{T}(t-\lambda T)x_{t-1}\right]+T\left[T^{-1}\sum_{t=1}^{T}x_{t-1}^2\right]$$

$$\sum_{t=1}^{T}t\breve{y}_t=\frac{1}{6}T(1-\lambda)[T^2(2\delta_b-\lambda^2\delta_b-\lambda\delta_b)+T(3\mu_b(1+\lambda)+3\delta_b)+\delta_b+3\mu_b]$$
$$+T^{\frac{3}{2}}\left[T^{-\frac{3}{2}}\sum_{t=1}^{T}tx_{t-1}\right]$$

$$\sum_{t=1}^{T}\breve{y}_t\widetilde{\Delta y}_t=\mu_b^2T(1-\lambda)+\frac{1}{2}\delta_b^2T(1-\lambda)(T-\lambda T+1)+\mu_b\sqrt{T}\left[T^{-\frac{1}{2}}\sum_{t=\lambda T+1}^{T}\Delta x_{t-1}\right]$$
$$+\delta_b T^{\frac{3}{2}}\left[T^{-\frac{3}{2}}\sum_{t=\lambda T+1}^{T}(\mathrm{t}-\lambda T)\Delta x_{t-1}\right]+\delta_b\sqrt{T}\left[T^{-\frac{1}{2}}\sum_{t=\lambda T+1}^{T}x_{t-1}\right]$$
$$+T\left[T^{-1}\sum_{t=1}^{T}x_{t-1}\Delta x_{t-1}\right]$$

Therefore, the scaling matrix is given by $\boldsymbol{\Gamma}=\mathrm{diag}\left(\sqrt{T},T^{\frac{3}{2}},T^{\frac{3}{2}},\sqrt{T}\right)$. Premultiplying (A.8) by $\boldsymbol{\Gamma}$ yields:

$$\boldsymbol{\Gamma}\left(\widehat{\breve{\boldsymbol{\beta}}}-\breve{\boldsymbol{\beta}}\right)=\left[\boldsymbol{\Gamma}^{-1}\left(\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t[\breve{\boldsymbol{x}}_t]'\right)\boldsymbol{\Gamma}^{-1}\right]^{-1}\left[\boldsymbol{\Gamma}^{-1}\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t e_t\right]$$

The first term $\boldsymbol{\Gamma}^{-1}\left(\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t[\breve{\boldsymbol{x}}_t]'\right)\boldsymbol{\Gamma}^{-1}$ is given by:

$$\boldsymbol{\Gamma^{-1}}\left(\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t[\breve{\boldsymbol{x}}_t]'\right)\boldsymbol{\Gamma^{-1}}$$
$$=\begin{bmatrix} 1 & T^{-2}\sum t & T^{-2}\sum \breve{y}_{t-1} & T^{-1}\sum \Delta\breve{y}_{t-1} \\ T^{-2}\sum t & T^{-3}\sum t^2 & T^{-3}\sum t\breve{y}_{t-1} & T^{-2}\sum t\Delta\breve{y}_{t-1} \\ T^{-2}\sum \breve{y}_{t-1} & T^{-3}\sum t\breve{y}_{t-1} & T^{-3}\sum \breve{y}_t^2 & T^{-2}\sum \breve{y}_{t-1}\widetilde{\Delta y}_{t-1} \\ T^{-1}\sum \Delta\breve{y}_{t-1} & T^{-2}\sum t\Delta\breve{y}_{t-1} & T^{-2}\sum \breve{y}_{t-1}\widetilde{\Delta y}_{t-1} & T^{-1}\sum \Delta\breve{y}_t^2 \end{bmatrix}$$

Where all summations run from 1 to T.

The elements $T^{-2}\sum t$ and $T^{-3}\sum t^2$ converge to $1/2$ and $1/3$, respectively.

As for the terms involving sums of $\Delta\breve{y}_{t-1}$ and $\breve{y}_{t-1}$, using Lemma A.1 and the appropriate scaling factor, it can be shown that they converge to the following:

$$T^{-1}\sum \Delta\breve{y}_{t-1} \rightarrow \delta_b(1-\lambda)$$

$$T^{-1}\sum \Delta\breve{y}_t^2 \rightarrow \delta_b^2(1-\lambda) + 2(\gamma_0 - \gamma_1)$$

$$T^{-2}\sum t\Delta\breve{y}_{t-1} \rightarrow \frac{1}{2}\delta_b(1-\lambda)^2$$

$$T^{-2}\sum \breve{y}_{t-1} \rightarrow \frac{1}{2}\delta_b(1-\lambda)^2$$

$$T^{-3}\sum \breve{y}_t^2 \rightarrow \frac{1}{6}(1-\lambda)(\delta_b(2\lambda^2-4\lambda)+2\lambda)$$

$$T^{-2}\sum \breve{y}_{t-1}\widetilde{\Delta y}_{t-1} \rightarrow \frac{1}{2}\delta_b^2(1-\lambda)^2$$

$$T^{-2}\sum t\breve{y}_{t-1} \rightarrow \frac{1}{6}(1-\lambda)(\delta_b(2\lambda^2-4\lambda)+2\lambda)$$

Hence, $\boldsymbol{\Gamma^{-1}}\left(\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t[\breve{\boldsymbol{x}}_t]'\right)\boldsymbol{\Gamma^{-1}} \longrightarrow \mathbf{V}$

The second term given by $\left[\boldsymbol{\Gamma^{-1}}\sum_{t=1}^{T}\breve{\boldsymbol{x}}_t e_t\right] = \boldsymbol{T}^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{\xi_t}$ where $\boldsymbol{\xi_t} = \begin{bmatrix} e_t \\ (t/T)e_t \\ \tilde{y}_{t-1}e_t \\ \widetilde{\Delta y}_{t-1}e_t \end{bmatrix}$, is a martingale difference sequence, with covariance matrix $\sigma^2\mathbf{V}_t$, where $T^{-1}\sum_{t=1}^{T}\mathbf{V_t} \rightarrow \mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2}\delta_b(1-\lambda)^2 & \delta_b(1-\lambda) \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{2}\delta_b(1-\lambda)^2 \\ \frac{1}{2}\delta_b(1-\lambda)^2 & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{6}(1-\lambda)(2\lambda^2\delta_b - 4\lambda\delta + 2\lambda) & \frac{1}{2}\delta_b^2(1-\lambda)^2 \\ \delta_b(1-\lambda) & \frac{1}{2}\delta_b(1-\lambda)^2 & \frac{1}{2}\delta_b^2(1-\lambda)^2 & \delta_b^2(1-\lambda) + 2(\gamma_0 - \gamma_1) \end{bmatrix}$$

Which does not depend on μ_b.

