Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Journal of Econometrics 115 (2003) 1–27

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

GLS detrending, efficient unit root tests and structural change

Pierre Perron[a,*], Gabriel Rodríguez[b]

[a] *Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA*
[b] *Département de Sciences Économiques, Université d'Ottawa, 200 Wilbrod St., P.O. Box 450, Station A, Ottawa, Ont., Canada, K1N-6N5*

Accepted 12 January 2002

Abstract

We extend the class of M-tests for a unit root analyzed by Perron and Ng (Rev. Econ. Studies 63 (1996) 435) and Ng and Perron (Econometrica 69 (2001) 1519) to the case where a change in the trend function is allowed to occur at an unknown time. These tests (M^{GLS}) adopt the GLS detrending approach developed by Elliott et al. (Econometrica 64 (1996) 813) (ERS) following the results of Dufour and King (J. Econometrics 47 (1991) 115). Following Perron (Econometrica 57 (1989) 1361), we consider two models: one allowing for a change in slope and the other for both a change in intercept and slope. We derive the asymptotic distributions of the tests as well as that of the feasible point optimal test (P_T^{GLS}) suggested by ERS. Also, we compute the non-centrality parameter used for the local GLS detrending that permits the test P_T^{GLS} to have 50% asymptotic power at that value. The asymptotic critical values of the tests are tabulated. We show that the M^{GLS} and P_T^{GLS} tests have an asymptotic power function close to the power envelope. A simulation study analyzes the size and power in finite samples under various methods to select the truncation lag for the autoregressive spectral density estimator. An empirical application is also provided.
© 2003 Elsevier Science B.V. All rights reserved.

JEL classification: C22

Keywords: Integrated process; Quasi-differencing; Change-point; Truncation lag; Information criteria

* Corresponding author. Tel.: 001-617-353-3026; fax: 001-617-353-4449.
E-mail addresses: perron@bu.edu (P. Perron), gabrielr@uottawa.ca (G. Rodríguez).

0304-4076/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0304-4076(03)00090-3

1. Introduction

Since the seminal paper of Nelson and Plosser (1982), the unit root hypothesis has received a lot of attention from both theoretical and empirical perspectives (e.g., Campbell and Perron (1991) and Stock (1994) for surveys). Using tests developed by Dickey and Fuller (1979), Nelson and Plosser (1982) argued that current shocks have permanent effects on the level of most macroeconomic series. This finding was supported by other approaches which found that a typical shock has both important transitory and permanent components (see, e.g., Campbell and Mankiw, 1987a, b; Shapiro and Watson, 1988; Clark, 1987; Cochrane, 1988).

In contrast to this literature, Perron (1989) argued, as an alternative to the unit root hypothesis, that macroeconomic fluctuations are most likely stationary if allowance is made for the trend function to exhibit occasional changes. Allowing for a single change in intercept and/or slope, he rejected the unit root hypothesis for 11 of the 14 series analyzed by Nelson and Plosser. As discussed in Banerjee et al. (1992) this finding may be important for the following reasons. First, it offers an alternative picture of the persistence in macroeconomics series. Second, this approach can provide a parsimonious model for a slowly changing trend component that may be useful as a data description. Third, the implications for inference in more complex models are very different.

Christiano (1992) criticized the results of Perron (1989) on the basis that the break points should not be treated as exogenous since the imposition of a given break date involves an issue of data mining. Accordingly, Zivot and Andrews (1992), Banerjee et al. (1992) and Perron (1997) considered unit root tests with unknown break points.

We continue to treat the potential break points as occurring at unknown times and contribute to this literature in two ways. First, we use the M^{GLS} tests analyzed by Perron and Ng (1996) and extend them to permit a one time change in the trend function. Second, following Elliott et al. (1996) (hereafter ERS) and the prior work of Dufour and King (1991), we use local to unity GLS detrending of the data. We consider two specific models: one with a break in the slope of the trend function and one with a break in both the intercept and slope. In this setup, there is no need to analyze the case where only a change in the intercept is allowed since the tests then have the same asymptotic distribution as the case where the deterministic components include a constant and a time trend which was analyzed in ERS (since a change in intercept is a special case of what they refer to as a "slowly evolving deterministic component").

The reasons for considering the M-tests, originally proposed by Stock (1999) and further analyzed by Perron and Ng (1996) is that these tests have much smaller size distortions than other classes of unit root tests when the errors have strong negative serial correlation. Also, using GLS detrending when constructing the M-tests allows substantial gains in power as showed by Ng and Perron (2001), similar to the DF^{GLS} test proposed by ERS.

Given that a uniformly most powerful test is not attainable, we follow ERS and derive a feasible point optimal test (P_T^{GLS}). The asymptotic power function of this test is derived and we use the associated power envelope to choose the non-centrality parameter ($\bar{c}$) to perform the GLS detrending such that the asymptotic power of the test is 50% against the local alternative $\bar{\alpha} = 1 + \bar{c}/T$. For our two models, we obtain $\bar{c} = -22.5$.

The rest of the paper is organized as follows. The model and some preliminary theoretical results are presented in Section 2. In Section 3, we derive the asymptotic distribution of the M^{GLS} and DF^{GLS} tests in both cases where the break point is known or unknown. Section 4 considers the asymptotic Gaussian power envelope and the limit distribution of the feasible point optimal test. The asymptotic critical values and the asymptotic power function of the various tests are presented in Section 5. Section 6 considers the size and power of the tests in finite samples using simulations. Section 7 presents an empirical application and Section 8 briefly concludes. An appendix contains technical derivations.

2. GLS detrending with structural change

The data generating process considered is of the form:

$$y_t = d_t + u_t, \quad t = 0, \ldots, T, \tag{1}$$

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

where $\{v_t\}$ is an unobserved stationary mean-zero process. We use the assumption that $u_0 = 0$ throughout, though the results generally hold for the weaker requirement that $E(u_0^2) < \infty$. The noise function is $v_t = \sum_{i=0}^{\infty} \gamma_i \eta_{t-i}$ with $\sum_{i=0}^{\infty} i|\gamma_i| < \infty$ and where $\{\eta_t\}$ is a martingale difference sequence. The process v_t has a non-normalized spectral density at frequency zero given by $\sigma^2 = \sigma_\eta^2 \gamma(1)^2$, where $\sigma_\eta^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{\infty} E(\eta_t^2)$. Furthermore, $T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \sigma W(r)$, where $\Rightarrow$ denotes weak convergence in distribution and $W(r)$ is the Wiener process defined on $C[0,1]$ the space of continuous functions on the interval $[0,1]$. In (1), $d_t = \psi' z_t$, where z_t is a set of deterministic components to be discussed below. For any series y_t, with deterministic components z_t, we define the transformed data $y_t^{\bar{\alpha}}$ and $z_t^{\bar{\alpha}}$ by

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

We let $\hat{\psi}$ be the estimate that minimizes

$$S^*(\psi, \bar{\alpha}, \delta) = \sum_{t=0}^{T} (y_t^{\bar{\alpha}} - \psi' z_t^{\bar{\alpha}})^2 \tag{3}$$

and denote the minimized value by $S(\bar{\alpha}, \delta)$.

2.1. The specifications of the deterministic components

Model I. Structural change in the slope: For this model, the set of deterministic components, z_t in (1), is given by

$$z_t = \{1, t, 1(t > T_B)(t - T_B)\}, \tag{4}$$

where 1(.) is the indicator function and T_B is the time of the change. Without loss of generality, we assume that $T_B = T\delta$ for some $\delta \in (0,1)$. In this case, $\hat{\psi}(\delta) = (\hat{\mu}_1, \hat{\beta}_1, \hat{\beta}_2)'$ is the vector of estimates that minimizes (3).

Model II. Structural change in intercept and slope: For Model II,

$$z_t = \{1, 1(t > T_B), t, 1(t > T_B)(t - T_B)\}. \tag{5}$$

In this case, the vector of coefficient estimates is $\hat{\psi}(\delta) = (\hat{\mu}_1, \hat{\mu}_2, \hat{\beta}_1, \hat{\beta}_2)'$. In this model, we have the same results as with Model 1 since the effect of $\hat{\mu}_2 - \mu_2$ is negligible in large samples. This is because the change in intercept is a special case of a slowly evolving deterministic component in condition B of ERS.

3. The tests and their asymptotic distributions

3.1. The tests

The M-tests, originally proposed by Stock (1999), and further analyzed by Perron and Ng (1996), exploit the feature that a series converges with different rates of normalization under the null and the alternative hypotheses. They are defined by

$$MZ_\alpha^{\mathrm{GLS}}(\delta) = (T^{-1}\tilde{y}_T^2 - s^2)\left(2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\right)^{-1}, \tag{6}$$

$$MSB^{\mathrm{GLS}}(\delta) = \left(T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2/s^2\right)^{1/2}, \tag{7}$$

$$MZ_t^{\mathrm{GLS}}(\delta) = (T^{-1}\tilde{y}_T^2 - s^2)\left(4s^2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\right)^{-1/2} \tag{8}$$

with $\tilde{y}_t = y_t - \hat{\psi}' z_t$ where $\hat{\psi}$ minimizes (3). The term s^2 is the autoregressive estimate of the spectral density at frequency zero of v_t, defined as

$$s^2 = s_{ek}^2/(1 - \hat{b}(1))^2, \tag{9}$$

with $s_{ek}^2 = (T-k)^{-1}\sum_{t=k+1}^{T}\hat{e}_{tk}^2$, $\hat{b}(1) = \sum_{j=1}^{k}\hat{b}_j$, and $\hat{b}_j$, $\{\hat{e}_{tk}\}$ obtained from the regression (see Perron and Ng (1998) for more details):

$$\Delta\tilde{y}_t = b_0\tilde{y}_{t-1} + \sum_{j=1}^{k} b_j\Delta\tilde{y}_{t-j} + e_{tk}. \tag{10}$$

The first statistic is a modified version of the Phillips and Perron (1988) Z_α test originally developed by Phillips (1987). The second statistic is a modified version of

Bhargava's (1986) R_1 statistic which builds upon the work of Sargan and Bhargava (1983). The third statistic is a modified version of the Phillips and Perron (1988) Z_t test. As Perron and Ng (1996) showed, the MSB and Z_α tests are related by $Z_t \approx MSB \times Z_\alpha$. This relation suggests the MZ_t test defined by (8) since it satisfies the relation $MZ_t = MSB \times MZ_\alpha$. Another test of interest is the so-called ADF test which is the t-statistic for testing $b_0=0$ in the regression (10), see Dickey and Fuller (1979) and Said and Dickey (1984). We denote this test by $ADF^{\mathrm{GLS}}(\delta)$. Our approach is an extension of Ng and Perron (2001) and Elliott et al. (1996) to the case where the trend function contains a structural change. In this case, the M^{GLS} tests will depend on the unknown break point δ.

3.2. Asymptotic distributions of the tests

We start with a statement of the limiting distribution of the various tests in the case where the break point is considered known.

