ELSEVIER Journal of Econometrics 80 (1997) 355–385

JOURNAL OF Econometrics

Further evidence on breaking trend functions in macroeconomic variables

Pierre Perron

Université de Montréal, C.R.D.E., Montréal, Qué., Canada H3C 3J7

Abstract

This study first reexamines the findings of Perron (1989) regarding the claim that most macroeconomic time series are best construed as stationary fluctuations around a deterministic trend function if allowance is made for the possibility of a shift in the intercept of the trend function in 1929 (a crash) and a shift in slope in 1973 (a slowdown in growth). Unlike that previous study, the date of possible change is not fixed a priori but is considered as unknown. We consider various methods to select the break points and the asymptotic and finite sample distributions of the corresponding statistics. A detailed discussion about the choice of the truncation lag parameter in the autoregression and of its effect on the critical values is also included. Most of the rejections reported in Perron (1989) are confirmed using this approach. Secondly, this paper investigates an international data set of post-war quarterly real GNP (or GDP) series for the G-7 countries. Our results are compared and contrasted to those of Banerjee et al. (1992) and Zivot and Andrews (1992). In contrast to the theoretical results contained in these papers, we derive the limiting distribution of the sequential test without trimming. © 1997 Elsevier Science S.A.

Keywords: Hypothesis testing, structural change; Stochastic trends; Deterministic trends; Simulation experiment; Unit root
JEL classification: C22; E30

I wish to thank Tim Vogelsang for computational assistance beyond the call of duty. This research was supported by grants from the Social Sciences and Humanities Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche du Québec. John Campbell, Robert Shiller and Charles Nelson kindly provided some of the data used. I also wish to thank Don Andrews, René Garcia, Robin Lumsdaine, Serena Ng, Jim Stock and Tim Vogelsang for comments on an earlier version. This is a substantially revised version of the 1990 Princeton University Econometric Research Program Memorandum No. 350.

PII S0304-4076(97)00049-3

1. Introduction

In a previous paper, Perron (1989), we argued that many macroeconomic time series could be represented as stationary fluctuations around a deterministic trend function if allowance is made for a possible change in its intercept in 1929 (a crash) and in its slope in 1973 (a slowdown in growth). The test statistics were constructed by adding dummy variables for different intercepts and slopes, extending the standard Dickey–Fuller procedure. The asymptotic distribution theory underlying the critical values obtained under the different models assumed that the dating of the break points was known a priori, or more precisely, that the dates chosen were uncorrelated with the data.

This postulate has been criticized, most notably by Christiano (1992) who argued that the choice of these dates had to be viewed, to a large extent, as being correlated with the data. This is an important problem because both the finite sample and asymptotic distributions of the statistics depend upon the extent of the correlation between the choice of the break points and the data. There is a sense, as argued before, in which the choice of these dates can be regarded as independent of the data. First, the dates used in the previous study were chosen ex-ante and not modified ex-post. Secondly, these dates are related to exogenous events for which economic theory would suggest the effects that actually happened; e.g., the stock market crash of 1929 with the ensuing dismantle of the economic organization and the exogenous sudden change in oil prices with the resulting alteration of international economic coordination and policies.

In the sense described above the choice of the dates can be viewed as uncorrelated with the data. There is, however, a validity to the argument that it is only ex-post (after looking at the data) that we can say that the changes that followed these exogenous events actually occurred as predicted by the theory. Furthermore, many other exogenous events did not have the major impact that some theories would have predicted. In this sense, the choice of the break points must be viewed as being correlated, at least to some extent, with the data. To what extent is a difficult and practically impossible question to answer. At the very least the choices were not perfectly correlated with the data as no attempts were systematically made to maximize the chances that the unit root be rejected nor to find where, according to some test criteria, were the most likely dates of change.

While we still believe that the assumption about the exogeneity of the choice of the break points is a good first approximation to the true extent of the correlation with the data, it is useful to investigate how robust the results are to different postulates. The aim of this paper is to take the extreme view where the choice of the break points is effectively made to be perfectly correlated with the data. This case is instructive to study because if one can still reject the unit-root hypothesis under such a scenario it must be the case that it would be rejected under a less stringent assumption.

We proceed as follows for the practical implementation. Again, as in the previous analysis, only one possible break point is allowed for any single series. This break point is first chosen such that the t-statistic for testing the null hypothesis of a unit root is smallest among all possible break points. We also consider choosing the break point that corresponds to a minimal t-statistic on the parameter of the change in the trend function. This allows the mild a priori imposition of a one-sided change which permits substantial gains in power. We also investigate various issues regarding the choice of the autoregressive truncation lag and its effect on the finite sample critical values.

Our paper is closely related to and complements those of Banerjee et al. (1992) and Zivot and Andrews (1992) in that similar procedures and series are analyzed. We extend their analysis in several directions. On a methodological level, we consider the asymptotic distribution of the sequential test based on the minimal value of the unit-root tests over possible break points. We show the results of Zivot and Andrews (1992) to be valid without any trimming at the end points. The proof, which is of interest in itself, is based on projection arguments and introduces a method that can be applied to a variety of frameworks. Concerning the empirical results, our analysis is more extensive and shows that alternative procedures can lead to conclusions that are less favorable to the unit root than suggested in these two studies. We pay particular attention to the importance of the selection of the truncation lag on the outcome of the tests.

The paper is organized as follows. Section 2 reviews the statistics involved. Section 3 discusses their asymptotic distribution under the null hypothesis of a unit root and Section 4 their finite sample distribution using simulation methods. Section 5 contains simulation experiments providing information about their size and power under various data-generating processes. Section 6 presents the empirical results for the Nelson–Plosser (1982) data set and Section 7 analyzes an international data set of post-war quarterly real GNP series. Section 8 offers concluding comments and a mathematical appendix contains the derivation of the limiting distributions.

2. The models and statistics

In this section, we briefly review the statistical procedures used to test for a unit root allowing for the presence of a change in the trend function occurring at most once. The reader is referred to Perron (1989) for details on the models. Throughout, T_b denotes the time at which the change in the trend function occurs. The first model allows only a change in the intercept under both the null and alternative hypotheses. Furthermore, this change is assumed to occur gradually and in a way that depends on the correlation structure of the noise function. This was termed the 'innovational outlier model', and the unit-root test is performed using the t-statistic for testing $\alpha = 1$ in the following regression:

$$y_t = \mu + \theta \mathrm{DU}_t + \beta t + \delta \mathrm{D}(T_\mathrm{b})_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + e_t, \tag{1}$$

where $DU_t = 1(t > T_\mathrm{b})$ and $\mathrm{D}(T_\mathrm{b})_t = 1(t = T_\mathrm{b} + 1)$ with $1(\cdot)$ the indicator function. Regression (1), like the others that follow, is estimated by OLS and is in the spirit of the Dickey–Fuller (1979) and Said–Dickey (1984) methodology whereby autoregressive moving average processes are approximated by autoregressive processes. Under the second model, both a change in the intercept and the slope are allowed at time T_b. The test is performed using the t-statistic for the null hypothesis that $\alpha = 1$ in the regression:

$$y_t = \mu + \theta \mathrm{DU}_t + \beta t + \gamma \mathrm{DT}_t + \delta \mathrm{D}(T_\mathrm{b})_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + e_t, \tag{2}$$

with $\mathrm{D}T_t = 1(t > T_\mathrm{b})t$. Under the third model, a change in the slope is allowed but both segments of the trend function are joined at the time of break. Here the change is presumed to occur rapidly and corresponds to the 'additive outlier model' in the terminology of Perron (1989). We use the following two-step procedure. First, the series is detrended using the following regression where $\mathrm{D}T_t^* = 1(t > T_\mathrm{b})(t - T_\mathrm{b})$:

$$y_t = \mu + \beta t + \gamma \mathrm{D}T_t^* + \tilde{y}_t. \tag{3a}$$

The test is then performed using the t-statistic for $\alpha = 1$ in the regression:

$$\tilde{y}_t = \alpha \tilde{y}_{t-1} + \sum_{i=1}^{k} c_i \Delta \tilde{y}_{t-i} + e_t. \tag{3b}$$

We denote by $t_{\hat{\alpha}}(i, T_\mathrm{b}, k)$ $(i = 1, 2, 3)$, the t-statistic for testing $\alpha = 1$ under model i with a break date T_b and truncation lag parameter k (using regressions (1), (2) and (3b) for $i = 1, 2,$ and 3, respectively). In these regressions, T_b and k are treated as unknown. We next describe various data-dependent methods to select these values endogeneously.

2.1. Methods to choose the break date T_b

We consider two methods to select T_b endogenously. First T_b is selected as the value which minimizes the t-statistic for testing $\alpha = 1$. We define the statistics as $t_{\hat{\alpha}}^*(i) = \mathrm{Min}_{T_\mathrm{b} \in (k+1, T)} t_{\hat{\alpha}}(i, T_\mathrm{b}, k)$ $(i = 1, 2, 3)$. The asymptotic distribution of $t_{\hat{\alpha}}^*(1)$ and $t_{\hat{\alpha}}^*(2)$ was studied by Zivot and Andrews (1992) under the condition that the range of possible values for the break point be restricted to some subset that excludes values at each end of the sample. In the next section, we show their result to remain valid even without trimming.

Secondly, T_b is chosen to minimize either $t_{\hat{\theta}}$, the t-statistic on the parameter associated with the change in the intercept (Model 1) or $t_{\hat{\gamma}}$, the t-statistic on the change in slope (Models 2 and 3). We denote the t-statistic on α (for a null hypothesis that $\alpha = 1$) obtained from such a procedure by $t^*_{\hat{\alpha},\theta}(1)$ for Model 1 and by $t^*_{\hat{\alpha},\gamma}(i)$ $(i = 2, 3)$ for Models 2 and 3. More precisely, $t^*_{\hat{\alpha},\theta}(1) = t_{\hat{\alpha}}(1, T^*_b, k)$, where T^*_b is such that $t_{\hat{\theta}}(T^*_b) = \mathrm{Min}_{T_b \in (k+1, T)} t_{\hat{\theta}}(T_b, k)$, where again different specifications about the choice of k will be analyzed. The statistics $t^*_{\hat{\alpha},\gamma}(i)$ $(i = 2, 3)$ are defined in an analogous fashion. This procedure allows the possibility of imposing the mild a priori restriction of a one-sided change, i.e. allowing the date of the change to be unknown but restricting the analysis to the cases of a 'crash' or a slowdown in growth. We also discuss the case where the break point is selected using the same procedure without any a priori assumption on the sign of the change. In this context the break date is selected using the maximum of the absolute value of $t_{\hat{\theta}}$ or $t_{\hat{\gamma}}$. The corresponding statistics are denoted by $t^*_{\hat{\alpha},|\theta|}(1)$ for Model 1 and $t^*_{\hat{\alpha},|\gamma|}(i)$ $(i = 2, 3)$ for Models 2 and 3.

2.2. Methods to select the truncation lag parameter k

There is now evidence that using data-dependent methods to select the truncation lag parameter k leads to test statistics having better properties (stable size and higher power) than if a fixed k is chosen a priori (unless, of course, one happens to select that value of k which is best), see Ng and Perron (1995), Perron and Vogelsang (1992) and Hall (1994). We consider two such data-dependent methods. The first is a general to specific recursive procedure based on the t-statistic on the coefficient associated with the last lag in the estimated autoregression. More specifically, the procedure selects that value of k, say k^*, such that the coefficient on the last lag in an autoregression of order k^* is significant and that the last coefficient in an autoregression of order greater than k^* is insignificant, up to a maximum order k max. In the simulations and empirical applications below, we use a two-sided 10% test based on the asymptotic normal distribution to assess the significance of the last lags. This procedure is denoted below as 't-sig'.

