DÉPARTEMENT DE SCIENCE ÉCONOMIQUE
DEPARTMENT OF ECONOMICS

CAHIERS DE RECHERCHE / WORKING PAPERS

0602E

Unit Root Tests and Structural Change when the Initial Observation is Drawn from its unconditional Distribution

by

Hui Liu and Gabriel Rodríguez

ISSN: 0225-3860

uOttawa

Faculté des sciences sociales
Faculty of Social Sciences

CP 450 SUCC. A
OTTAWA (ONTARIO)
CANADA K1N 6N5

P.O. BOX 450 STN. A
OTTAWA, ONTARIO
CANADA K1N 6N5

CAHIER DE RECHERCHE #0602E
Département de science économique
Faculté des sciences sociales
Université d'Ottawa

WORKING PAPER #0602E
Department of Economics
Faculty of Social Sciences
University of Ottawa

Unit Root Tests and Structural Change when the Initial Observation is Drawn from its Unconditional Distribution[1]

by

Hui Liu
Department of Economics
University of Ottawa

and

Gabriel Rodríguez[2]
Department of Economics
University of Ottawa

March 8, 2006

[1]This paper is drawn from the second chapter of the PhD Dissertation of H. Liu at the Department of Economics of the University of Ottawa. We thank participants to the 38^{th} Annual Meeting of the Canadian Economic Association (CEA at Toronto, June 2004), Simon Power and Mahmoud Zarepour when Liu presented an earlier version of this paper at the Department of Economics of the University of Ottawa. We also thank the Editor Stéphane Grégoir and an anonymous referee for very useful insights and comments. Useful conversations with Pierre Perron are also acknowledged. Rodríguez thanks Financial support from the Faculty of Social Sciences of the University of Ottawa.

[2]Address for Correspondence: Gabriel Rodríguez, Department of Economics, University of Ottawa, P.O. Box 450, Station A, Ottawa, Ontario, Canada, K1N 6N5. E-mail address: gabrielr@uottawa.ca, Telephone: +613-562-5800 (1750), Fax: +613-562-5999.

Abstract

Following Elliott (1999) and Perron and Rodríguez (2003), we develop unit root tests in the context of structural change models using GLS detrended data (Elliott, Rothenberg and Stock, 1996) when the initial observation is drawn from its unconditional distribution. We derive the limiting distributions of the M-tests (Stock, 1999; Perron, and Ng, 1996), the ADF statistic and a feasible optimal point test from which we derive the power envelope. Asymptotic power functions are calculated and compared with the case where the initial condition is not random. Finite sample size and power simulations under various forms of error processes are performed using different lag selection methods and two different methods to select the break point. Empirical applications are also provided.

Keywords: Initial Condition, Feasible Optimal Point Test, GLS Detrending, Power Envelope, Unit Root, Structural Change.

JEL: C2, C5.

Résumé

Suivant Elliott (1999) et Perron et Rodríguez (2003), nous dérivons des tests pour racine unitaire dans le cas où la fonction de tendance peut avoir une rupture à une date inconnue. Ces tests utilisent la méthode des moindres carrés généralisés (MCG) pour éliminer les composantes déterministes, tel que proposé par Elliott, Rothenberg et Stock (1996). Nous considérons le cas où la condition initielle est obtenue à partir de sa distribution non conditionnelle. Nous dérivons les distributions asymptotiques de M-tests (Stock, 1999; Perron and Ng, 1996), du test ADF et celle d'une version réalisable du test optimal en un point. Ce test nous permet de dériver l'enveloppe de puissance. Nous calculons les fonctions de puissance asymptotique et nous les comparons au cas où la condition initielle n'est pas aléatoire. En utilisant des simulations, nous évaluons le niveau et la puissance des tests en échantillon finis et nous étudions plusieurs méthodes pour sélectionner le retard nécessaire pour calculer l'estimateur de la densité spectrale, ainsi que deux méthodes pour sélectionner le point de rupture. Une application à des séries des salaires réels et aux prix des actions ordinaires aux Etats-Unis est aussi considérée à la fin.

Mots-clés: Condition initielle, test optimal en un point, MCG, Enveloppe de puissance, Racine unitaire, Changement structurel.

Code JEL: C2, C5.

1 Introduction

Since the seminal work of Nelson and Plosser (1982), abundant research has contributed to the discussion between stationarity in levels against stationarity in first differences of the data. Perron (1989) suggested that allowing for a (known) break point in the trend function of the time series, strong rejection of the null hypothesis of a unit root is possible. Zivot and Andrews (1992), Christiano (1992), Banerjee, Lunsdaine and Stock (1992), and Perron (1997) contributed to the debate deriving limiting distribution of the statistics when the break point is considered unknown.

In the literature of efficient unit root tests without structural change, Elliott, Rothenberg and Stock (1996) -hereafter ERS (1996)- have proposed the use of GLS detrended data which allows to improve the power of the statistics. Ng and Perron (2001) have applied this approach to the so-called M statistics (Stock, 1999). On the other hand, in the literature of unit roots with structural change, Perron and Rodríguez (2003) have extended the M-statistics, the ADF statistic (Dickey and Fuller, 1979; Said and Dickey, 1984), and the feasible point optimal test (ERS, 1996) to the case of an unknown structural break. They use two time series of the Nelson-Plosser data, real wages and stock prices, and find a rejection for most of the test statistics analyzed.

However, only a few papers have explored the impact of the initial condition on unit root testing in large samples. Elliott (1999), in a model where no structural change is allowed, found a loss of power when the initial observation is drawn from its unconditional distribution under the alternative hypothesis. He derived the point optimal test under the new initial condition and showed that power envelope shifted down from the one corresponding to the fixed initial value. On the other hand, Müller and Elliott (2003) treated a variety of initial conditions under the alternative hypothesis as nuisance parameters and derived a family of point optimal tests over a weighting function of different initial conditions. They also related unit root test statistics that do not have optimality properties to this family of optimal tests, in order to understand what implicit assumptions these statistics make on the initial condition. They found that many of the test statistics used in the literature can be closely related to the optimal test families but with very different weights for the initial condition.

This paper follows the research lines of Elliott (1999) and Perron and Rodríguez (2003). Under the alternative hypothesis of stationarity with a broken trend, we consider an unknown break point and assume that the initial value is drawn from its unconditional distribution. We evaluate the performance of unit root statistics in both large and small samples. By doing so, we achieve a deeper understanding on the impact of the initial condition in unit root testing in the context of structural change models.

The rest of this paper is organized as follows. Section 2 and 3 derives the limiting distributions of the different statistics using GLS-detrended data. They are the M-tests (M^{GLS}), the augmented Dickey-Fuller statistic (ADF^{GLS}), and the feasible point optimal statistic ($P_{T\xi}^{GLS}$) statistics under the new initial condition assumption. Section 4 calculates the asymptotic critical values, the power envelope and the asymptotic power functions. Section 5 presents the finite sample evaluations in terms of size and power performance of different statistics. Section 6 gives an empirical application and the last section concludes. All the proofs are presented in the Appendix.

2 The Models and Asymptotic Theory

2.1 The models

For the purpose of comparison, this paper considers the same models and statistics as in Perron and Rodríguez (2003) but with a different assumption concerning the initial condition to be established below. Hence, the data generating process is,

$$y_t = d_t + u_t \tag{1}$$

$$u_t = \alpha u_{t-1} + v_t \tag{2}$$

$$\alpha = 1 + cT^{-1} \tag{3}$$

where $d_t = \varphi' z_t$, and z_t includes the deterministic components. Model I contains a break in slope, therefore the deterministic components are $z_t = \{1, t, 1(t > T_B)(t - T_B)\}$ where $\mathbf{1}(\cdot)$ is the indicator function, T_B is the break point, and the set of estimated coefficients is denoted as $\widehat{\psi}' = (\hat{\mu}_1, \hat{\beta}_1, \hat{\beta}_2)$. Model II contains a break in both intercept and slope, where the deterministic components are given by $z_t = \{1, 1(t > T_B), t, 1(t > T_B)(t - T_B)\}$ and the set of estimated coefficients is therefore denoted as $\widehat{\psi}' = (\hat{\mu}_1, \hat{\mu}_2, \hat{\beta}_1, \hat{\beta}_2)$. Equation (3) represents the local to unity framework examined in Phillips (1987) and Chan and Wei (1987), where the parameter c measures the deviation from unity. When $c = 0$, we are under the null hypothesis, while when $c < 0$, we are under the alternative hypothesis.

Instead of assuming $u_0 = 0$ as the initial condition (ERS, 1996, and Perron and Rodríguez, 2003), we assume that the initial observation under the alternative hypothesis is drawn from its unconditional distribution. This would result in an unconditional variance of $\sigma^2/(1-\alpha^2)$ for $\alpha < 1$ where σ^2 is the variance of v_t. This alternative assumption made on the initial condition involves the unknown parameter of interest α. It does not disappear asymptotically so the likelihood ratio statistic would differ from the optimal test derived in ERS (1996) and Perron and Rodríguez (2003) and affect the asymptotic power. Therefore, we use the following condition adopted from Elliott (1999):

Assumption A (Initial condition assumption**).** *We assume that u_0 is zero when $\alpha = 1$, so $u_1 = v_1$, while u_1 has mean zero and variance $\sigma^2/(1-\alpha^2)$ when $\alpha < 1$.*

Under this assumption, the initial observation does not disappear at the convergence rate of $T^{1/2}$ under the alternative hypothesis and will have an effect on the limiting distribution[3].

Assumption B (see Elliott, 1999, and Davidson, 1994 for a general treatment**).** The innovations $\{v_t\}$ satisfy (a) $E[v_t] = 0$ and $\widehat{\gamma}_v(j) \to^P \gamma_v(j)$, for fixed j, where $\gamma_v(j) = E[v_t v_{t-j}]$ and $\widehat{\gamma}_v(j) = T^{-1}\sum_{t=j+1}^{T} v_t v_{t-j}$; (b) the functional central limiting theorem can be applied to the partial sums $S_t = \sum_{j=1}^{t} v_j$ and $T^{-1/2}S_{[Tr]} \Rightarrow N(0, \sigma^2 r) \equiv \sigma W(r)$, where $\Rightarrow$ signifies weak convergence, $W(r)$ is a standard Wiener process on the interval [0, 1], $\sigma^2 = \lim_{T\to\infty} T^{-1}E(S_T^2)$ is the non-normalized spectral density at frequency zero[4]. σ^2 is a finite positive number such that $\sigma^2 = \sum_{j=-\infty}^{\infty} \gamma_v(j)$[5].

For the GLS detrending approach, we use the same notation as in Perron and Rodríguez (2003). That is, we first transform the data using $y_t^{\bar{\alpha}} = [\left(1-\bar{\alpha}^2\right)^{1/2} y_1,\ (1-\bar{\alpha}L)y_t], z_t^{\bar{\alpha}} = [\left(1-\bar{\alpha}^2\right)^{1/2} z_1,\ (1-\bar{\alpha}L)z_t]$, and $u_t^{\bar{\alpha}} = [\left(1-\bar{\alpha}^2\right)^{1/2} u_1,\ (1-\bar{\alpha}L)u_t]$, for $t = 2, 3, ..., T$ and $\bar{\alpha} = 1 + \bar{c}T^{-1}$. Notice that the treatment of the first observation changes compared to the case analyzed by ERS (1996) and Perron and Rodríguez (2003).

Then the set of coefficients related to the deterministic component is

[3] See Lemma A.1 in the Appendix.

[4] According to Elliott (1999), the process v_t is a potentially serially correlated stationary process. Under these assumptions with v_t Gaussian and further that the initial value $u_0 = 0$ under the null and alternative hypothesis, ERS (1996) show that no uniformly most powerful test against the relevant stationary alternative exists and derive the asymptotic power function for the most powerful test against a sequence of local alternatives of α to 1. Perron and Rodríguez (2003) proceeded with the same way. We will refer to this as the fixed initial condition case.

[5] A number of consistent estimators of σ^2 are available. See Stock (1994) for a discussion and review.

estimated by the following OLS regression:

$$y_t^{\bar{\alpha}} = \widehat{\psi}' z^{\bar{\alpha}} + u^{\bar{\alpha}}. \tag{4}$$

The limiting distribution of $\Lambda^{-1}(\widehat{\psi} - \psi)$ with $\Lambda = diag(T^{1/2}, T^{-1/2}, T^{-1/2})$ is stated in Lemma A.3 in the Appendix. At last, the resulting time series after eliminating the deterministic component is defined by

$$\tilde{y}_t = y_t - \widehat{\psi}' z_t. \tag{5}$$

2.2 The Statistics and their Limiting Distributions

In this section two types of statistics are analyzed. One is the widely used ADF statistic (Dickey and Fuller, 1979; Said and Dickey, 1984) which tests whether $\widehat{\alpha}_0 = 0$ in the following augmented regression:

$$\Delta\tilde{y}_t = \widehat{\alpha}_0 \tilde{y}_{t-1} + \sum_{j=1}^{k} \widehat{b}_j \Delta\tilde{y}_{t-j} + \hat{e}_{tk} \tag{6}$$

where the lagged first differences are used to account for the serial correlation in the time series.