Hence, $\mathbf{\Gamma^{-1}}\left(\sum_{t=1}^{T} \breve{\boldsymbol{x}}_t e_t\right) \rightarrow^{\boldsymbol{d}} \boldsymbol{N}(\mathbf{0}, \sigma^2\mathbf{V})$**,** and applying Slutsky's theorem, it follows $\mathbf{\Gamma}\left(\widehat{\breve{\boldsymbol{\beta}}} - \breve{\boldsymbol{\beta}}\right) \rightarrow^{\boldsymbol{d}} \boldsymbol{N}[\mathbf{0}, \sigma^2\mathbf{V^{-1}}]$.

Therefore:

$T^{\frac{3}{2}}\left(\hat{\phi}_1^* - \phi_1^*\right) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_1]$

$\sqrt{T}\left(\hat{\phi}_2^* - \phi_2^*\right) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_2]$

Next, consulting $[\boldsymbol{H}']^{\mathbf{-1}} = \begin{bmatrix} 1 & 0 & \mu^* - \delta^* & \mu^* - 2\delta^* \\ 0 & 1 & \delta^* & \delta^* \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, we find that the linear transformation need to go from ϕ_1^* and ϕ_2^* to ρ_1 and ρ_2, respectively is given by $\begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \phi_1^* \\ \phi_2^* \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ which is estimated by $\begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1^* \\ \hat{\phi}_2^* \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

Writing $\begin{bmatrix} \hat{\rho}_1 - \rho_1 \\ \hat{\rho}_2 - \rho_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \hat{\phi}_1^* - \phi_1^* \\ \hat{\phi}_2^* - \phi_2^* \end{bmatrix}$ and premultiplying by $\begin{bmatrix} T^{\frac{3}{2}} \\ \sqrt{T} \end{bmatrix}$, we find that the limiting distribution is left unchanged, hence:

$T^{\frac{3}{2}}(\hat{\rho}_1 - \rho_1) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_1]$

$\sqrt{T}(\hat{\rho}_2 - \rho_2) \rightarrow^{\boldsymbol{d}} N[0, \sigma^2\theta_2]$

Where

$\theta_1 = [\boldsymbol{V}]^{-1}{}_{3,3}$ (Element 3,3 of the respective matrix)

$\theta_2 = [\boldsymbol{V}]^{-1}{}_{4,4}$ (Element 4,4 of the respective matrix)

Which completes the proof.

Part 2 – models B and C

Under the alternative hypothesis of exactly one unit root, the error process given in equation $(A.2)$ can be rewritten as $x_t = x_{t-1} + \zeta\Delta x_{t-1} + e_t \Leftrightarrow \Delta x_{t-1} = \psi(L)e_t = u_t \Leftrightarrow x_t = \sum_{j=1}^{t} u_j$,

Now, consider model A. If x_t has a unit root, then the stochastic trend $\sum_{j=1}^{t} u_j$ asymptotically dominates the level shift, thus, the limiting distribution of the OLS estimator under model A does not depend on the break parameter or the break fraction.

As for models B and C, let's consider the data generating process,

$(A.9)\quad y_t = \delta_b DT_t + x_t$

$(A.10)\quad x_t = \sum_{j=1}^{t} u_j,\quad x_0 = 0$

We estimate the equation

$(A.11)\quad \Delta^2 y_t = \mu^* + \delta^* t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$

$\Leftrightarrow y_t = \mu^* + \delta^* t + \omega_1 y_{t-1} + \omega_2 \Delta y_{t-1} + e_t$

$\Leftrightarrow y_t = \boldsymbol{dD_t} + \omega_1 y_{t-1} + \omega_2 \Delta y_{t-1} + e_t$

Where:

$\omega_1 = \rho_1 + 1, \omega_2 = -\rho_2 - 1, \boldsymbol{D_t} = (1,t)', \boldsymbol{d} = (\mu^*, \delta^*)$

Then, apply the Frisch-Waugh-Lovell theorem and estimate the equation bellow, where the "~" indicates the residuals from the regression of that variable on a constant and a linear trend. The OLS estimators $\widehat{\omega}_1$ and $\widehat{\omega}_2$ are numerically equal in (A.11) and (A.12).

$(A.12)\quad \tilde{y}_t = \omega_1 \tilde{y}_{t-1} + \omega_2 \widetilde{\Delta y}_{t-1} + e_t$

The OLS estimation error is then,

$$(A.13)\quad \begin{bmatrix} \widehat{\omega_1} - \omega_1 \\ \widehat{\omega_2} - \omega_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} \tilde{y}_{t-1}^2 & \sum_{t=1}^{T} \tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1} \\ \sum_{t=1}^{T} \tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1} & \sum_{t=1}^{T} \widetilde{\Delta y}_{t-1}^2 \end{bmatrix}^{\boldsymbol{-1}} \begin{bmatrix} \sum_{t=1}^{T} \tilde{y}_{t-1} e_t \\ \sum_{t=1}^{T} \widetilde{\Delta y}_{t-1} e_t \end{bmatrix}$$

To prove the limiting distribution of the OLS estimator above, we will make use of the step function $X_T(r)$:

$$(A.14)\quad X_T(r) = T^{-1} \sum_{j=1}^{[Tr]} u_j$$

Appealing to Donsker's invariance theorem, under the assumptions made for the error sequence e_t (independence and identically distributed), the following holds:

$$(A.15)\quad \frac{1}{\sqrt{T}}\left[\frac{X_T(\cdot)}{\sigma}\right] \to^d W(\cdot)$$

Where $W(\cdot)$ is a standard Brownian motion.

The connection between the sample moments and the function $X_T(r)$ is given by:

$$(A.16)\quad X_T(r) = T^{-\frac{1}{2}}\sigma^{-1}S_{[Tr]}$$

The objective is now to express the sample moments, as functions of $X_T(r)$, σ and δ_b, in the following way. Take for example:

$$(A.17)\quad \sum_{t=1}^{T}\tilde{y}_{t-1}e_t$$

$$= \sum_{t=1}^{T}\left\{\delta_b DT_{t-1}e_t + x_{t-1}e_t - e_t\boldsymbol{D}_t'\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}_t\boldsymbol{D}_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}_t(\delta_b DT_{t-1} + x_{t-1})\right\}$$

$$= \underbrace{\sum_{t=\lambda T+1}^{T}\delta_b(t-1-\lambda T)e_t}_{O_p\left(T^{\frac{3}{2}}\right)} + \underbrace{\sum_{t=1}^{T}x_{t-1}e_t}_{O_p(T)} - \underbrace{\sum_{t=1}^{T}e_t\boldsymbol{D}_t'\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}_t\boldsymbol{D}_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}_t\, x_{t-1}}_{O_p(T)}$$

$$\underbrace{-\sum_{t=1}^{T}e_t\boldsymbol{D}_t'\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{D}_t\boldsymbol{D}_t'\right)^{-1}\frac{1}{T}\sum_{t=\lambda T+1}^{T}\boldsymbol{D}_t\delta_b\,(t-1-\lambda T)}_{O_p\left(T^{\frac{3}{2}}\right)}$$