Theorem 1. *Let y_t be generated by* (1) *with $\alpha=1+c/T$, MZ_α^{GLS}, MSB^{GLS} and MZ_t^{GLS} be defined by* (6)–(8) *with data obtained from local* GLS *detrending* ($\tilde{y}_t$) *at $\bar{\alpha}=1+\bar{c}/T$, and ADF^{GLS} be the t-statistic for testing $b_0=0$ in the regression* (10). *Also, s^2 is a consistent estimate of σ^2. For Models I and II, we have*

$$MZ_\alpha^{\mathrm{GLS}}(\delta) \Rightarrow \frac{0.5K_1(c,\bar{c},\delta)}{K_2(c,\bar{c},\delta)} \equiv H^{MZ_\alpha^{\mathrm{GLS}}}(c,\bar{c},\delta),$$

$$MSB_\alpha^{\mathrm{GLS}}(\delta) \Rightarrow (K_2(c,\bar{c},\delta))^{1/2} \equiv H^{MSB^{\mathrm{GLS}}}(c,\bar{c},\delta),$$

$$MZ_t^{\mathrm{GLS}}(\delta) \Rightarrow \frac{0.5K_1(c,\bar{c},\delta)}{(K_2(c,\bar{c},\delta))^{1/2}} \equiv H^{MZ_t^{\mathrm{GLS}}}(c,\bar{c},\delta),$$

$$ADF^{\mathrm{GLS}}(\delta) \Rightarrow \frac{0.5K_1(c,\bar{c},\delta)}{(K_2(c,\bar{c},\delta))^{1/2}} \equiv H^{ADF^{\mathrm{GLS}}}(c,\bar{c},\delta),$$

where

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

$$K_2(c,\bar{c},\delta) = \int_0^1 V_{c\bar{c}}^{(1)}(r,\delta)^2\,\mathrm{d}r - 2\int_\delta^1 V_{c\bar{c}}^{(2)}(r,\delta)\,\mathrm{d}r,$$

and $V_{c\bar{c}}^{(1)}(r,\delta) = W_c(r) - rb_3$, $V_{c\bar{c}}^{(2)}(r,\delta) = b_4(r-\delta)[W_c(r) - rb_3 - (1/2)(r-\delta)b_4]$ with $W_c(r)$ the Ornstein–Uhlenbeck process that is the solution to the stochastic differential equation $\mathrm{d}W_c(r) = cW_c(r)\,\mathrm{d}r + \mathrm{d}W(r)$ with $W_c(0)=0$. Also, b_3, b_4 are defined by $b_3 = (\lambda_1 b_1 + \lambda_2 b_2)$ and $b_4 = (\lambda_2 b_1 + \lambda_3 b_2)$ where $b_1 = (1-\bar{c})W_c(1) + \bar{c}^2\int_0^1 rW_c(r)\,\mathrm{d}r$, $b_2=(1-\bar{c}+\delta\bar{c})W_c(1)+\bar{c}^2\int_\delta^1 W_c(r)(r-\delta)\,\mathrm{d}r - W_c(\delta)$, $\lambda_1=d/\Theta$, $\lambda_2=-m/\Theta$, $d=1-\delta-\bar{c}+2\bar{c}\delta-\bar{c}\delta^2-\bar{c}^2\delta+\bar{c}^2\delta^2+(\bar{c}^2/3)(1-\delta^3)$, $m=1-\delta-\bar{c}+\bar{c}\delta-(\bar{c}^2/2)\delta+(\bar{c}^2/2)\delta^3+(\bar{c}^2/3)(1-\delta^3)$, $a=1-\bar{c}+\bar{c}^2/3$, $\Theta = ad - m^2$ and $\lambda_3 = a/\Theta$.

In practice, it is usually the case that an investigator wants to treat the break point as unknown. In this case, an estimate of the break point is needed. A method suggested by Zivot and Andrews (1992) is to consider estimating δ as the break point that yields the minimal value of the statistics, i.e. using $\inf_\delta J^{\mathrm{GLS}}(\delta)$ where $J = MZ_\alpha$, MSB, MZ_t, and ADF. Using the continuous mapping theorem and arguments as in Perron (1997), we have, assuming no shift in the trend function under the null hypothesis:

$$\inf_{\delta\in[0,1]} J^{\mathrm{GLS}}(\delta) \Rightarrow \inf_{\delta\in[0,1]} H^{J^{\mathrm{GLS}}}(c,\bar{c},\delta), \tag{11}$$

for $J = MZ_\alpha,\ MSB,\ MZ_t$, and ADF with the functions $H(\cdot)$ defined in Theorem 1. Note that no truncation for the range of possible break points needs to be imposed. As discussed in Vogelsang and Perron (1998), the implied estimate of δ is not consistent for the true value of the break point when the data generating process contains a break. These authors also note that the tests statistic are not invariant (even asymptotically) to values of the coefficients of the change in the trend. Nevertheless, they argue that, in typical sample sizes, this is not a problem unless the changes are extremely large. Thus, these tests can still be used with the critical values derived assuming no shift under the null hypothesis.

An alternative method to select the break date, as used in Perron (1997), is to choose it such that the absolute value of the t-statistic on the coefficient of the change in slope is maximized. This procedure has been used by many authors, e.g. Christiano (1992), Banerjee et al. (1992), Perron (1997) and Vogelsang and Perron (1998). Consider, for example Model I where the deterministic component is given by $d_t = \mu_1 + \beta_1 t + \beta_2(t - T_B)1(t > T_B)$. Let $\hat{\beta}_2(\delta)$ be the GLS estimate of β_2 and $t_{\hat{\beta}_2}(\delta)$ be its associated t-statistic. The break point can be selected using the estimate $\hat{\delta}=\arg\max_{\delta\in[\varepsilon,1-\varepsilon]} |t_{\hat{\beta}_2}(\delta)|$, where ε is some small number imposing a trimming on the possible values of the break dates. We shall use $\varepsilon = 0.15$ throughout. As discussed in Vogelsang and Perron (1998), if under the null hypothesis we have $\beta_2 \neq 0$ and the true break point given by $T_B^0/T = \delta^0$, then $\hat{\delta}$ is a consistent estimate of δ^0 and the limiting distributions of the test statistics correspond to those in the case where the break date is known, i.e. the limit distributions given in Theorem 1 evaluated at δ^0. In practice, one can simply evaluate these limit distributions at the estimated value $\hat{\delta}$.

When, under the null hypothesis, $\beta_2 = 0$ in which case there is no change in the slope of the trend function, it is easy to show (using the results of Lemma A.2 in Appendix A) that $t_{\hat{\beta}_2}(\delta) \Rightarrow b_4/(\lambda_3^{1/2})$, where b_4 and λ_3 are defined in Theorem 1. We then have

$$\hat{\delta} = \arg \max_{\delta\in[\varepsilon,1-\varepsilon]} |t_{\hat{\beta}_2}(\delta)| \Rightarrow \arg \max_{\delta\in[\varepsilon,1-\varepsilon]} |b_4/(\lambda_3^{1/2})| \equiv \delta^*. \tag{12}$$

Hence, the limiting distributions of the statistics are given by

$$J^{\mathrm{GLS}}(\hat{\delta}) \Rightarrow H^{J^{\mathrm{GLS}}}(c,\bar{c},\delta^*), \tag{13}$$

for $J = MZ_\alpha,\ MSB,\ MZ_t$, and ADF with the functions $H(\cdot)$ defined in Theorem 1.

In practice, it is difficult to know if there is a change in slope since any test of such hypothesis would depend on whether a unit root is present or not. Hence, a conservative procedure is to use the critical values corresponding to the case where it is assumed that no break is present, i.e. (13). This is the procedure we use in the following.

4. Feasible point optimal test and the power envelope

Elliott et al. (1996), following Dufour and King (1991), have considered the issue of developing tests with optimality properties under Gaussian errors. The case where the break point is assumed known follows closely their analysis. While a uniformly most powerful test is not attainable, it is possible to define a point optimal test against the alternative $\alpha=\bar{\alpha}$. If v_t is i.i.d., this is provided by the likelihood ratio statistic, which simplifies, under normality, to $L(\delta)=S(\bar{\alpha},\delta)-S(1,\delta)$, where $S(\bar{\alpha},\delta)$ and $S(1,\delta)$ are the sums of squared errors from a GLS regression with $\alpha=\bar{\alpha}$ and $\alpha=1$, respectively. Varying the value of $\bar{\alpha}$, gives a family of point optimal tests and the Gaussian power envelope for testing $\alpha=1$. To allow for serial correlation in the errors v_t, ERS proposed a feasible point optimal test (P_T^{GLS}) defined by

$$P_T^{\mathrm{GLS}}(c,\bar{c},\delta)=\{S(\bar{\alpha},\delta)-\bar{\alpha}S(1,\delta)\}/s^2. \tag{14}$$

The next theorem provides the limiting distribution of the P_T^{GLS} test.

Theorem 2. *Let y_t be generated by* (1) *with $\alpha=1+c/T$. Let P_T^{GLS} be defined by* (14) *with data obtained from local* GLS *detrending ($\tilde{y}_t$) at $\bar{\alpha}=1+\bar{c}/T$. Also, let s^2 be a consistent estimate of σ^2. The limit distribution of the P_T^{GLS} test under Models I and II is given by*

$$\begin{aligned} P_T^{\mathrm{GLS}}(c,\bar{c},\delta) &\Rightarrow M(c,0,\delta)-M(c,\bar{c},\delta)-2\bar{c}\int_0^1 W_c(r)\,\mathrm{d}W(r) \\ &\quad+(\bar{c}^2-2\bar{c}c)\int_0^1 W_c(r)^2\,\mathrm{d}r-\bar{c} \\ &\equiv H^{P_T^{\mathrm{GLS}}}(c,\bar{c},\delta). \end{aligned} \tag{15}$$

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

$$\begin{bmatrix} W(1)+(c-\bar{c})\displaystyle\int_0^1 W_c(r)\,\mathrm{d}r \\ \qquad -\bar{c}\displaystyle\int_0^1 r\,\mathrm{d}W(r)-(c-\bar{c})\bar{c}\int_0^1 rW_c(r)\,\mathrm{d}r \\ (1+\delta\bar{c})\left([W(1)-W(\delta)]+(c-\bar{c})\displaystyle\int_\delta^1 W_c(r)\,\mathrm{d}r\right) \\ \qquad -\bar{c}\displaystyle\int_\delta^1 r\,\mathrm{d}W(r)-(c-\bar{c})\bar{c}\int_\delta^1 rW_c(r)\,\mathrm{d}r \end{bmatrix}$$

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

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

The asymptotic expression (15) for the P_T^{GLS} test allows us to define the asymptotic power envelope for the two models. It is given by $\pi(c,\delta)=\Pr[H^{P_T^{\mathrm{GLS}}}(c,c,\delta) < b^{P_T^{\mathrm{GLS}}}(c,\delta)]$, where $b^{P_T^{\mathrm{GLS}}}(c,\delta)$ is such that $\Pr[H^{P_T^{\mathrm{GLS}}}(0,c,\delta)<b^{P_T^{\mathrm{GLS}}}(c,\delta)]=\upsilon$, with υ the size of the test. Note that, in general, a different power envelope exists for each values of δ.