Said and Dickey (1984) use yet a different method in their empirical application. It is based on testing whether additional lags are jointly significant using an 'F-test' on the estimated coefficients. First a maximum value of k, k max, is specified and the autoregressions with k max and (k max $- 1$) lags are estimated. A 10% one-tailed F-test is used to assess whether the coefficient on the k maxth lag is significant and if so, the value of k chosen is this maximum value. If not, the model is estimated with (k max $- 2$) lags. The lag(k max $- 1$) is deemed significant if either the F-test for (k max $- 2$) versus (k max $- 1$) lags or the F-test for (k max $- 2$) versus k max lags are significant based on the 10% critical values of the chi-square distribution. This is repeated by lowering k until a rejection that additional lags are insignificant occurs or some lower bound is attained. In the empirical applications, the lower bound is set to $k = 1$. This procedure is denoted below as 'F-sig'.

We choose these 'general to specific' procedures rather than methods based on information criteria, such as AIC, because the latter tend to select very parsimonious models leading to tests with sometimes serious size distortions and/or power losses with data in the class of ARMA processes. Indeed, Ng and Perron (1995) show that using an information criterion leads to a selected value of k that increases to infinity, as T increases, only at the very slow rate $\log(T)$. This is consistent with many empirical results showing that using the AIC leads to very small values of k being selected (typically 0 or 1) and that often times the estimated residuals exhibit serial correlation (see Perron, 1994).

3. The asymptotic distribution of the statistics

In this section, we consider the limiting distribution of the statistics. To simplify the derivations we suppose the data-generating process to be a random walk,

$$y_t = y_{t-1} + e_t, \quad (t = 1, 2, \dots, T), \tag{4}$$

where the errors e_t are martingale differences (y_0 is some fixed value), and consider the statistics constructed with $k = 0$. Using arguments in Ng and Perron (1995), we can then state that the resulting limiting distribution remains the same when additional correlation is present and the statistics are constructed with one of the data-dependent methods to select k. This holds provided $k\max^3/T \to 0$ as $T \to \infty$. This is the same strategy as used by Zivot and Andrews (1992) and Banerjee et al. (1992). All statistics are asymptotically invariant to a change in intercept. Vogelsang and Perron (1994) show that they are not asymptotically invariant to a change in slope but that the asymptotic distribution corresponding to a zero change in slope is a better approximation to the finite sample distribution for values typically encountered in practice. The following Theorem concerning the asymptotic distribution of $t_{\hat{\alpha}}^*(i)$ $(i = 1, 2, 3)$ is proved in the appendix.

Theorem 1. *Let* $\{y_t\}_0^T$ *be generated by* (4) *and denote by* '$\Rightarrow$' *weak convergence in distribution from the space* D[0, 1] *to the space* C[0, 1] *using the uniform metric on the space of functions on* [0, 1]. *Then:*

(a) *for* $i = 1, 2$:

$$\inf_{T_b \in [1, T]} t_{\hat{\alpha}}(i, T_b, k = 0) \Rightarrow \inf_{\lambda \in [0, 1]} \int_0^1 W_i(r, \lambda)\, \mathrm{d}W(r) \Big/ \left[\int_0^1 W_i(r, \lambda)^2\, \mathrm{d}r \right]^{1/2},$$

(b) *for Model* 3:

$$\inf_{T_b \in [1,T]} t_{\hat{\alpha}}(i, T_b, k = 0) \Rightarrow \inf_{\lambda \in [0,1]} \left[\int_0^1 W_3(r,\lambda)\,\mathrm{d}W(r) - a \int_\lambda^1 (r-\lambda) W_0(r)\,\mathrm{d}r \int_0^1 W_3(r,\lambda)^2\,\mathrm{d}r \right] \Big/ \left[\int_0^1 W_3(r,\lambda)^2\,\mathrm{d}r \right]^{1/2},$$

where $a = (\lambda^3(1-\lambda)^3/3)^{-1}$, $W_0(r)$ *and* $W_i(r,\lambda)$ *are residuals from a projection of a standard Wiener process* $W(r)$ *onto the subspace generated by the functions* $\{1, r\}$ $(i = 0)$, $\{1, r, \mathrm{d}u(r,\lambda)\}$ $(i = 1)$, $\{1, r, \mathrm{d}u(r,\lambda), \mathrm{d}t^*(r,\lambda)\}$ $(i = 2)$ and $\{1, r, \mathrm{d}t^*(r,\lambda)\}$ $(i = 3)$, *with* $\mathrm{d}u(r,\lambda) = 1(r > \lambda)$ *and* $\mathrm{d}t^*(r,\lambda) = 1(r > \lambda)(r - \lambda)$.

Theorem 1 differs from the results in Zivot and Andrews (1992) in two respects. First there is no need to have the hybrid metric considered in that paper. The weak convergence results hold under the uniform metric. This is achieved using arguments in Gregory and Hansen (1996) so that there is no need for a weak convergence result for the sample moments of DU_t or DT_t^*. The most important and novel aspect in which our result differs is that we do not require that the possible range of values for the break point be restricted to exclude the end points. To achieve this, our proof is rather different and somewhat more involved and is based on projection arguments. The intuition is quite simple. With a break at either end points, the regressions indeed exhibit perfect multicollinearity but the coefficient on the lagged-dependent variable, α, is a linear combination of the parameter vector that is identifiable and estimable and its t-statistic is also well defined. In such cases, the regressions become equivalent to ones without dummy and the standard limiting distribution of Dickey and Fuller (1979) applies. This last result is important because it makes it unnecessary to use an arbitrary trimming near the end points, such as the 15% exclusion on both sides suggested by Banerjee et al. (1992). The arguments in the proof of Theorem 1 can also be applied to other context such as the cointegration tests with regime shifts considered by Gregory and Hansen (1996).

We used simulations to obtain the percentage points of the asymptotic distributions described above. These were based on 10,000 replications using partial sums of i.i.d. N(0, 1) random variables to approximate the Wiener process and 1000 steps to compute the integrals. The critical values are presented in the rows labelled '$T = \infty$' in Table 1.

This relaxation of the need for trimming at the end points does not appear to be possible for the tests whereby the break point is chosen with respect to the t-statistic on the coefficient of the intercept or slope change. The asymptotic distributions of $t^*_{\alpha,\theta}(1)$ and $t^*_{\alpha,|\theta|}(1)$ assuming the break point to be in some compact subset was derived in Banerjee et al. (1992). Similar asymptotic results are in Vogelsang and Perron (1994) for $t^*_{\alpha,\gamma}(i)$ and $t^*_{\alpha,|\gamma|}(i)$ $(i = 2, 3)$. The critical values are reproduced in Table 1.

Table 1
Finite sample and asymptotic distributions

		1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
(a) *Model* 1, $t_{\alpha}^{*}(1)$, *Choosing* T_b *minimizing* $t_{\hat{\alpha}}$										
$T = 60$	k(F-sig)	−5.83	−5.49	−5.21	−4.91	−3.91	−3.00	−2.70	−2.41	−1.96
	k(t-sig)	−5.92	−5.58	−5.23	−4.92	−3.91	−3.00	−2.74	−2.55	−2.25
$T = 80$	k(F-sig)	−5.77	−5.35	−5.15	−4.84	−3.87	−2.96	−2.70	−2.41	−2.12
	k(t-sig)	−5.77	−5.31	−5.09	−4.84	−3.88	−2.95	−2.73	−2.55	−2.22
$T = 100$	k(F-sig)	−5.70	−5.35	−5.09	−4.82	−3.89	−3.00	−2.74	−2.46	−2.22
	k(t-sig)	−5.70	−5.36	−5.10	−4.82	−3.87	−3.05	−2.75	−2.46	−2.22
$T = \infty$		−5.41	−5.02	−4.80	−4.58	−3.75	−2.99	−2.77	−2.56	−2.32
(b) *Model* 1, $t_{\alpha,\theta}^{*}(1)$, *Choosing* T_b *minimizing* $t_{\hat{\theta}}$										
$T = 60$	k(F-sig)	−5.58	−5.15	−4.88	−4.47	−3.33	−1.60	−0.84	−0.05	0.56
	k(t-sig)	−5.70	−5.21	−4.92	−4.53	−3.32	−1.79	−1.14	−0.35	0.42
$T = 80$	k(F-sig)	−5.50	−5.11	−4.85	−4.53	−3.33	−1.86	−1.06	−0.32	0.67
	k(t-sig)	−5.59	−5.09	−4.83	−4.54	−3.33	−1.92	−1.19	−0.46	0.34
$T = 100$	k(F-sig)	−5.42	−5.03	−4.80	−4.47	−3.33	−1.92	−1.33	−0.77	0.02
	k(t-sig)	−5.43	−5.05	−4.83	−4.50	−3.34	−2.02	−1.38	−0.84	−0.05
$T = \infty$		−5.15	−4.87	−4.64	−4.37	−3.39	−2.27	−1.85	−1.38	−0.70
(c) *Model* 1, $t_{\alpha,\lvert\theta\rvert}^{*}(1)$, *Choosing* T_b *maximizing* $\lvert t_{\hat{\theta}}\rvert$										
$T = 60$	k(F-sig)	−5.77	−5.42	−5.13	−4.80	−3.70	−1.87	−1.19	−0.39	0.24
	k(t-sig)	−5.85	−5.51	−5.18	−4.83	−3.70	−2.14	−1.34	−0.55	0.05
$T = 80$	k(F-sig)	−5.75	−5.26	−5.06	−4.77	−3.71	−2.14	−1.42	−0.79	0.11
	k(t-sig)	−5.66	−5.29	−5.04	−4.78	−3.72	−2.28	−1.67	−0.96	−0.06
$T = 100$	k(F-sig)	−5.69	−5.34	−5.03	−4.75	−3.74	−2.33	−1.80	−1.20	−0.18
	k(t-sig)	−5.68	−5.36	−5.05	−4.77	−3.71	−2.40	−1.88	−1.21	−0.34
$T = \infty$		−5.34	−5.08	−4.84	−4.59	−3.74	−2.71	−2.35	−2.01	−1.54
(d) *Model* 2, $t_{\alpha}^{*}(2)$, *Choosing* T_b *minimizing* $t_{\hat{\alpha}}$										
$T = 70$	k(F-sig)	−6.22	−5.81	−5.52	−5.22	−4.21	−3.28	−3.00	−2.76	−2.54
	k(t-sig)	−6.32	−5.90	−5.59	−5.29	−4.24	−3.32	−3.08	−2.85	−2.67
$T = 100$	k(F-sig)	−6.07	−5.72	−5.48	−5.17	−4.17	−3.29	−3.05	−2.83	−2.58
	k(t-sig)	−6.21	−5.86	−5.55	−5.25	−4.22	−3.35	−3.13	−2.85	−2.63
$T = \infty$		−5.57	−5.30	−5.08	−4.82	−3.98	−3.25	−3.06	−2.91	−2.72
(e) *Model* 2, $t_{\alpha,\gamma}^{*}(2)$, *Choosing* T_b *minimizing* $t_{\hat{\gamma}}$										
$T = 70$	k(F-sig)	−5.77	−5.32	−4.95	−4.51	−2.92	−1.37	−0.93	−0.54	−0.02
	k(t-sig)	−5.77	−5.38	−4.98	−4.55	−3.04	−1.53	−1.10	−0.71	−0.27
$T = 100$	k(F-sig)	−5.50	−5.16	−4.85	−4.47	−2.91	−1.50	−1.11	−0.73	−0.30
	k(t-sig)	−5.36	−5.23	−4.91	−4.47	−2.99	−1.55	−1.19	−0.78	−0.38
$T = \infty$		−5.28	−4.95	−4.62	−4.28	−2.94	−1.64	−1.33	−0.98	−0.59
(f) *Model* 2, $t_{\alpha,\lvert\gamma\rvert}^{*}(2)$, *Choosing* T_b *maximizing* $\lvert t_{\hat{\gamma}}\rvert$										
$T = 70$	k(F-sig)	−6.01	−5.56	−5.25	−4.88	−3.64	−2.17	−1.82	−1.37	−0.76
	k(t-sig)	−6.07	−5.61	−5.33	−4.94	−3.72	−2.28	−1.89	−1.50	−0.85