The other statistics are the so-called M-tests proposed by Stock (1999), and further analyzed by Ng and Perron (1996). This class of tests include a modified version of Phillips-Perron's (1988) Z_α test; Sargan and Bhargava's (1983) and Bhargava's (1986); and the modified Phillips-Perron's (1988) Z_{t_α} tests. Using the above defined $\tilde{y}_t$ series, the M-tests can be written as:

$$MZ_\alpha^{GLS} = (\tilde{y}_T^2 - T^{-1}s^2)(2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2)^{-1}, \tag{7}$$

$$MSB^{GLS} = (T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2/s^2)^{1/2}, \tag{8}$$

$$MZ_t^{GLS} = (T^{-1}\tilde{y}_T^2 - s^2)(4s^2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2)^{-1/2}. \tag{9}$$

Perron and Ng (1996) showed that the main advantage of the M-tests is that they have less size distortions when the error term contains negative moving average dynamics and in other cases they still have acceptable size distortions. In Equations (7)-(9), an estimator of the spectral density at frequency zero (s^2) is required. We use the following consistent autoregressive estimate s^2 proposed by Perron and Ng (1998):

$$s^2 = s_{ek}^2/[1-\hat{\beta}(1)]^2,$$

where $s_{ek}^2 = (T-k)^{-1}\sum_{t=k+1}^{T}\hat{e}_{tk}^2$, $\hat{\beta}(1) = \sum_{j=1}^{k}\hat{\beta}_j$, $\{\hat{e}_{tk}\}$ and $\hat{\beta}_j$ are obtained from the augmented regression (6).

Assuming a known break date at T_B and $\delta = T_B/T$, the following theorem states the limiting distribution of the unit root statistics when the initial condition is defined as in *condition A*, and $\{v_t\}$ satisfies the *condition B*.

Theorem 1 *Suppose $\{y_t\}$ is generated by (1) to (3), the initial value is given by the Condition A, $\{v_t\}$ satisfies the Condition B, GLS-detrending is applied using $\bar{\alpha} = 1 + \bar{c}T^{-1}$, $\delta = T_B/T$ is the break point, s^2 is a consistent estimate of the spectral density at the frequency zero σ^2,then the M^{GLS} and ADF^{GLS} statistics for Model I and II have the following limiting distributions:*

$$\begin{aligned}
MZ_\alpha^{GLS}(\delta) &\Rightarrow \frac{0.5g_1(c,\overline{c},\delta)}{g_2(c,\overline{c},\delta)} \equiv J^{MZ_\alpha^{GLS}}(c,\overline{c},\delta) \\
MSB^{GLS}(\delta) &\Rightarrow (g_2(c,\overline{c},\delta))^{1/2} \equiv J^{MSB^{GLS}}(c,\overline{c},\delta) \\
MZ_t^{GLS}(\delta) &\Rightarrow \frac{0.5g_1(c,\overline{c},\delta)}{(g_2(c,\overline{c},\delta))^{1/2}} \equiv J^{MZ_t^{GLS}}(c,\overline{c},\delta) \\
ADF^{GLS}(\delta) &\Rightarrow \frac{0.5g_1(c,\overline{c},\delta)}{(g_2(c,\overline{c},\delta))^{1/2}} \equiv J^{ADF^{GLS}}(c,\overline{c},\delta)
\end{aligned}$$

where $g_1(c,\overline{c},\delta) = V_{c\overline{c}}^{(1)}(1,\delta)^2 - 2V_{c\overline{c}}^{(2)}(1,\delta) - 1$, $g_2(c,\overline{c},\delta) = \int_0^1 V_{c\overline{c}}^{(1)}(r,\delta)dr - 2\int_\delta^1 V_{c\overline{c}}^{(2)}(r,\delta)dr$, $V_{c\overline{c}}^{(1)}(r,\delta) = W_c(r) - b_4 - rb_5$, $V_{c\overline{c}}^{(2)}(r,\delta) = b_6(r-\delta)[W_c(r) - rb_5 - b_4 - (1/2)(r-\delta)b_6]$. The elements b_4, b_5, b_6 are calculated using

$$\begin{aligned}
&\begin{bmatrix} \bar{c}^2 - 2\bar{c} & \frac{1}{2}\bar{c}^2 - \bar{c} & -\bar{c}(1-\delta) + \frac{1}{2}\bar{c}^2(1-\delta)^2 \\ \frac{1}{2}\bar{c}^2 - \bar{c} & 1 + \frac{1}{3}\bar{c}^2 - \bar{c} & m \\ -\bar{c}(1-\delta) + \frac{1}{2}\bar{c}^2(1-\delta)^2 & m & d \end{bmatrix} \\
&\times \begin{pmatrix} \sigma b_1 \\ \sigma b_2 \\ \sigma b_3 \end{pmatrix} \\
= \;&\sigma \begin{pmatrix} b_4 \\ b_5 \\ b_6 \end{pmatrix}
\end{aligned}$$

where $b_1 = -2\bar{c}\xi - \bar{c}(c-\bar{c})\int_0^1 W_c(r) - \bar{c}W(1)$, $b_2 = (1-\bar{c})W(1) + c\int_0^1 W_c(r) - \bar{c}(c-\bar{c})\int_0^1 rW_c(r)$, $b_3 = (1-\bar{c}+\delta\bar{c})W(1) - W(\delta) - \bar{c}(c-\bar{c})\int_\delta^1 rW_c(r) + \delta\bar{c}(c-\bar{c})\int_\delta^1 W_c(r) + c\int_\delta^1 W_c(r)$, $m = 1-\delta-\bar{c}+\bar{c}\delta - \frac{1}{2}\bar{c}^2\delta + \frac{1}{3}\bar{c}^2 + \frac{1}{6}\bar{c}^2\delta^3$, $d = 1-\delta-\bar{c}(1-\delta)^2 + \frac{1}{3}\bar{c}^2(1-\delta)^3$, and $\xi \sim N(0, \frac{1}{-2c})$ under the alternative hypothesis, with $W_c(r)$ the Ornstein-Uhlenbeck process and the solution to the stochastic differential equation $dW_c(r) = cW_c(r)dr + dW(r)$ with $W_c(0) = 0$.

3 The Feasible Point Optimal Test and its Asymptotic Distribution

ERS (1996) showed that no uniformly optimal tests exist for unit root testing. Based on Dufour and King (1981), they developed a feasible point optimal test, which has a power function tangent to the power envelope at one point of the alternative hypothesis. The statistic is defined as

$$P_T^{GLS} = \{S(\bar{\alpha}) - \bar{\alpha}S(1)\}/s^2 \qquad (10)$$

where $S(\bar{\alpha}) = [y^{\bar{\alpha}} - \widehat{\psi}' z^{\bar{\alpha}}]'[y^{\bar{\alpha}} - \widehat{\psi}' z^{\bar{\alpha}}]$ and $S(1) = [y^1 - \widehat{\psi}' z^1]'[y^1 - \widehat{\psi}' z^1]$ are the squared sum of residuals under the alternative and the null hypothesis respectively.

Perron and Rodríguez (2003) extended the P_T^{GLS} test to the case of an unknown structural break. We derive the limiting distribution of this statistic (denoted here by $P_{T\xi}^{GLS}$) considering the effect of the new initial condition. The results are summarized in the following theorem.

Theorem 2 *Suppose $\{y_t\}$ is generated by (1) to (3), the initial value is given by the Condition A, $\{v_t\}$ satisfies the Condition B, GLS-detrending is applied using $\bar{\alpha} = 1 + \bar{c}T^{-1}$, $\delta = T_B/T$ is the break point, s^2 is a consistent estimate of the spectral density at the frequency zero σ^2, then the $P_{T\xi}^{GLS}$ test for Model I and II has the following limiting distribution*

$$\begin{aligned} P_{T\xi}^{GLS}(c,\bar{c},\delta) &\Rightarrow -2\bar{c}\xi^2 - 2\bar{c}\int_0^1 W_c(r)dW(r) + (\bar{c}^2 - 2\bar{c}c)\int_0^1 W_c^2(r)dr \\ &\quad + \bar{M}(c,0,\delta) - \bar{M}(c,\bar{c},\delta) - \bar{c} \\ &\equiv J^{P_{T\xi}^{GLS}}(c,\bar{c},\delta) \end{aligned}$$

where $\bar{M}(c,\bar{c},\delta)=\bar{A}(c,\bar{c},\delta)'\bar{B}\,(\bar{c},\delta)\,\bar{A}(c,\bar{c},\delta)$, $\bar{A}(c,\bar{c},\delta)$, and $\bar{B}\,(\bar{c},\delta)$ are defined in the appendix.

Unlike Perron and Rodríguez (2003), the term $\xi(\cdot)$, representing the effect of the initial condition, enters the limiting distribution of the statistic.

4 Selecting the Break Point and Asymptotic Results

4.1 Selecting the Break Point

Two data-dependent methods to estimate the break point endogenously are used in order to circumvent the data mining problem caused by pre-examination of the data. One is the so-called infimum method proposed by Zivot and Andrews (1992). The other is the supremum method suggested by Perron (1997).

Zivot and Andrews (1992) proposed selecting the break point that gives the strongest rejection against the null hypothesis of $\alpha = 1$. Then, the break point δ^* can be selected using

$$J(c,\bar{c}) = \inf_{\delta\in[0,1]} J\,(c,\bar{c},\delta)$$

where J represents the asymptotic distributions of the M^{GLS}, and the ADF^{GLS} statistics derived in Theorem 1.

The selection of δ^* for $P_{T\xi}^{GLS}$ is slightly different. According to Perron and Rodríguez (2003), the feasible point optimal statistic is defined by

$$P_{T\xi}^{GLS}(c,\bar{c}) = \{\inf_{\delta\in[\varepsilon,1-\varepsilon]} S(\bar{\alpha},\delta) - \inf_{\delta\in[\varepsilon,1-\varepsilon]} \bar{\alpha}S(1,\delta)\}/s^2 \qquad (12)$$

where a truncation ϵ is needed for critical values to be bounded and $\epsilon = 0.15$ is used throughout the paper, see Perron and Rodríguez (2003). These authors recognize that when the break point is unknown "things are different. The principle is, however, still the same" (page 8, Perron and Rodríguez, 2003). Their argument, which was not clearly explained, is concerned to the fact that the infimum of the GLS squared sum of residuals are constructed in the same way as ERS (1996) suggested. Notice that this criterium may be considered as not optimal if we consider that there does not exist a break point under the null hypothesis. The reason is clear because we have an unidentified parameter under the null hypothesis; see Andrews (1993). In this case we should take the Sup, Mean or Exp (see Andrews and Ploberger, 1994) of the likelihood ratio statistic for a specific break point. However, if we consider the existence of a break point under the null hypothesis, the procedure used by Perron and Rodríguez (2003) appears to be optimal because it is invariant with respect to all nuisance parameters of the model, including the break point. In other words, the minimization of the GLS square sum of residuals is performed taking into account of all corresponding parameters of the model[6].

Applying (12) to Theorem 2, we get

$$\begin{aligned} P_{T\xi}^{GLS}(c,\bar{c}) &\Rightarrow -2\bar{c}\int_0^1 W_c(r)dW(r) + (\bar{c}^2 - 2\bar{c}c)\int_0^1 W_c^2(r)dr \\ &\quad + \sup \bar{M}(c,0,\delta) - \sup \bar{M}(c,\bar{c},\delta) - \bar{c} - 2\bar{c}\xi^2 \\ &\equiv J^{P_{T\xi}^{GLS}}(c,\bar{c}). \end{aligned} \quad (13)$$

In the supremum method, Perron (1997) recommends to choose the δ^* related to the largest absolute value of the $t-$statistic associated with the parameter of break on slope. After selecting δ^*, we calculate the $M^{GLS}(\delta^*)$, and $ADF^{GLS}(\delta^*)$ statistics. There is no feasible optimal point test available using this method to select the break point.

4.2 Asymptotic Critical Values, Power Envelope, and Asymptotic Power Functions

Under the null hypothesis $c = 0$, we use $T = 1000$, and 10,000 replications to simulate asymptotic critical values for $\bar{c} = -1$ to -70 ($\bar{\alpha} = 0.999$ to 0.930), then we let $c = \bar{c}$ to calculate maximal power at each c and obtain the power envelope. As suggested by ERS (1996), we choose the value of $\bar{c}$ that gives 50% power as the one used for GLS-detrending and we select $\bar{c} = -24$. Intuitively, lower $\bar{c}$ in this paper (compared to $\bar{c} = -22.5$ in Perron and Rodríguez, 2003) tells us that it may take longer to reach the same percentage of power. In other words, the power envelope shifts down from the previous one in Perron and Rodríguez (2003), and the loss of power is caused by the relaxation of assumption for the initial observation. We graph power envelopes for both cases in Figure 1.

Next, we use the critical values when $\bar{c} = -24$, $T = 1000$ and 10,000 replications to calculate the asymptotic power for each statistics and the

[6] We thank important discussions with Pierre Perron concerning the adequacy of the constructed power envelope in Perron and Rodríguez (2003). Our conclusion is that, even in the case where their power envelope is not completely right, the correct or "optimal" power envelope should be very close, and the statistics proposed have power very close to this power envelope. In practical terms, there are no big problems.

results are graphed in Figures 1 and 2. We can see that the power curve for each test lies under the power envelope, but not far from it. Using the infimum method to choose break point sometimes gives a slightly higher power than supremum method, although it does not necessarily give a consistent estimate of the true break point (see Vogelsang and Perron, 1998). The results from Perron and Rodríguez (2003) are also graphed in the same figure as a comparison. We can see that the power of each test has dropped due to the change in the assumption of the initial condition.