Therefore,

$$(A.18)\quad T^{-\frac{3}{2}}\sum_{t=1}^{T}\tilde{y}_{t-1}e_t$$

$$= \delta_b\sigma\left(X_T(1) - X_T(\lambda) - \int_\lambda^1 X_T(r)dr\right) - \lambda\delta_b\sigma\big(X_T(1) - X_T(\lambda)\big)$$

$$- \begin{bmatrix}\sigma X_T(1) & \sigma\left(X_T(1) - \int_0^1 X_T(r)dr\right)\end{bmatrix}\begin{bmatrix}-\delta_b\dfrac{\lambda-1}{T(T-1)}[T^2(\lambda^2-\lambda) + T(2\lambda-1)+1] \\ \delta_b\dfrac{\lambda-1}{T^2-1}[T^2(2\lambda^2-\lambda-1)+3\lambda T+1)]\end{bmatrix}$$

$$+ o_p(1)$$

Letting T grow large, it follows:

$$(A.19)\quad T^{-\frac{3}{2}}\sum_{t=1}^{T}\tilde{y}_{t-1}e_t \to^d \sigma\delta_b(\lambda^3-\lambda^2-2\lambda+2)W(1)-\sigma\delta_b(1-\lambda)W(\lambda) - \sigma\delta_b\left(\int_\lambda^1 W(r)dr+(\lambda-1)^2(1+2\lambda)\int_0^1 W(r)dr\right)$$

Likewise, $\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}e_t$ can be decomposed as:

$$(A.20)\quad \sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}e_t$$

$$= \underbrace{\sum_{\boldsymbol{t=[\lambda T]+1}}^{T}\delta_b e_t}_{O_p(\sqrt{T})} + \underbrace{\sum_{t=1}^{T}u_{t-1}e_t}_{O_p(\sqrt{T})} \underbrace{-\sum_{t=1}^{T}e_t\boldsymbol{D}_t'\left(\frac{1}{T}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}\boldsymbol{D}_t\boldsymbol{D}_t'\right)^{\boldsymbol{-1}}\frac{1}{T}\sum_{\boldsymbol{t=[\lambda T]+1}}^{\boldsymbol{T}}\boldsymbol{D}_t\delta_b}_{O_p(\sqrt{T})}$$

$$\underbrace{-\sum_{t=1}^{T}e_t\boldsymbol{D}_t'\left(\frac{1}{T}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}\boldsymbol{D}_t\boldsymbol{D}_t'\right)^{\boldsymbol{-1}}\frac{1}{T}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}\boldsymbol{D}_t u_{t-1}}_{o_p(1)}$$

Consider the term $\sum_{t=1}^{T}u_{t-1}e_t$. When multiplied by $\frac{1}{\sqrt{T}}$, it is equal to $\sqrt{T}$ times the sample mean of a martingale difference sequence whose variance is given by:

$E(u_{t-1}e_t)^2 = E(u_{t-1}^2e_t^2) = \sigma^2\gamma_0$, where $\gamma_0 = \lim_{T\to\infty}E(u_{t-1}^2)$.

Thus, it satisfies the usual central limit theorem for martingale difference sequences.

$$(A.21)\quad \sqrt{T}\sum_{t=1}^{T}u_{t-1}e_t \to^d N(0,\sigma^2\gamma_0)=\sigma\sqrt{\gamma_0}W(1)$$

Therefore,

$$(A.22)\quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}e_t = \delta_b\sigma\big(X_T(1)-X_T(\lambda)\big)+\sigma\sqrt{\gamma_0}\,X_T(1)$$

$$-\left[\sigma X_T(1)\quad \sigma\left(X_T(1)-\int_0^1 X_T(r)dr\right)\right]\begin{bmatrix}\delta_b\dfrac{\lambda-1}{T-1}[T(3\lambda-1)+1]\\ -6\delta_b\lambda T^2\dfrac{\lambda-1}{T^2-1}\end{bmatrix}$$

Again, letting T grow large, this weakly converges to:

$$(A.23)\quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}e_t \to^d \sigma\big(\delta_b(2+2\lambda-3\lambda^2)-\sqrt{\gamma_0}\,\big)W(1)-\delta_b\sigma W(\lambda) - 6\lambda(1-\lambda)\delta_b\sigma\int_0^1 W(r)dr$$

Proceeding in the same way as before, we obtain that:

$$(i)\ T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2} \longrightarrow^{p} \frac{1}{3}\delta_b^2(1-\lambda)^3$$

$$(ii)\ T^{-1}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{2} \longrightarrow^{p} -\lambda\delta_b^2(3\lambda^3-6\lambda^2+4\lambda-1)+\gamma_0$$

$$(iii)\ T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1} \longrightarrow^{p} \frac{1}{2}\lambda^2\delta_b^2(\lambda-1)^2(2\lambda-1)$$

Next, define the (2×2) diagonal matrix $\mathbf{\Gamma}_{\boldsymbol{T}} = diag\left(T^{-\frac{3}{2}}, T^{-\frac{1}{2}}\right)$. Then multiplying the OLS estimator in $(A.13)$ by $\mathbf{\Gamma}_{\boldsymbol{T}}^{-\mathbf{1}}$ yields:

$$(A.35)\quad \begin{bmatrix} T^{\frac{3}{2}}(\widehat{\omega_1}-\omega_1) \\ T^{\frac{1}{2}}(\widehat{\omega_2}-\omega_2) \end{bmatrix}$$

$$= \begin{bmatrix} T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^{2} & T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1} \\ T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1} & T^{-1}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{2} \end{bmatrix}^{-\mathbf{1}} \begin{bmatrix} T^{-\frac{3}{2}}\sum_{t=1}^{T}\tilde{y}_{t-1}e_t \\ T^{-\frac{1}{2}}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}e_t \end{bmatrix}$$

Finally, using $(A.29), (A.34)$ and Lemma A.3, we have:

$$(A.24)\quad T^{\frac{3}{2}}(\widehat{\omega_1}-\omega_1) \rightarrow^{d} \theta^{-1}(\mathrm{D}_1-\mathrm{D}_2)$$

$$(A.25)\quad T^{\frac{1}{2}}(\widehat{\omega_2}-\omega_2) \rightarrow^{d} \theta^{-1}(\mathrm{D}_3-\mathrm{D}_4)$$