When δ is unknown, things are rather different. The principle is, however, the same. To maximize the likelihood function under the null and alternative hypotheses, the estimate of δ must be chosen to minimize the sum of squares residuals $S(1,\delta)$ and $S(\bar{\alpha},\delta)$, respectively. Hence, the corresponding asymptotic version of the feasible point optimal test is then

$$P_{T,*}^{\mathrm{GLS}}(c,\bar{c})=\left\{\inf_{\delta\in[\varepsilon,1-\varepsilon]} S(\bar{\alpha},\delta)-\inf_{\delta\in[\varepsilon,1-\varepsilon]}\bar{\alpha}S(1,\delta)\right\}\Big/ s^2.$$

Note that a trimming ε is necessary otherwise the critical values become unbounded. The reason is similar to that encountered in the context of tests for structural change (see, e.g., Andrews, 1993). We use ε=0.15 throughout. In the case of this feasible point optimal test, there is a problem of which method to choose to select the break date to construct the estimate s^2. Based on finite sample properties assessed via simulations we opted for evaluating s^2 at the break point $\hat{\delta}$, say, which minimizes the sum of squared residuals under the alternative, i.e. we select $\hat{\delta}=\arg\min_{\delta\in[\varepsilon,1-\varepsilon]} S(\bar{\alpha},\delta)$. Using Theorem 2, we have

$$\begin{aligned} P_{T,*}^{\mathrm{GLS}}(c,\bar{c}) \Rightarrow & \sup_{\delta\in[\varepsilon,1-\varepsilon]} M(c,0,\delta)-\sup_{\delta\in[\varepsilon,1-\varepsilon]} M(c,\bar{c},\delta) \\ & -2\bar{c}\int_0^1 W_c(r)\,\mathrm{d}W(r)+(\bar{c}^2-2\bar{c}c)\int_0^1 W_c(r)^2\,\mathrm{d}r-\bar{c} \\ \equiv & H_*^{P_T^{\mathrm{GLS}}}(c,\bar{c}). \end{aligned} \tag{16}$$

The asymptotic Gaussian power envelope is then defined as $\pi^*(c)=\Pr[H_*^{P_T^{\mathrm{GLS}}}(c,c) < b_*^{P_T^{\mathrm{GLS}}}(c)]$, where, with υ the size of the test, $b_*^{P_T^{\mathrm{GLS}}}(c)$ is such that $\Pr[H_*^{P_T^{\mathrm{GLS}}}(0,c) < b_*^{P_T^{\mathrm{GLS}}}(c)]=\upsilon$. Furthermore, the power envelope allows us to find the "optimal" non-centrality parameter $\bar{c}$ for our models. ERS recommended to choose the value $\bar{c}$ such that the asymptotic power of the test is 50%, i.e. $\bar{c}$ is such that $\Pr[H_*^{P_T^{\mathrm{GLS}}}(\bar{c},\bar{c}) < b_*^{P_T^{\mathrm{GLS}}}(\bar{c})]=0.5$. Using simulations, we found that $\bar{c}=-22.5$ and we use this value in the rest of the paper.

5. Critical values and asymptotic power functions

In this section, we obtain the asymptotic critical values for the tests assuming $\bar{c} = -22.5$ is used to detrend the data. We simulate directly the asymptotic distributions using 1000 steps to approximate the Wiener process on $[0, 1]$ as the partial sums of i.i.d. $N(0, 1)$ random variables. The limiting distributions are tabulated for the null hypothesis $c = 0$. For the finite sample distributions, we use $T = 100$ with data generated by a random walk with zero initial condition and i.i.d. $N(0, 1)$ errors. Here k is set to 0 which is equivalent to using the true value of σ^2; the effects of selecting k are investigated in the next section. In all cases, 10,000 replications are used. The results are in the first three columns of Table 1a and Table 1b for, respectively, the case where the break point is selected by minimizing the tests and when the break point is selected maximizing the absolute value of the t-statistic on the coefficient of the change in slope. In general, the approximation to the finite sample distribution is adequate but somewhat less good for Model II which contains a change in intercept that is asymptotically negligible.

The asymptotic power functions of the tests are defined by $\pi^*_{J^{\mathrm{GLS}}}(c, \bar{c}) = \Pr[\inf_{\delta \in [0,1]} H^{J^{\mathrm{GLS}}}(c, \bar{c}, \delta) < b^{J^{\mathrm{GLS}}}(\bar{c})]$ or $\pi^*_{J^{\mathrm{GLS}}}(c, \bar{c}) = \Pr[H^{J^{\mathrm{GLS}}}(c, \bar{c}, \delta^*) < b_*^{J^{\mathrm{GLS}}}(\bar{c})]$ for $J = MZ_\alpha$, MSB, MZ_t, ADF and with $H^i(c, \bar{c})$ defined in Theorem 1 and δ^* defined by (12). The constants $b^{J^{\mathrm{GLS}}}(\bar{c})$ and $b_*^{J^{\mathrm{GLS}}}(\bar{c})$ are such that $\Pr[\inf_{\delta \in [0,1]} H^{J^{\mathrm{GLS}}}(0, \bar{c}, \delta) < b^{J^{\mathrm{GLS}}}(\bar{c})] = \upsilon$, and $\Pr[H^{J^{\mathrm{GLS}}}(0, \bar{c}, \delta^*) < b_*^{J^{\mathrm{GLS}}}(\bar{c})] = \upsilon$, the size of the tests. The asymptotic power functions are shown in Fig. 1 where the solid line is the power envelope. The M^{GLS} tests, and especially the P_T^{GLS} test, have asymptotic power functions very close to the power envelope both when the break point is selected by minimizing the tests and when it is selected maximizing the absolute value of the t-statistic of the coefficient on the change in slope. This is also true of the ADF^{GLS} test since it is asymptotically equivalent to the MZ_t^{GLS} test. Hence, in terms of asymptotic power, all tests considered are basically equivalent.

6. Size and power of the tests in finite samples

6.1. The size issue, the selection of k and information criteria

All tests require the estimation of the augmented autoregression (10). Ng and Perron (2001) recommended using GLS detrended data with the same non-centrality parameter $\bar{c}$ for constructing s^2 and the tests. We follow their suggestion and, in subsequent results, $\bar{c} = -22.5$ is used to detrend the data when constructing the tests and when estimating the autoregression (10) to construct s^2.

In our simulations and empirical applications, we consider three data dependent methods to select the order of the autoregression. The first is the standard Bayesian Information Criterion (BIC). We follow the recommendation of Ng and Perron (2001) by confining the search for the best value of k in a range $[0, k_{\max}]$. Also, all regressions are estimated using the same number of effective observations, $T^* = T - k_{\max}$. The

Table 1
(a) Critical values; P_T test, and M and ADF tests choosing T_B minimizing the statistics

Test	Size	$T=\infty$	$k=0$		Model I						Model II					
			$T=100$		$T=100$			$T=200$			$T=100$			$T=200$		
			Model I	Model II	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig
MZ_α	0.01	−40.54	−41.90	−44.97	−29.44	−41.3	−185.7	−33.2	−38.3	−69.6	−31.3	−65.2	−293.9	−33.5	−40.0	−82.1
	0.025	−35.08	−36.71	−39.99	−26.4	−34.2	−113.5	−29.4	−31.1	−51.7	−28.1	−39.1	−163.7	−30.5	−33.1	−62.5
	0.05	−31.04	−32.86	−35.99	−23.5	−28.6	−69.8	−25.4	−28.2	−42.8	−25.9	−33.7	−99.4	−27.1	−29.9	−49.9
	0.10	−26.57	−28.42	−31.38	−21.3	−24.2	−43.7	−22.2	−24.6	−34.9	−23.0	−26.9	−60.5	−23.7	−26.5	−39.5
	0.20	−22.07	−23.60	−26.48	−18.2	−20.2	−30.5	−18.7	−20.6	−26.3	−19.9	−22.6	−39.0	−20.2	−22.2	−30.1
MSB	0.01	0.110	0.109	0.105	0.130	0.110	0.052	0.122	0.114	0.085	0.125	0.087	0.041	0.121	0.110	0.078
	0.025	0.118	0.116	0.111	0.137	0.120	0.066	0.129	0.126	0.098	0.132	0.112	0.055	0.127	0.122	0.089
	0.05	0.126	0.123	0.117	0.145	0.130	0.084	0.139	0.132	0.107	0.137	0.121	0.072	0.135	0.129	0.099
	0.10	0.135	0.132	0.125	0.151	0.142	0.106	0.149	0.141	0.119	0.146	0.135	0.090	0.143	0.137	0.112
	0.20	0.148	0.144	0.136	0.163	0.155	0.127	0.161	0.154	0.137	0.157	0.146	0.113	0.155	0.148	0.128
MZ_t	0.01	−4.49	−4.56	−4.74	−3.83	−4.54	−9.59	−4.06	−4.36	−5.89	−3.93	−5.59	−12.11	−4.07	−4.44	−6.41
	0.025	−4.17	−4.27	−4.46	−3.63	−4.13	−7.53	−3.79	−3.94	−5.08	−3.75	−4.41	−9.03	−3.86	−4.06	−5.56
	0.05	−3.91	−4.04	−4.23	−3.42	−3.76	−5.89	−3.53	−3.74	−4.61	−3.56	−4.07	−7.01	−3.64	−3.85	−4.98
	0.10	−3.62	−3.75	−3.94	−3.24	−3.44	−4.65	−3.30	−3.49	−4.17	−3.36	−3.65	−5.48	−3.42	−3.60	−4.41
	0.20	−3.29	−3.41	−3.62	−2.99	−3.15	−3.87	−3.04	−3.19	−3.61	−3.13	−3.34	−4.40	−3.15	−3.30	−3.86
ADF	0.01	−4.49	−4.70	−4.87	−4.54	−5.00	−5.00	−4.47	−4.72	−4.89	−4.74	−5.19	−5.18	−4.48	−4.88	−4.99
	0.025	−4.17	−4.37	−4.54	−4.25	−4.67	−4.68	−4.11	−4.32	−4.46	−4.42	−4.77	−4.88	−4.22	−4.42	−4.49
	0.05	−3.91	−4.11	−4.27	−3.91	−4.32	−4.42	−3.78	−4.06	−4.19	−4.14	−4.44	−4.63	−3.90	−4.17	−4.31
	0.10	−3.62	−3.83	−3.99	−3.67	−3.91	−4.13	−3.49	−3.70	−3.88	−3.84	−4.19	−4.34	−3.62	−3.85	−4.07
	0.20	−3.29	−3.48	−3.65	−3.31	−3.54	−3.73	−3.18	−3.37	−3.51	−3.50	−3.72	−3.93	−3.29	−3.49	−3.69
P_T	0.01	6.71	7.86	8.87	9.47	6.96	1.85	8.11	7.23	4.05	9.94	8.18	2.12	9.16	8.80	4.59
	0.025	7.82	8.96	9.55	10.53	8.27	3.15	9.39	8.82	5.44	10.91	9.31	3.08	10.60	9.46	5.76
	0.05	8.84	9.77	10.53	11.68	9.57	4.65	10.60	9.49	6.90	12.04	10.35	5.28	11.50	10.42	7.66
	0.10	10.29	11.53	11.99	13.01	11.46	7.22	12.29	11.12	8.31	13.83	11.77	7.43	13.08	11.79	9.32
	0.20	12.28	13.50	14.08	15.12	13.67	10.08	14.62	13.23	10.89	15.79	14.06	10.20	15.06	14.04	11.50

(b) Critical values; M and ADF tests choosing T_B maximizing $|t_{\hat{\beta}_2}|$