5 Finite Sample Results

We simulate finite sample critical values and evaluate the performance in terms of size and power of the test statistics. In the following, we use five data dependent methods to choose k.

5.1 The Selection of k

We first use Akaike and Bayesian Information Criteria (AIC and BIC hereafter). They take the form of $\text{AIC}(k) = \arg\min_{k\in[0,k_{\max}]}\{\ln(s_{ek}^2) + \frac{2k}{T^*}\}$ and $\text{BIC}(k) = \arg\min_{k\in[0,k_{\max}]}\{\ln(s_{ek}^2) + \frac{\ln(T^*)k}{T^*}\}$, where $T^* = T - k_{\max}$, and $k_{\max}$ should be large enough to account for the potential presence of serial correlation. We use $k_{\max} = int[12 \times (T/100)^{1/4}]$ as recommended by Ng and Perron (2001). The shortcoming of these information criteria is that when there are strong negative MA components in the error term, they tend to select a smaller k than that is necessary for unit root tests to have good size.

To account for the size distortions mentioned above, Ng and Perron (2001) have proposed modified versions of the AIC and the BIC (MAIC and MBIC) by using a penalty factor $\hat{\tau}_T(k)$ to correct underfitting. The MAIC and MBIC are defined respectively, as $\text{MAIC}(k) = \arg\min_{k\in[0,k_{\max}]}\{\ln(s_{ek}^2) + 2(\hat{\tau}_T(k)+k)/T^*\}$ and $MBIC(k) = \arg\min_{k\in[0,k_{\max}]}\{\ln(s_{ek}^2) + \ln T^*(\hat{\tau}_T(k) + k)/T^*\}$, where $\hat{\tau}_T(k) = (s_{ek}^2)^{-1}\widehat{\alpha}_0^2\sum_{t=k_{\max}+1}^{T}\widetilde{y}_{t-1}^2$, $\widehat{\alpha}_0$ is estimated using the augmented autoregression (6) and $s_{ek}^2 = (T-k_{\max})^{-1}\sum_{t=k_{\max}+1}^{T}\hat{e}_{tk}^2$. Ng and Perron (2001) showed that when a strong negative MA error exists, $\hat{\tau}_T(k)$ increases as k decreases. Therefore, $\hat{\tau}_T(k)$ can be used as a penalty factor for small k.

The last method is the so-called sequential $t-$test or recursive method proposed by Campbell and Perron (1991). To apply this method, we start from augmented regression (6) with $k_{\max} = int[(4 \times (T/100)^{1/4}]$. If the t-statistic associated with the $k_{\max}th$ lag is significant ($p-$value less than

0.1), then $k = k_{\max}$ is chosen. Otherwise we re-estimate the regression with $k_{\max} - 1$ lags, and so on, until we find the lag that has a significant t-statistic. Note that if $k = 0$ and no rejection is found, we select $k = 0$. This method has less size distortion than AIC and BIC when there is a strong negative MA error, but tends to overparameterize in the other cases.

5.2 Size and Power in Finite Samples

The performance of AIC is very poor and hence, not included. Furthermore, in order to save space, the results for the MBIC are also not included[7]. The critical values for model I and II, using k selected by the methods BIC, MAIC and t-sig and 1000 replications, are tabulated in Table 1 to 4. Table 1 and 2 give the critical values using the infimum method to choose break point, whereas Table 3 and 4 calculate the critical values using the supremum method.

Based on these critical values, we calculate finite sample size and power when $T = 100$, 200. The term v_t follows AR(1) and MA(1) processes, i.e., $v_t = \rho v_{t-1} + e_t$ and $v_t = e_t + \theta e_{t-1}$, respectively, and where $e_t \sim i.i.d.\ N(0, 1)$. The results are presented in Tables 5 to 8. We summarize the following two points from these results. First, when comparing different methods of selecting k, the MAIC has much acceptable size distortions than the BIC when there are strong negative moving average errors. For example, when $\theta = -0.8$, $T = 200$ in Table 6, the size distortions for MAIC are 0.140, 0.137, 0.140, 0.167 and 0.103, respectively. Using BIC, they are 0.874, 0.873, 0.871, 0.911, and 0.849, respectively. Second, when comparing different statistics, all tests except the ADF^{GLS} statistic have low power when there is strong negative autoregressive errors. For example, when $\rho = -0.8$, $T = 200$ in Table 6 and 8, the power for M^{GLS} and $P_{T\xi}^{GLS}$ tests are 2% at the most.

Overall, the results are very similar to those reported by Perron and Rodríguez (2003). However, unlike them, we found that power decreases for $T = 200$ when the errors are $i.i.d.$ and the MAIC is used to select the lag length. Given the fact that the term $\sum_{t=k_{\max}+1}^{T} \widetilde{y}_{t-1}^2$ enters in the formula of $\hat{\tau}_T(k)$, the loss of power is due to the role of the initial condition on the MAIC and consequently, in the selection of the lag length. Notice that in the case where the errors are $i.i.d.$, a $k = 0$ is sufficient. However, the initial condition increases $\hat{\tau}_T(k)$ which is a penalty factor. Given this fact, a higher k is selected, and power decreases consequently[8].

[7] All these results are available upon request.

[8] More insights concerning the role of the initial condition on the selection of the lag length are beyond the scope of this paper. It is the topic of a future research.

6 Empirical Applications

In a similar way as Perron and Rodríguez (2003), two time series from the Nelson-Plosser data set are examined. They are real wages (1900-1970) and common stock prices (1870-1970). A common characteristic that appears by observing the data is that they both exhibit a change in level and slope (see Figures 3 and 4). Therefore, the model II is used to test whether the null hypothesis of a unit root is rejected or not. The test results using information criteria to select lag k are tabulated in Tables 9a,b. We find that the break points selected are the same as those in Perron and Rodríguez (2003) and they are associated with major events. The fitted real wages series with a break in 1938 is graphed in Figure 3, and the stock prices series with a break in 1931 is graphed in Figure 4.

When using information criteria BIC and MAIC, most test statistics can reject the null hypothesis of a unit root at least at 5% of significance level. But comparing the results from those of Perron and Rodríguez (2003), there exist less evidences against the unit root null hypothesis. For example, the null hypothesis of a unit root for real wages is rejected at 1% of significance level in Perron and Rodríguez (2003) but at 2.5% of significance level in our case when using the supremum method and the MAIC for the MZ_t statistic. When using infimum method and ADF statistic, the null hypothesis of a unit root for the real wages is rejected at 5% for BIC and 2.5% for MAIC in Perron and Rodríguez (2003). In the present study, the null hypothesis is rejected at 10.0% when BIC is used and is rejected at 5.0% of significance level when MAIC is used. These evidences indicate that the power of these unit root tests may have decreased due to the new assumption on the initial condition.

The results for sequential t-statistic method are summarized in Tables 10a,b. When using infimum method, the null hypothesis of a unit root is rejected in all cases at least at 10.0%. When using the supremum method, the null hypothesis is not rejected for the time series of real wages.

In conclusion, the empirical evidence suggests rejection of the null hypothesis of a unit root but this evidence is weak in comparison to the results found in Perron and Rodríguez (2003).

7 Concluding Remarks

Changing the initial condition in the DGP has different impact under the null and alternative hypothesis in unit root testing (Elliott, 1999; Müller and Elliott, 2003). Under the null, the initial value change is equivalent to a mean shift. Therefore invariance method can be applied for various initial conditions and the tests are not changed. But this is not true under the alternative hypothesis, where invariant tests will have a different distribution as the initial condition changes, and hence, impacts on power performance are expected.

This paper examines M^{GLS}, ADF^{GLS}, $P_{T\xi}^{GLS}$ tests in the context of structural change when the initial observation is drawn from its unconditional distribution, in comparison with zero or fixed initial values as dealt with in Perron and Rodríguez (2003). As a result, we find asymptotic power loss. Consequently, one should be cautious when using unit root tests for the time series believed to start from an unconditional distribution. The finite sample size and power performance are also studied when the term v_t follows AR(1) and MA(1) processes, i.e., $v_t = \rho v_{t-1} + e_t$ and $v_t = e_t + \theta e_{t-1}$, respectively, and where $e_t \sim i.i.d.\ N(0,1)$. The performances are quite different when different procedures are used to select the order of autoregressions, but they are consistent with what standard literature predicts.

References

[1] Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* **61**, 821-856.

[2] Andrews, D. W. K., and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present only under the Alternative ," *Econometrica* **62**, 1383-1414.

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

[4] Banerjee, A., R. Lumsdaine and J. H. Stock (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis: Theory and International Evidence," *Journal of Business and Economics Statistics* **10**, 271-287.

[5] Campbell, J. Y. and P. Perron (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots" in Blanchard, O.J., Fischer, S. eds., *NBER Macroeconomics Annual* **6**, 141-201.

[6] Chan, N. H. and C. Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics* **15**, 1050-63.

[7] Christiano, L. J. (1992), "Searching for a Break in GNP," *Journal of Business and Economics Statistics* **10**, 237-250.

[8] Davidson J. (1994), "Stochastic Limit Theory," Oxford University Press.

[9] Dickey, D. A. and W. A. Fuller (1979), "Distribution of the Estimator for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* **74**, 427-431.

[10] Dufour, J. -M. and King, M. (1991), "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary Errors," *Journal of Econometrics* **47**, 115-143.

[11] Elliott, G. (1999), "Efficient Tests for a Unit Root When the Initial Observations is Drawn from its Unconditional Distribution," *International Economic Review* **40**, 767-783.

[12] Elliott, G., T. Rothenberg and J. H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica* **64**, 813-839.

[13] Müller, U. K. and G. Elliott (2003), "Tests for Unit Roots and the Initial Condition," *Econometrica* **71,** 1269-1286.

[14] Nelson, C. R. and C. I. Plosser (1982), "Trends and Random Walks in Macroeconomics Time Series: Some Evidence and Implications," *Journal of Monetary Economics* **10**, 139-162.

[15] Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association* **90**, 268-281.

[16] Ng, S. and Perron, P. (2001), " Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica* **69**, 1519-1554.

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

[18] Perron, P. (1997), "Further Evidence of Breaking Trend Functions in Macroeconomics Variables," *Journal of Econometrics* **80**, 355-385.

[19] Perron, P. and S. Ng (1996), "Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," *Review of Economics Studies* **63**, 435-463.

[20] Perron, P. and S. Ng (1998), "An Autoregressive Spectral Density Estimator at Frequency Zero for Nonstationarity Tests," *Econometric Theory* **14**, 560-603.

[21] Perron, P. and G. Rodríguez (2003), "GLS Detrending, Efficient Unit Root Tests and Structural Change," *Journal of Econometrics* **115**, 1-27.

[22] Phillips, P. C. B. (1987), "Time Series Regression with Unit Roots," *Econometrica* **55**, 277-302.

[23] Phillips, P. C. B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika* **75**, 335-346.

[24] Said, S. E. and Dickey, D. A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **71**, 599-608.

[25] Sargan, J. D. and Bhargava, A. (1983), "Testing Residuals from Least Squares Regression for being Generated by the Gaussian Random Walk" *Econometrica* **51**, 153-174.

[26] Stock, J. H. (1999), "A Class of Tests for Integration and Cointegration," in Engle, R.F. and H. White (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, 137-167.

[27] Vogelsang, T. J and P. Perron (1998), "Additional Tests for a Unit Root Allowing the Possibility of Breaks in the Trend Function," *International Economic Review* **39**, 1073-1100.

[28] Zivot, E. and D. W. K. Andrews (1992), "Further Evidence on the Great Crash, The Oil-Price Shock and the Unit Root Hypothesis," *Journal of Business and Economics Statistics* **10**, 251-270.

Appendix

Throughout, we use the next Lemma which is the same as Lemma 2 of Elliott (1999).

Lemma A.1. If u_t is given by equation (2), v_t satisfies Assumption B, $T(\alpha-1) = c < 0$, and the initial condition is drawn according to Assumption A, then

$$\begin{aligned} T^{-1/2}(u_{[Tr]} - u_1) &\Rightarrow \sigma[W_c(r) + (e^{cr} - 1)\xi \\ &\equiv \sigma N_c(r), \end{aligned}$$

where $W_c(r) = c\int_0^r e^{c(r-\lambda)}W(\lambda)d\lambda + W(r)$ is an Ornstein Uhlenbeck process, $W(r)$ is a standard Brownian motion, $\xi \sim N[0, (-2c)^{-1}]$ and is independent of $W_c(r)$, $[.]$ indicates the greatest lesser integer function, and σ is the spectral density of v_t at the frequency zero (scaled by 2π). Further, as $c < 0$ tends to zero, this is continuous in c and converges to $\sigma W(r)$.

Proof of Lemma A.1. The proof appears in Elliott (1999), and therefore, it is omitted.

Lemma A.2. Suppose y_t is generated by (1) to (3), the deterministic components given by Model I, the initial condition is defined by Assumption 1, and $\delta = T_B/T$ is the break point, then we have,

$$\begin{aligned} T^{-1/2}(\widehat{\mu}_1 - \mu_1) &\Rightarrow \sigma b_4, \\ T^{1/2}(\widehat{\beta}_1 - \beta_1) &\Rightarrow \sigma b_5, \\ T^{1/2}(\widehat{\beta}_2 - \beta_2) &\Rightarrow \sigma b_6. \end{aligned}$$

where the definitions of b_4, b_5, b_6 are given in the following proof.