Where

$$(i)\ \theta = \frac{1}{12}\delta_b^2(\lambda-1)^3\left(4\gamma_0-\lambda\delta_b^2(12\lambda^6-24\lambda^5+15\lambda^4-15\lambda^3+24\lambda^2-16\lambda+4)\right)$$

$$(ii)\ \mathrm{G}_1 = \left(\gamma_0-\lambda\delta_b^2(3\lambda^3-6\lambda^2+4\lambda-1)\right) \times\Bigg(\sigma\delta_b(\lambda^3-\lambda^2-2\lambda+2)W(1)-\sigma\delta_b(1-\lambda)W(\lambda) - \sigma\delta_b\left(\int_\lambda^1 W(r)dr+(\lambda-1)^2(1+2\lambda)\int_0^1 W(r)dr\right)\Bigg)$$

$$(iii)\ \mathrm{G}_2 = \frac{1}{2}\lambda^2\delta_b^2(\lambda-1)^2(2\lambda-1)$$
$$\times\left(\sigma\left(\delta_b(2+2\lambda-3\lambda^2)-\sqrt{\gamma_0}\,\right)W(1)-\delta_b\sigma W(\lambda)-6\lambda(1-\lambda)\delta_b\sigma\int_0^1 W(r)dr\right)$$

$$(iv)\ \mathrm{G}_3 = \frac{1}{6}\delta_b^2(\lambda-1)^2(3\lambda^2-6\lambda^3)$$
$$\times\left(\sigma\delta_b(\lambda^3-\lambda^2-2\lambda+2)W(1)-\sigma\delta_b(1-\lambda)W(\lambda)\right.$$
$$\left.-\sigma\delta_b\left(\int_\lambda^1 W(r)dr+(\lambda-1)^2(1+2\lambda)\int_0^1 W(r)dr\right)\right)$$

$$(v)\ \mathrm{G}_4 = \frac{1}{6}\delta_b^2(\lambda-1)^2(3\lambda^2-6\lambda^3)$$
$$\times\left(\sigma\left(\delta_b(2+2\lambda-3\lambda^2)-\sqrt{\gamma_0}\,\right)W(1)-\delta_b\sigma W(\lambda)-6\lambda(1-\lambda)\delta_b\sigma\int_0^1 W(r)dr\right)$$

Since, $\rho_1 = \omega_1 - 1$ and $\rho_2 = \omega_2 - 1$, we verify that $T^{\frac{3}{2}}(\widehat{\omega}_1 - \omega_1) = T^{\frac{3}{2}}(\widehat{\omega}_1 - 1 + 1 - \omega_1) = T^{\frac{3}{2}}((\widehat{\omega}_1 - 1) - (\omega_1 - 1)) = T^{\frac{3}{2}}(\hat{\rho}_1 - \rho_1)$. Using the same argument, we verify $\sqrt{T}(\widehat{\omega}_2 - \omega_2) = \sqrt{T}(\hat{\rho}_2 - \rho_2)$, meaning that the distribution is left unchanged by this transformation, and the desired result follows.

Proof of Theorem 2

We start by remembering the data generating process.

$$y_t = \boldsymbol{\beta^{i\prime}} \boldsymbol{z_t^i}(\lambda) + x_t$$

Where $\boldsymbol{z_t^i}(\lambda)$ are as defined in equations (4) and,

$$\boldsymbol{\beta^{A\prime}} = (\mu^A, \delta^A, \mu_b^A)', \quad \boldsymbol{\beta^{B\prime}} = (\mu^B, \delta^B, \delta_b^B)', \quad \boldsymbol{\beta^{C\prime}} = (\mu^C, \delta^C, \mu_b^C, \delta_b^C)'$$

The test equation is given by:

$(A.26)$

$$\Delta^2 y_t = \mu^* + \delta^* t + \theta_1 DT_t + \theta_2 DU_t + \theta_3 \Delta DU_t + \theta_4 \Delta^2 DU_t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$$

$t = 3, \ldots T, i = A, B, C$

But, as mentioned in the text, since ΔDU_t and $\Delta^2 DU_t$ are asymptotically negligible, we derive the limiting results from the equation below:

$$\Delta^2 y_t = \mu^* + \delta^* t + \theta_1 DT_t + \theta_2 DU_t + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$$

$$\Leftrightarrow \Delta^2 y_t = \boldsymbol{\pi}^{i'} \boldsymbol{z_t^i}(\lambda) + \rho_1 y_{t-1} + \rho_2 \Delta y_{t-1} + e_t$$

Where $\boldsymbol{z_t^i}(\lambda)$ is given in equation (4) and $\boldsymbol{\pi}^i = (\mu^*, \delta^*, \theta_1, \theta_2)'$.

Next, designate with a "~" the residuals of the projection of the indicated variable onto the space spanned by $\boldsymbol{z_t^i}(\lambda)$. For instance:

$$(A.27) \quad \tilde{y}_{t-1}^i = y_{t-1}^i - \boldsymbol{z_t^i}(\lambda)' \left(\sum_{t=1}^{T} \boldsymbol{z_t^i}(\lambda) \boldsymbol{z_t^i}(\lambda)' \right)^{-1} \sum_{t=1}^{T} \boldsymbol{z_t^i}(\lambda)\, y_{t-1}^i$$

Additionally, define the diagonal matrix $\mathbf{K}_{\boldsymbol{T}}^{\boldsymbol{i}}$, $s.t$ $\mathbf{K}_{\boldsymbol{T}}^{\boldsymbol{i}} \boldsymbol{z_t^i}(\lambda) \rightarrow \boldsymbol{Z^i}(\lambda, r)$. For instance, for model A, $\mathbf{K}_{\boldsymbol{T}}^{\boldsymbol{A}} = diag(1, T^{-1}, 1)$.

The next lemma provides the limiting representations for the sample moments.