Table 1 (continued)

		1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
$T = 100$	k(F-sig)	−5.72	−5.37	−5.14	−4.84	−3.54	−2.11	−1.76	−1.42	−0.89
	k(t-sig)	−5.86	−5.49	−5.19	−4.88	−3.60	−2.23	−1.87	−1.49	−0.95
$T = \infty$		−5.57	−5.20	−4.91	−4.59	−3.47	−2.15	−1.86	−1.59	−1.30
(g) *Model* 3, $t_{\hat{\alpha}}^*(3)$, *Choosing* T_b *minimizing* $t_{\hat{\alpha}}$										
$T = 100$	k(F-sig)	−5.41	−4.99	−4.74	−4.44	−3.36	−2.53	−2.34	−2.21	−2.08
	k(t-sig)	−5.45	−5.11	−4.83	−4.48	−3.44	−2.60	−2.39	−2.22	−2.06
$T = 150$	k(F-sig)	−5.19	−4.85	−4.59	−4.31	−3.32	−2.47	−2.29	−2.11	−1.96
	k(t-sig)	−5.28	−4.96	−4.65	−4.38	−3.33	−2.50	−2.30	−2.13	−1.93
$T = 200$	k(F-sig)	−5.19	−4.84	−4.59	−4.30	−3.30	−2.46	−2.26	−2.09	−1.96
	k(t-sig)	−5.28	−4.96	−4.65	−4.38	−3.32	−2.48	−2.27	−2.10	−1.90
$T = \infty$		−4.91	−4.62	−4.36	−4.07	−3.13	−2.32	−2.12	−1.96	−1.78
(h) *Model* 3, $t_{\hat{\alpha},\gamma}^*(3)$, *Choosing* T_b *minimizing* $t_{\hat{\gamma}}$										
$T = 100$	k(F-sig)	−5.02	−4.69	−4.40	−3.99	−2.76	−1.76	−1.46	−1.12	−0.79
	k(t-sig)	−5.26	−4.82	−4.44	−4.07	−2.83	−1.76	−1.45	−1.12	−0.83
$T = 150$	k(F-sig)	−4.89	−4.54	−4.27	−3.93	−2.74	−1.70	−1.33	−1.01	−0.64
	k(t-sig)	−5.00	−4.63	−4.36	−3.99	−2.78	−1.72	−1.40	−1.07	−0.49
$T = 200$	k(F-sig)	−4.75	−4.43	−4.13	−3.79	−2.69	−1.53	−1.23	−0.90	−0.59
	k(t-sig)	−4.77	−4.50	−4.22	−3.83	−2.72	−1.57	−1.24	−0.96	−0.56
$T = \infty$		−4.67	−4.36	−4.08	−3.77	−2.65	−1.57	−1.22	−0.90	−0.49
(i) *Model* 3, $t_{\hat{\alpha},\lvert\gamma\rvert}^*(3)$, *Choosing* T_b *maximizing* $\lvert t_{\hat{\gamma}}\rvert$										
$T = 100$	k(F-sig)	−5.29	−4.87	−4.57	−4.27	−3.15	−2.19	−1.99	−1.69	−1.40
	k(t-sig)	−5.38	−5.02	−4.67	−4.36	−3.24	−2.28	−2.04	−1.75	−1.46
$T = 150$	k(F-sig)	−5.15	−4.77	−4.49	−4.21	−3.15	−2.16	−1.89	−1.59	−1.19
	k(t-sig)	−5.23	−4.91	−4.57	−4.28	−3.18	−2.19	−1.92	−1.63	−1.30
$T = 200$	k(F-sig)	−5.02	−4.75	−4.41	−4.10	−3.07	−2.11	−1.86	−1.63	−1.29
	k(t-sig)	−5.02	−4.75	−4.41	−4.17	−3.11	−2.15	−1.91	−1.68	−1.26
$T = \infty$		−4.87	−4.58	−4.34	−4.04	−3.08	−2.14	−1.87	−1.61	−1.30

4. Finite sample critical values

In this section we report simulation experiments to evaluate the finite sample distributions of the statistics under the null hypothesis of a unit root. We consider the leading case of a random walk where the data are generated by (4) with $y_0 = 0$ and $e_t \sim$ i.i.d. N(0, 1). This allows us to assess the effects of the data dependent methods to select the truncation lag. Given the nature of the data sets analyzed in later sections, we present critical values for the following sample sizes. For Model 1, $T = 60$, 80 and 100; for Model 2, $T = 70$ and 100; and for Model 3, $T = 100$, 150 and 200. The results were obtained using 2000 replications. The program was coded using the C language and N(0, 1) random deviates were obtained from the routine RAN1 of Press et al. (1986). For purely computational reasons, k max is set to 5. The results are presented in Table 1.

We first give some remarks on the finite sample distributions when k is fixed. The results are not reported but are available in the working paper version. In all cases, the critical values are fairly stable as k changes provided that k is held fixed when minimizing over T_b. In those cases where k is fixed, the asymptotic distribution is a good approximation to the finite sample distribution. Upon comparison with the results in Perron (1989), it is readily seen that the critical values are much lower when T_b is allowed to be data dependent than when it is considered fixed. For example, consider $t_\alpha^*(1)$ with $T = 100$ and $k = 0$, the 5% critical value is -4.93 when minimizing over T_b as opposed to -3.76 when the date of the break is considered fixed at mid-sample.

The critical values for the test constructed with k chosen according to recursive F-tests on the coefficients of the lagged first differences are presented in the rows labelled k (F-sig). For example, the 5% point with $T = 100$ is -5.09. The critical values for the test constructed with k chosen according to a t-test on the last included lag in the autoregression are presented in the rows labelled k (t-sig). The resulting values are close to the values obtained using F-sig.[1] In those cases where a data-dependent method is used to select k, the asymptotic approximation is not as good, indicating that the use of the asymptotic critical values would lead to tests that are liberal in finite samples. Comparing the distribution of the statistics $t_\alpha^*(i)$ ($i = 1, 2, 3$), it is interesting to note that the highest critical values (in the left tail of the distribution) occur for Model 3. This is contrary to the fixed T_b case where the highest critical values correspond to Model 1.

Consider now the critical values of the statistic $t_{\alpha,\theta}^*(1)$ and $t_{\alpha,\gamma}^*(i)$ ($i = 2, 3$) where T_b is chosen to minimize $t_{\hat\theta}$ or $t_{\hat\gamma}$, the t-statistic on the change in intercept or slope. The corresponding critical values are now smaller in absolute value. This is due to the a priori imposition of a one-sided change in the intercept of the trend function.

Consider now the statistics based upon choosing the break date maximizing the absolute value of $t_{\hat\theta}$ or $t_{\hat\gamma}$, $t_{\alpha,|\theta|}^*(1)$ and $t_{\alpha,|\gamma|}^*(i)$ ($i = 2, 3$). These statistics like $t_\alpha^*(i)$ do not impose any a priori condition on the sign of the change. We remark that for Models 1 and 3, the critical values in the left tail of the distribution are essentially the same between $t_\alpha^*(1)$ and $t_{\alpha,|\theta|}^*(1)$ and between $t_\alpha^*(3)$ and $t_{\alpha,|\gamma|}^*(3)$. Hence for Models 1 and 3, these two statistics are likely to have similar

[1] The simulated critical values involving a test of significance on the lagged first differences of the data are for tests of size 10%. We chose this value on the principle that it is safer to include extra lags to achieve the correct size in finite samples (at the expense of a loss in power). However, critical values with 5% tests were also computed and are not included since they are very similar to those with the 10% tests.

properties. Things are different for Model 2. The critical values in the left tail of the distribution are smaller (in absolute value) for $t^*_{\alpha,|\gamma|}(2)$ compared to $t^*_\alpha(2)$. Hence, one could expect the former to provide a more powerful test.

5. Finite sample size and power simulations

We now discuss finite sample size and power simulations. The aim is to determine the following: how size and power are affected by the choice of k in the presence of more general error processes, and by different values of the change in intercept and slope. Finally how power varies across procedures for choosing T_b. The focus of the simulations is placed on Models 1 and 3. The data generating process (DGP) used for Model 1 is of the form

$$y_t = \theta DU_t + \delta D(T_b)t + \alpha y_{t-1} + \sum_{i=1}^{4} \phi(i)\Delta y_{t-i} + (1+\psi L)e_t, \tag{5}$$

where $e_t \sim$ i.i.d. N(0, 1) and $y_0 = e_0 = 0$. For Model 3, the DGP is of the form

$$y_t = \gamma DT_t^* + \tilde{y}_t; \quad \tilde{y}_t = \alpha\tilde{y}_{t-1} + \sum_{i=1}^{4} \phi(i)\Delta\tilde{y}_{t-i} + (1+\psi L)e_t. \tag{6}$$

For the size simulations, $\alpha = 1$ and for power α is set to 0.8. The sample size for all simulations is $T = 100$ and 1000 replications are used. Regressions were run for fixed $k = 0, 1, \ldots, 5$ and for k(F-sig) and k(t-sig) with $k\max = 5$. For fixed k, the 5% asymptotic critical values were used, and for k(F-sig) and k(t-sig), the appropriate 5% finite sample critical values for $T = 100$ were used. When the change in intercept or slope is non zero, the break date is $T_b = 50$ (at mid-sample). For Model 1, we used values of δ (under the null) and θ (under the alternative) of 0, 2, 5 and 10. For Model 3, we used values of γ at 0, 0.1, 0.3, 0.5 and 1. Seven different error specifications were used: (1) $\phi(i) = 0$ ($i = 1, \ldots, 4$) and $\psi = 0$; (2) $\phi(1) = 0.6$, $\phi(i) = 0$ ($i = 2, 3, 4$) and $\psi = 0$; (3) $\phi(1) = -0.6$, $\phi(i) = 0$ ($i = 2, 3, 4$) and $\psi = 0$; (4) $\phi(1) = 0.4$, $\phi(2) = 0.2$ and $\phi(3) = \phi(4) = \psi = 0$; (5) $\phi(1) = 0.3$, $\phi(2) = 0.3$, $\phi(3) = 0.24$, $\phi(4) = 0.14$ and $\psi = 0$; (6) $\phi(i) = 0$ ($i = 1, 2, 3, 4$) and $\psi = 0.5$; (7) $\phi(i) = 0$ ($i = 1, 2, 3, 4$) and $\psi = -0.4$. Experiment (1) has i.i.d. errors. This specification is used to isolate the effects of choosing k too large. Experiment (2) has positive correlation in the errors and is quite common in empirical data. Experiment (3) has negative correlation in the errors. Experiments (4) and (5) have higher-order correlation and are useful in isolating the effects of picking k too small. Finally, experiments (6) and (7) have MA(1) errors.