Proof of Lemma A.2. In matrix notation, we have:

$$\Lambda^{-1}[\hat{\psi}(\delta) - \psi] = \left[\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right]^{-1}\left[\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime}\right]$$

where

$$\begin{aligned} z^{\bar{\alpha}\prime} &= \begin{bmatrix} \left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} & -\bar{c}T^{-1} & \cdots & -\bar{c}T^{-1} \\ \left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} & \cdots & -\bar{c}T^{-1}(t-1) + 1 & \cdots \\ 0 & \cdots & -\bar{c}T^{-1}(t-1) + \delta\bar{c} + 1 & \cdots \end{bmatrix}, \\ u^{\bar{\alpha}\prime} &= \left[\left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} u_1, (c-\bar{c})T^{-1}u_1 + v_2, \cdots, (c-\bar{c})T^{-1}u_{T-1} + v_T\right], \\ \Lambda &= diag(T^{1/2}, T^{-1/2}, T^{-1/2}). \end{aligned}$$

Let D be a 3×3 matrix which is the limiting distribution of $\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime} \Lambda$. We know that the limiting distributions of the elements $D_{11},\ D_{12}, D_{22}$ are the same as those in Elliott (1999); the limiting distributions of D_{23}, D_{33} have been calculated in Perron and Rodríguez (2003). That is, $D_{11} = \bar{c}^2 - 2\bar{c}$, $D_{12} = \frac{1}{2}\bar{c}^2 - \bar{c}$, $D_{22} = 1 + \frac{1}{3}\bar{c}^2 - \bar{c}$, $D_{23} = 1 - \delta - \bar{c} + \bar{c}\delta - \frac{1}{2}\bar{c}^2\delta + \frac{1}{3}\bar{c}^2 + \frac{1}{6}\bar{c}^2\delta^3 \equiv m$, $D_{33} = 1 - \delta - \bar{c}\left(1-\delta\right)^2 + \frac{1}{3}\bar{c}^2\left(1-\delta\right)^3 \equiv d$. Therefore we only need to calculate the limiting distribution of the term D_{13} which is given by:

$$
\begin{aligned}
D_{13} &= \left[\begin{array}{cccccc} \left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} & -\bar{c}/T & \cdots & -\bar{c}/T & -\bar{c}/T & \cdots & -\bar{c}/T \end{array}\right] \\
&\quad \times \left[\begin{array}{ccccccc} 0 & \cdots & 1 & 1-\bar{c}/T & 1-2\bar{c}/T & \cdots & 1-\left(T-T_B-1\right)\bar{c}/T \end{array}\right]' \\
&= -\frac{\bar{c}}{T}\left[\sum_{t=T_B+1}^{T}\left[1-\frac{\bar{c}}{T}\left(t-T_B-1\right)\right]\right] \\
&\Rightarrow -\bar{c}\left(1-\delta\right) + \frac{1}{2}\bar{c}^2\left(1-\delta\right)^2.
\end{aligned}
$$

Next, we calculate the limiting distribution of $\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime}$. We have

$$
\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime} = \left[\begin{array}{lll} T^{1/2} & 0 & 0 \\ 0 & T^{-1/2} & 0 \\ 0 & 0 & T^{-1/2} \end{array}\right]
$$

$$
\times \left[\begin{array}{llll} \left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} & -\bar{c}T^{-1} & \cdots & -\bar{c}T^{-1} \\ \left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} & \cdots & -\bar{c}T^{-1}\left(t-1\right)+1 & \cdots \\ 0 & \cdots & -\bar{c}T^{-1}\left(t-1\right)+\delta\bar{c}+1 & \cdots \end{array}\right]
$$

$$
\times \left[\left(-2\bar{c}/T - \bar{c}^2/T^2\right)^{1/2} u_1, \left(c-\bar{c}\right)T^{-1}u_1 + v_2, \cdots \left(c-\bar{c}\right)T^{-1}u_{T-1} + v_T,\right]'
$$

$$
= \left[\begin{array}{l} T^{\frac{1}{2}}\left(-\frac{2\bar{c}}{T} - \frac{\bar{c}^2}{T^2}\right)u_1 + T^{\frac{1}{2}}\sum_{t=2}^{T}\left(-\bar{c}T^{-1}\right)\left[\left(c-\bar{c}\right)T^{-1}u_{t-1}+v_t\right] \\ T^{-\frac{1}{2}}\left(-\frac{2\bar{c}}{T} - \frac{\bar{c}^2}{T^2}\right)u_1 + T^{-\frac{1}{2}}\sum_{t=2}^{T}\left[-\bar{c}T^{-1}\left(t-1\right)+1\right]\left[\left(c-\bar{c}\right)T^{-1}u_{t-1}+v_t\right] \\ T^{-\frac{1}{2}}\sum_{t=T_B+1}^{T}\left[-\bar{c}T^{-1}\left(t-1\right)+\delta\bar{c}+1\right]\left[\left(c-\bar{c}\right)T^{-1}u_{t-1}+v_t\right] \end{array}\right].
$$

The first element of the above 3×1 vector may be expressed as:

$$
\begin{aligned}
&= -2\overline{c}T^{-1/2}u_1 - \overline{c}^2T^{-3/2}u_1 + T^{1/2}\sum_{t=2}^{T}(-\overline{c}T^{-1})(c-\overline{c})T^{-1}u_{t-1} \\
&\quad +T^{1/2}\sum_{t=2}^{T}(-\overline{c}T^{-1})v_t \\
&= -2\overline{c}T^{-1/2}u_1 - \overline{c}^2T^{-3/2}u_1 - T^{-3/2}\overline{c}(c-\overline{c})\sum_{t=2}^{T}u_{t-1} \\
&\quad -\overline{c}T^{-1/2}\sum_{t=2}^{T}v_t \\
&\Rightarrow \sigma\{-2\overline{c}\xi - \overline{c}(c-\overline{c})\int_0^1 W_c(r) - \overline{c}W(1)\} \\
&\equiv \sigma b_1. \qquad (1)
\end{aligned}
$$

By the same token, the second element may be written as

$$
\begin{aligned}
&= -2T^{-3/2}\bar{c}u_1 - \bar{c}^2T^{-5/2}u_1 + \\
&\quad T^{-1/2}\sum_{t=2}^{T}\{-\bar{c}T^{-1}(t-1)(c-\bar{c})T^{-1}u_{t-1} - \bar{c}T^{-1}(t-1)v_t \\
&\quad +(c-\bar{c})T^{-1}u_{t-1} + v_t\} \\
&= -2T^{-3/2}\bar{c}u_1 - \bar{c}^2T^{-5/2}u_1 - \bar{c}(c-\bar{c})T^{-5/2}\sum_{t=2}^{T}(t-1)u_{t-1} \\
&\quad -\bar{c}T^{-3/2}\sum_{t=2}^{T}(t-1)v_t + (c-\bar{c})T^{-3/2}\sum_{t=2}^{T}u_{t-1} + T^{-1/2}\sum_{t=2}^{T}v_t \\
&\Rightarrow \sigma\{-\bar{c}(c-\bar{c})\int_0^1 rW_c(r) - \bar{c}[W(1) - \int_0^1 W_c(r)] \\
&\quad +(c-\bar{c})\int_0^1 W_c(r) + W(1)\} \\
&\equiv \sigma\{(1-\bar{c})W(1) + c\int_0^1 W_c(r) - \bar{c}(c-\bar{c})\int_0^1 rW_c(r)\} \\
&\equiv \sigma b_2. \qquad (2)
\end{aligned}
$$

Finally, the third element is given by:

$$
\begin{aligned}
&= T^{-1/2}\sum_{t=T_B+1}^{T}\{-\bar{c}T^{-1}(t-1)(c-\bar{c})T^{-1}u_{t-1}-\bar{c}T^{-1}(t-1)v_t \\
&\quad +\delta\bar{c}(c-\bar{c})T^{-1}u_{t-1}+\delta\bar{c}v_t+(c-\bar{c})T^{-1}u_{t-1}+v_t\} \\
&= T^{-1/2}\sum_{t=T_B+1}^{T}\{-\bar{c}T^{-1}(c-\bar{c})T^{-1}tu_{t-1}+\bar{c}T^{-1}(c-\bar{c})T^{-1}u_{t-1} \\
&\quad -\bar{c}T^{-1}tv_t+\bar{c}T^{-1}v_t+\delta\bar{c}(c-\bar{c})T^{-1}u_{t-1}+\delta\bar{c}v_t \\
&\quad +(c-\bar{c})T^{-1}u_{t-1}+v_t\} \\
&\Rightarrow \sigma\{(1-\bar{c}+\delta\bar{c})W(1)-W(\delta)-\bar{c}(c-\bar{c})\int_{\delta}^{1}rW_c(r) \\
&\quad +\delta\bar{c}(c-\bar{c})\int_{\delta}^{1}W_c(r)+c\int_{\delta}^{1}W_c(r)\} \\
&\equiv \sigma b_3. \qquad (3)
\end{aligned}
$$

Therefore, using D_{ij}, for $i,j=1,2,3$ and (1), (2), and (3), we have:

$$
\begin{aligned}
&\Lambda^{-1}\begin{pmatrix}\hat{\mu}_1-\mu_1\\ \hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{pmatrix} \\
&= \left(\Lambda z^{\bar{\alpha}}z^{\bar{\alpha}\prime}\Lambda\right)^{-1}\Lambda z^{\bar{\alpha}}u^{\bar{\alpha}\prime} \\
&\Rightarrow \begin{bmatrix}\bar{c}^2-2\bar{c} & \frac{1}{2}\bar{c}^2-\bar{c} & -\bar{c}(1-\delta)+\frac{1}{2}\bar{c}^2(1-\delta)^2\\ \frac{1}{2}\bar{c}^2-\bar{c} & 1+\frac{1}{3}\bar{c}^2-\bar{c} & m\\ -\bar{c}(1-\delta)+\frac{1}{2}\bar{c}^2(1-\delta)^2 & m & d\end{bmatrix}^{-1} \\
&\quad \times\begin{pmatrix}\sigma b_1\\ \sigma b_2\\ \sigma b_3\end{pmatrix} \\
&\equiv \sigma\begin{pmatrix}b_4\\ b_5\\ b_6\end{pmatrix},
\end{aligned}
$$

which completes the proof.■

Lemma A.3. Suppose $\{y_t\}$ is generated by (1) to (3), the set of deterministic components is given by that of Model II, the initial condition is defined in Condition A, δ is the break point. Let $\hat{\psi}(\delta)$ be the estimates of the coefficients of (4), then the results of Lemma A.3 still hold with the addition that $\hat{\mu}_2-\mu_2\Rightarrow\lim_{T\to\infty}\kappa(\bar{c},\delta)v_{T_B}\equiv v^*$, where $\kappa(\bar{c},\delta)$ is defined in the proof.

Proof of Lemma A.3. We have

$$\Lambda^{-1}\begin{pmatrix} \hat{\mu}_1-\mu_1 \\ \hat{\mu}_2-\mu_2 \\ \hat{\beta}_1-\beta_1 \\ \hat{\beta}_2-\beta_2 \end{pmatrix} = \left(\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)^{-1}\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime},$$

where $\Lambda = diag\left(T^{1/2}, 1, T^{-1/2}, T^{-1/2}\right)$, and

$$z^{\bar{\alpha}\prime} = \begin{bmatrix} \left(-2\bar{c}/T-\bar{c}^2/T^2\right)^{1/2} & -\frac{\bar{c}}{T} & \cdot & \cdots & & \cdot & -\frac{\bar{c}}{T} \\ 0 & \cdots & 1 & -\frac{\bar{c}}{T} & & \cdots & -\frac{\bar{c}}{T} \\ \left(-2\bar{c}/T-\bar{c}^2/T^2\right)^{1/2} & 1-\frac{\bar{c}}{T} & \cdots & 1-\frac{(t-1)\bar{c}}{T} & & \cdots & 1-\frac{(T-1)\bar{c}}{T} \\ 0 & \cdots & \cdots & 1-\frac{(t-1)\bar{c}}{T}+\delta\bar{c} & & \cdots & 1-\frac{(T-1)\bar{c}}{T}+\delta\bar{c} \end{bmatrix}.$$