Lemma A.2

$$(i) \quad T^{-\frac{3}{2}} y_{t-1} \rightarrow^d \sigma[rV(r) + V_r(r) + W_r(r)] = \sigma \Upsilon_1$$

$$(ii) \quad T^{-\frac{1}{2}} \Delta y_{t-1} \rightarrow^d \sigma[V(r) + W(r)] = \sigma \Upsilon_2$$

$$(iii) \quad T^{-2} \sum_{t=1}^{T} \tilde{y}_{t-1}^i e_t \rightarrow^d \sigma^2 \xi_1^i(\lambda)$$

$$\xi_1^i(\lambda) = \left\{ \int_0^1 \Upsilon_1 dW(r) - \int_0^1 \boldsymbol{Z^i}(\lambda, r)' dW(r) \left(\int_0^1 \boldsymbol{Z^i}(\lambda, r) \boldsymbol{Z^i}(\lambda, r)' dr \right)^{-1} \int_0^1 \boldsymbol{Z^i}(\lambda, r)\, \Upsilon_1 dr \right\}$$

$$(iv)\ \ T^{-1}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{i}e_t \rightarrow^d \sigma^2\xi_2^i(\lambda)$$

$$\xi_2^i(\lambda)=\left\{\int_0^1 \Upsilon_2 dW(r)-\int_0^1 \boldsymbol{Z^i}(\lambda,r)'dW(r)\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\}$$

$$(v)\ \ T^{-4}\sum_{t=1}^{T}\left(\tilde{y}_{t-1}^{i}\right)^2 \rightarrow^d \sigma^2\eta_1^i(\lambda)$$

$$\eta_1^i(\lambda)=\int_0^1\left\{\Upsilon_1-\boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_1 dr\right\}^2 dr$$

$$(vi)\ \ T^{-2}\sum_{t=1}^{T}\left(\widetilde{\Delta y}_{t-1}^{i}\right)^2 \rightarrow^d \sigma^2\eta_2^i(\lambda)$$

$$\eta_2^i(\lambda)=\int_0^1\left\{\Upsilon_2-\boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\}^2 dr$$

$$(vii)\ \ T^{-3}\sum_{t=1}^{T}\left(\tilde{y}_{t-1}^{i}\widetilde{\Delta y}_{t-1}^{i}\right) \rightarrow^d \sigma^2\eta_3^i(\lambda)$$

$$\eta_3^i(\lambda)=\int_0^1\left\{\Upsilon_1-\boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_1 dr\right\}$$
$$\times\left\{\Upsilon_2-\boldsymbol{Z^i}(\lambda,r)'\left(\int_0^1 \boldsymbol{Z^i}(\lambda,r)\boldsymbol{Z^i}(\lambda,r)'dr\right)^{-1}\int_0^1 \boldsymbol{Z^i}(\lambda,r)\,\Upsilon_2 dr\right\}dr$$

Both (i) and (ii) are derived in Taylor & Rodrigues (2004).

The proof can be done with (i) and (ii) and by expressing the sample moments as functions of partial sums that weakly converge to functions of Wiener processes, and functions of the deterministic vector $\mathbf{K}_{\boldsymbol{T}}^{\boldsymbol{i}}\boldsymbol{z}_{\boldsymbol{t}}^{\boldsymbol{i}}(\lambda)=\boldsymbol{Z_t}$.

Finally, under the null hypothesis, the test statistic of interest is given by:

$$(A.28)\quad F_{\hat{\rho}}^{i}(\lambda) =$$

$$(2s^2)^{-1}\left[\textstyle\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}e_t \quad \sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{i}e_t\right]\begin{bmatrix}\sum_{t=1}^{T}\left(\tilde{y}_{t-1}^{i}\right)^2 & \sum_{t=1}^{T}\tilde{y}_{t-1}^{i}\widetilde{\Delta y}_{t-1}^{i}\\ \sum_{t=1}^{T}\tilde{y}_{t-1}^{i}\widetilde{\Delta y}_{t-1}^{i} & \sum_{t=1}^{T}\left(\widetilde{\Delta y}_{t-1}^{i}\right)^2\end{bmatrix}\begin{bmatrix}\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}e_t\\ \sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{i}e_t\end{bmatrix}$$

Let $\boldsymbol{D}_T^{\boldsymbol{i}} = diag(T^{-2}, T^{-1})$ be a diagonal scaling matrix. Then:

$$(A.29)\quad F_{\hat{\rho}}^{i}(\lambda) = (2s^2)^{-1}\left[T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}e_t \quad T^{-1}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{i}e_t\right]\times$$

$$\times\begin{bmatrix}T^{-4}\sum_{t=1}^{T}\left(\tilde{y}_{t-1}^{i}\right)^2 & T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}\widetilde{\Delta y}_{t-1}^{i}\\ T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}\widetilde{\Delta y}_{t-1}^{i} & T^{-2}\sum_{t=1}^{T}\left(\widetilde{\Delta y}_{t-1}^{i}\right)^2\end{bmatrix}\begin{bmatrix}T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^{i}e_t\\ T^{-1}\sum_{t=1}^{T}\widetilde{\Delta y}_{t-1}^{i}e_t\end{bmatrix}$$

Therefore, using Lemma A.2 and the fact that $s^2 \to^p \sigma^2$, after simple algebra, expression $(A.29)$ has the following distributional limit.

$$(A.30)\quad F_{\hat{\rho}}^{i}(\lambda) \to^d \frac{1}{2}\left(\eta_1^i(\lambda)\eta_2^i(\lambda) - \left[\eta_3^i(\lambda)\right]^2\right)^{-1}$$

$$\times\left(\eta_2^i(\lambda)\left[\xi_1^i(\lambda)\right]^2 - 2\eta_3^i(\lambda)\xi_1^i(\lambda)\xi_2^i(\lambda) + \eta_1^i(\lambda)\left[\xi_2^i(\lambda)\right]^2\right)$$

As intended.

Proof of Theorem 3

In order to prove Theorem 3, it needs to be established that taking the supremum of the sequence of test statistics preserves the convergence of the functionals that compose the argument of the function. To do that, we closely follow the work of Zivot & Andrews (1992) and divide the problem in three levels. In each level, the purpose is to show the continuity of the functions involved, and thus, by the continuous mapping theorem (CMT), it follows that convergence is preserved in each successive level.

Remember that $y_t = \sum_{j=1}^{t} \sum_{k=1}^{j} e_k = \sum_{j=1}^{t} S_j$, $\Delta y_t = \sum_{k=1}^{t} e_k = S_t$, where $S_h = \sum_{k=1}^{h} e_k$ and $S_0 = 0$. Next, define the step function $X_T(r) = \sigma^{-1} T^{-\frac{1}{2}} S_{[rT]}$. Donsker's theorem establishes $X_T(\cdot) \Rightarrow W(\cdot)$.