Due to space constraints, we only include, in Table 2, the full set of results for $t^*_\alpha(3)$ and some for $t^*_{\alpha,\gamma}(3)$ with t-sig (the full set of results is available on request). We begin by summarizing results pertaining to the choice of k. When k is chosen less than the true order of the process, substantial size distortions often occur. In most cases the exact size is much greater than the nominal size. If k is chosen at least as big as the true order of the process, the exact size is rarely greater than the nominal size. However, power is lost if the lag structure is over parameterized. When the $k(t\text{-sig})$ or $k(\text{F-sig})$ procedure is used to pick k, the exact size is close to the nominal size in all cases except when there is a negative MA component as in experiment (7). In this case the exact size is substantially inflated above the nominal size. Power using $k(t\text{-sig})$ or $k(\text{F-sig})$ is generally quite good. It is greater than when k is larger than the true order of the process and is nearly as high as when k is set to the true order in the case of autoregressive errors. Overall, the $k(t\text{-sig})$ and $k(\text{F-sig})$ procedures have good size and power properties and clearly dominate using a fixed k. The results indicate that tests based on the $k(t\text{-sig})$ procedure are slightly more powerful than those based on $k(\text{F-sig})$.

Now consider how a change in intercept or slope affects the exact size. The tests $t_\alpha^*(1)$ and $t_{\alpha,\theta}^*(1)$ become oversized as δ increases. For example, consider experiment (1) for $t_\alpha^*(1)$ with $k(t\text{-sig})$; when $\delta = 0$ the size of the test is 0.047, when $\delta = 2$ it is 0.053, when $\delta = 5$ it is 0.096 and it rises to 0.486 when δ is as big as 10. The results in Tables 2 for $t_\alpha^*(3)$ concerning models with a change in slope γ show that changes in γ do not affect the size of the tests for the range of values considered. For $t_{\alpha,\gamma}^*(3)$, there are slight distortions in some cases as γ increases. Additional simulations revealed that larger values of γ induce substantial size distortions. The reader is referred to Vogelsang and Perron (1994) for a more detailed analysis on this issue. It is important to note here, though, that the magnitude of δ and γ where size distortions become a problem are of the order of 5 to 10 times the standard deviation of the errors for δ and at least 2 times the standard deviation of the errors for γ. For most macroeconomic time series (including those analyzed in later sections) intercept shifts are less than 5 standard deviations and slope changes are less than 0.7 standard deviations. Therefore, distortions caused by large changes are not a problem in practice but care should be used if a series is suspected to have a very large intercept or slope change.

We conclude by noting the effect on power of imposing the mild a priori condition on the sign of the change, i.e. comparing $t_\alpha^*(1)$ versus $t_{\alpha,\theta}^*(1)$ and $t_\alpha^*(3)$ versus $t_{\alpha,\gamma}^*(3)$. It is seen that power is higher when this condition is imposed and there is indeed a non-zero change in the trend function.

6. Empirical results for the Nelson–Plosser data set

Table 3 presents the empirical results for the Nelson–Plosser (1982) series for which, in Perron (1989), Model 1 was the specification of interest. A rejection of the unit root was claimed for all these series except the Consumer Price Index, Velocity and Interest Rate. Results are presented for both cases where the

Table 2
Finite sample size and power simulations; Model 3, $t_\alpha^*(3)$ (DGP: $y_t = \gamma DT_t^* + \tilde{y}_t$; $\tilde{y}_t = \alpha\tilde{y}_{t-1} + \sum_{i=1}^{4} \phi(i)\Delta\tilde{y}_{t-i} + (1 + \psi L)\, e_t$, $e_t \sim$ i.i.d. N(0, 1); $T = 100$, $T_b = 50$; 2000 replications; 5% nominal size; k max = 5)

k	Size ($\alpha = 1.0$)					Power ($\alpha = 0.8$)				
	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 1$
(1) $\phi(i) = 0.0$ ($i = 1, \ldots, 4$), $\psi = 0.0$										
0	0.049	0.053	0.055	0.047	0.036	0.358	0.365	0.344	0.331	0.325
1	0.044	0.049	0.048	0.042	0.037	0.287	0.299	0.283	0.277	0.277
2	0.045	0.046	0.048	0.041	0.040	0.203	0.215	0.207	0.199	0.205
3	0.038	0.039	0.042	0.040	0.047	0.160	0.177	0.169	0.165	0.167
4	0.035	0.037	0.039	0.036	0.041	0.122	0.129	0.134	0.130	0.134
5	0.035	0.035	0.039	0.038	0.039	0.110	0.123	0.125	0.116	0.117
F-sig	0.050	0.054	0.058	0.055	0.050	0.235	0.256	0.244	0.231	0.233
t-sig	0.049	0.051	0.058	0.050	0.045	0.257	0.278	0.270	0.259	0.258
$t_{\alpha,\gamma}^*(3)$	0.051	0.064	0.093	0.092	0.094	0.292	0.419	0.417	0.413	0.398
(2) $\phi(1) = 0.6$, $\psi = \phi(i) = 0.0$ ($i = 2, 3, 4$)										
0	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
1	0.058	0.060	0.056	0.054	0.062	0.908	0.903	0.904	0.902	0.901
2	0.046	0.048	0.049	0.049	0.055	0.753	0.758	0.761	0.753	0.756
3	0.045	0.046	0.040	0.041	0.045	0.586	0.592	0.600	0.594	0.593
4	0.037	0.040	0.034	0.038	0.044	0.405	0.426	0.424	0.417	0.417
5	0.033	0.033	0.034	0.037	0.045	0.289	0.302	0.306	0.305	0.305
F-sig	0.049	0.051	0.047	0.047	0.054	0.676	0.679	0.679	0.688	0.693
t-sig	0.049	0.047	0.038	0.046	0.049	0.760	0.773	0.774	0.778	0.785
$t_{\alpha,\gamma}^*(3)$	0.044	0.048	0.055	0.067	0.079	0.781	0.828	0.840	0.841	0.835
(3) $\phi(1) = -0.6$, $\psi = \phi(i) = 0.0$ ($i = 2, 3, 4$)										
0	0.858	0.874	0.873	0.858	0.848	0.997	0.997	0.998	0.998	0.998
1	0.051	0.048	0.043	0.038	0.034	0.131	0.132	0.117	0.114	0.113
2	0.046	0.040	0.040	0.037	0.034	0.090	0.100	0.096	0.094	0.098
3	0.044	0.045	0.041	0.037	0.039	0.084	0.098	0.094	0.099	0.099
4	0.034	0.030	0.033	0.032	0.033	0.063	0.074	0.082	0.083	0.077
5	0.033	0.035	0.038	0.037	0.039	0.056	0.073	0.073	0.075	0.070
F-sig	0.037	0.040	0.044	0.042	0.044	0.091	0.104	0.104	0.104	0.100
t-sig	0.039	0.039	0.042	0.034	0.037	0.090	0.105	0.105	0.104	0.097
$t_{\alpha,\gamma}^*(3)$	0.046	0.062	0.082	0.076	0.073	0.113	0.192	0.180	0.178	0.167
(4) $\phi(1) = 0.4$, $\phi(2) = 0.2$, $\psi = \phi(3) = \phi(4) = 0.0$										
0	0.004	0.004	0.005	0.004	0.004	0.001	0.000	0.000	0.000	0.000
1	0.009	0.008	0.008	0.007	0.006	0.432	0.439	0.428	0.421	0.424
2	0.048	0.051	0.050	0.049	0.048	0.756	0.764	0.765	0.763	0.760
3	0.040	0.042	0.047	0.044	0.051	0.598	0.611	0.611	0.607	0.602
4	0.038	0.039	0.043	0.044	0.046	0.413	0.432	0.438	0.436	0.421

Table 2 (Continued)
(DGP: $y_t = \gamma DT_t^* + \tilde{y}_t$; $\tilde{y}_t = \alpha\tilde{y}_{t-1} + \sum_{i=1}^{4}\phi(i)\Delta\tilde{y}_{t-i} + (1+\psi L)\ e_t$, $e_t \sim$ i.i.d. N(0, 1); $T = 100$, $T_b = 50$; 2000 replications; 5% nominal size; k max = 5)

k	Size ($\alpha = 1.0$)					Power ($\alpha = 0.8$)				
	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 1$	$\gamma = 0$	$\gamma = 0.1$	$\gamma = 0.3$	$\gamma = 0.5$	$\gamma = 1$
5	0.042	0.043	0.040	0.040	0.050	0.300	0.314	0.318	0.311	0.313
F-sig	0.040	0.048	0.049	0.047	0.050	0.582	0.593	0.600	0.591	0.593
t-sig	0.038	0.040	0.038	0.040	0.044	0.607	0.625	0.626	0.624	0.620
$t^*_{\hat{\alpha},\gamma}(3)$	0.037	0.043	0.049	0.056	0.069	0.659	0.733	0.755	0.737	0.737
(5) $\phi(1) = 0.3$, $\phi(2) = 0.3$, $\phi(3) = 0.25$, $\phi(4) = 0.14$, $\psi = 0.0$										
0	0.108	0.107	0.106	0.105	0.100	0.000	0.000	0.000	0.000	0.000
1	0.001	0.001	0.001	0.001	0.000	0.033	0.036	0.033	0.033	0.034
2	0.002	0.002	0.001	0.001	0.000	0.566	0.568	0.569	0.571	0.578
3	0.033	0.033	0.031	0.031	0.034	0.877	0.881	0.878	0.876	0.873
4	0.071	0.073	0.078	0.077	0.080	0.904	0.898	0.900	0.897	0.899
5	0.051	0.051	0.054	0.051	0.051	0.762	0.774	0.771	0.770	0.769
F-sig	0.048	0.046	0.050	0.048	0.051	0.857	0.863	0.858	0.858	0.856
t-sig	0.038	0.037	0.039	0.037	0.037	0.855	0.859	0.864	0.860	0.859
$t^*_{\hat{\alpha},\gamma}(3)$	0.022	0.020	0.025	0.025	0.025	0.864	0.893	0.882	0.893	0.894
(6) $\psi = 0.5$, $\phi(i) = 0.0$ ($i = 1, \ldots, 4$)										
0	0.002	0.003	0.003	0.003	0.001	0.008	0.010	0.007	0.007	0.005
1	0.150	0.142	0.160	0.160	0.156	0.548	0.552	0.566	0.545	0.542
2	0.021	0.021	0.028	0.028	0.020	0.096	0.100	0.110	0.103	0.103
3	0.052	0.050	0.057	0.061	0.059	0.190	0.202	0.215	0.203	0.202
4	0.034	0.035	0.033	0.033	0.034	0.101	0.105	0.117	0.111	0.113
5	0.038	0.037	0.039	0.035	0.041	0.102	0.111	0.128	0.119	0.117
F-sig	0.053	0.055	0.062	0.065	0.061	0.206	0.213	0.222	0.210	0.209
t-sig	0.067	0.066	0.070	0.069	0.064	0.244	0.258	0.271	0.257	0.262
$t^*_{\hat{\alpha},\gamma}(3)$	0.056	0.067	0.092	0.104	0.111	0.254	0.339	0.373	0.365	0.364
(7) $\psi = -0.4$, $\phi(i) = 0.0$ ($i = 1, \ldots, 4$)										
0	0.734	0.738	0.739	0.717	0.691	0.996	0.996	0.996	0.995	0.995
1	0.202	0.210	0.206	0.189	0.174	0.751	0.763	0.727	0.720	0.715
2	0.079	0.087	0.075	0.074	0.074	0.395	0.406	0.383	0.385	0.376
3	0.044	0.048	0.050	0.046	0.046	0.252	0.263	0.254	0.251	0.245
4	0.036	0.037	0.034	0.033	0.036	0.159	0.173	0.163	0.164	0.156
5	0.031	0.031	0.033	0.032	0.035	0.131	0.144	0.139	0.138	0.135
F-sig	0.145	0.148	0.156	0.148	0.149	0.456	0.476	0.452	0.451	0.442
t-sig	0.235	0.247	0.257	0.251	0.248	0.631	0.658	0.646	0.643	0.647
$t^*_{\hat{\alpha},\gamma}(3)$	0.174	0.236	0.287	0.272	0.273	0.587	0.690	0.673	0.665	0.659

Note: All entries refer to the statistic $t^*_{\hat{\alpha}}(3)$ except those in the rows labelled $t^*_{\hat{\alpha},\gamma}(3)$ which correspond to this statistic constructed using the t-sig procedure.