As before, we first calculate the limiting distribution of $D = z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda$, then that of $\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime}$. From the proof of Lemma A.3, we know that $D_{11} \Rightarrow \bar{c}^2 - 2\bar{c}$, $D_{13} \Rightarrow \frac{\bar{c}^2}{2} - \bar{c}$, $D_{14} \Rightarrow -\bar{c}(1-\delta) + \frac{\bar{c}^2}{2}(1-\delta)^2$, $D_{33} \Rightarrow 1 + \frac{1}{3}\bar{c}^2 - \bar{c} \equiv a$, $D_{34} \Rightarrow m, D_{44} \Rightarrow d$. We only need to calculate the terms D_{12}, D_{22}, D_{23}, D_{24}, which may be found in the following way:

$$\begin{aligned} D_{12} &= T^{1/2}\{-\frac{\bar{c}}{T} + \frac{\bar{c}^2}{T^2} + \cdots + \frac{\bar{c}^2}{T^2}\} \\ &= T^{\frac{1}{2}}\{-\frac{\bar{c}}{T} + \frac{\bar{c}^2}{T^2}(T-T_B)\} \\ &\Rightarrow 0, \end{aligned}$$

$$\begin{aligned} D_{22} &= 1 + \frac{\bar{c}^2}{T^2} + \cdots + \frac{\bar{c}^2}{T^2} \\ &= 1 + \frac{\bar{c}^2}{T^2}(T-T_B) \\ &\Rightarrow 1, \end{aligned}$$

$$\begin{aligned} D_{23} &= T^{-1/2}\{1-(T_B-1)\frac{\bar{c}}{T} - \frac{\bar{c}}{T}\sum_{t=T_B+1}^{T}[1-(t-1)\frac{\bar{c}}{T}]\} \\ &= T^{-1/2}\{1-(T_B-1)\frac{\bar{c}}{T} - \frac{\bar{c}}{T}(T-T_B) \\ &\quad + \frac{\bar{c}^2}{T^2}\sum_{t=T_B+1}^{T}(t-1)\} \\ &\Rightarrow 0, \end{aligned}$$

$$
\begin{aligned}
D_{24} &= T^{-1/2}\{1-(T_B-1)\frac{\bar{c}}{T}+\delta\bar{c} \\
&\quad -\frac{\bar{c}}{T}\sum_{t=T_B+1}^{T}[1-(t-1)\frac{\bar{c}}{T}+\delta\bar{c}]\} \\
&= T^{-1/2}\{1+\frac{\bar{c}}{T}-\frac{\bar{c}}{T}(T-T_B)+\frac{\bar{c}^2}{T^2}\sum_{t=T_B+1}^{T}(t-1) \\
&\quad -\frac{\delta\bar{c}^2}{T}(T-T_B)\} \\
&\Rightarrow 0.
\end{aligned}
$$

For the limiting distribution of $\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime}$, the first, third and fourth elements are already calculated in the proof for Lemma A.3. We only need to calculate the second element which is:

$$
\begin{aligned}
& \begin{bmatrix} 0 & \cdots & 1 & -\frac{\bar{c}}{T} & \cdots & -\frac{\bar{c}}{T} \end{bmatrix} \\
& \times \left[\left(-\frac{2\bar{c}}{T}-\frac{\bar{c}^2}{T^2}\right)^{1/2} u_1, \cdots (c-\bar{c})\, T^{-1} u_{t-1}+v_t, \cdots\right]' \\
= \; & (c-\bar{c})\, T^{-1} u_{T_B-1} + v_{T_B} - \frac{\bar{c}}{T}\sum_{t=T_B+1}^{T}\left[(c-\bar{c})\, T^{-1} u_{t-1}+v_t\right] \\
\Rightarrow \; & v_{T_B}.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
& \left(\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)^{-1}\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime} \\
\Rightarrow \; & \begin{bmatrix}
\bar{c}^2-2\bar{c} & 0 & \frac{\bar{c}^2}{2}-\bar{c} & -\bar{c}(1-\delta)+\frac{\bar{c}^2}{2}(1-\delta)^2 \\
0 & 1 & 0 & 0 \\
\frac{\bar{c}^2}{2}-\bar{c} & 0 & a & m \\
-\bar{c}(1-\delta)+\frac{\bar{c}^2}{2}(1-\delta)^2 & 0 & m & d
\end{bmatrix}^{-1} \\
& \times \begin{bmatrix} \sigma b_1 \\ v_{T_B} \\ \sigma b_2 \\ \sigma b_3 \end{bmatrix}.
\end{aligned}
$$

According to matrix algebra, the second element of the resulting matrix is equal to:

$$
\kappa(\bar{c},\delta)\, v_{T_B} = \frac{\kappa^*(\bar{c},\delta)\, v_{T_B}}{(\bar{c}^2-2\bar{c})\, ad-\left(\frac{\bar{c}^2}{2}-\bar{c}\right)^2 d},
$$

with $\kappa^*(\bar{c},\delta) = \left(\bar{c}^2 - 2\bar{c}\right) ad + 2m\left(\bar{c}^2 - 2\bar{c}\right)[-\bar{c}(1-\delta) + \frac{\bar{c}^2}{2}(1-\delta)^2] - [-\bar{c}(1-\delta) + \frac{\bar{c}^2}{2}(1-\delta)^2]^2 a - m^2\left(\bar{c}^2 - 2\bar{c}\right) - d(\frac{\bar{c}^2}{2} - \bar{c})^2$. That is, $\hat{\mu}_2 - \mu_2 \Rightarrow \lim_{t\to\infty} \kappa(\bar{c},\delta)\, v_{T_B} \equiv v^*$.■

Proof of Theorem 1. We only show the proof of MZ_{α}^{GLS} for Model I in detail. The proof for Model II and other statistics follows analogously. First, we calculate the limiting distribution of $T^{-1}\tilde{y}_T^2$ as follows:

$$\begin{aligned}
T^{-1}\tilde{y}_T^2 \;=\;& T^{-1}\left\{u_T - [(\widehat{\mu}_1 - \mu_1) + (\widehat{\beta}_1 - \beta_1)T + (\widehat{\beta}_2 - \beta_2)1(\cdot)(T - T\delta)]\right\}^2 \\
& T^{-1}\{u_T^2 + (\widehat{\mu}_1 - \mu_1)^2 + (\widehat{\beta}_1 - \beta_1)^2 T^2 \\
& +(\widehat{\beta}_2 - \beta_2)^2 1(\cdot)(T - T\delta)^2 - 2u_T(\widehat{\mu}_1 - \mu_1) \\
& +2(\widehat{\beta}_1 - \beta_1)T(\widehat{\beta}_2 - \beta_2)1(\cdot)(T - T\delta) \\
& -2u_T(\widehat{\beta}_1 - \beta_1)T - 2u_T(\widehat{\beta}_2 - \beta_2)1(\cdot)(T - T\delta) \\
& +2(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_1 - \beta_1)T \\
& +2(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_2 - \beta_2)1(\cdot)(T - T\delta)\}.
\end{aligned}$$

It is worth to note that the limiting distributions of some terms have been found already in Perron and Rodríguez (2003). The other terms are calculated as follow:

1. $T^{-1}(\widehat{\mu}_1 - \mu_1)^2 = [T^{-1/2}(\widehat{\mu}_1 - \mu_1)]^2 \Rightarrow \sigma^2 b_4^2$.
2. $-2T^{-1}u_T(\widehat{\mu}_1 - \mu_1) = -2(T^{-1/2}u_T)[T^{-1/2}(\widehat{\mu}_1 - \mu_1)] \Rightarrow -2\sigma^2 W_c(1) b_4$.
3. $2T^{-1}(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_1 - \beta_1)T = 2[T^{-1/2}(\widehat{\mu}_1 - \mu_1)][T^{1/2}(\widehat{\beta}_1 - \beta_1)] \Rightarrow 2\sigma^2 b_4 b_5$.
4. $2T^{-1}(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_2 - \beta_2)(T - T\delta) = 2(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_2 - \beta_2) - 2(\widehat{\mu}_1 - \mu_1)(\widehat{\beta}_2 - \beta_2)\delta = 2[T^{-1/2}(\widehat{\mu}_1 - \mu_1)][T^{1/2}(\widehat{\beta}_2 - \beta_2)] - 2T^{-1/2}(\widehat{\mu}_1 - \mu_1)T^{1/2}(\widehat{\beta}_2 - \beta_2)\delta] \Rightarrow 2\sigma^2 b_4 b_6(1 - \delta)$.

Therefore,

$$
\begin{aligned}
T^{-1}\tilde{y}_T^2 &\Rightarrow \sigma^2 W_c^2(1) + \sigma^2 b_4^2 + \sigma^2 b_5^2 + \sigma^2 b_6^2(1-\delta)^2 \\
&\quad -2\sigma^2 W_c(1) b_4 + 2\sigma^2 b_5 b_6 (1-\delta) - 2\sigma^2 b_5 W_c(1) \\
&\quad -2\sigma^2 b_6 W_c(1)(1-\delta) + 2\sigma^2 b_4 b_5 + 2\sigma^2 b_4 b_6 (1-\delta) \\
&= \sigma^2\{[W_c^2(1) - 2b_5 W_c(1) + b_5^2 + b_4^2 - 2b_4 W_c(1) + 2b_4 b_5] \\
&\quad -[2b_6(1-\delta)W_c(1) - 2b_5 b_6(1-\delta) - b_6^2(1-\delta)^2 - 2b_4 b_6\,(1-\delta)] \\
&= \sigma^2\{[W_c(1) - b_4 - b_5]^2 - 2[b_6(1-\delta)][W_c(1) - b_4 - b_5 \\
&\quad -\frac{1}{2}(1-\delta)b_6] \\
&\equiv \sigma^2 \left\{ V_{c\bar{c}}^{(1)}(1,\delta)^2 - 2V_{c\bar{c}}^{(2)}(1,\delta) \right\}. \qquad (4)
\end{aligned}
$$

where $V_{c\bar{c}}^{(1)}(1,\delta) = W_c(1) - b_4 - b_5$, and $V_{c\bar{c}}^{(2)}(1,\delta) = [b_6(1-\delta)][W_c(1) - b_4 - b_5 - \frac{1}{2}(1-\delta)b_6]$.
Next we calculate the limiting distribution of $T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2$. Using the above results and by the Continuous Mapping Theorem (CMT), we have

$$
2T^{-2}\sum_{t=1}^{T}\tilde{y}_t^2 \Rightarrow 2\sigma^2\{\int_0^1 V_{c\bar{c}}^{(1)}(r,\delta)^2 dr - 2\int_\delta^1 V_{c\bar{c}}^{(2)}(r,\delta)dr\}. \qquad (5)
$$

Then by substituting (4), (5) into (7) and by using the fact that s^2 is a consistent estimate of σ^2, the proof is complete.■

Proof of Theorem 2. Here we only give the proof for Model I. Defining $\bar{M}_T\,(c,\bar{c},\delta) = \left(u^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)\left(\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)^{-1}(\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime})$, we have $S\,(\bar{\alpha},\delta) = u^{\bar{\alpha}} u^{\bar{\alpha}\prime} - \bar{M}_T\,(c,0,\delta)$ and $S\,(1) = u^1 u^{1\prime} - \bar{M}_T\,(0,0,\delta)$. By definition, $u^{\bar{\alpha}} = [(1-\bar{\alpha}^2)^{1/2} u_{1,}(1-\bar{\alpha}L)u_2,\cdot\cdot\cdot,(1-\bar{\alpha}L)_T]$, and $u^1 = [0_,(1-L)u_2,\cdot\cdot\cdot,(1-L)_T]$. Simple algebra shows that

$$
\begin{aligned}
&u^{\bar{\alpha}} u^{\bar{\alpha}\prime} - u^1 u^{1\prime} \\
\Rightarrow\; &-2\bar{c}\xi^2 + \left(\bar{c}^2 - 2c\bar{c}\right)\int_0^1 W_c^2\,(r)\,dr - 2\bar{c}\int_0^1 W_c\,(r)\,dr.
\end{aligned}
$$

Notice that all elements of $\bar{M}_T\,(c,\bar{c},\delta) = \left(u^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)\left(\Lambda z^{\bar{\alpha}} z^{\bar{\alpha}\prime}\Lambda\right)^{-1}(\Lambda z^{\bar{\alpha}} u^{\bar{\alpha}\prime})$

have already been calculated. Therefore,

$$\begin{aligned}
& \bar{M}_T\left(c, \bar{c}, \delta\right) \\
= & \left[\begin{array}{lll} \sigma b_1 & \sigma b_2 & \sigma b_3 \end{array}\right] \\
& \times \left[\begin{array}{lll} \bar{c}^2 - 2\bar{c} & \frac{1}{2}\bar{c}^2 - \bar{c} & -\bar{c}\left(1-\delta\right) + \frac{1}{2}\bar{c}^2\left(1-\delta\right)^2 \\ \frac{1}{2}\bar{c}^2 - \bar{c} & 1 + \frac{1}{3}\bar{c}^2 - \bar{c} & m \\ -\bar{c}\left(1-\delta\right) + \frac{1}{2}\bar{c}^2\left(1-\delta\right)^2 & m & d \end{array}\right]^{-1} \\
& \times \left[\begin{array}{l} \sigma b_1 \\ \sigma b_2 \\ \sigma b_3 \end{array}\right].
\end{aligned}$$

The calculation of $\bar{M}_T\left(0, 0, \delta\right)$ follows the arguments of Perron and Rodríguez (2003) since the initial value is not changed under the null. Using infimum method to estimate the break point, the limiting distribution of the feasible point optimal statistic is given by

$$\begin{aligned}
P_{T\xi}^{GLS}(c, \bar{c}) \quad \Rightarrow & \sup_{\delta \in [\varepsilon, 1-\varepsilon]} \bar{M}(c, 0, \delta) - \sup_{\delta \in [\varepsilon, 1-\varepsilon]} \bar{M}(c, \bar{c}, \delta) \\
& -2\bar{c}\xi^2 - 2\bar{c}\int_0^1 W_c(r)dW(r) + (\bar{c}^2 - 2\bar{c}c)\int_0^1 W_c(r)^2 dr - \bar{c} \\
\equiv & \ J_{P_{T\xi}}^{GLS}(c, \bar{c}).
\end{aligned}$$

and this completes the proof.■

Table 1. Critical Values for M_{ξ}^{GLS} and ADF_{ξ}^{GLS}, and $P_{T\xi}^{GLS}$ tests choosing T_B minimizing the statistics; Model I ($\bar{c} = -24$ when constructing the tests and s^2)