The sketch of the proof begins by expressing the F-statistic in terms of the partial sum S_t, the vector of deterministic regressors $\boldsymbol{Z_t}\left(\lambda, \frac{t}{T}\right)$, and the vector $T^{-\frac{1}{2}} \sum_{t=1}^{T} e_t \boldsymbol{Z_t}\left(\lambda, \frac{t}{T}\right)$. In what follows, we will only consider model A, since the proof is analogous for models B and C.

$$(A.31) \quad F_{\hat{\rho}}(\lambda) = \left[\sum_{i=1}^{T} \tilde{y}_{t-1}^2 \sum_{i=1}^{T} \widetilde{\Delta y}_{t-1}^2 - \left(\sum_{t=1}^{T} \tilde{y}_{t-1} \widetilde{\Delta y}_{t-1} \right)^2 \right]^{-1}$$

$$\times \left[\left(\sum_{t=1}^{T} \tilde{y}_{t-1} e_t \right)^2 \sum_{i=1}^{T} \widetilde{\Delta y}_{t-1}^2 - 2 \sum_{t=1}^{T} \tilde{y}_{t-1} e_t \sum_{t=1}^{T} \tilde{y}_{t-1} \widetilde{\Delta y}_{t-1} \sum_{t=1}^{T} \widetilde{\Delta y}_{t-1} e_t \right.$$

$$\left. + \left(\sum_{t=1}^{T} \widetilde{\Delta y}_{t-1} e_t \right)^2 \sum_{i=1}^{T} \tilde{y}_{t-1}^2 \right] \times [2 s^2(\lambda)]^{-1}$$

Where "$\sim$ " represents, as before, the residuals from the projection of the variable indicated, onto the vector space spanned by $\boldsymbol{z_t}(\lambda)'$. Denote $\mathbf{K}_{\boldsymbol{T}} \boldsymbol{z_t}(\lambda) = \boldsymbol{Z_t}\left(\lambda, \frac{t}{T}\right) = \boldsymbol{Z_t}$. The supremum over all values of λ, in a closed subset of (0,1), of expression (A.13) can be written as:

$$(A.32) \quad \sup_{\lambda \in [0,1]} F_{\hat{\rho}}(\lambda) = \sup_{\lambda \in [0,1]} f\left(\sigma X_T(r), \boldsymbol{Z_t}, T^{-\frac{1}{2}} \sum_{t=1}^{T} e_t \boldsymbol{Z_t}\left(\lambda, \frac{t}{T}\right), \sigma^2, \sigma_T^2 \right)$$

$$= g\big(h(m_1, m_2, m_3)\big)$$

Each function will be defined later.

The sample moments are given next:

$$(A.33) \quad T^{-4}\sum_{i=1}^{T}\tilde{y}_{t-1}^{2}$$

$$= T^{-1}\sum_{t=1}^{T}\left\{T^{-1}\sum_{j=1}^{t}T^{-\frac{1}{2}}S_j - \boldsymbol{Z}_{\boldsymbol{t}}'\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{Z}_{\boldsymbol{t}}\boldsymbol{Z}_{\boldsymbol{t}}'\right)^{-1}T^{-1}\sum_{t=1}^{T}\boldsymbol{Z}_{\boldsymbol{t}}T^{-1}\sum_{t=1}^{t}T^{-\frac{1}{2}}S_{s-1}\right\}^2$$

$$= \int_0^1\left\{\int_0^r \sigma X_T(u)du - \boldsymbol{Z}_{\boldsymbol{t}}'\left(\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\boldsymbol{Z}_{\boldsymbol{t}}'dr\right)^{-1}\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\int_0^r \sigma X_T(u)du\,ds\right\}^2 dr + o_p(1)$$

$$= A_1[\sigma X_T, \boldsymbol{Z}_{\boldsymbol{t}}](\lambda) + o_p(1)$$

$$(A.34) \quad T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}\,\widetilde{\Delta y}_{t-1}$$

$$= \int_0^1\left\{\sigma\int_0^r X_T(s)ds - \boldsymbol{Z}_{\boldsymbol{t}}'\left(\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\boldsymbol{Z}_{\boldsymbol{t}}'dr\right)^{-1}\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\int_0^r \sigma X_T(u)du\,ds\right\}$$

$$\times\left\{\sigma X_T(r) - \boldsymbol{Z}_{\boldsymbol{t}}'\left(\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\boldsymbol{Z}_{\boldsymbol{t}}'dr\right)^{-1}\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\,\sigma X_T(s)ds\right\}dr$$

$$= A_3[\sigma X_T, \boldsymbol{Z}_{\boldsymbol{t}}](\lambda) + o_p(1)$$

$$(A.35) \quad T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}e_t$$

$$= \sigma^2\int_0^1\int_0^r X_T(s)ds\Delta W_T(r) - T^{-\frac{1}{2}}\sum_{t=1}^{T}e_t\boldsymbol{Z}_{\boldsymbol{t}}'\left(\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\boldsymbol{Z}_{\boldsymbol{t}}'dr\right)^{-1}\int_0^1 \boldsymbol{Z}_{\boldsymbol{t}}\int_0^r \sigma X_T(u)du\,ds$$

$$= A_4\left[\sigma X_T, \boldsymbol{Z}_{\boldsymbol{t}}, T^{-\frac{1}{2}}\sum_{t=1}^{T}e_t\,\boldsymbol{Z}_{\boldsymbol{t}}, \sigma_T^2\right](\lambda) + o_p(1)$$

The functionals $T^{-2}\sum_{i=1}^{T}\widetilde{\Delta y}_{t-1}^{2} = A_2[\sigma X_T(r), \boldsymbol{Z_t}](\lambda) + o_p(1)$, and $T^{-1}\sum_{t=1}^{T}\tilde{y}_{t-1}e_t = A_5\left[\sigma X_T, \boldsymbol{Z_t}, T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t\,\boldsymbol{Z_t}, \sigma_T^2\right](\lambda) + o_p(1)$ are the same as $H_1(\cdot)$ and $H_2(\cdot)$, respectively, in Zivot & Andrews (1992).

Lemma A.3. $T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t\,\boldsymbol{Z_t} \Rightarrow \sigma\int_0^1 \boldsymbol{Z}(\cdot, r)dW(r)$

Proof. $T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t\,\boldsymbol{Z_t} = \left(T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t\,, T^{-\frac{1}{2}}\sum_{t=1}^{T}\frac{t}{Te_t}, T^{-\frac{1}{2}}\sum_{t=[\lambda T]+1}^{T} e_t\right)' =$ $\left(\sigma X_T(1), \sigma\left(X_T(1) - \int_0^1 X_T(r)\,dr\right), \sigma\big(X_T(1) - X_T(\lambda)\big)\right)'$. By the CMT, we have joint convergence to the vector $\left(\sigma W(1), \sigma\left(W(1) - \int_0^1 W(r)\,dr\right), \sigma\big(W(1) - W(\lambda)\big)\right)'$

Lemma A.4. $\mathrm{s}^2(\lambda) \Rightarrow \sigma^2 1(\lambda \in \Lambda)$, and $\sigma_T^2 \Rightarrow \sigma^2$. $1(\lambda \in \Lambda)$ is the indicator function equal to 1 for all $\lambda \in \Lambda$.