Table 3
Empirical results, Nelson–Plosser data; Model 1, k max = 10

Series	Sample	T	T_b	k	$t_{\hat{\theta}}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	$t_{\hat{\alpha}}^*(1)$, p-values			$t_{\hat{\alpha},\theta}^*(1)$, p-values		
								(asy)	(F-sig)	(t-sig)	(asy)	(F-sig)	(t-sig)
Real GNP	1909–1970	62	1928	9	−5.13	0.190	−5.93	<0.01	<0.01		<0.01	<0.01	
			1928	8	−4.79	0.267	−5.50	<0.01		0.03	<0.01		0.02
Nominal GNP[a]	1909–1970	62	1928	11	−6.34	0.404	−8.16	<0.01		<0.01	<0.01		<0.01
			1928	15	−5.94	0.497	−6.21	<0.01	<0.01		<0.01	<0.01	
Real per capita GNP	1909–1970	62	1928	9	−3.73	0.313	−4.81	0.06	0.12		0.03	0.06	
			1928	7	−3.31	0.484	−4.51	0.13		0.21	0.07		0.10
Industrial production	1860–1970	111	1928	8	−5.18	0.272	−6.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Employment	1890–1970	81	1928	8	−3.42	0.586	−5.14	0.02	0.05		0.01	0.02	
			1928	7	−3.11	0.650	−4.91	0.04		0.09	0.02		0.04
GNP deflator	1889–1970	82	1928	5	−3.28	0.783	−4.14	0.29	0.35	0.35			
			1919	5	−3.51	0.886	−3.24				0.58		0.54
			1919	9	−3.61	0.829	−3.87				0.27	0.28	
CPI	1860–1970	111	1939	5	2.00	0.948	−3.09	0.88	0.88	0.88			
			1919	5	−3.12	0.982	−1.16				0.98	0.96	0.96
Wages	1900–1970	71	1929	7	−4.32	0.619	−5.41	<0.01		0.02	<0.01		0.01
			1929	9	−4.10	0.635	−4.62	0.10	0.16		0.05	0.08	
Money stock	1889–1970	82	1929	7	−2.80	0.783	−4.69	0.08	0.14		0.04	0.07	
			1927	6	−2.50	0.831	−4.30	0.21		0.28			
			1928	6	−2.63	0.824	−4.28				0.12		0.15
Velocity	1869–1970	102	1949	8	2.95	0.830	−2.81	0.95	0.94				
			1946	0	3.24	0.858	−3.29	0.81		0.81			
			1880	5	−2.74	0.928	−1.62				0.96	0.93	
			1880	0	−2.46	0.897	−2.43				0.87		0.83
Interest rate	1900–1970	71	1965	3	3.86	0.934	−1.35	>0.99	>0.99				
			1963	3	3.44	0.928	−1.35	>0.99		>0.99			
			1920	0	−4.16	1.058	1.16				>0.99		>0.99
			1918	0	−3.59	1.079	2.08				>0.99	>0.99	

[a] For Nominal GNP, k max = 15 (See footnote 3).

truncation lag is selected using the F-sig or t-sig methods and for both ways of selecting the break point. Here, k max is specified to be 10.[2] The statistics of most interest are the estimates of α and their t-statistic as well as six p-values in the last columns (reported to the nearest 1%). The first set corresponds to $t_{\alpha}^{*}(1)$ and the second to $t_{\alpha,\theta}^{*}(1)$. The first p-value in each set is based on the asymptotic distribution. It is included because it may be more robust, for example, to the presence of additional correlation than those based on finite sample distributions. The second and third p-values correspond to the F-sig and t-sig methods, respectively. The critical values used correspond to samples of size 60, 80 or 100 whichever is closest to the actual sample size.

The empirical results show that the unit root hypothesis can be rejected at the 5% significance level or better, under either scenario about the choice of k, for Real GNP, Nominal GNP,[3] Industrial Production and Nominal Wages. For the Employment series, the finite sample p-value is 0.05 with F-sig and 0.09 with t-sig (the corresponding asymptotic p-values are 0.02 and 0.04, respectively). Hence, the unit root is also rejected for the Employment series. The Real per capita GNP and Money Stock series present a more ambiguous case. When k is chosen with the F-sig procedure, the p-value for the Real per capita GNP series is 0.12 using the finite sample distribution and 0.06 using the asymptotic distribution. The corresponding figures are 0.14 and 0.08 for the Money Stock series. These values are marginal for a rejection at the 10% level.[4]

The unit-root hypothesis cannot be rejected for the Consumer Price Index (CPI), Velocity and Interest Rate series under any procedure. The choices of T_b and k obtained using the data-dependent methods for choosing k are different but yield the same qualitative results. The only series offering a markedly different picture from the fixed T_b case is the GNP Deflator. With k chosen according to either method the p-value is 0.35 (0.29 using the asymptotic distribution). Hence, for this series, the rejection of the unit root reported in Perron (1989) is not robust to correlation between the choice of T_b and the data.

---

[2] The choice of k max is somewhat arbitrary. On the one hand, one would like a large value to have as unrestricted a procedure as possible. On the other hand, a large value of k max yields problems of multicollinearity in the data and also a substantial loss of power. The choice of k max was also set such that the estimated autoregressions did not show any sign of remaining correlation in the residuals as indicated by the Box–Pierce statistic. Most of the results are robust to alternative choices for k max.

[3] For the Nominal GNP series, k max was found binding in the sense that $k = 10$ was selected. Hence k max was increased to 15 which again was found to be binding. We did not increase k max further given the relatively few number of observations. Nevertheless, the conclusion is robust to basically any value of the truncation lag parameter k chosen.

[4] It is of interest to note, that applying the tests to the Friedman and Schwartz (1982) Real per Capita GNP (using the sample 1909–1970) allows an overwhelming rejection of the unit root under either method. See the working paper version for more details.

A comment is warranted about the choice of T_b selected according to these procedures. Except for the CPI, Velocity and Interest Rate series (for which the unit root is not rejected), the value T_b is either 1929 (for Nominal Wage and Money Stock) or 1928 (for the other series). While 1928 does not exactly correspond to the date specified in Perron (1989), the economic interpretation remains the same. The selection of 1928 is due to the presence of the dummy variable $D(T_b)_t$ in regression (1). Hence, 1928 is often chosen because the dummy variable takes value 1 in 1929 and offers some additional fit to the 1929 crash over what the change in the intercept can do alone.

Now consider the results when T_b is chosen to minimize t_{θ}, the t-statistic on the change in intercept, i.e. imposing the one-sided restriction of a crash. When a rejection of the unit-root hypothesis occurred using $t_{\alpha}^{*}(1)$, it does so again here and more strongly, given that the tests have higher power. As was the case earlier, the unit root cannot be rejected for the GNP Deflator series. The results offer, however, a different picture for three series. First, for the Employment series, the unit root can be rejected at the 5% level (using any procedure) instead of 10% with the statistic $t_{\alpha}^{*}(1)$. More interestingly, the unit root can now be rejected at the 10% level for the Real per capita GNP and Money Stock series. For example, the p-values under the F-sig procedure are 0.06 and 0.07, respectively.

We now turn to the analysis of the Common Stock Price and Real Wage series where Model 2 is specified, i.e. allowing both a change in the intercept and the slope of the trend function. The procedures used and the presentation of the estimation results in Table 4 follow our previous analysis except that k max is now 5. Consider first the case where T_b is chosen to minimize the t-statistic on α. The date of break selected for the Common Stock Price series is 1928 (consistent with the imposition of 1929 as the break date in Perron, 1989). Both methods to choose the truncation lag yield the same model and test statistic with an asymptotic p-value of 0.02 and finite sample p-value of 0.04 for F-sig and 0.06 for t-sig. Similar results hold for the Real Wage series. The break date is 1939; the asymptotic p-value of the test is 0.03 and the finite sample ones are 0.07 with F-sig and 0.08 with t-sig.

Consider now results obtained when T_b is chosen maximizing $t_{\hat{\gamma}}$ or $|t_{\hat{\gamma}}|$, the t-statistic on the coefficient of the slope change. The results are quite interesting in that the unit root is strongly rejected using either method to select the truncation lag even without the a priori imposition on the sign of the change in slope. The selected break date is still 1939 for the Real Wage series but now 1936 for Common Stock Price.

To compare our results with those of Zivot and Andrews (1992), note first the methodological differences involved. First, we retained the one time dummy $D(T_b)_t$ in regressions (1) and (2); we consider the F-sig procedure to select the truncation lag as well as the t-sig procedure; we consider k max $= 5$ instead of 10 for the Real Wages and Common Stock Price series; and we also consider the

Table 4
Empirical results, Nelson–Plosser data; Model 2, k max = 5

Series	Sample	T	T_b	k	$\hat{\theta}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	$t^*_{\alpha}(2)$, p-values			$t^*_{\alpha,\lvert\gamma\rvert}(2)$, p-values			$t^*_{\alpha,\gamma}(2)$, p-values		
										(asy)	(F-sig)	(t-sig)	(asy)	(F-sig)	(t-sig)	(asy)	(F-sig)	(t-sig)
Common stock prices	1871–1970	100	1928	1	−1.36	0.0065	0.0141	0.716	−5.50	0.02	0.04	0.06						
					(−4.92)	(4.43)	(4.61)											
			1936	3	−2.10	0.0094	0.0268	0.553	−5.49				0.01	0.02	0.02	<0.01	0.01	0.01
					(−4.86)	(4.93)	(4.91)											
Real wages	1900–1970	71	1939	3	−0.25	0.0086	0.0047	0.390	−5.41	0.03	0.07	0.08	0.02	0.04	0.04	<0.01	0.02	0.02
					(−2.65)	(5.26)	(3.38)											

case where the break date is selected using a test of significance on the coefficient of the change in slope. For the series Real GNP, Nominal GNP, Industrial Production, Nominal Wages and Common Stock Prices our results agree with those of Zivot and Andrews (1982), namely a rejection of the unit root. Our results also show these rejections to be robust to alternative specifications for choosing the break date and the truncation lag (except for Nominal Wage using $t_{\alpha}^{*}(1)$ and F-sig). For the Employment series, our results allow a rejection at the 10% level using the finite sample critical values for the t-sig method when the break is selected minimizing the unit-root statistic (basically due to the inclusion of the one-time dummy $D(T_b)_t$ in (1)). However, our results show that a stronger rejection, at the 5% level, is possible using the F-sig method and that this rejection becomes even stronger if the mild a priori restriction of a one-sided change is imposed. For the Real per capita GNP and Money Stock series, the results with $t_{\alpha}^{*}(1)$ and the t-sig method are similar to those in Zivot and Andrews (1992), namely p-values of 0.21 and 0.28. Using the F-sig procedure, the p-values are substantially reduced to 0.12 and 0.14, respectively. Imposing the sign of the change a priori allows a rejection at the 10% level for both series using F-sig and for Real per capita GNP using t-sig. The difference for the Real Wage series is due to the different choice of k max. Our results agree for non-rejections for GNP Deflator, CPI, Velocity and Interest Rate.