Test	Size	$T = \infty$	$T = 100$	$T = 100$			$T = 200$		
			$k = 0$	BIC	MAIC	t-sig	BIC	MAIC	t-sig
$MZ_{\alpha\xi}$	.01	-43.210	-30.422	-43.919	-30.876	-261.061	-39.131	-35.537	-60.412
	.025	-37.283	-27.589	-35.735	-27.956	-146.495	-33.697	-31.174	-51.096
	.05	-33.366	-25.507	-30.209	-25.379	-75.534	-30.185	-28.461	-43.068
	.10	-28.788	-22.994	-25.817	-22.492	-50.765	-27.080	-24.545	-32.537
	.20	-24.105	-20.016	-21.408	-19.368	-34.711	-22.068	-20.815	-26.767
MSB_{ξ}	.01	0.107	0.127	0.106	0.125	0.043	0.113	0.118	0.090
	.025	0.115	0.133	0.118	0.132	0.058	0.120	0.126	0.098
	.05	0.121	0.138	0.127	0.139	0.081	0.127	0.131	0.107
	.10	0.130	0.145	0.138	0.147	0.098	0.135	0.142	0.123
	.20	0.142	0.155	0.150	0.159	0.118	0.149	0.152	0.135
$MZ_{t\xi}$	.01	-4.621	-3.874	-4.674	-3.894	-11.408	-4.421	-4.213	-5.468
	.025	-4.300	-3.691	-4.217	-3.701	-8.525	-4.067	-3.923	-5.014
	.05	-4.064	-3.545	-3.863	-3.541	-6.143	-3.876	-3.769	-4.601
	.10	-3.766	-3.367	-3.577	-3.340	-5.033	-3.657	-3.455	-4.020
	.20	-3.449	-3.132	-3.240	-3.075	-4.129	-3.299	-3.205	-3.607
ADF_{ξ}	.01	-4.621	-4.919	-5.122	-4.980	-5.297	-4.889	-4.790	-5.067
	.025	-4.300	-4.562	-4.804	-4.538	-4.984	-4.530	-4.396	-4.675
	.05	-4.064	-4.304	-4.488	-4.230	-4.655	-4.256	-4.172	-4.432
	.10	-3.766	-3.997	-4.176	-3.865	-4.292	-4.046	-3.819	-4.149
	.20	-3.449	-3.673	-3.755	-3.532	-3.929	-3.625	-3.479	-3.746
$P_{T\xi}$	.01	6.967	7.600	7.379	9.948	1.662	8.054	9.021	5.340
	.025	8.065	9.149	8.950	11.340	2.935	9.260	10.008	6.399
	.05	9.340	10.470	10.430	12.118	4.509	10.161	10.846	7.681
	.10	10.866	12.135	12.016	13.479	7.129	11.755	12.962	10.003
	.20	13.110	14.401	14.413	16.004	9.934	14.088	15.129	11.962

Table 2. Critical Values for M_ξ^{GLS} and ADF_ξ^{GLS}, and $P_{T\xi}^{GLS}$ tests choosing T_B minimizing the statistics; Model II ($\bar{c} = -24$ when constructing the tests and s^2)

Test	Size	$T = \infty$	$T = 100$	$T = 100$			$T = 200$		
			$k = 0$	BIC	MAIC	t-sig	BIC	MAIC	t-sig
$MZ_{\alpha\xi}$	.01	-43.210	-32.049	-92.017	-33.065	-529.130	-44.641	-36.079	-70.335
	.025	-37.283	-29.436	-43.719	-30.396	-220.145	-34.854	-32.217	-58.017
	.05	-33.366	-27.214	-36.687	-27.039	-119.088	-32.066	-29.773	-49.142
	.10	-28.788	-24.566	-30.082	-24.686	-73.058	-28.546	-26.219	-36.865
	.20	-24.105	-21.600	-24.610	-21.238	-47.797	-23.533	-21.852	-30.140
MSB_ξ	.01	0.107	0.123	0.073	0.122	0.030	0.105	0.117	0.084
	.025	0.115	0.129	0.106	0.127	0.047	0.117	0.124	0.092
	.05	0.121	0.134	0.116	0.133	0.064	0.124	0.129	0.100
	.10	0.130	0.141	0.127	0.141	0.082	0.131	0.136	0.115
	.20	0.142	0.150	0.141	0.151	0.101	0.144	0.148	0.127
$MZ_{t\xi}$	.01	-4.621	-3.977	-6.763	-4.034	-16.265	-4.716	-4.233	-5.914
	.025	-4.300	-3.805	-4.653	-3.878	-10.489	-4.164	-3.971	-5.367
	.05	-4.064	-3.660	-4.254	-3.653	-7.710	-3.963	-3.854	-4.913
	.10	-3.766	-3.475	-3.836	-3.480	-6.014	-3.749	-3.596	-4.274
	.20	-3.449	-3.255	-3.465	-3.216	-4.832	-3.403	-3.277	-3.848
ADF_ξ	.01	-4.621	-5.109	-5.281	-5.072	-5.518	-5.065	-4.811	-5.187
	.025	-4.300	-4.758	-5.068	-4.685	-5.227	-4.661	-4.509	-4.841
	.05	-4.064	-4.492	-4.739	-4.381	-4.945	-4.385	-4.243	-4.570
	.10	-3.766	-4.183	-4.402	-4.103	-4.576	-4.121	-3.964	-4.260
	.20	-3.449	-3.845	-3.987	-3.716	-4.163	-3.754	-3.593	-3.894
$P_{T\xi}$	.01	6.967	9.136	7.824	10.761	2.276	8.964	10.195	5.307
	.025	8.065	10.409	9.353	12.174	3.540	10.357	11.109	6.926
	.05	9.340	11.600	11.087	13.250	4.853	11.388	11.911	8.345
	.10	10.866	13.208	13.061	14.572	7.415	12.617	13.761	10.457
	.20	13.110	15.232	15.255	16.739	10.394	14.842	15.904	12.765

Table 3. Critical Values for M_ξ^{GLS} and ADF_ξ^{GLS} tests choosing T_B maximizing $|t_{\widehat{\beta}_2}|$; Model I
($\bar{c} = -24$ when constructing the tests and s^2)

Test	Size	$T = \infty$	$T = 100$	$T = 100$			$T = 200$		
			$k = 0$	BIC	MAIC	t-sig	BIC	MAIC	t-sig
$MZ_{\alpha\xi}$	.01	-42.432	-30.876	-40.049	-30.424	-187.405	-38.111	-34.804	-57.572
	.025	-36.810	-28.678	-33.988	-27.585	-99.541	-33.506	-30.694	-50.102
	.05	-32.689	-25.717	-29.498	-24.982	-66.232	-29.851	-27.964	-40.333
	.10	-28.266	-23.204	-24.967	-21.966	-42.880	-26.705	-24.048	-31.524
	.20	-23.748	-19.986	-20.751	-18.829	-30.353	-21.764	-20.405	-25.718
MSB_ξ	.01	0.107	0.126	0.110	0.127	0.051	0.114	0.118	0.092
	.025	0.115	0.131	0.121	0.132	0.070	0.121	0.126	0.099
	.05	0.122	0.138	0.129	0.139	0.086	0.129	0.133	0.111
	.10	0.131	0.145	0.139	0.149	0.107	0.136	0.143	0.125
	.20	0.143	0.156	0.153	0.160	0.126	0.149	0.154	0.137
$MZ_{t\xi}$	.01	-4.579	-3.928	-4.440	-3.862	-9.653	-4.355	-4.171	-5.342
	.025	-4.263	-3.750	-4.100	-3.675	-7.008	-4.059	-3.899	-4.970
	.05	-4.019	-3.575	-3.784	-3.514	-5.751	-3.845	-3.734	-4.473
	.10	-3.737	-3.381	-3.506	-3.291	-4.600	-3.634	-3.415	-3.926
	.20	-3.423	-3.124	-3.836	-3.008	-3.875	-3.274	-3.162	-3.560
ADF_ξ	.01	-4.579	-4.981	-5.103	-4.897	-5.205	-4.788	-4.606	-4.941
	.025	-4.263	-4.615	-4.713	-4.435	-4.928	-4.506	-4.324	-4.597
	.05	-4.019	-4.366	-4.427	-4.163	-4.543	-4.226	-4.081	-4.388
	.10	-3.737	-3.960	-4.076	-3.814	-4.226	-3.989	-3.757	-4.079
	.20	-3.423	-3.621	-3.681	-3.458	-3.840	-3.567	-3.419	-3.689

Table 4. Critical Values for M_ξ^{GLS} and ADF_ξ^{GLS} tests choosing T_B maximizing $|t_{\widehat{\beta}_2}|$; Model II ($\bar{c} = -24$ when constructing the tests and s^2)

Test	Size	$T=\infty$	$T=100$	$T=100$			$T=200$		
			$k=0$	BIC	MAIC	t-sig	BIC	MAIC	t-sig
$MZ_{\alpha\xi}$	.01	-42.432	-28.651	-38.583	-27.367	-108.695	-34.731	-30.784	-54.661
	.025	-36.810	-26.634	-32.090	-24.442	-78.069	-31.155	-27.036	-40.413
	.05	-32.689	-23.638	-27.184	-21.780	-49.497	-26.721	-24.692	-33.122
	.10	-28.266	-20.865	-22.605	-19.560	-36.566	-22.881	-21.167	-27.747
	.20	-23.748	-17.970	-18.664	-16.975	-25.375	-19.313	-17.837	-21.795
MSB_ξ	.01	0.107	0.130	0.113	0.133	0.067	0.119	0.127	0.095
	.025	0.115	0.136	0.123	0.142	0.079	0.126	0.135	0.110
	.05	0.122	0.143	0.134	0.149	0.100	0.136	0.141	0.122
	.10	0.131	0.153	0.147	0.158	0.115	0.146	0.152	0.134
	.20	0.143	0.165	0.162	0.169	0.138	0.159	0.164	0.150
$MZ_{t\xi}$	.01	-4.578	-3.763	-4.385	-3.684	-7.367	-4.158	-3.909	-5.206
	.025	-4.265	-3.608	-3.977	-3.483	-6.240	-3.909	-3.675	-4.494
	.05	-4.012	-3.421	-3.660	-3.290	-4.972	-3.640	-3.491	-4.050
	.10	-3.737	-3.208	-3.326	-3.102	-4.253	-3.351	-3.219	-3.687
	.20	-3.423	-2.963	-3.028	-2.882	-3.555	-3.096	-2.957	-3.282
ADF_ξ	.01	-4.579	-4.674	-4.823	-4.529	-4.987	-4.631	-4.352	-4.681
	.025	-4.263	-4.403	-4.483	-4.121	-4.600	-4.286	-4.130	-4.434
	.05	-4.019	-4.062	-4.220	-3.784	-4.289	-4.030	-3.803	-4.163
	.10	-3.737	-3.750	-3.845	-3.548	-3.987	-3.724	-3.520	-3.801
	.20	-3.423	-3.468	-3.512	-3.275	-3.627	-3.381	-3.223	-3.476

Table 5. Size and Power when using Infimum Method for Model I; T=100
($\bar{c} = -24$ when constructing the tests and s^2)

		Size					Power				
	Criteria	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}	$P_{T\xi}$	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}	$P_{T\xi}$
$i.i.d.$	BIC	0.051	0.050	0.051	0.050	0.051	0.318	0.299	0.324	0.510	0.381
	MAIC	0.051	0.051	0.050	0.050	0.051	0.508	0.499	0.502	0.448	0.495
	t-sig	0.050	0.050	0.051	0.051	0.051	0.180	0.181	0.180	0.433	0.164
MA(1) Errors											
$\theta = -0.8$	BIC	0.937	0.935	0.938	0.976	0.932	0.996	0.996	0.996	1.000	0.995
	MAIC	0.399	0.397	0.400	0.378	0.339	0.827	0.829	0.825	0.829	0.707
	t-sig	0.107	0.108	0.107	0.798	0.088	0.297	0.299	0.297	0.961	0.263
$\theta = -0.4$	BIC	0.401	0.396	0.395	0.487	0.402	0.887	0.884	0.886	0.913	0.884
	MAIC	0.135	0.135	0.136	0.100	0.111	0.556	0.552	0.550	0.457	0.488
	t-sig	0.040	0.040	0.040	0.253	0.038	0.136	0.136	0.136	0.737	0.121
$\theta = 0.4$	BIC	0.214	0.212	0.213	0.095	0.201	0.729	0.720	0.732	0.524	0.735
	MAIC	0.100	0.105	0.094	0.006	0.048	0.134	0.129	0.137	0.065	0.101
	t-sig	0.066	0.066	0.066	0.061	0.064	0.189	0.190	0.187	0.404	0.191
$\theta = 0.8$	BIC	0.526	0.526	0.521	0.118	0.448	0.729	0.728	0.728	0.407	0.707
	MAIC	0.223	0.231	0.220	0.008	0.105	0.338	0.336	0.335	0.046	0.265
	t-sig	0.089	0.090	0.087	0.049	0.091	0.282	0.283	0.281	0.227	0.299
AR(1) Errors											
$\rho = -0.8$	BIC	0.018	0.017	0.018	0.047	0.013	0.040	0.039	0.040	0.456	0.033
	MAIC	0.000	0.000	0.000	0.043	0.000	0.008	0.007	0.008	0.330	0.010
	t-sig	0.020	0.020	0.019	0.041	0.015	0.072	0.073	0.072	0.386	0.057
$\rho = -0.4$	BIC	0.142	0.138	0.145	0.168	0.136	0.510	0.510	0.511	0.596	0.510
	MAIC	0.062	0.064	0.064	0.050	0.058	0.353	0.339	0.356	0.335	0.335
	t-sig	0.048	0.049	0.048	0.085	0.038	0.150	0.150	0.150	0.434	0.138
$\rho = 0.4$	BIC	0.155	0.152	0.149	0.053	0.131	0.538	0.527	0.543	0.296	0.562
	MAIC	0.103	0.110	0.098	0.013	0.055	0.074	0.073	0.075	0.021	0.048
	t-sig	0.076	0.079	0.075	0.046	0.059	0.170	0.170	0.167	0.262	0.165
$\rho = 0.8$	BIC	0.293	0.303	0.282	0.060	0.186	0.350	0.347	0.351	0.131	0.346
	MAIC	0.326	0.339	0.308	0.049	0.189	0.383	0.390	0.387	0.104	0.338
	t-sig	0.145	0.149	0.143	0.061	0.079	0.117	0.117	0.114	0.132	0.100