Using Lemma A.2., Lemma A.3., and Donsker's theorem it follows that

$$\left(\sigma X_T(\cdot), \boldsymbol{Z_t}(\cdot,\cdot), T^{-\frac{1}{2}}\sum_{t=1}^{T} e_t\,\boldsymbol{Z_t}\left(\cdot, \frac{t}{T}\right), \sigma_T^2, s^2(\cdot)\right)'$$

$$\Rightarrow \left(\sigma W(\cdot), \boldsymbol{Z}(\cdot,\cdot), \sigma\int_0^1 \boldsymbol{Z}(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot)\right)'$$

$W(\cdot)$ is a Wiener process with support on $\mathbb{C}[0,1]$, and is continuous with probability 1.

Lemma A.5 The functions A_1 to A_5 are continuous with probability 1.

Proof. Take for example $A_1[\sigma W(\cdot), Z](\lambda)$ defined in (A.33). To prove continuity of A_1 at (W, Z), it needs to be shown that every function that composes it, is continuous and bounded over $\lambda \in \Lambda$, where Λ is a closed subset of $(0,1)$.

Let $W(r)$ and $\widetilde{W}(r)$ be two Wiener processes s.t $\sup_{r\in[0,1]} \left|W(r) - \widetilde{W}(r)\right| < \varepsilon$, for $\varepsilon > 0$.

Then,

$$\left|\int_0^r W(u)du - \int_0^r \widetilde{W}(u)du\right| \leq \left|\int_0^r W(u) - \widetilde{W}(u)du\right| \leq \sup_{0\leq u\leq r} \left|W(r) - \widetilde{W}(r)\right| < \varepsilon$$

Thus, the map $(W, Z) \rightarrow \int_0^r W(u)du$ is continuous

Next, consider $\boldsymbol{ZZ'} = \begin{bmatrix} 1 & r & du(\lambda, r) \\ r & r^2 & rdu(\lambda, r) \\ du(\lambda, r) & rdu(\lambda, r) & du^2(\lambda, r) \end{bmatrix}$. Integrating over [0,1] yields,

$$\int_0^1 ZZ'dr = \begin{bmatrix} 1 & \frac{1}{2} & 1-\lambda \\ \frac{1}{2} & \frac{1}{3} & (1-\lambda^2)/2 \\ 1-\lambda & (1-\lambda^2)/2 & 1-\lambda \end{bmatrix}.$$

From the proof of the previous theorem, we have $\boldsymbol{Z_T Z_T'} \rightarrow \boldsymbol{ZZ'}$. For a fixed T, write,

$\sup_{r\in[0,1]} |\boldsymbol{Z_T Z_T'} - \boldsymbol{ZZ'}| < \varepsilon$, for $\varepsilon > 0$. Then,

$$\left|\int_0^1 \boldsymbol{Z_T Z_T'} dr - \int_0^1 \boldsymbol{ZZ'} dr\right| \leq \left|\int_0^1 \boldsymbol{Z_T Z_T'} - \boldsymbol{ZZ'} dr\right| \leq \int_0^1 |\boldsymbol{Z_T Z_T'} - \boldsymbol{ZZ'}| dr$$

$$\leq \sup_{r\in[0,1]} |\boldsymbol{Z_t Z_t'} - \boldsymbol{ZZ'}| < \varepsilon$$

Thus, by the CMT, the integral preserves the convergence of $\boldsymbol{Z_t Z_t'}$ to $\boldsymbol{ZZ'}$.

Now, it needs to be established that $\int_0^1 \boldsymbol{ZZ}'dr$ is bounded over $\lambda \in (0,1)$. Note that as long as $\inf_{\lambda\in\Lambda} \det\left(\int_0^1 \boldsymbol{ZZ}'dr\right) > 0$, the matrix is bounded.

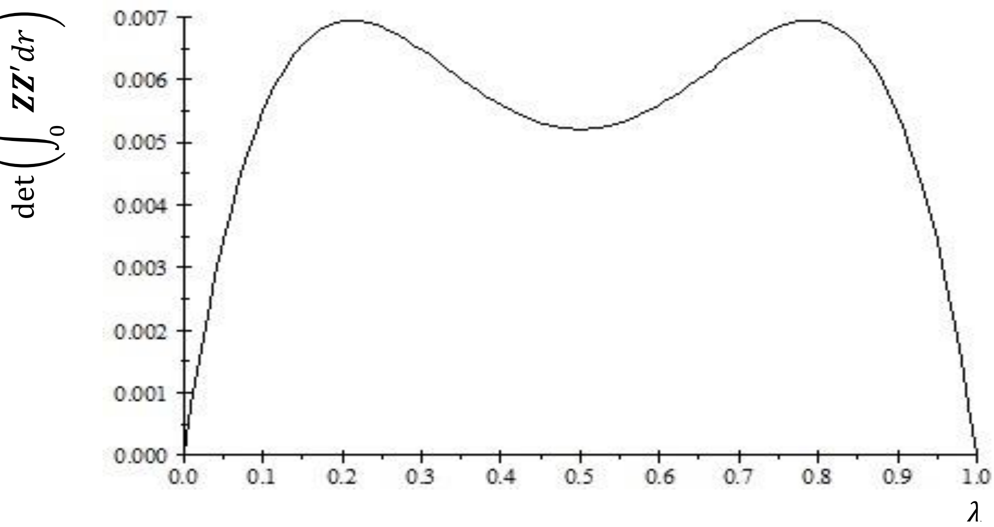


Figure 5: det $\left(\int_0^1 \boldsymbol{ZZ}'dr\right)$ as a function of λ.

Furthermore, it can be shown that $\int_0^1 \boldsymbol{ZZ}'dr$ is definite positive, and symmetric, with probability 1, thus, it follows that $\max|a_{ii}| \geq \max|a_{ij}|$, that is, the greatest element in the diagonal is greater than any off-diagonal element. Since the diagonal elements are bounded over $\lambda \in (0,1)$, it follows that $\int_0^1 ZZ'dr$ is bounded.

Because $\inf_{\lambda\in\Lambda} \det\left(\int_0^1 \boldsymbol{ZZ}'dr\right) > 0$, it follows that the map $\int_0^1 \boldsymbol{ZZ}'dr \rightarrow \left(\int_0^1 \boldsymbol{ZZ}'dr\right)^{-1}$ is continuous, hence the map $(W, \boldsymbol{Z}) \rightarrow \left(\int_0^1 \boldsymbol{ZZ}'dr\right)^{-1}$ is also continuous.

Applying the same arguments, it can be shown that the remaining functions are all continuous at $(W, \boldsymbol{Z})$ with probability 1.

Now define $h[m_1, m_2, m_3] = m_1^{-1} m_2 m_3^{-1}$, where each function $m(\cdot)$ is defined in accordance with (A.31). $h(\cdot)$ is continuous if $\inf_{\lambda\in\Lambda}|m_1| > 0$ and $\sigma^2 > 0$.