Test	Size	$T=\infty$	$k=0$		Model I						Model II					
			$T=100$		$T=100$			$T=200$			$T=100$			$T=200$		
			Model I	Model II	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig	*MAIC*	*BIC*	t-sig
MZ_α	0.01	−39.97	−32.03	−30.88	−27.2	−39.8	−139.1	−32.7	−37.1	−64.3	−27.0	−34.6	−130.5	−31.2	−32.8	−62.1
	0.025	−34.68	−28.79	−27.75	−25.7	−31.7	−87.0	−28.3	−30.5	−49.9	−24.9	−29.5	−86.5	−27.2	−30.1	−48.5
	0.05	−30.54	−26.12	−25.56	−22.6	−27.1	−58.2	−24.8	−27.5	−39.6	−22.9	−25.9	−55.3	−24.5	−27.5	−40.6
	0.10	−26.30	−23.23	−22.86	−20.6	−23.4	−37.0	−21.8	−24.3	−32.3	−20.7	−23.4	−36.6	−21.4	−23.8	−30.1
	0.20	−21.82	−19.91	−19.72	−17.4	−19.6	−26.5	−18.2	−20.4	−24.8	−17.4	−19.5	−25.8	−18.2	−20.2	−24.4
MSB	0.01	0.111	0.124	0.126	0.135	0.111	0.059	0.122	0.116	0.088	0.134	0.119	0.061	0.125	0.123	0.089
	0.025	0.119	0.131	0.132	0.138	0.124	0.075	0.132	0.127	0.099	0.140	0.128	0.076	0.133	0.128	0.101
	0.05	0.127	0.137	0.138	0.147	0.135	0.092	0.140	0.133	0.111	0.145	0.137	0.094	0.141	0.134	0.110
	0.10	0.137	0.145	0.146	0.154	0.144	0.115	0.150	0.143	0.124	0.154	0.145	0.116	0.150	0.143	0.128
	0.20	0.150	0.156	0.157	0.166	0.158	0.136	0.164	0.155	0.140	0.166	0.158	0.137	0.163	0.156	0.141
MZ_t	0.01	−4.46	−3.98	−3.91	−3.69	−4.44	−8.34	−4.02	−4.30	−5.66	−3.66	−4.15	−7.97	−3.91	−4.04	−5.56
	0.025	−4.14	−3.77	−3.70	−3.56	−3.96	−6.58	−3.74	−3.87	−4.99	−3.51	−3.83	−6.58	−3.67	−3.86	−4.88
	0.05	−3.89	−3.59	−3.54	−3.35	−3.67	−5.37	−3.49	−3.71	−4.44	−3.35	−3.59	−5.19	−3.47	−3.67	−4.47
	0.10	−3.59	−3.38	−3.35	−3.16	−3.39	−4.29	−3.27	−3.46	−3.98	−3.19	−3.37	−4.28	−3.24	−3.43	−3.86
	0.20	−3.27	−3.13	−3.11	−2.92	−3.09	−3.61	−2.99	−3.16	−3.49	−3.91	−3.08	−3.58	−2.99	−3.14	−3.46
ADF	0.01	−4.46	−4.67	−4.59	−4.40	−4.91	−4.87	−4.41	−4.67	−4.83	−4.31	−4.85	−4.86	−4.25	−4.44	−4.66
	0.025	−4.14	−4.33	−4.29	−4.18	−4.58	−4.57	−4.00	−4.23	−4.36	−4.11	−4.40	−4.48	−3.96	−4.20	−4.24
	0.05	−3.89	−4.06	−4.04	−3.83	−4.25	−4.33	−3.73	−4.00	−4.08	−3.83	−4.16	−4.25	−3.67	−3.94	−4.05
	0.10	−3.59	−3.78	−3.76	−3.55	−3.86	−3.99	−3.45	−3.67	−3.79	−3.59	−3.84	−4.02	−3.42	−3.63	−3.76
	0.20	−3.27	−3.44	−3.43	−3.23	−3.48	−3.64	−3.14	−3.34	−3.47	−3.24	−3.49	−3.65	−3.12	−3.31	−3.43

$\bar{c} = -22.5$ when constructing the tests and s^2.

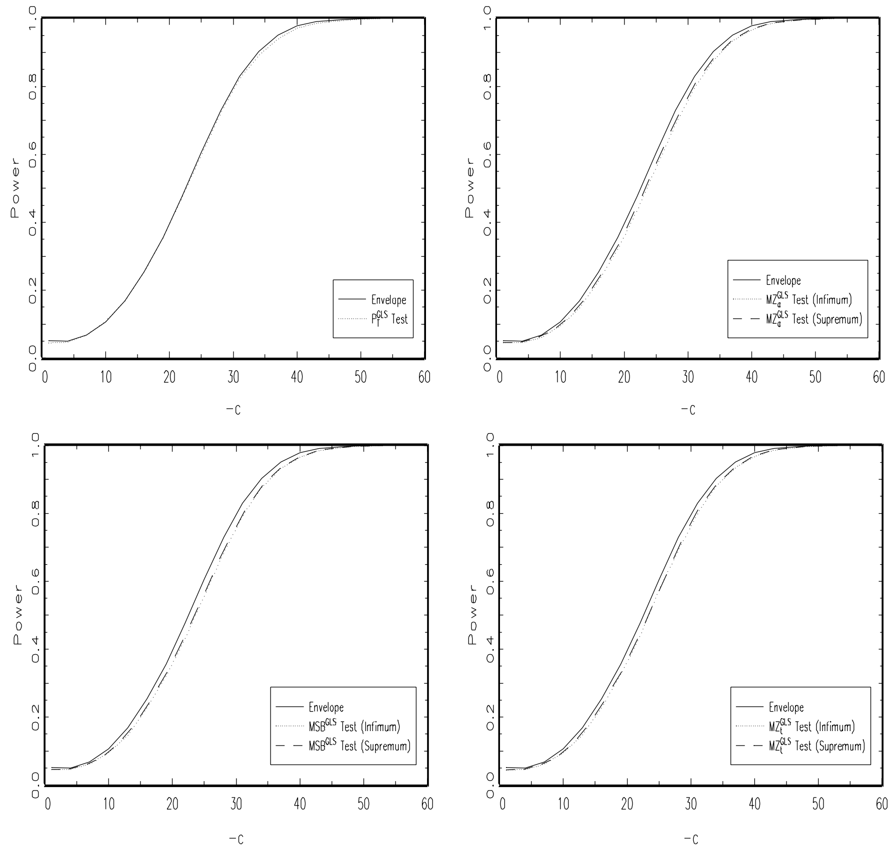

Fig. 1. Gaussian local power envelope and the local asymptotic power functions of the tests.

BIC is then defined as $k_{\text{bic}} = \arg\min_{k\in[0,k_{\max}]} \{\log(s_{ek}^2) + (\ln(T^*)k)/T^*\}$ with $s_{ek}^2 = T^{*-1}\sum_{t=k_{\max}+1}^{T} \hat{e}_{tk}^2$ with $\hat{e}_{tk}$ obtained from (10) estimated from $t = k_{\max}+1$ to T. We also consider the Modified Akaike Information Criterion (*MAIC*), advocated by Ng and Perron (2001), defined by $k_{\text{maic}} = \arg\min_{k\in[0,k_{\max}]} \{\log(s_{ek}^2) + (2(\hat{\tau}_T(k)+k))/T^*\}$, where $\hat{\tau}_T(k) = (s_{ek}^2)^{-1}\hat{b}_0^2 \sum_{t=k_{\max}+1}^{T} \tilde{y}_{t-1}^2$ with $\hat{b}_0$ obtained from (10). As shown in Ng and Perron (2001), the *MAIC* works as well as standard information criteria when the extent of correlation is mild but provides unit root tests having better finite sample size with a negative *MA* component. We also consider the sequential t-test, denoted t-sig, for the significance of the last lag, considered in Ng and Perron (1995), with a two-tailed 10% size. When using an information criterion, we set $k_{\max} = int[10 * (T/100)\hat{}(1/4)]$ and for the t-sig method, we set $k_{\max} = int[4 * (T/100)\hat{}(1/4)]$. Note that once the order k is selected, at k^* say, s^2 is constructed using all possible observations, i.e. estimating (10) from $t = k^*+1$ to T.

6.2. Critical values with data-dependent methods to select k

While the asymptotic distribution is a good approximation to the finite sample distribution for any of the test considered when k is fixed, significant differences can occur when using a data-dependent method to select k, especially in the context of tests involving breaks at unknown dates (e.g., Perron, 1997). We present in Tables 2 and 3, for both Models I and II, the finite sample critical values of the tests when using either of the three data-dependent methods to select k. Two sample sizes are considered, $T = 100$ and 200. These were obtained from simulations with 1000 replications from the data-generating process defined by (1) with $d_t = 0$, $\alpha = 1$ and $v_t \sim$ i.i.d. $N(0, 1)$.

The results show substantial differences, especially when using the method t-sig which yields finite sample critical values much smaller than the asymptotic ones (hence, using the latter would imply liberal size distortions). When using the $MAIC$, the result is opposite, namely finite sample critical values that are higher than the asymptotic ones (hence, using the latter would imply conservative tests). With the BIC, the difference are not so large but still important. For these reasons, we recommend the use of these finite sample critical values "adjusted for the effect of using a data-dependent method to select k". The simulations and empirical applications below make use of these instead of the asymptotic ones.

6.3. Size and power of the tests

We now consider the size and power of the tests in finite samples using the various data-dependent methods to select the truncation lag described above. Our simulations are based on 1000 replications of the DGP defined by (1) with $d_t = 0$. We consider pure $MA(1)$ processes, i.e. with $v_t = (1 + \theta L)e_t$ and pure $AR(1)$ processes, i.e. $(1 - \rho L)v_t = e_t$, where $e_t \sim$ i.i.d. $N(0, 1)$. For both the $MA(1)$ and $AR(1)$ cases, we consider θ and ρ in the range $[-0.8, 0.8]$. We consider the sample sizes $T = 100$ and 200. The power is evaluated at $\bar{\alpha} = 1 + \bar{c}/T$ for $\bar{c} = -22.5$ which implies that the asymptotic power is 50%. All results presented are for 5% nominal size tests.

The results for the case where the break point is chosen by minimizing the tests are presented in Table 2 for $T = 100$ and Table 3 for $T = 200$. With i.i.d. errors, as expected, the power of the tests when the sequential t-sig is used to select k is low. With the $MAIC$ or BIC, the power is indeed close to the asymptotic value of 50%. For the ADF^{GLS} test, the power is high for all methods to choose k. Given these results, we shall not discuss further the behavior of the tests with the sequential t-sig method.