7. Results with an international data set for postwar real GNP

This Section analyzes an international data set of post-war quarterly real GNP or GDP series. The countries analyzed and the type and sampling period of the series are the following: USA (GNP; 1947:1–1991:3); Canada (GDP; 1947:1–1989:1); Japan (GNP; 1957:1–1988:4); France (GDP; 1965:1–1988:3); Germany (GNP; 1960:1–1986:2); Italy (GDP; 1960:1–1985:1); and the United Kingdom (GDP; 1957:1–1986:3). The data for USA are from Citibase and for Canada from the Cansim data bank. For Japan and France they are from the IFS data tape. The remaining series (UK, Germany and Italy) are from Data Resources Inc. and are those used in Campbell and Mankiw (1989). All series are seasonally adjusted and at annual rates, except for the USA and the United Kingdom which are at quarterly rates. The plots of the logarithm of most series are presented in Fig. 1. In these graphs the dashed line is the estimated trend function allowing a one-time change in slope. The date of the change varies for each series and was selected using the test $t_{\alpha}^{*}(3)$.

The results pertaining to the statistic $t_{\alpha}^{*}(3)$ are presented in Table 5. Using the asymptotic critical values, the unit root is rejected, at close to the 5% level, for all series except Italy (for Canada this rejection is not robust when using F-sig). Using the finite sample critical values, the results are not, in general, as sharp. For Japan, the unit root is strongly rejected (p-values of 0.02 and 0.03). The

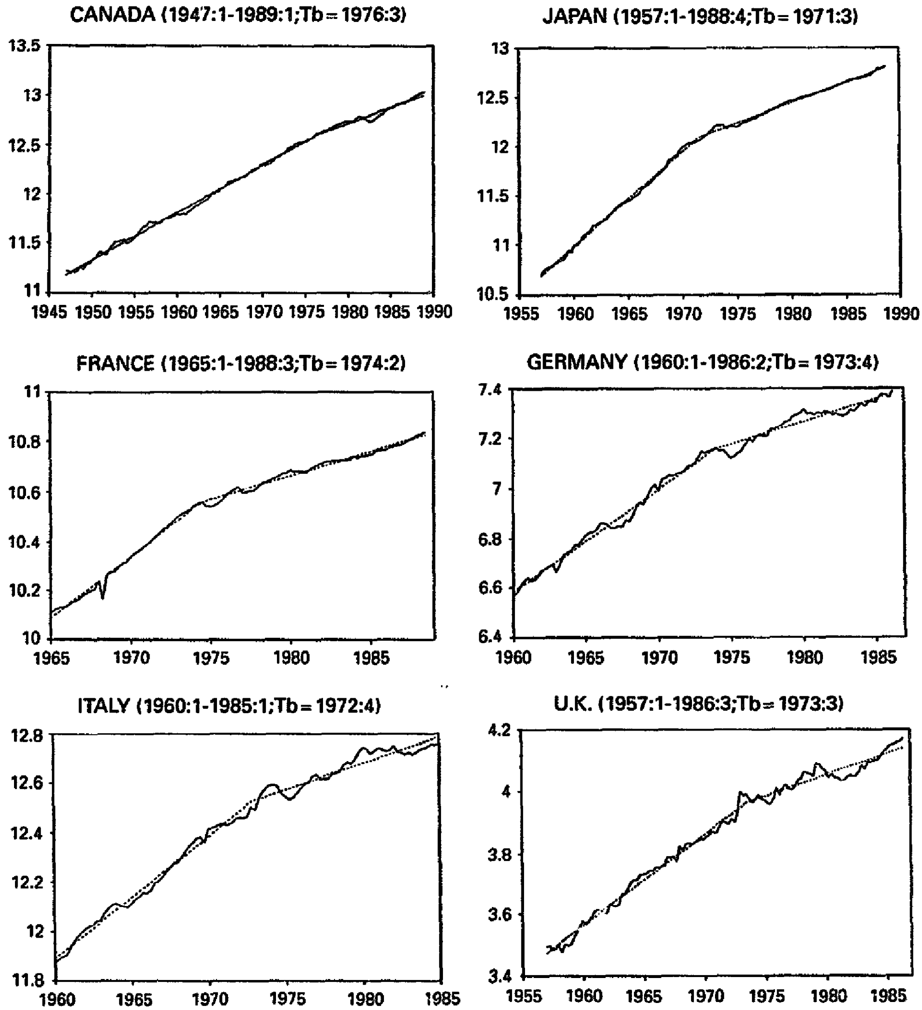

Fig. 1. Log real GNP for selected countries.

results are not as clear for most of the other series but some interesting cases still emerge. For the United Kingdom, using the F-sig and t-sig procedures, the p-values are 0.07 and 0.08 respectively, allowing a rejection of the unit root at the 10% level. The results for Canada, France, Germany and the United States are similar in terms of the t-statistics obtained. They range from -4.22 to -4.33 with finite sample p-values between 0.12 and 0.14 (this excludes the case of Canada with F-sig). While the unit root cannot be rejected at the 10% level, the results are not very much at odds with the hypothesis that the series can be construed as stationary fluctuations around a breaking trend function. Such is

Table 5
Empirical results, Model 3, international data set

Series	Sample	T	T_b	k	$\hat{\beta}$	$\hat{\gamma}$	$t_{\hat{\gamma}}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	$t_{\hat{\alpha}}^*(3)$, p-values			$t_{\hat{\alpha},\gamma}^*(3)$, p-values		
										(asy)	(F-sig)	(t-sig)	(asy)	(F-sig)	(t-sig)
USA; GNP	47:1–91:3	179	71:2	2	0.0088	−0.0024	−13.03	0.900	−4.22	0.07	0.12	0.14			
(k max = 5)			69:2	2	0.0089	−0.0024	−13.34	0.900	−4.16				0.04	0.05	0.06
Canada; GDP	47:1–89:1	169	76:3	3	0.0122	−0.0044	−21.18	0.818	−4.26	0.06		0.13			
(k max = 5)			76:3	4	0.0122	−0.0044	−21.18	0.830	−3.78	0.19	0.26				
			77:1	3	0.0121	−0.0045	−21.24	0.817	−4.25				0.03		0.06
			77:1	4	0.0121	−0.0045	−21.24	0.829	−3.77				0.10	0.13	
Japan; GNP	57:1–88:4	128	71:3	4	0.0243	−0.0142	−76.11	0.650	−5.14	<0.01	0.02		<0.01	<0.01	
(k max = 5)			71:4	4	0.0241	−0.0141	−75.19	0.659	−5.11	<0.01		0.03			
			71:3	3	0.0243	−0.0142	−76.11	0.685	−4.99				<0.01		0.02
France[a]; GDP	65:1–88:3	95	74:2	3	0.0127	−0.0080	−46.93	0.679	−4.33	0.05	0.12	0.13			
(k max = 10)			74:1	2	0.0129	−0.0081	−47.76	0.705	−4.26				0.03		0.07
			74:1	3	0.0129	−0.0081	−47.76	0.685	−4.16				0.04	0.07	
Germany; GNP	60:1–86:2	106	73:4	7	0.0104	−0.0058	−21.92	0.624	−4.33	0.05	0.12	0.13			
(k max = 10)			73:1	4	0.0106	−0.0058	−22.55	0.705	−4.00				0.06		0.11
			73:1	6	0.0106	−0.0058	−22.55	0.681	−3.82				0.09	0.14	
Italy; GDP	60:1–85:1	101	72:4	3	0.0125	−0.0071	−20.91	0.755	−3.90	0.15	0.25				
(k max = 10)			73:2	1	0.0123	−0.0071	−21.37	0.814	−3.31	0.40		0.56			
			73:3	1	0.0122	−0.0072	−21.51	0.817	−3.26				0.25	0.29	0.33
UK; GDP	57:1–86:3	118	73:3	7	0.0074	−0.0040	−18.96	0.552	−4.62	0.02	0.07	0.08			
(k max = 10)			73:2	7	0.0074	−0.0040	−18.98	0.554	−4.61				0.01	0.03	0.04

[a]A dummy taking value 1 in 1968:2 was included for France to take account of the general strike in May 1968.

not the case, however, with the GDP series from Italy. Here the p-values are large enough to cast little doubt on the unit root.

It is interesting to look at the estimated change in the slope of the trend function and the dating of the break implied by the estimation procedure. The estimated percentage decrease in the rates of growth are: USA, 37%; Canada, 36%; Japan, 58%; France, 66%; Germany, 56%; Italy, 57%; UK, 54%. These figures are indeed quite large and suggest, besides the unit-root issue, that an important structural change has occurred. The break dates are different for each country but are all close to the year 1973, associated with the first oil price shock. They vary between 1971:2 (USA) and 1976:3 (Canada). It is to be noted, however, that the method used here is not directly geared at providing a consistent estimate of the date of change. Hence, the break dates should be viewed as approximate.

As discussed, using $t^*_{\alpha,\gamma}(3)$, which select T_b based on the parameter of the change in slope, is likely to allow tests with greater power. p-values pertaining to this test are presented in the last columns of the table. Indeed, it appears more powerful. Using the t-sig procedure, the p-values for the null hypothesis of a unit root are at most 0.11 for all countries except Italy. Using the F-sig procedure, the p-values are smaller than 0.10 for USA, Japan, France and the United Kingdom; they are 0.13 and 0.14 for Canada and Germany, respectively. These results show that a simple imposition of a one-sided downward change in slope (still with an unknown break point) is enough to warrant rejection of the unit root hypothesis at close to the 10% level for all countries except Italy.

We view these results, especially given the small span of the data, as substantial evidence against the unit root. It is indeed somewhat revealing to consistently obtain such p-values given the relatively low power of unit-root tests using data over a short span (see Perron, 1991). Given that the statistical procedure used is one where an extreme assumption is made about the correlation of the choice of the break point and the data (yielding a procedure with low power compared to the case where T_b is assumed fixed), we view these results as consistent with the hypothesis that the series are best characterized as stationary fluctuations around a breaking trend function with a change in slope near 1973.

To compare our results with those of Banerjee et al. (1992) (BLS), we first note the main differences in the studies. First, the data used are slightly different in terms of both the sources and the horizon. Second, they use the one-step innovational outlier method which does not allow for a change in slope under the null hypothesis. Third, they use a fixed value of the truncation lag set at 4 for all countries and they note that the results are robust when setting this fixed value to 8 or when using an information criterion (AIC or Schwartz) to select the order. Using these different specifications they found little evidence against the unit root for all countries except Japan.