Lemma A.6. $h[m_1, m_2, m_3]$ is continuous.

Proof. Write $m_1 = A_1A_2 - (A_3)^2$, where the functions A_j form the continuous map $(W, Z) \to \Lambda$. Now, $\inf_{\lambda\in\Lambda}|A_1A_2 - (A_3)^2| \geq \inf_{\lambda\in\Lambda}|A_1| \inf_{\lambda\in\Lambda}|A_2| + \sup_{\lambda\in\Lambda}|(A_3)^2|$. The results in Zivot & Andrews (1992) establish $\inf_{\lambda\in\Lambda}|A_2| > 0$.

Following a similar argument, consider those realizations of W $s.t. \inf_{\lambda\in\Lambda}|A_1| = 0$. Then, since A_1 is continuous on a compact set Λ, from the extreme value theorem it follows that, for a given $\lambda_0 \in [0,1]$ random variable, we have $A_1(\lambda_0) = 0$ and $\lambda_0 \in \Lambda$ hence, taking the limit of $(A.33)$, this implies

$$(A.36) \int_0^r W(s)ds = \boldsymbol{Z}(\lambda_0, r)' \left(\int_0^1 \boldsymbol{Z}(\lambda_0, u) \boldsymbol{Z}(\lambda_0, u)' du \right)^{-1} \int_0^1 \boldsymbol{Z}(\lambda_0, u) \int_0^u W(s)ds\, du$$

This equation implies that the distribution of the left-hand side, which is $N[0, (1/3)r^3]$, is equal to the distribution of the right-hand side, for every $r \in [0,1]$.

We start by deriving the distribution of $\int_0^1 \boldsymbol{Z}(\lambda_0, u) \int_0^u W(s)ds\, du$, where $\int_0^u W(s)ds \sim N[0, (1/3)u^3]$.

Define the random variable $Y = \int_0^1 \boldsymbol{Z}(\lambda_0, u) \int_{s=0}^u W(s)ds\, du$ and $Y = \int_0^1 \boldsymbol{Z}(\lambda_0, u) \int_{t=0}^u W(t)dt\, du$. Then, the expectation of Y is zero, and its variance is given by $E[YY'] = E\left[\int_0^1 \int_{t=0}^u \int_{s=0}^u \boldsymbol{Z}(\lambda_0, u) W(s) W(t)\, \boldsymbol{Z}(\lambda_0, u)' ds\, dt\, du\right]$

$$= \begin{bmatrix} \frac{1}{12} & \frac{1}{15} & \frac{1}{12}(\lambda_0-1)^2\sum_{j=0}^{3}\lambda_0{}^j \\ \frac{1}{15} & \frac{1}{18} & \frac{1}{15}(\lambda_0-1)^2\sum_{j=0}^{4}\lambda_0{}^j \\ \frac{1}{12}(\lambda_0-1)^2\sum_{j=0}^{3}\lambda_0{}^j & \frac{1}{15}(\lambda_0-1)^2\sum_{j=0}^{4}\lambda_0{}^j & \frac{1}{12}(\lambda_0-1)^2\sum_{j=0}^{3}\lambda_0{}^j \end{bmatrix} = \Sigma(\lambda_0)$$

From which it follows $\int_0^1 \boldsymbol{Z}(\lambda_0, u)\int_0^u W(s)ds\, du \sim \boldsymbol{N}[\mathbf{0}, \boldsymbol{\Sigma}(\lambda_0)]$. Now write the right-hand side of equation $(A.36)$ as,

$$(A.37) \quad \left(1, r, du(\lambda_0, r)\right)C\left(\int W, \lambda_0\right),$$

where $C(\int W, \lambda_0) \sim \boldsymbol{N}[\mathbf{0}, \mathbf{A^{-1}}(\lambda_0)\boldsymbol{\Sigma}(\lambda_0)\mathbf{A^{-1}}(\lambda_0)]$, and the matrix $\mathbf{A^{-1}}(\lambda_0) = \left(\int_0^1 \boldsymbol{Z}(\lambda_0, u)\boldsymbol{Z}(\lambda_0, u)' du\right)^{-1}$.

Now, consider setting $0 \le r < \inf(\lambda: \lambda \in \Lambda)$, and let $\inf(\lambda: \lambda \in \Lambda)$ get closer and closer to zero. Then, as $\inf(\lambda: \lambda \in \Lambda)$ (and thus r) approaches zero, the distribution of the left-hand side of equation $(A.36)$ collapses to zero, but the variance of the distribution on the right-hand side goes to infinity, which is a contradiction. In fact, there is no $\lambda_0 \in \Lambda$ that satisfies equation (A.36). Therefore, we conclude that, with W-probability 1, $\inf_{\lambda \in \Lambda}|A_1| > 0$, and the desired result follows.

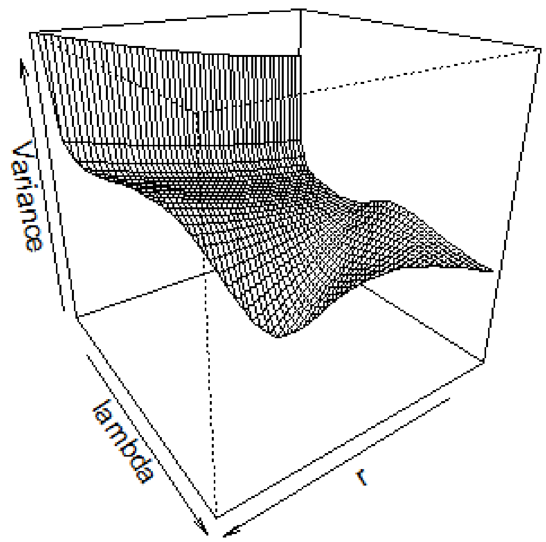


Figure 6: Variance of $\left(1, r, du(\lambda_0, r)\right)C(\int W, \lambda_0)$

Finally, the continuity of the function $g\big(h(\cdot)\big) = \sup_{\lambda\in[0,1]} h(\cdot)$ is established in the following way. From the results built thus far, we have $h_T(\cdot) \Rightarrow h(\cdot)$ over $\lambda \in \Lambda$ Therefore, for a fixed T, we can find $> 0, s.t \sup_{\lambda\in\Lambda} | h(\lambda) - h_T(\lambda)| < \varepsilon$. Then,

$$(A.38) \quad |\sup_{\lambda\in\Lambda} h(\lambda) - \sup_{\lambda\in\Lambda} h_T(\lambda)| \leq \sup_{\lambda\in\Lambda} | h(\lambda) - h_T(\lambda)| < \varepsilon$$

And thus, by the CMT, convergence is preserved under the supremum function, from which it follows the claim of Theorem 3.