Consider now the case where the errors have a negative MA component. For all tests, the use of the BIC to select k implies tests with severe size distortions with exact sizes above 90% with $T = 100$ (80% with $T = 200$) when the MA component is -0.8 and about 40% with $T = 100$ (23% with $T = 200$) when it is -0.4. On the other hand, the $MAIC$ allows the M^{GLS} and the P_T^{GLS} tests to have much less size distortions. The exact size is about 33% with $T = 100$ (10% with $T = 200$) when the coefficient is -0.8 and about 14% with $T = 100$ (9% with $T = 200$) when it is -0.4. But even

Table 2
Size and power; choosing T_B minimizing the tests; Model I; $T = 100$ ($\bar{c} = -22.5$ when constructing s^2 for the M^{GLS} and P_T^{GLS} tests; 5% nominal size tests)

	Criteria	Size					Power				
		MZ_α	MSB	MZ_t	P_T	ADF	MZ_α	MSB	MZ_t	P_T	ADF
i.i.d.	*BIC*	0.051	0.051	0.051	0.050	0.050	0.348	0.319	0.347	0.375	0.462
	MAIC	0.050	0.051	0.050	0.050	0.050	0.481	0.484	0.468	0.486	0.459
	t-sig	0.050	0.051	0.051	0.051	0.051	0.131	0.131	0.130	0.153	0.421
Moving-average processes											
$\theta = -0.8$	*BIC*	0.930	0.931	0.930	0.922	0.969	1.000	1.000	1.000	0.999	1.000
	MAIC	0.334	0.334	0.329	0.286	0.353	0.778	0.783	0.776	0.669	0.809
	t-sig	0.074	0.074	0.074	0.071	0.717	0.293	0.293	0.295	0.250	0.938
$\theta = -0.4$	*BIC*	0.391	0.385	0.385	0.368	0.447	0.874	0.873	0.874	0.865	0.903
	MAIC	0.145	0.146	0.141	0.138	0.124	0.506	0.505	0.497	0.475	0.461
	t-sig	0.025	0.025	0.025	0.026	0.228	0.096	0.096	0.096	0.100	0.683
$\theta = 0.4$	*BIC*	0.226	0.224	0.227	0.201	0.076	0.696	0.687	0.698	0.683	0.494
	MAIC	0.109	0.108	0.098	0.066	0.017	0.166	0.169	0.160	0.162	0.100
	t-sig	0.051	0.053	0.051	0.056	0.058	0.157	0.157	0.158	0.199	0.378
$\theta = 0.8$	*BIC*	0.481	0.482	0.478	0.364	0.107	0.703	0.699	0.707	0.679	0.400
	MAIC	0.185	0.191	0.176	0.091	0.011	0.336	0.337	0.327	0.301	0.084
	t-sig	0.074	0.076	0.074	0.081	0.052	0.233	0.231	0.232	0.269	0.215
Autoregressive processes											
$\rho = -0.8$	*BIC*	0.010	0.010	0.010	0.007	0.041	0.030	0.029	0.031	0.026	0.410
	MAIC	0.003	0.002	0.003	0.002	0.043	0.015	0.017	0.015	0.016	0.359
	t-sig	0.016	0.016	0.015	0.016	0.043	0.054	0.054	0.054	0.067	0.384
$\rho = -0.4$	*BIC*	0.129	0.120	0.128	0.115	0.135	0.475	0.469	0.476	0.465	0.549
	MAIC	0.064	0.063	0.063	0.063	0.059	0.360	0.365	0.356	0.374	0.396
	t-sig	0.030	0.030	0.030	0.035	0.071	0.116	0.116	0.116	0.134	0.420
$\rho = 0.4$	BIC	0.146	0.146	0.146	0.112	0.038	0.496	0.477	0.499	0.502	0.275
	MAIC	0.122	0.131	0.113	0.083	0.022	0.122	0.124	0.116	0.083	0.058
	t-sig	0.057	0.058	0.057	0.057	0.047	0.147	0.146	0.147	0.162	0.273
$\rho = 0.8$	*BIC*	0.290	0.298	0.278	0.183	0.060	0.346	0.341	0.343	0.331	0.107
	MAIC	0.342	0.360	0.321	0.204	0.074	0.404	0.415	0.390	0.355	0.141
	t-sig	0.136	0.139	0.134	0.084	0.064	0.108	0.108	0.107	0.119	0.118

when using the *MAIC*, the ADF^{GLS} test still suffers from high size distortions (with $\theta = -0.8$, the size is 18% when $T = 200$).

When the errors have a positive moving average coefficient, the M^{GLS} tests and the P_T^{GLS} tests are liberal, while the ADF^{GLS} test has the correct size (especially with $T = 200$). With a negative autoregressive coefficient, the M^{GLS} tests and the P_T^{GLS} test

Table 3
Size and power; choosing T_B minimizing the tests; Model I; $T = 200$ ($\bar{c} = -22.5$ when constructing s^2 for the M^{GLS} and P_T^{GLS} tests; 5% nominal size tests)

	Criteria	Size					Power				
		MZ_α	MSB	MZ_t	P_T	ADF	MZ_α	MSB	MZ_t	P_T	ADF
i.i.d.	BIC	0.051	0.051	0.050	0.051	0.051	0.512	0.494	0.505	0.488	0.501
	$MAIC$	0.050	0.050	0.051	0.050	0.051	0.495	0.492	0.505	0.482	0.476
	t-sig	0.051	0.051	0.051	0.050	0.051	0.235	0.230	0.239	0.289	0.430
Moving-average processes											
$\theta = -0.8$	BIC	0.814	0.809	0.814	0.783	0.860	0.992	0.991	0.992	0.989	0.997
	$MAIC$	0.102	0.099	0.104	0.076	0.178	0.366	0.366	0.365	0.272	0.510
	t-sig	0.278	0.277	0.277	0.296	0.539	0.655	0.655	0.656	0.628	0.931
$\theta = -0.4$	BIC	0.239	0.230	0.235	0.218	0.225	0.817	0.808	0.815	0.797	0.820
	$MAIC$	0.091	0.091	0.094	0.074	0.074	0.453	0.458	0.460	0.418	0.427
	t-sig	0.034	0.034	0.035	0.043	0.110	0.297	0.287	0.295	0.342	0.605
$\theta = 0.4$	BIC	0.144	0.144	0.141	0.130	0.089	0.702	0.687	0.698	0.673	0.552
	$MAIC$	0.102	0.104	0.101	0.083	0.046	0.514	0.511	0.510	0.468	0.356
	t-sig	0.068	0.067	0.069	0.072	0.050	0.329	0.324	0.331	0.380	0.401
$\theta = 0.8$	BIC	0.296	0.290	0.291	0.241	0.076	0.694	0.683	0.693	0.665	0.392
	$MAIC$	0.181	0.182	0.181	0.120	0.027	0.447	0.445	0.453	0.382	0.155
	t-sig	0.149	0.151	0.146	0.156	0.064	0.515	0.509	0.517	0.550	0.313
Autoregressive processes											
$\rho = -0.8$	BIC	0.006	0.006	0.006	0.006	0.045	0.045	0.037	0.044	0.040	0.463
	$MAIC$	0.001	0.001	0.001	0.000	0.047	0.033	0.033	0.035	0.032	0.395
	t-sig	0.016	0.016	0.016	0.014	0.046	0.061	0.059	0.061	0.068	0.411
$\rho = -0.4$	BIC	0.052	0.049	0.049	0.037	0.048	0.444	0.424	0.434	0.417	0.458
	$MAIC$	0.047	0.045	0.046	0.036	0.045	0.400	0.403	0.408	0.381	0.425
	t-sig	0.034	0.034	0.033	0.041	0.043	0.201	0.194	0.199	0.228	0.418
$\rho = 0.4$	BIC	0.106	0.105	0.104	0.090	0.041	0.585	0.562	0.582	0.560	0.411
	$MAIC$	0.105	0.103	0.108	0.083	0.054	0.494	0.493	0.506	0.465	0.356
	t-sig	0.068	0.068	0.067	0.065	0.044	0.267	0.263	0.268	0.320	0.367
$\rho = 0.8$	BIC	0.146	0.141	0.138	0.106	0.057	0.405	0.394	0.393	0.371	0.229
	$MAIC$	0.164	0.175	0.156	0.108	0.062	0.399	0.397	0.400	0.369	0.229
	t-sig	0.091	0.094	0.091	0.070	0.053	0.212	0.204	0.211	0.228	0.206

are very conservative and, hence, show basically no power. The ADF has the correct size and power is good. When the autoregressive coefficient is positive, the M^{GLS} tests and the P_T^{GLS} test are liberal. The ADF^{GLS} has better size but no power.

The results show that the ADF^{GLS} with k chosen using the $MAIC$ has better overall properties unless there is a negative MA component in the errors, in which case the M^{GLS} and the P_T^{GLS} tests are superior.

Table 4
Size and power; choosing T_B maximizing $|t_{\hat{\beta}_2}|$; Model I; $T=100$ ($\bar{c}=-22.5$ when constructing s^2 for the M^{GLS} and P_T^{GLS} tests; 5% nominal size tests)

	Criteria	Size				Power			
		MZ_α	MSB	MZ_t	ADF	MZ_α	MSB	MZ_t	ADF
i.i.d.	*BIC*	0.050	0.051	0.051	0.051	0.411	0.414	0.402	0.478
	MAIC	0.051	0.050	0.050	0.050	0.503	0.475	0.499	0.465
	t-sig	0.051	0.051	0.051	0.051	0.150	0.151	0.154	0.447
Moving-average processes									
$\theta=-0.8$	*BIC*	0.928	0.930	0.926	0.966	1.000	1.000	1.000	1.000
	MAIC	0.289	0.286	0.289	0.304	0.669	0.669	0.669	0.698
	t-sig	0.070	0.070	0.070	0.689	0.250	0.250	0.250	0.925
$\theta=-0.4$	*BIC*	0.399	0.405	0.397	0.430	0.870	0.870	0.870	0.895
	MAIC	0.138	0.131	0.136	0.113	0.479	0.466	0.475	0.433
	t-sig	0.025	0.025	0.025	0.206	0.099	0.099	0.100	0.668
$\theta=0.4$	*BIC*	0.216	0.225	0.211	0.072	0.707	0.711	0.703	0.494
	MAIC	0.064	0.065	0.066	0.010	0.156	0.139	0.161	0.107
	t-sig	0.053	0.053	0.054	0.055	0.195	0.199	0.195	0.389
$\theta=0.8$	*BIC*	0.390	0.397	0.380	0.089	0.683	0.686	0.682	0.386
	MAIC	0.089	0.088	0.087	0.008	0.297	0.281	0.294	0.091
	t-sig	0.079	0.080	0.078	0.045	0.260	0.262	0.263	0.212
Autoregressive processes									
$\rho=-0.8$	*BIC*	0.006	0.006	0.006	0.041	0.028	0.028	0.027	0.398
	MAIC	0.002	0.003	0.002	0.041	0.015	0.017	0.016	0.335
	t-sig	0.016	0.016	0.016	0.040	0.062	0.062	0.062	0.377
$\rho=-0.4$	*BIC*	0.121	0.120	0.121	0.128	0.474	0.475	0.468	0.546
	MAIC	0.063	0.057	0.063	0.053	0.368	0.352	0.370	0.391
	t-sig	0.032	0.035	0.033	0.066	0.127	0.128	0.127	0.415
$\rho=0.4$	*BIC*	0.132	0.136	0.127	0.033	0.525	0.530	0.520	0.287
	MAIC	0.078	0.084	0.080	0.014	0.080	0.077	0.079	0.047
	t-sig	0.055	0.056	0.054	0.040	0.156	0.158	0.157	0.281
$\rho=0.8$	*BIC*	0.206	0.217	0.197	0.037	0.353	0.356	0.344	0.108
	MAIC	0.225	0.240	0.210	0.048	0.372	0.362	0.365	0.127
	t-sig	0.087	0.091	0.087	0.042	0.117	0.119	0.118	0.114

The results for the case where the break is selected by maximizing the absolute value of the t-statistic on the coefficient of the change in slope are in Table 4 for $T=100$ and Table 5 for $T=200$. They show properties with basically similar qualitative features with size distortions being slightly smaller.