After several investigations using both types of methods applied to both data set,[5] it turns out that the major factor responsible for the conflicting results is the method to choose the truncation lag. For example, our data-dependent methods select $k = 4$ only for Japan for which we both reject the unit root. For the other countries the implied value of the selected truncation lag is different (in no case do our methods select $k = 8$ either). We believe our methods to select the truncation lag to be better for the purpose of the unit-root tests for the following reasons documented in Ng and Perron (1994). First, fixing k to some arbitrary value can involve serious size distortions and/or power losses because the actual correlation structure of the data is not only unknown but is likely to be different across countries. However, even for data-dependent methods that implies asymptotically valid unit-root tests, there are important differences between methods based on a general to specific approach and methods based on information criteria. In the context of a model where the noise component is an ARMA process, Ng and Perron (1995) show that the latter implies a sequence of selected values for k that increases with the sample size at a logarithmic rate, a very slow rate. The finite sample implication of this result is that methods based on information criteria will tend to select very low autoregressive orders. These implied parsimonious autoregressions will often not be enough to capture important serial correlation in the data and can lead to tests with size distortions and/or power losses. These theoretical issues are consistent with the empirical results of BLS, who report values of k at 0 or 1 for all countries when using an information criterion. In no cases do our methods select such low values (except for Italy where we both agree for a non-rejection).

8. Concluding comments

This paper documents the robustness of the results presented in Perron (1989). Unlike this previous study, we analyzed the case where the break date is explicitly correlated with the data and provided critical values to carry inference under a variety of procedures. This work is not intended as a substitute for the statistical procedures presented in that earlier paper but rather as a complement. Indeed, a case can often be made for using critical values that are based on the assumption of no correlation between the choice of the break point and the data. On the one hand, it may represent a close approximation to the actual extent of the correlation. On the other hand, each investigator may differ as to the amount of a priori information he or she is willing to incorporate into the analysis.

---

[5] My thanks to Robin Lumsdaine for correspondance on this issue.

Another issue concerns the power of the tests. There appears to be a clear trade-off between power and the amount of a priori information one is willing to incorporate with respect to the choice of the break point. The presumption is clearly that a procedure imposing no such a priori information, as the ones presented in this paper, has relatively low power. In this respect, the rejection of the unit-root hypothesis, even when assuming a perfect correlation between the choice of the break point and the data, is quite strong.

Appendix. Proof of Theorem 1

To simplify cross-references, we adopt the notation of Zivot and Andrews (1992), henceforth referred to as $Z-A$. Let $S_t = \sum_{j=1}^{t} e_j$ $(S_0 = 0)$ and $X_T(r) = \sigma^{-1}T^{-1/2}S_{[Tr]}$, $(j-1)/T \leqslant r < j/T$ (for $j = 1, \ldots, T$), where $\sigma^2 = \lim_{T\to\infty} T^{-1}E(S_T^2)$ and $[\cdot]$ denotes the integer part of the argument. Since $\{e_t\}$ is i.i.d. with finite variance, we have $X_T(r) \Rightarrow W(r)$, where $\Rightarrow$ denotes weak convergence in distribution (from the space D[0, 1] to the space C[0, 1] using the uniform metric on the space of functions on [0, 1]) with W(r) a standard Wiener process on [0, 1]. Also, $\sigma_T^2 \equiv T^{-1}\sum_1^T e_t^2 \to_p \sigma^2$ where $\to_p$ denotes convergence in probability. Omitting the one-time dummy variable $\mathrm{D}(T_\mathrm{b})_t$ (since it is asymptotically negligible), we consider the following regressions:

$$y_t = \beta^i(\lambda)z_{tT}^i(\lambda) + \alpha^i(\lambda)y_{t-1} + e_t, \quad (t = 1, \ldots, T), \tag{A.1}$$

for models $i = 1, 2$. The vector $z_{tT}^i(\lambda)$ encompasses the deterministic components of the model and depends explicitly on λ, the break fraction, and T, the sample size. For example, $z_{t,T}^1(\lambda)' = (1, t, DU_t(\lambda))$. Let $Z_T^i(\lambda, r) = \delta_T^i z_{[Tr],T}^i(\lambda)$ be a rescaled version with δ_T^i a diagonal matrix of weights. For example, $\delta_T^1 = \mathrm{diag}(1, T^{-1}, 1)$. We also define the limiting functions $Z^1(\lambda, r) = (1, r, du(\lambda, r))'$ where $du(\lambda, r) = 1(r > \lambda)$, and $Z^2(\lambda, r) = (1, r, du(\lambda, r), dt^*(\lambda, r))'$ where $dt^*(\lambda, r) = 1(r > \lambda)(r - \lambda)$. Note that, as argued in $Z - A$, we do not have $Z_T^i(\lambda, r) \Rightarrow Z^i(\lambda, r)$ $(i = 1, 2)$ as $T \to \infty$, using the uniform metric on the space of functions on D[0, 1]. The proof nevertheless remains valid without the need to introduce another metric to guarantee such convergence results. For simplicity, we henceforth drop the superscript denoting the model. Note that the following proof is valid, with trivial modifications, for a large class of deterministic components including higher-order polynomials, multiple structural changes and other types of discrete shifts.

It is convenient to first transform (A.1) as follows. Let $Pz_T(\lambda) = [Pz_{1,T}(\lambda), \ldots, Pz_{T,T}(\lambda)]$ be the linear map projecting onto the space spanned by the columns of $z_T(\lambda)' = (z_{1,T}(\lambda), \ldots, z_{T,T}(\lambda))$. By definition $Pz_T(\lambda) = z_T(\lambda)(z_T(\lambda)'z_T(\lambda))^- z_T(\lambda)'$ where $(\cdot)^-$ denotes a g-inverse. Premultiplying by $Mz_T(\lambda) \equiv (I - Pz_T(\lambda))$, (A.1) can be written, in matrix notation, as

$$Mz_T(\lambda)Y = \alpha(\lambda)Mz_T(\lambda)Y_{-1} + Mz_T(\lambda)e, \tag{A.2}$$

where $Y' = (y_1, \dots, y_T)$, $Y'_{-1} = (y_0, \dots, y_{T-1})$ and $e' = (e_1, \dots, e_T)$. The t-statistic of interest can be written as

$$\inf_{\lambda\in[0,1]} t_{\hat\alpha}(\lambda) = \inf_{\lambda\in[0,1]}[T^{-2}Y'_{-1}Mz_T(\lambda)Y_{-1}]^{-1/2}[T^{-1}Y'_{-1}Mz_T(\lambda)e]/s_T(\lambda),$$

where $s_T^2(\lambda) = T^{-1}(Y - \hat\alpha(\lambda)Y_{-1})'Mz_T(\lambda)(Y - \hat\alpha(\lambda)Y_{-1})$ with $\hat\alpha(\lambda)$ the OLS estimate of α in (A.2). We have

$$\begin{aligned}
T^{-2}Y'_{-1}Mz_T(\lambda)Y_{-1} &= T^{-2}\sum_{t=1}^{T} \times\left\{y_{t-1} - z_{t,T}(\lambda)'\left[\sum_{s=1}^{T} z_{s,T}(\lambda)z_{s,T}(\lambda)'\right]^{-} \times \sum_{s=1}^{T} z_{s,T}(\lambda)y_{s-1}\right\}^2 \\
&= T^{-1}\sum_{t=1}^{T}\left\{T^{-1/2}S_{t-1} - z_{t,T}(\lambda)'\delta_T \times\left[T^{-1}\sum_{s=1}^{T}\delta_T z_{s,T}(\lambda)z_{s,T}(\lambda)'\delta_T\right]^{-} T^{-1}\sum_{s=1}^{T}\delta_T z_{s,T}(\lambda)S_{s-1}\right\}^2 + o_{p\lambda}(1) \\
&= \int_0^1\left\{\sigma X_T(r) - Z_T(\lambda,r)'\left[\int_0^1 Z_T(\lambda,s)Z_T(\lambda,s)'\,ds\right]^{-1}\int_0^1 Z_T(\lambda,s)\sigma X_T(s)ds\right\}^2 dr \\
&\quad + o_{p\lambda}(1) \\
&= \sigma^2\int_0^1\{X_T(r) - Pz_T(\lambda)X_T(r)\}^2\,dr + o_{p\lambda}(1),
\end{aligned} \tag{A.3}$$

($o_{p\lambda}(1)$ denotes a random variable that converges in probability to 0 uniformly in λ) and:

$$Pz_T(\lambda)X_T(r) = Z_T(\lambda,r)'\left[\int_0^1 Z_T(\lambda,s)Z_T(\lambda,s)'\,ds\right]^{-1}\int_0^1 Z_T(\lambda,s)X_T(s)ds.$$

Also, using developments as in $Z - A$,

$$\begin{aligned}
T^{-1}Y'_{-1}Mz_T(\lambda)e &= T^{-1}\sum_{t=1}^{T}\left\{y_{t-1} - z_{t,T}(\lambda)'\left[\sum_{s=1}^{T} z_{s,T}(\lambda)z_{s,T}(\lambda)'\right]^{-} \times \sum_{s=1}^{T} z_{s,T}(\lambda)y_{s-1}\right\}e_t + o_{p\lambda}(1) \\
&= \sigma^2\int_0^1 X_T(r)\,dX_T(r) - \sigma^2\int_0^1 P_T z(\lambda)X_T(r)\,dX_T(r) + o_{p\lambda}(1).
\end{aligned} \tag{A.4}$$

We can, therefore, express the t-statistic as a composite functional:

$$\inf_{\lambda\in[0,1]} t_{\hat{\alpha}}(\lambda) = g\left(X_T(r), \int_0^1 X_T(r)\,\mathrm{d}X_T(r), Pz_T(\lambda)X_T(r), \int_0^1 Pz_T(\lambda)X_T(r)\,\mathrm{d}X_T(r), s_T(\lambda) \right) + \mathrm{o}_{\mathrm{p}\lambda}(1),$$

where

$$g = h^*\left[h\left[H_1[X_T(r), Pz_T(\lambda)X_T(r)], H_2\left[\int_0^1 X_T(r)\,\mathrm{d}X_T(r), \int_0^1 Pz_T(\lambda)X_T(r)\,\mathrm{d}X_T(r) \right], s_T(\lambda) \right]\right],$$

with $h^*(m) = \inf_{\lambda\in[0,1]} m(\lambda)$ for any real function $m = m(\cdot)$ on [0, 1]; and for any real functions $m_1(\cdot)$, $m_2(\cdot)$, $m_3(\cdot)$ on $[0,1]$, $h[m_1(\lambda), m_2(\lambda), m_3(\lambda)] = m_1(\lambda)^{-1/2}m_2(\lambda)/m_3(\lambda)$. The functionals H_1 and H_2 are defined by (A.3) and (A.4). The weak convergence results for each of the elements are contained in the following lemma.

Lemma A.1. The following convergence results hold jointly:

(a) $X_T(r) \Rightarrow W(r)$;

(b) $\int_0^1 X_T(r)\,\mathrm{d}X_T(r) \Rightarrow \int_0^1 W(r)\,\mathrm{d}W(r)$;

(c) $Pz_T(\lambda)X_T(r) \Rightarrow Pz(\lambda)W(r) \equiv Z(\lambda,r)'\left[\int_0^1 Z(\lambda,s)Z(\lambda,s)'\,\mathrm{d}s\right]^{-1}\int_0^1 Z(\lambda,s)W(s)\,\mathrm{d}s$;

(d) $\int_0^1 Pz_T(\lambda)X_T(r)\,\mathrm{d}X_T(r) \Rightarrow \int_0^1 Pz(\lambda)W(r)\,\mathrm{d}W(r)$;

(e) $s_T^2(\lambda) = \sigma^2 + \mathrm{o}_{\mathrm{p}\lambda}(1)$.

Parts (a) and (b) are standard results, and part (e) follows using (c) and (d) and the fact that $T^{-1}\sum_1^T e_T^2 \to_p \sigma^2$. To prove part (c), we start with the following Lemma which follows from Theorem 5.5 of Billingsley (1968).