Table 6. Size and Power when using Infimum Method for Model I; T=200
($\bar{c} = -24$ when constructing the tests and s^2)

		Size					Power				
	Criteria	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}	$P_{T\xi}$	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}	$P_{T\xi}$
$i.i.d.$	BIC	0.050	0.050	0.050	0.050	0.050	0.506	0.497	0.495	0.517	0.486
	MAIC	0.051	0.050	0.051	0.050	0.051	0.431	0.423	0.420	0.372	0.423
	t-sig	0.050	0.051	0.051	0.051	0.051	0.200	0.204	0.213	0.432	0.242
MA(1) Errors											
$\theta = -0.8$	BIC	0.874	0.873	0.871	0.911	0.849	0.996	0.996	0.996	0.999	0.988
	MAIC	0.140	0.137	0.140	0.167	0.103	0.394	0.390	0.391	0.421	0.281
	t-sig	0.425	0.427	0.432	0.706	0.418	0.813	0.813	0.816	0.973	0.755
$\theta = -0.4$	BIC	0.236	0.238	0.228	0.241	0.213	0.849	0.842	0.845	0.861	0.809
	MAIC	0.097	0.095	0.088	0.058	0.076	0.416	0.410	0.406	0.326	0.363
	t-sig	0.056	0.058	0.058	0.123	0.061	0.337	0.338	0.346	0.633	0.384
$\theta = 0.4$	BIC	0.177	0.175	0.173	0.104	0.156	0.719	0.708	0.711	0.567	0.699
	MAIC	0.103	0.099	0.099	0.037	0.080	0.450	0.443	0.441	0.219	0.378
	t-sig	0.064	0.066	0.068	0.060	0.064	0.338	0.342	0.347	0.378	0.373
$\theta = 0.8$	BIC	0.338	0.338	0.328	0.097	0.264	0.730	0.726	0.723	0.393	0.695
	MAIC	0.189	0.195	0.188	0.006	0.114	0.430	0.423	0.426	0.083	0.366
	t-sig	0.053	0.053	0.054	0.030	0.055	0.267	0.269	0.279	0.190	0.310
AR(1) Errors											
$\rho = -0.8$	BIC	0.005	0.005	0.005	0.053	0.004	0.027	0.025	0.029	0.475	0.021
	MAIC	0.001	0.001	0.001	0.040	0.000	0.013	0.013	0.013	0.276	0.010
	t-sig	0.011	0.012	0.012	0.047	0.010	0.055	0.055	0.055	0.373	0.051
$\rho = -0.4$	BIC	0.053	0.051	0.050	0.063	0.047	0.438	0.429	0.427	0.473	0.402
	MAIC	0.044	0.043	0.044	0.041	0.041	0.346	0.339	0.336	0.310	0.302
	t-sig	0.034	0.036	0.034	0.049	0.033	0.161	0.166	0.167	0.385	0.1830
$\rho = 0.4$	BIC	0.115	0.111	0.108	0.057	0.097	0.571	0.558	0.562	0.401	0.551
	MAIC	0.109	0.108	0.104	0.041	0.090	0.454	0.448	0.439	0.246	0.404
	t-sig	0.054	0.052	0.052	0.047	0.052	0.269	0.273	0.276	0.350	0.315
$\rho = 0.8$	BIC	0.145	0.149	0.138	0.066	0.102	0.403	0.398	0.393	0.234	0.380
	MAIC	0.153	0.154	0.136	0.042	0.093	0.365	0.371	0.357	0.162	0.345
	t-sig	0.091	0.096	0.092	0.051	0.061	0.200	0.200	0.204	0.192	0.215

Table 7. Size and Power when using Supremum Method for Model I; T=100
($\bar{c} = -24$ when constructing the tests and s^2)

		Size				Power			
	Criteria	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}	$MZ_{\alpha\xi}$	MSB_{ξ}	$MZ_{t\xi}$	ADF_{ξ}
$i.i.d.$	BIC	0.050	0.050	0.050	0.050	0.335	0.336	0.381	0.510
	MAIC	0.051	0.049	0.050	0.051	0.495	0.465	0.490	0.445
	t-sig	0.050	0.050	0.050	0.051	0.169	0.170	0.169	0.451
MA(1) Errors									
$\theta = -0.8$	BIC	0.931	0.931	0.935	0.969	0.993	0.993	0.994	1.000
	MAIC	0.333	0.328	0.334	0.329	0.707	0.706	0.706	0.711
	t-sig	0.080	0.080	0.080	0.770	0.266	0.266	0.265	0.934
$\theta = -0.4$	BIC	0.395	0.401	0.401	0.466	0.876	0.875	0.883	0.906
	MAIC	0.109	0.102	0.111	0.087	0.490	0.481	0.489	0.412
	t-sig	0.038	0.039	0.038	0.234	0.121	0.121	0.121	0.704
$\theta = 0.4$	BIC	0.190	0.192	0.202	0.088	0.718	0.714	0.734	0.519
	MAIC	0.047	0.048	0.047	0.005	0.098	0.082	0.097	0.071
	t-sig	0.066	0.068	0.064	0.060	0.194	0.194	0.193	0.423
$\theta = 0.8$	BIC	0.440	0.444	0.451	0.098	0.703	0.705	0.708	0.397
	MAIC	0.103	0.101	0.102	0.003	0.238	0.224	0.239	0.035
	t-sig	0.092	0.092	0.091	0.050	0.299	0.302	0.299	0.223
AR(1) Errors									
$\rho = -0.8$	BIC	0.012	0.011	0.012	0.042	0.030	0.029	0.030	0.417
	MAIC	0.000	0.000	0.000	0.038	0.006	0.005	0.008	0.279
	t-sig	0.014	0.014	0.014	0.042	0.057	0.057	0.057	0.362
$\rho = -0.4$	BIC	0.135	0.133	0.140	0.158	0.497	0.494	0.504	0.582
	MAIC	0.057	0.052	0.056	0.050	0.322	0.295	0.324	0.326
	t-sig	0.041	0.042	0.041	0.088	0.137	0.138	0.136	0.425
$\rho = 0.4$	BIC	0.134	0.137	0.134	0.047	0.537	0.535	0.558	0.296
	MAIC	0.053	0.053	0.048	0.006	0.044	0.036	0.046	0.019
	t-sig	0.065	0.065	0.064	0.042	0.165	0.165	0.164	0.285
$\rho = 0.8$	BIC	0.193	0.203	0.196	0.036	0.330	0.330	0.342	0.123
	MAIC	0.205	0.211	0.198	0.034	0.327	0.316	0.326	0.095
	t-sig	0.087	0.088	0.085	0.038	0.103	0.104	0.103	0.132

Table 8. Size and Power when using Supremum Method for Model I; T=200
($\bar{c} = -24$ when constructing the tests and s^2)

		Size				Power			
	Criteria	$MZ_{\alpha\xi}$	MSB_ξ	$MZ_{t\xi}$	ADF_ξ	$MZ_{\alpha\xi}$	MSB_ξ	$MZ_{t\xi}$	ADF_ξ
$i.i.d.$	BIC	0.049	0.052	0.051	0.051	0.506	0.523	0.505	0.510
	MAIC	0.048	0.050	0.051	0.050	0.423	0.421	0.410	0.366
	t-sig	0.050	0.051	0.050	0.051	0.240	0.241	0.237	0.422
				MA(1) Errors					
$\theta = -0.8$	BIC	0.861	0.862	0.859	0.900	0.989	0.989	0.989	0.994
	MAIC	0.105	0.104	0.105	0.138	0.283	0.283	0.283	0.329
	t-sig	0.414	0.415	0.416	0.672	0.756	0.756	0.756	0.945
$\theta = -0.4$	BIC	0.228	0.238	0.229	0.232	0.826	0.827	0.823	0.832
	MAIC	0.084	0.086	0.079	0.062	0.376	0.377	0.370	0.308
	t-sig	0.060	0.061	0.059	0.113	0.386	0.387	0.380	0.599
$\theta = 0.4$	BIC	0.166	0.171	0.163	0.094	0.711	0.716	0.708	0.561
	MAIC	0.096	0.098	0.091	0.035	0.386	0.387	0.374	0.207
	t-sig	0.066	0.069	0.066	0.052	0.365	0.366	0.363	0.373
$\theta = 0.8$	BIC	0.289	0.293	0.284	0.088	0.699	0.705	0.697	0.374
	MAIC	0.118	0.126	0.115	0.007	0.364	0.359	0.360	0.091
	t-sig	0.054	0.057	0.053	0.030	0.289	0.291	0.289	0.179
				AR(1) Errors					
$\rho = -0.8$	BIC	0.005	0.005	0.005	0.045	0.018	0.020	0.019	0.427
	MAIC	0.000	0.000	0.000	0.040	0.013	0.012	0.012	0.275
	t-sig	0.011	0.011	0.011	0.037	0.048	0.048	0.048	0.341
$\rho = -0.4$	BIC	0.050	0.051	0.049	0.053	0.428	0.441	0.427	0.452
	MAIC	0.043	0.044	0.041	0.040	0.324	0.327	0.311	0.303
	t-sig	0.033	0.034	0.033	0.037	0.182	0.187	0.182	0.363
$\rho = 0.4$	BIC	0.103	0.108	0.102	0.049	0.565	0.578	0.565	0.390
	MAIC	0.096	0.097	0.089	0.038	0.402	0.405	0.390	0.227
	t-sig	0.051	0.053	0.052	0.044	0.301	0.306	0.299	0.349
$\rho = 0.8$	BIC	0.113	0.126	0.109	0.052	0.377	0.385	0.376	0.220
	MAIC	0.105	0.113	0.099	0.039	0.342	0.347	0.338	0.170
	t-sig	0.067	0.073	0.063	0.036	0.201	0.205	0.200	0.186

Table 9a. Empirical Results using Informatiioin Criteria to select lag k and Infimum Method to choose Break Point T_B

Series	T	Criterial	$MZ_{\alpha\xi}$	k	T_B	$MZ_{t\xi}$	k	T_B	ADF_ξ	k	T_B	$P_{T\xi}$	k	T_B	$\hat{\alpha}$
Stock Prices	100	BIC	-49.89^b	1	1941	-4.95^b	1	1941	-5.25^b	1	1937	8.92^b	1	1931	0.65
		$MAIC$	-49.22^a	1	1937	-4.93^a	1	1937	-5.25^a	1	1937	13.26^d	2	1931	0.65
		$MBIC$	-49.22^a	1	1837	-4.93^a	1	1937	-5.25^a	1	1937	13.26^d	2	1931	0.65
Real Wages	71	BIC	-39.12^c	1	1938	-4.37^c	1	1938	-4.69	1	1938	9.43^c	1	1940	0.61
		$MAIC$	-39.12^a	1	1938	-4.37^a	1	1938	-4.67^d	1	1938	11.28^b	1	1940	0.61
		$MBIC$	-39.12^a	1	1938	-4.37^a	1	1938	-4.67^c	1	1938	11.28^b	1	1940	0.61

We use a, b, c, d to represent rejection at 1%, 2.5%, 5%, 10% of significance level.

Table 9b. Empirical Results using Information Criteria to select Lag k and Supremum Method to choose Break Point T_B

Series	T	Criterial	$MZ_{\alpha\xi}$	$MZ_{t\xi}$	ADF_ξ	k	T_B	$\hat{\alpha}$
Stock Prices	100	BIC	-33.10^b	-4.05^b	-4.32^c	1	1931	0.73
		$MAIC$	-20.25^d	-3.21^d	-3.38	2	1931	0.77
		$MBIC$	-20.25^c	-3.21^c	-3.38	2	1931	0.77
Real Wages	71	BIC	-27.94^c	-3.67^c	-3.86^d	1	1933	0.69
		$MAIC$	-27.94^a	-3.67^b	-3.86^c	1	1933	0.69
		$MBIC$	-27.94^a	-3.67^b	-3.86^c	1	1933	0.69

We use a, b, c, d to represent rejection at 1%, 2.5%, 5%, 10% of significance level.