Table 5
Size and power; choosing T_B maximizing $|t_{\hat{\beta}_2}|$; Model I; $T = 200$ ($\bar{c} = -22.5$ when constructing s^2 for the M^{GLS} and P_T^{GLS} tests; 5% nominal size tests)

	Criteria	Size				Power			
		MZ_α	MSB	MZ_t	ADF	MZ_α	MSB	MZ_t	ADF
i.i.d.	*BIC*	0.051	0.051	0.051	0.051	0.518	0.504	0.513	0.520
	MAIC	0.051	0.050	0.051	0.050	0.505	0.484	0.515	0.478
	t-sig	0.051	0.050	0.051	0.051	0.275	0.267	0.279	0.483
Moving-average processes									
$\theta = -0.8$	*BIC*	0.796	0.793	0.796	0.856	0.991	0.989	0.991	0.997
	MAIC	0.076	0.072	0.079	0.139	0.277	0.272	0.281	0.422
	t-sig	0.290	0.286	0.288	0.538	0.625	0.620	0.624	0.915
$\theta = -0.4$	*BIC*	0.226	0.220	0.224	0.223	0.822	0.808	0.818	0.826
	MAIC	0.079	0.077	0.081	0.068	0.436	0.424	0.438	0.406
	t-sig	0.038	0.037	0.040	0.116	0.325	0.318	0.327	0.633
$\theta = 0.4$	*BIC*	0.142	0.138	0.138	0.088	0.703	0.683	0.696	0.567
	MAIC	0.094	0.092	0.094	0.042	0.479	0.472	0.481	0.342
	t-sig	0.068	0.070	0.069	0.057	0.365	0.351	0.364	0.422
$\theta = 0.8$	*BIC*	0.261	0.256	0.249	0.070	0.684	0.669	0.681	0.400
	MAIC	0.132	0.128	0.128	0.018	0.403	0.388	0.407	0.148
	t-sig	0.148	0.147	0.146	0.069	0.544	0.535	0.542	0.350
Autoregressive processes									
$\rho = -0.8$	*BIC*	0.006	0.005	0.006	0.046	0.045	0.041	0.045	0.471
	MAIC	0.000	0.000	0.000	0.037	0.034	0.031	0.035	0.376
	t-sig	0.014	0.014	0.015	0.051	0.061	0.062	0.062	0.428
$\rho = -0.4$	*BIC*	0.047	0.045	0.042	0.046	0.454	0.426	0.448	0.477
	MAIC	0.039	0.036	0.041	0.043	0.399	0.387	0.412	0.421
	t-sig	0.037	0.035	0.035	0.052	0.214	0.200	0.214	0.449
$\rho = 0.4$	*BIC*	0.101	0.095	0.096	0.041	0.594	0.563	0.584	0.427
	MAIC	0.087	0.083	0.088	0.048	0.480	0.462	0.490	0.353
	t-sig	0.065	0.064	0.063	0.043	0.308	0.298	0.308	0.407
$\rho = 0.8$	*BIC*	0.119	0.116	0.110	0.046	0.400	0.381	0.384	0.229
	MAIC	0.118	0.128	0.121	0.053	0.382	0.364	0.387	0.224
	t-sig	0.070	0.067	0.066	0.043	0.218	0.212	0.215	0.227

7. Empirical applications

Among the macroeconomic time series considered by Nelson and Plosser (1982), Perron (1989) argued that two of them were likely affected by a significant change in slope and intercept for the samples analyzed, namely the Real Wages and Stock

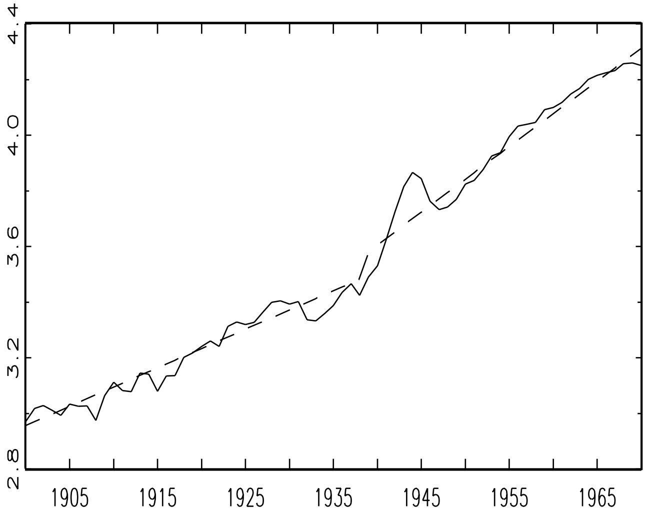

Fig. 2. Logarithm of real wages (1900–1970).

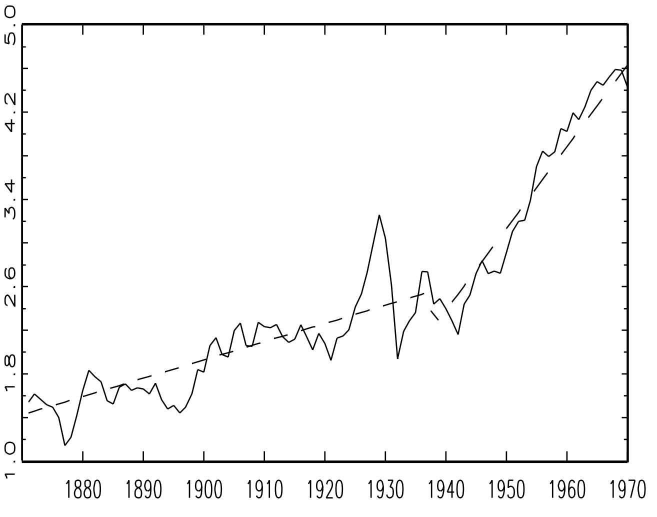

Fig. 3. Logarithm of common stock prices (1871–1970).

Prices series. The series are presented in Figs. 2 and 3. We re-evaluate the claim made by Perron (1989) that these series are trend-stationary if allowance is made for such a change in slope and intercept using our new tests. We applied the MZ_t^{GLS}, P_T^{GLS} and ADF^{GLS} tests using the *BIC* and *MAIC* criteria to select the autoregressive order (imposing a minimal value of 1).

The results are presented in Table 6 for the case where the break date is selected minimizing the test. Using the *BIC* to select k, all tests point to a strong rejection at the

Table 6
Empirical results for the real wages and stock prices series choosing the break point minimizing the tests

Serie	T	Criteria	MZ_α	k	T_B	MZ_t	k	T_B	P_T	k	T_B	ADF	k	T_B	$\hat{\alpha}$
Stock prices	100	*BIC*	-48.4^{b}	1	1941	-4.9^{b}	1	1941	8.3^{b}	1	1931	-5.1^{b}	1	1937	0.666
		MAIC	-47.7^{a}	1	1937	-4.8^{a}	1	1937	12.5^{d}	1	1931	-5.1^{a}	1	1937	0.666
Real wages	71	*BIC*	-38.4^{c}	1	1938	-4.3^{c}	1	1938	10.3^{c}	1	1940	-4.6^{c}	1	1938	0.619
		MAIC	-38.4^{a}	1	1938	-4.3^{a}	1	1938	10.3^{b}	1	1940	-4.6^{b}	1	1938	0.619

Notes: (1) For the applications, we impose a minimal value $k = 1$; (2) the superscripts a, b, c and d denote significance levels at the 1.0%, 2.5%, 5.0%, and 10.0%, respectively.

Table 7
Empirical results for the real wages and stock prices series choosing the break point maximizing $|t_{\hat{\beta}_2}|$

Serie	T	Criteria	MZ_α	MZ_t	ADF	k	T_B	$\hat{\alpha}$
Stock prices	100	*BIC*	-31.9^{b}	-3.9^{b}	-4.1^{c}	1	1931	0.753
		MAIC	-21.2^{d}	-3.2^{d}	-3.2	1	1931	0.793
Real wages	71	*BIC*	-27.7^{c}	-3.6^{c}	-3.8^{d}	1	1933	0.697
		MAIC	-27.7^{a}	-3.6^{a}	-3.8^{c}	1	1933	0.697

Notes: (1) For the applications, we impose a minimal value $k = 1$; (2) the superscripts a, b, c and d denote significance levels at the 1.0%, 2.5%, 5.0%, and 10.0%, respectively.

2.5% significance level for the Stock Prices series with the break date selected between 1931 and 1941 depending on the specification used. With the *MAIC* to select k, there is a rejection at the 1% level (except with the test P_T^{GLS}). For the Real Wages series, there is a rejection at least at the 1% or 2.5% significance level using the criterion *MAIC* to select k and at the 5% level using the *BIC*. The break date is selected at 1938 or 1940 depending on the specification used. The estimated trend function is plotted in Figs. 2 and 3 using $T_B = 1937$ for the Stock Prices series and $T_B = 1938$ for the Real Wages series.

Table 7 presents the results of the tests when the break date is selected by maximizing the absolute value of the t-statistic on the coefficient of the slope change. For Stock Prices, the break date selected is 1931 and there is a rejection at the 2.5% or 5% significance level using *BIC* but there is little evidence against the unit root using the *MAIC*. For the Real Wages series, the break date selected is 1933. The tests show a rejection at least at the 5% significance level (except for the ADF^{GLS} test using *BIC* to select the autoregressive order).

8. Conclusions

We considered tests for the null hypothesis of a unit root in the presence of a one time change in the trend function and followed Elliott et al. (1996) and Dufour and King (1991) by detrending the data using a local to unity GLS approach. The extensions of the ADF and the P_T as well as of the various M tests suggested by Perron and Ng (1996) were studied. We also investigated the properties of the tests when the break point is selected either by minimizing the tests or by maximizing the absolute value of the t-statistic on the coefficient of the change in slope. All tests have a local asymptotic power function that lies close to the Gaussian power envelope though our simulations reveal that, in finite sample, the latter method yields tests with less power. Hence, for applications we recommend using either the ADF^{GLS} or the M^{GLS} or P_T^{GLS} tests with the break point selected by minimizing the tests. The main difference among the tests is that the ADF^{GLS} has worse size distortions in the negative MA case but better power in the negative AR case; the M^{GLS} and P_T^{GLS} tests have good size overall but very little power in the negative AR case. The choice between the two depends on the investigator's assessment of the likely importance of one or the other class of processes in the data considered. Our experiments also suggest that the use of the $MAIC$ to select the autoregressive truncations lag leads to tests with better overall properties.

Acknowledgements

This paper is drawn from chapter 1 of Gabriel Rodríguez's Ph.D. Dissertation at the Université de Montréal (Rodríguez, 1999). We wish to thank Alain Guay for useful comments when this paper was presented at the 38th annual meeting of the Société Canadienne de Sciences Economiques. Perron acknowledges financial support from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR).

Appendix A.

Throughout, we use the following lemma which is by now standard.

Lemma A.1. *Let* $\{u_t\}$ *be a near-integrated series generated by* (2). *Then, we have*: (a) $T^{-1/2}u_{[Tr]} \Rightarrow \sigma W_c(r)$; (b) $T^{-3/2}\sum_{t=1}^T u_t \Rightarrow \sigma \int_0^1 W_c(r)\,\mathrm{d}r$; (c) $T^{-2}\sum_{t=1}^T u_t^2 \Rightarrow \sigma^2 \int_0^1 W_c^2(r)\,\mathrm{d}r$; (d) $T^{-1}\sum_{t=1}^T u_{t-1}v_t \Rightarrow \sigma^2\{\int_0^1 W_c(r)\,\mathrm{d}W(r)+\gamma\}$ *with* $\gamma=(\sigma^2-\sigma_v^2)/2\sigma^2$.

We start with results concerning the limit of the estimates of the coefficients of the trend function obtained from (3). For Model I, we have

Lemma A.2. *Suppose that* y_t *is generated by* (1) *with* $\alpha = 1 + c/T$ *and* $\{z_t\}$ *is given by* (4). *Let* $\hat{\psi}(\delta)$ *be the GLS estimates, from minimizing* (3), *of the coefficients of the trend function obtained using* $\bar{\alpha} = 1 + \bar{c}/T$. *Then, with terms as defined in Theorem* 1:

$$\hat{\mu}_1 - \mu_1 \Rightarrow v_1,$$

$$T^{1/2}(\hat{\beta}_1 - \beta_1) \Rightarrow \sigma(\lambda_1 b_1 + \lambda_2 b_2) \equiv \sigma b_3,$$
$$T^{1/2}(\hat{\beta}_2 - \beta_2) \Rightarrow \sigma(\lambda_2 b_1 + \lambda_3 b_2) \equiv \sigma b_4.$$

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

$$\hat{\psi}(\delta) - \psi = [(\Delta z - \bar{c}T^{-1}z_{-1})'(\Delta z - \bar{c}T^{-1}z_{-1})]^{-1} \times [(\Delta z - \bar{c}T^{-1}z_{-1})'(\Delta u - \bar{c}T^{-1}u_{-1})], \tag{A.1}$$

where

$$\Delta z = (z_1, z_2 - z_1, \dots, z_T - z_{T-1}),$$
$$z_{-1} = (0, z_1, z_2, \dots, z_{T-1}),$$
$$\Delta u = (u_1, u_2 - u_1, \dots, u_T - u_{T-1}),$$
$$u_{-1} = (0, u_1, u_2, \dots, u_{T-1}).$$

Now define the scaling matrix $\Upsilon_T = diag(1, T^{1/2}, T^{1/2})$, we can write expression (A.1) as

$$\Upsilon_T(\hat{\psi}(\delta) - \psi) = \Gamma_T(\delta)^{-1}\Psi_T(\delta), \tag{A.2}$$

where

$$\Gamma_T(\delta) = \Upsilon_T^{-1}[(\Delta z - \bar{c}T^{-1}z_{-1})'(\Delta z - \bar{c}T^{-1}z_{-1})]\Upsilon_T^{-1},$$

and

$$\Psi_T(\delta) = \Upsilon_T^{-1}[(\Delta z - \bar{c}T^{-1}z_{-1})'(\Delta u - \bar{c}T^{-1}u_{-1})].$$

We first consider the limit of each element of the matrix $\Gamma_T(\delta)$ denoted Γ_{ij} $(i,j{=}1,2,3)$. We let $\Delta z_{(i)}$ and $z_{-1(i)}$ be the ith element of the vectors Δz and z_{-1}, respectively. We have

$$\Gamma_{11} = (\Delta z_{(1)} - \bar{c}T^{-1}z_{-1(1)})'(\Delta z_{(1)} - \bar{c}T^{-1}z_{-1(1)}) \Rightarrow 1,$$

$$\Gamma_{12} = T^{-1/2}(\Delta z_{(1)} - \bar{c}T^{-1}z_{-1(1)})'(\Delta z_{(2)} - \bar{c}T^{-1}z_{-1(2)}) \Rightarrow 0,$$

$$\Gamma_{13} = T^{-1/2}(\Delta z_{(1)} - \bar{c}T^{-1}z_{-1(1)})'(\Delta z_{(3)} - \bar{c}T^{-1}z_{-1(3)}) \Rightarrow 0,$$

$$\Gamma_{22} = T^{-1}(\Delta z_{(2)} - \bar{c}T^{-1}z_{-1(2)})'(\Delta z_{(2)} - \bar{c}T^{-1}z_{-1(2)}) \Rightarrow 1 - \bar{c} + \bar{c}^2/3 \equiv a,$$

$$\Gamma_{23} = T^{-1}(\Delta z_{(2)} - \bar{c}T^{-1}z_{-1(2)})'(\Delta z_{(3)} - \bar{c}T^{-1}z_{-1(3)})$$
$$\Rightarrow 1 - \delta - \bar{c} + \bar{c}\delta - (\bar{c}^2/2)\delta + (\bar{c}^2/2)\delta^3 + (\bar{c}^2/3)(1 - \delta^3) \equiv m,$$

$$\Gamma_{33} = T^{-1}(\Delta z_{(3)} - \bar{c}T^{-1}z_{-1(3)})'(\Delta z_{(3)} - \bar{c}T^{-1}z_{-1(3)})$$
$$\Rightarrow 1 - \delta - \bar{c} + 2\bar{c}\delta - \bar{c}\delta^2 - \bar{c}^2\delta + \bar{c}^2\delta^2 + (\bar{c}^2/3)(1-\delta^3) \equiv d.$$

We next consider the limit of each element of the vector $\Psi_T(\delta)$, denoted Ψ_i $(i=1,2,3)$. We have

$$\Psi_1 = (\Delta z_{(1)} - \bar{c}T^{-1}z_{-1(1)})'(\Delta u - \bar{c}T^{-1}u_{-1}) \Rightarrow v_1,$$

$$\Psi_2 = T^{-1/2}(\Delta z_{(2)} - \bar{c}T^{-1}z_{-1(2)})'(\Delta u - \bar{c}T^{-1}u_{-1})$$
$$\Rightarrow \sigma\left[W_c(1)(1-\bar{c}) + \bar{c}^2\int_0^1 rW_c(r)\,\mathrm{d}r\right] \equiv \sigma b_1,$$

$$\Psi_3 = T^{-1/2}(\Delta z_{(3)} - \bar{c}T^{-1}z_{-1(3)})'(\Delta u - \bar{c}T^{-1}u_{-1})$$
$$\Rightarrow \sigma\left[W_c(1)(1-\bar{c}+\delta\bar{c}) + \bar{c}^2\int_\delta^1 W_c(r)(r-\delta)\,\mathrm{d}r - W_c(\delta)\right] \equiv \sigma b_2.$$

Hence, using the symmetry of $\Gamma_T(\delta)$,

$$\Upsilon_T(\hat{\psi}(\delta) - \psi) \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & m \\ 0 & m & d \end{bmatrix}^{-1} \begin{bmatrix} v_1 \\ \sigma b_1 \\ \sigma b_2 \end{bmatrix}. \qquad \square$$

The proof of the lemma follows upon solving for the inverse. For Model II, we have

Lemma A.3. *Suppose that y_t is generated by* (1) *with $\alpha = 1 + c/T$ and $\{z_t\}$ is given by* (5). *Let $\hat{\psi}(\delta)$ be the* GLS *estimates, from minimizing* (3), *of the coefficients of the trend function obtained using $\bar{\alpha} = 1 + \bar{c}/T$. Then, the result of Lemma A.2 still apply with the addition that $\hat{\mu}_2 - \mu_2 \Rightarrow \lim_{T\to\infty} v_{[T\delta]+1} \equiv v^*$.*

The proof of Lemma A.3 is basically the same as that of Lemma A.2 and, hence, omitted.

Proof of Theorem 1. The proof uses the results of Lemmas A.2 and A.3. We show the proof only for Model I and for the $MZ_\alpha^{\mathrm{GLS}}(\delta)$ test, the proof for the other Model and tests follows analogously. We first have

$$T^{-1}\tilde{y}_T^2 = T^{-1}\{y_T - (\hat{\mu}_1 + \hat{\beta}_1 T + \hat{\beta}_2(T - T\delta))\}^2$$
$$= T^{-1}\{u_T - [(\hat{\mu}_1 - \mu_1) + (\hat{\beta}_1 - \beta_1)T + (\hat{\beta}_2 - \beta_2)(T - T\delta)]\}^2.$$

After some algebra, we obtain: $T^{-1}u_T^2 \Rightarrow \sigma^2 W_c(1)^2$; $2T^{-1}u_T(\hat{\mu}_1 - \mu_1) \Rightarrow 0$,

$$2T^{-1}u_T(\hat{\beta}_1 - \beta_1)T \Rightarrow 2\sigma^2 b_3 W_c(1),$$

$$2T^{-1}u_T(\hat{\beta}_2 - \beta_2)(T - T\delta) \Rightarrow 2\sigma^2 b_4 W_c(1)(1 - \delta),$$

$$2T^{-1}(\hat{\mu}_1 - \mu_1)(\hat{\beta}_1 - \beta_1)T \Rightarrow 0,$$

$$T^{-1}(\hat{\beta}_1 - \beta_1)^2 T^2 \Rightarrow \sigma^2 b_3^2,$$

$$2T^{-1}(\hat{\mu}_1 - \mu_1)(\hat{\beta}_2 - \beta_2)(T - T\delta) \Rightarrow 0,$$

$$2T^{-1}(\hat{\beta}_1 - \beta_1)T(\hat{\beta}_2 - \beta_2)(T - T\delta) \Rightarrow 2\sigma^2 b_3 b_4(1 - \delta),$$

$$T^{-1}(\hat{\beta}_2 - \beta_2)^2(T - T\delta)^2 \Rightarrow \sigma^2 b_4^2(1 - \delta)^2.$$

Using these results, we have

$$T^{-1}\tilde{y}_T^2 \Rightarrow \sigma^2\{V_{c\bar{c}}^{(1)}(1,\delta)^2 - 2V_{c\bar{c}}^{(2)}(1,\delta)\}, \tag{A.3}$$

where $V_{c\bar{c}}^{(1)}(1,\delta) = W_c(1) - b_3$, and

$$V_{c\bar{c}}^{(2)}(1,\delta) = b_4(1 - \delta)[W_c(1) - b_3 - (1/2)(1 - \delta)b_4].$$

Consider now the term $2T^{-2}\sum_{t=1}^{T}\tilde{y}_t^2$, defined by

$$\begin{aligned}2T^{-2}\sum_{t=1}^{T}\tilde{y}_t^2 &= 2T^{-2}\sum_{t=1}^{T}\{y_t - [\hat{\mu}_1 + \hat{\beta}_1 t + \hat{\beta}_2 1(t > T\delta)(t - T\delta)]\}^2\\ &= 2T^{-2}\sum_{t=1}^{T}\{u_t - [(\hat{\mu}_1 - \mu_1) + (\hat{\beta}_1 - \beta_1)t\\ &\quad + (\hat{\beta}_2 - \beta_2)1(t > T\delta)(t - T\delta)]\}^2.\end{aligned}$$