Table 10a. Empirical Results using Recursive Method to select Lag k and Infimum Method to choose T_B

Series	T	$MZ_{\alpha\xi}$	k	T_B	$MZ_{t\xi}$	k	T_B	ADF	k	T_B	$P_{T\xi}$	k	T_B	$\hat{\alpha}$
Stock Prices	100	-143.10^c	3	1948	-8.43^c	3	1948	-5.25^b	1	1936	5.94^d	3	1930	0.65
Real Wages	71	-11628.50^a	4	1941	-76.24^a	4	1941	-4.69^d	1	1938	4.05^c	3	1940	0.61

We use a, b, c, d to represent rejection at 1%, 2.5%, 5%, 10% of significance level.

Table 10b. Empirical Results using Recursive Method to select Lag k and supremum Method to choose T_B

Series	T	$MZ_{\alpha\xi}$	$MZ_{t\xi}$	ADF_ξ	k	T_B	$\hat{\alpha}$
Stock Prices	100	-49.87^c	-4.98^c	-3.99^d	3	1930	0.73
Real Wages	71	-27.94	-3.67	-3.86	1	1933	0.69

We use a, b, c, d to represent rejection at 1%, 2.5%, 5%, 10% of significance level.

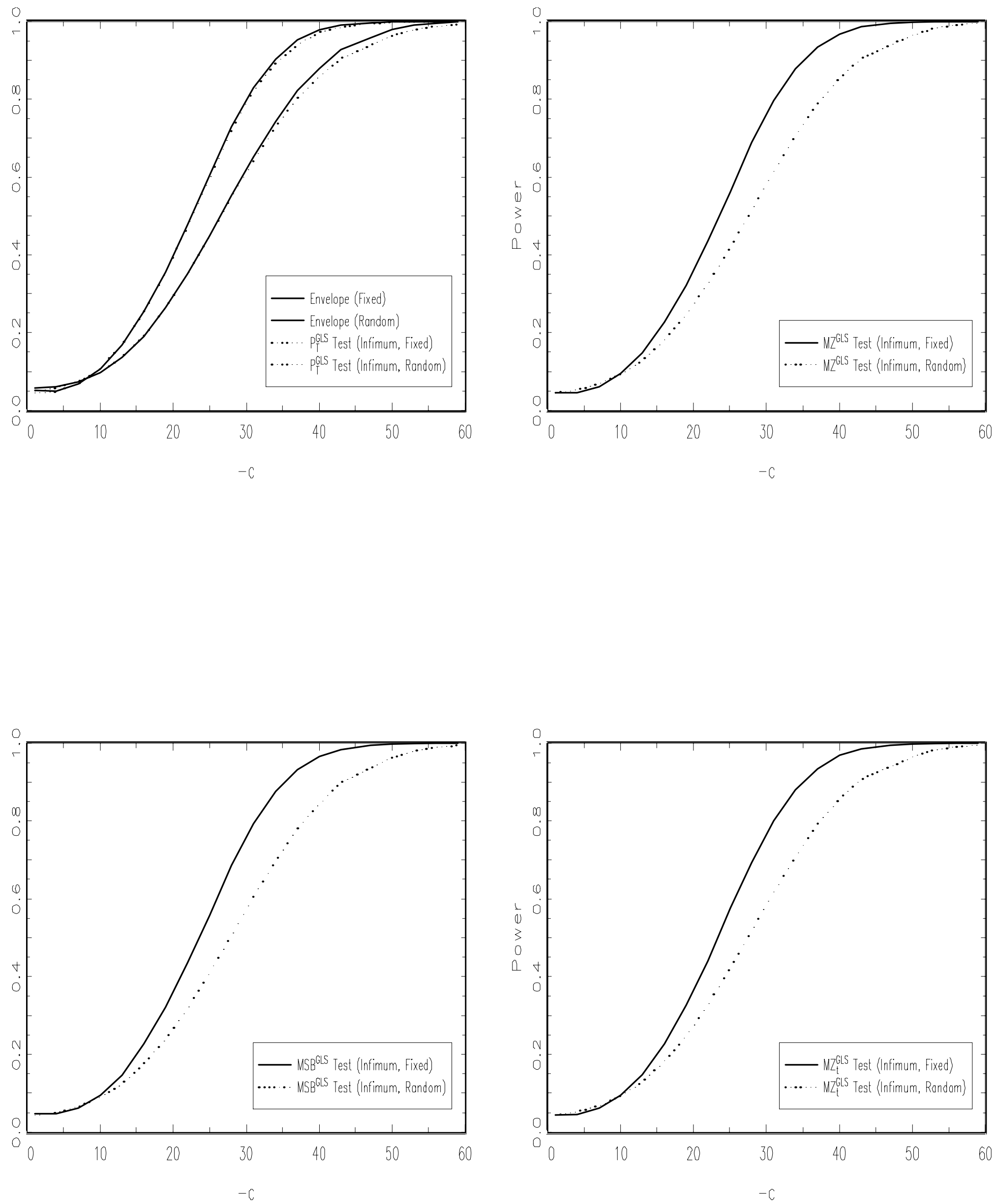

Figure 1. Gaussian Power Envelope and Asymptotic Power Functions; Infimum Method and Fixed and Random Initial Condition.

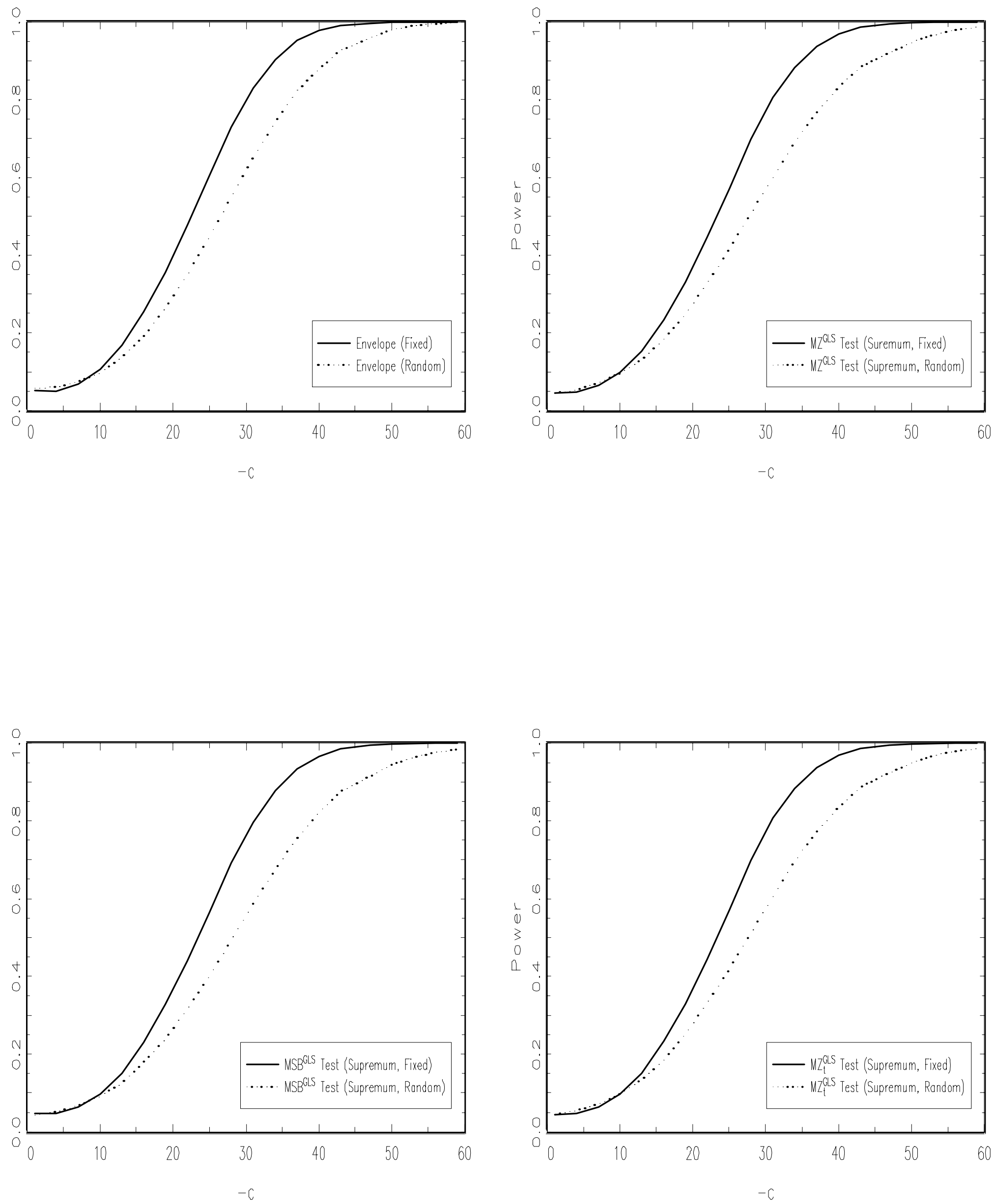

Figure 2. Gaussian Power Envelope and Asymptotic Power Functions; Supremum Method and Fixed and Random Initial Condition.

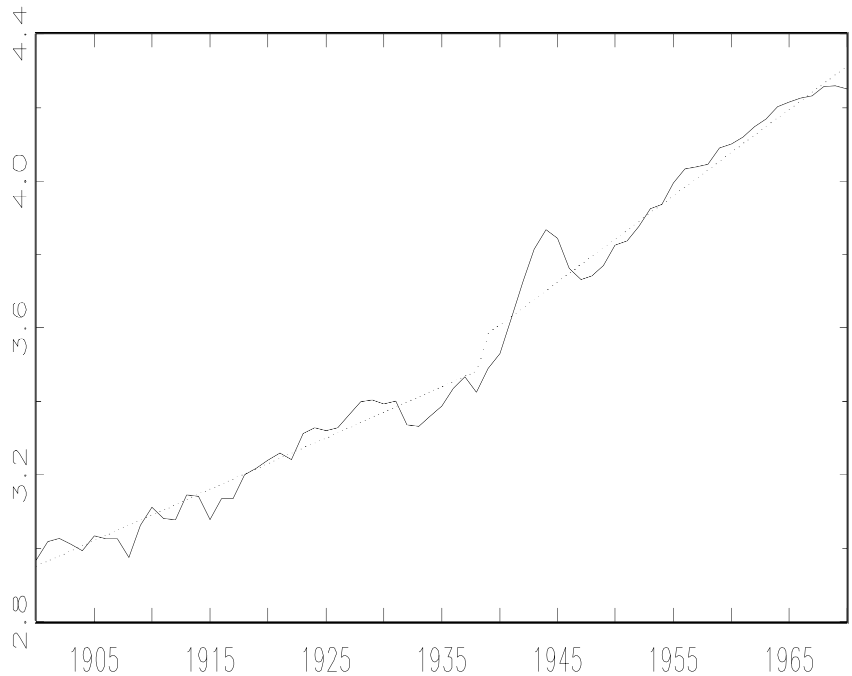

Figure 3. Logarithmes of Real Wages with a broken time trend; 1900-1970.

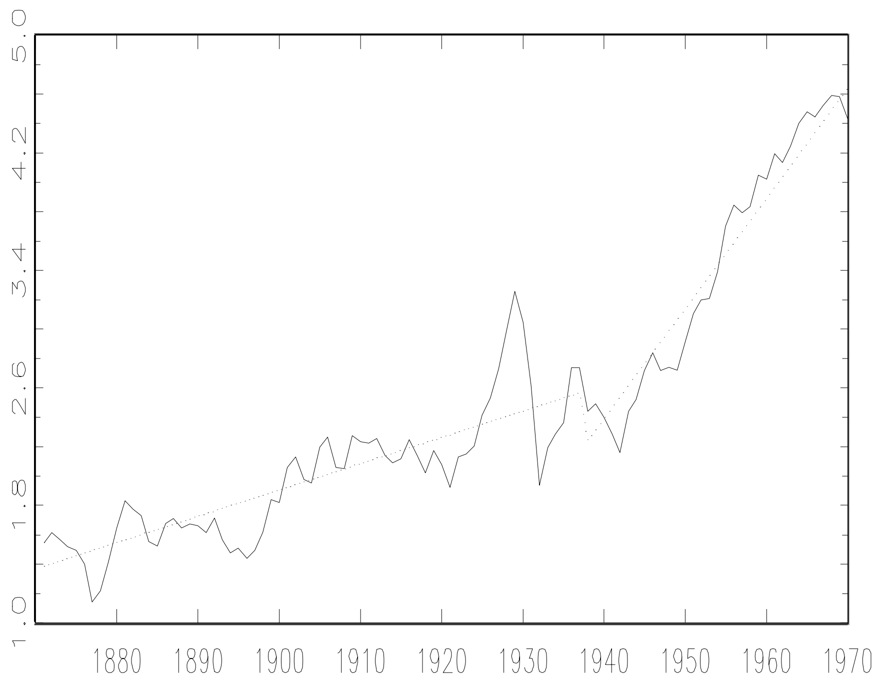

Figure 4. Logarithmes of Stock Prices with a broken time trend; 1871-1971.