After some algebra, we obtain

$$2T^{-2}\sum_{t=1}^{T}u_t^2 \Rightarrow 2\sigma^2\int_0^1 W_c(r)^2\,\mathrm{d}r,$$

$$4T^{-2}(\hat{\mu}_1 - \mu_1)\sum_{t=1}^{T}u_t \Rightarrow 0,$$

$$4T^{-2}(\hat{\beta}_1 - \beta_1)\sum_{t=1}^{T}tu_t \Rightarrow 4\sigma^2\int_0^1 rb_3 W_c(r)\,\mathrm{d}r,$$

$$4T^{-2}(\hat{\beta}_2 - \beta_2)\sum_{t=1}^{T}1(t > T\delta)(t - T\delta)u_t \Rightarrow 4\sigma^2\int_\delta^1 b_4 W_c(r)(r - \delta)\,\mathrm{d}r,$$

$$2T^{-1}(\hat{\mu}_1-\mu_1)^2 \Rightarrow 0,$$

$$4T^{-2}(\hat{\mu}_1-\mu_1)(\hat{\beta}_1-\beta_1)\sum_{t=1}^{T} t \Rightarrow 0,$$

$$2T^{-2}(\hat{\beta}_1-\beta_1)^2\sum_{t=1}^{T} t^2 \Rightarrow 2\sigma^2\int_0^1 b_3^2 r^2\,\mathrm{d}r,$$

$$4T^{-2}(\hat{\mu}_1-\mu_1)(\hat{\beta}_2-\beta_2)\sum_{t=1}^{T} 1(t>T\delta)(t-T\delta) \Rightarrow 0,$$

$$4T^{-2}(\hat{\beta}_1-\beta_1)(\hat{\beta}_2-\beta_2)\sum_{t=1}^{T} t1(t>T\delta)(t-T\delta) \Rightarrow 4\sigma^2\int_\delta^1 b_3b_4r(r-\delta)\,\mathrm{d}r,$$

$$2T^{-2}(\hat{\beta}_2-\beta_2)^2\sum_{t=1}^{T} 1(t>T\delta)(t-T\delta)^2 \Rightarrow 2\sigma^2\int_\delta^1 b_4^2(r-\delta)^2\,\mathrm{d}r.$$

Using these results we have

$$2T^{-2}\sum_{t=1}^{T}\tilde{y}_t^2 \Rightarrow 2\sigma^2\left\{\int_0^1 V_{c\bar{c}}^{(1)}(r,\delta)^2\,\mathrm{d}r - 2\int_\delta^1 V_{c\bar{c}}^{(2)}(r,\delta)\,\mathrm{d}r\right\}. \tag{A.4}$$

Using (A.3), (A.4) and the fact that s^2 is a consistent estimate of σ^2, the proof is complete. □

Proof of Theorem 2. We first give the proof for Model I. Defining

$$M_T(c,\bar{c},\delta)=(u^{\alpha\prime}z^\alpha)(z^{\alpha\prime}z^\alpha)^{-1}(z^{\alpha\prime}u^\alpha),$$

we have $S(\bar{\alpha},\delta)=u^{\bar{\alpha}\prime}u^{\bar{\alpha}}-M_T(c,\bar{c},\delta)$ and $S(1)=u^{1\prime}u^1-M_T(c,0,\delta)$. Using the fact that

$$u_t^{\bar{\alpha}}=v_t+T^{-1}(c-\bar{c})u_{t-1}$$

for $t=2,\dots,T$ and $u_1^{\bar{\alpha}}=v_1$, we have

$$s^2P_T^{\mathrm{GLS}}(c,\bar{c},\delta)=M_T(c,0,\delta)-M_T(c,\bar{c},\delta)-2\bar{c}T^{-1}\sum_{t=2}^{T}u_{t-1}v_t$$
$$+(\bar{c}^2-2\bar{c}c)T^{-2}\sum_{t=2}^{T}u_{t-1}^2-cT^{-1}u^{1\prime}u^1+o_p(1). \tag{A.5}$$

Note that $T^{-1}\sum_{t=2}^{T}u_{t-1}v_t \Rightarrow \sigma^2\int_0^1 W_c(r)\,\mathrm{d}W(r)$ and $T^{-2}\sum_{t=2}^{T}u_{t-1}^2 \Rightarrow \sigma^2\int_0^1 W_c(r)^2\,\mathrm{d}r$. Consider now the limit of $M_T(c,\bar{c},\delta)$. Using the scaling matrix $D_T=diag(1,T^{-1/2},T^{-1/2})$, we have $M_T(c,\bar{c},\delta)=(u^{\alpha\prime}z^\alpha D_T)(D_Tz^{\alpha\prime}z^\alpha D_T)^{-1}(D_Tz^{\alpha\prime}u^\alpha)$. The first term is given by

$$D_T u^{\alpha\prime} z^{\alpha} = \begin{bmatrix} v_1 + T^{-1}(c-\bar{c})u_0 - \bar{c}T^{-1}\sum_{t=2}^{T}[v_t + (c-\bar{c})T^{-1}u_{t-1}] \\ T^{-1/2}\sum_{t=2}^{T}(v_t + (c-\bar{c})T^{-1}u_{t-1})(-T^{-1}\bar{c}t + 1) + o_p(1) \\ T^{-1/2}\sum_{t=T_B+1}^{T}(v_t + (c-\bar{c})T^{-1}u_{t-1})(-T^{-1}\bar{c}t + 1 + \delta\bar{c}) \end{bmatrix}$$

and its limit is

$$\sigma \begin{bmatrix} v_1/\sigma \\ W(1) + (c-\bar{c})\int_0^1 W_c(r)\,\mathrm{d}r \\ \qquad -\bar{c}\int_0^1 r\,\mathrm{d}W(r) - (c-\bar{c})\bar{c}\int_0^1 rW_c(r)\,\mathrm{d}r \\ (1+\delta\bar{c})\left([W(1) - W(\delta)] + (c-\bar{c})\int_\delta^1 W_c(r)\,\mathrm{d}r\right) \\ \qquad -\bar{c}\int_\delta^1 r\,\mathrm{d}W(r) - (c-\bar{c})\bar{c}\int_\delta^1 rW_c(r)\,\mathrm{d}r \end{bmatrix}.$$

The term $D_T z^{\alpha\prime} z^{\alpha} D_T$ is given by

$$\begin{bmatrix} 1 + \sum_{t=2}^{T}(-\bar{c}/T)^2 & T^{-1/2}\left[1 + \sum_{t=2}^{T}(-\bar{c}/T)(-\bar{c}t/T + 1)\right] & T^{-1/2}\left[\sum_{t=T_B+1}^{T}(-\bar{c}/T)(-\bar{c}t/T + 1 + \delta\bar{c})\right] \\ & T^{-1}\left[1 + \sum_{t=2}^{T}(-\bar{c}t/T + 1)^2\right] & T^{-1}\sum_{t=T_B+1}^{T}(-\bar{c}t/T + 1)(-\bar{c}t/T + 1 + \delta\bar{c}) \\ & & T^{-1}\sum_{t=T_B+1}^{T}(-\bar{c}t/T + 1 + \delta\bar{c})^2 \end{bmatrix}$$

and its limit is

$$\begin{bmatrix} 1 & 0 & 0 \\ & \bar{c}^2/3-\bar{c}+1 & (1-\delta)(1-\bar{c})+\bar{c}^2(2+\delta^3-3\delta)/6 \\ & & \bar{c}^2(1-\delta^3)/3-\bar{c}(1-\delta^2)(1+\delta\bar{c})+(1-\delta)(1+\delta\bar{c})^2 \end{bmatrix}.$$

Simple algebra shows that $M_T(c,\bar{c},\delta) \Rightarrow v_1^2+\sigma^2 M(c,\bar{c},\delta)$ and the result of the theorem follows using (A.5). The proof for Model II is entirely analogous and, hence, omitted. The result stated in (16) follows using the fact that $S(\bar{\alpha},\delta)=u^{\bar{\alpha}\prime}u^{\bar{\alpha}}-M_T(c,\bar{c},\delta)$ which depends on δ only through $M_T(c,\bar{c},\delta)$ which enters with a negative sign, hence taking the supremum instead of the infimum. □

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Banerjee, A., Lumsdaine, R., Stock, J.H., 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.

Bhargava, A., 1986. On the theory of testing for unit root in observed time series. Review of Economic Studies 53, 369–384.

Campbell, J.Y., Mankiw, N.G., 1987a. Permanent and transitory components in macroeconomics fluctuations. American Economic Review: Papers and Proceedings 77, 111–117.

Campbell, J.Y., Mankiw, N.G., 1987b. Are output fluctuations transitory? Quarterly Journal of Economics 102, 857–880.

Campbell, J.Y., Perron, P., 1991. Pitfalls and opportunities: what macroeconomists should know about unit roots. In: Blanchard, O.J., Fischer, S. (Eds.), NBER Macroeconomics Annual 6, 141–201.

Christiano, L.J., 1992. Searching for a break in GNP. Journal of Business and Economics Statistics 10, 237–250.

Clark, P.K., 1987. The cyclical component of U.S. economic activity. Quarterly Journal of Economics 102, 798–814.

Cochrane, J.H., 1988. How big is the random walk in GNP? Journal of Political Economy 96, 893–920.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimator for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Dufour, J.-M., King, M., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary errors. Journal of Econometrics 47, 115–143.

Elliott, G., Rothenberg, T., Stock, J.H., 1996. Efficient tests for an autoregressive unit root. Econometrica 64, 813–839.

Nelson, C.R., Plosser, C.I., 1982. Trends and random walks in macroeconomics time series: some evidence and implications. Journal of Monetary Economics 10, 139–162.

Ng, S., Perron, P., 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.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.

Perron, P., 1997. Further evidence of breaking trend functions in macroeconomic variables. Journal of Econometrics 80, 355–385.

Perron, P., Ng, S., 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Review of Economics Studies 63, 435–463.

Perron, P., Ng, S., 1998. An autoregressive spectral density estimator at frequency zero for nonstationarity tests. Econometric Theory 14, 560–603.

Phillips, P.C.B., 1987. Time series regression with unit roots. Econometrica 55, 277–302.

Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. Biometrika 75, 335–346.

Rodríguez, G., 1999. Unit root, outliers and cointegration analysis with macroeconomic applications. Unpublished Ph.D. Dissertation, Département de Sciences Économiques, Université de Montréal.

Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599–608.

Sargan, J.D., Bhargava, A., 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk. Econometrica 51, 153–174.

Shapiro, M., Watson, M., 1988. Sources and business cycles fluctuations. In: Fischer, S. (Ed.), NBER Macroeconomics Annual, Vol. 3, pp. 111–148.

Stock, J.H., 1994. Unit roots, structural breaks and trends. In: McFaden, D.L., Engle, R.F. (Eds.), Handbook of Econometrics, Vol. IV. Elsevier Science, Amsterdam, pp. 2740–2841.

Stock, J.H., 1999. A class of tests for integration and cointegration. In: Engle, R.F., White, H. (Eds.), Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, Oxford, pp. 137–167.

Vogelsang, T.J., Perron, P., 1998. Additional tests for a unit root allowing the possibility of breaks in the trend function. International Economic Review 39, 1073–1100.

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 and Economics Statistics 10, 251–270.