Lemma A.2. $Pz_T(\lambda)X_T(r) \Rightarrow Pz(\lambda)W(r)$ if $X_T(r) \Rightarrow W(r)$ and for any sequence of functions $\{v_T(s)\}$ $(0 \leqslant s \leqslant 1)$ approaching $v(s)$, we have:

$$Pz_T(v_T(s)) \to Pz(v(s)), \tag{A.5}$$

where

$$Pz_T(v_T(s)) = Z_T(\lambda, r)'\left[\int_0^1 Z_T(\lambda, s)Z_T(\lambda, s)'\,ds\right]^{-1}\int_0^1 Z_T(\lambda, s)v_T(s)\,ds,$$

and

$$Pz(v(s)) = Z(\lambda, r)'\left[\int_0^1 Z(\lambda, s)Z(\lambda, s)'\,ds\right]^{-1}\int_0^1 Z(\lambda, s)v(s)\,ds.$$

We prove (A.5) in two steps. First, let

$$Pz(v_T(s)) = Z(\lambda, r)'\left[\int_0^1 Z(\lambda, s)Z(\lambda, s)'\,ds\right]^{-1}\int_0^1 Z(\lambda, s)v_T(s)\,ds.$$

By the properties of projections in Hilbert spaces (e.g., Brockwell and Davis, 1991, p. 52):

$$Pz(v_T(s)) \to Pz(v(s)) \text{ if } v_T(s) \to v(s). \tag{A.6}$$

Now let

$$Pz_T(v(s)) = Z_T(\lambda, r)'\left[\int_0^1 Z_T(\lambda, s)Z_T(\lambda, s)'\,ds\right]^{-1}\int_0^1 Z_T(\lambda, s)v(s)\,ds.$$

We need the following Lemma stated in Parthasarathy (1977, proposition 41.19).

Lemma A.3. Let $S_1 \subset S_2 \subset \cdots$ be an increasing sequence of subspaces in a Hilbert space $\mathscr{H}$ and let $S_\infty = \bigcup_i S_i$. Then $\lim_{T\to\infty} P(S_T)(x) = P(S_\infty)(x)$ for all x, where $P(S_T)(x)$ is the projection of x on the subspace S_T.

Lemma A.3 applied to our problem implies that

$$Pz_T(\lambda)(v(s)) \to Pz(\lambda)(v(s)), \tag{A.7}$$

since we can take $\mathscr{H} = \mathrm{D}[0, 1]$ in which case $Z_T(\lambda, r) \in \mathrm{D}[0, 1]$ and $Z(\lambda, r) \in \mathrm{C}[0, 1] \subset \mathrm{D}[0, 1]$.

Next, we use the result that if for some sequence of random variables $\{X_T\}$ and $\{Y_T\}$ we have $X_T \Rightarrow X$ and $\|X_T - Y_T\| \to 0$ (under some P-measure), then $Y_T \Rightarrow X$ (under the same P-measure) (e.g., Billingsley, 1968, Theorem 4.1; Parthasarathy, 1977, Corollary 51.3). Let $X = Pz(v(s))$, $X_T = Pz_T(v(s))$ and $Y_T = Pz_T(v_T(s))$. Given (A.7), we only need to show that $\|Pz_T(v(s)) - Pz_T(v_T(s))\| \to 0$. This follows easily since

$$\begin{aligned}\|Pz_T(v(s)) - Pz_T(v_T(s))\|^2 &= \|Pz_T(v(s) - v_T(s))\|^2 \\ &= \|v(s) - v_T(s)\|^2\end{aligned}$$

$$- \|v(s) - v_T(s) - Pz_T(v(s) - v_T(s))\|^2$$
$$\leqslant \|v(s) - v_T(s)\|^2 \to 0.$$

This completes the proof of part (c). To prove part (d), note that we have

$$\int_0^1 P_T z(\lambda) X_T(r)\,\mathrm{d}X_T(r)$$
$$= \int_0^1 Z_T(\lambda, r)'\,\mathrm{d}X_T(r)\left[\int_0^1 Z_T(\lambda, s) Z_T(\lambda, s)'\,\mathrm{d}s\right]^{-1}\int_0^1 Z_T(\lambda, s) X_T(s)\,\mathrm{d}s.$$

For concreteness consider model 2 where $Z_T(\lambda, s) \equiv (Z_{1,T}(s), Z_{2,T}(\lambda, s))'$ with $Z_{1,T}(s) \equiv (1, [Ts]/T)'$, $Z_{2,T}(\lambda, s) \equiv (1([Ts]/T > \lambda),\ 1([Ts]/T > \lambda)([Ts]/T - \lambda))'$ and $Z(\lambda, s) \equiv (Z_1(s), Z_2(s, \lambda))'$ with $Z_1(s) \equiv (1, s)'$, $Z_2(\lambda, s) \equiv (\mathrm{d}u(\lambda, s), \mathrm{d}t^*(\lambda, s))'$. Also define $Z^*_{2,T}(s) = (1, [Ts]/T - \lambda)'$ and $Z^*_2(s) = (1, (s - \lambda))'$. Using arguments as in Gregory and Hansen (1996),

$$\int_0^1 Z_T(\lambda, s) Z_T(\lambda, s)'\,\mathrm{d}s$$
$$= \begin{bmatrix} \int_0^1 Z_{1,T}(s) Z_{1,T}(s)'\,\mathrm{d}s & \int_0^1 Z_{1,T}(s) Z_{2,T}(\lambda, s)'\,\mathrm{d}s \\ \int_0^1 Z_{2,T}(\lambda, s) Z_{1,T}(s)'\,\mathrm{d}s & \int_0^1 Z_{2,T}(\lambda, s) Z_{2,T}(\lambda, s)'\,\mathrm{d}s \end{bmatrix}$$
$$= \begin{bmatrix} \int_0^1 Z_{1,T}(s) Z_{1,T}(s)'\,\mathrm{d}s & \int_\lambda^1 Z_{1,T}(s) Z^*_{2,T}(s)'\,\mathrm{d}s \\ \int_\lambda^1 Z^*_{2,T}(s) Z_{1,T}(s)'\,\mathrm{d}s & \int_\lambda^1 Z^*_{2,T}(s) Z^*_{2,T}(s)'\,\mathrm{d}s \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} \int_0^1 Z_1(s) Z_1(s)'\,\mathrm{d}s & \int_\lambda^1 Z_1(s) Z^*_2(s)'\,\mathrm{d}s \\ \int_\lambda^1 Z^*_2(s) Z_1(s)'\,\mathrm{d}s & \int_\lambda^1 Z^*_2(s) Z^*_2(s)'\,\mathrm{d}s \end{bmatrix}$$
$$= \begin{bmatrix} \int_0^1 Z_1(s) Z_1(s)'\,\mathrm{d}s & \int_0^1 Z_1(s) Z_2(\lambda, s)'\,\mathrm{d}s \\ \int_0^1 Z_2(\lambda, s) Z_1(s)'\,\mathrm{d}s & \int_0^1 Z_2(\lambda, s) Z_2(\lambda, s)'\,\mathrm{d}s \end{bmatrix} = \int_0^1 Z(\lambda, s) Z(\lambda, s)'\,\mathrm{d}s.$$

Note that the result does not require that $Z_T(\lambda, s) \Rightarrow Z(\lambda, s)$ under the uniform metric. Similarly, we have

$$\int_0^1 Z_T(\lambda, s) X_T(s)\,\mathrm{d}s \Rightarrow \int_0^1 Z(\lambda, s) X(s)\,\mathrm{d}s.$$

Finally,

$$\begin{aligned}\int_0^1 Z_T(\lambda, r)'\,\mathrm{d}X_T(r) &= T^{-1/2}\sum_1^T Z_T(\lambda, t/T)e_t \\ &= \left(T^{-1/2}\sum_1^T e_t,\ T^{-3/2}\sum_1^T te_t,\ T^{-1/2}\sum_{T_b+1}^T e_t,\ T^{-3/2}\sum_{T_b+1}^T (t-T_b)e_t\right) \\ &\Rightarrow \left(W(1),\ \int_0^1 r\,\mathrm{d}W(r),\ W(1)-W(\lambda),\ \int_\lambda^1 (r-\lambda)\,\mathrm{d}W(r)\right) \\ &= \int_0^1 Pz(\lambda)W(r)\,\mathrm{d}W(r).\end{aligned}$$

This completes the proof of part (d).

To complete the proof of the main result, we need to show continuity of the various functionals. Continuity of h^* and h is proved in $Z-A$.

Lemma A.2. The functions H_1 and H_2 defined by (A.3) *and* (A.4) *are continuous at* $(W(r), Pz(\lambda)W(r))$ *and* $(\int_0^1 W(r)\mathrm{d}W(r), \int_0^1 Pz(\lambda)W(r)\mathrm{d}W(r))$ *with W-probability one.*

Proof. Since H_1 and H_2 are continuous functions of their respective elements, the proof follows if each of the elements is bounded over $[0, 1]$ with W-probability one. $W(\cdot)$ is bounded with W-probability one and so is $\int_0^1 W(r)\mathrm{d}W(r)$ as discussed in $Z-A$. Using arguments similar to those in $Z-A$, $\int_0^1 Pz(\lambda)W(r)\mathrm{d}W(r)$ will be continuous if $Pz(\lambda)W(r)$ is continuous, i.e. if $\sup_{\lambda\in[0,1]}|Pz(\lambda)W(r)| < \infty$. We note that $Pz(\cdot)$ is a linear operator that maps an element on $C[0, 1]$ (the Wiener process $W(r)$ which is continuous) to a subspace defined by the functions $Z(\lambda, r)$. Continuity of $Pz(\lambda)W(r)$ follows since a linear projection map is bounded and continuous (see, e.g., Ash, 1972, p. 130 and p. 148). □

It is useful to illustrate this result by an example. Consider Model 1 where $Z(\lambda, r) = (1, r, du(\lambda, r))$. Note that

$$\int_0^1 Z(\lambda, s)Z(\lambda, s)'\,\mathrm{d}s = \begin{bmatrix} 1 & 1/2 & (1-\lambda) \\ 1/2 & 1/3 & (1-\lambda^2)/2 \\ (1-\lambda) & (1-\lambda^2)/2 & (1-\lambda) \end{bmatrix}.$$

If $\lambda = 0$, $\int_0^1 Z(0, s)Z(0, s)'\mathrm{d}s = \mathrm{A}$ and if $\lambda = 1$, $\int_0^1 Z(1, s)Z(1, s)'\mathrm{d}s = B$, where

$$A = \begin{bmatrix} 1 & 1/2 & 1 \\ 1/2 & 1/3 & 1/2 \\ 1 & 1/2 & 1 \end{bmatrix} \quad \text{and} \quad B \equiv \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1/3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A and B are obviously singular, but a common g-inverse is given by

$$G = 12\begin{bmatrix} 1/3 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the choice of g-inverse leaves a projection map unchanged, we have for $\lambda = 0, 1$:

$$Pz(\lambda)W(r) = Z^{\perp}(r)'\left[\int_0^1 Z^{\perp}(s)Z^{\perp}(s)'\,\mathrm{d}s\right]^{-1}\int_0^1 Z^{\perp}(s)W(s)\,\mathrm{d}s,$$

where $Z^{\perp}(r)' = (1, r)$, in which case the limiting distribution of $t_{\hat{\alpha}}(\lambda)$ $(\lambda = 0, 1)$ reduces to that in the case where no dummy for structural change is included. The proof for Model 3 follows similar arguments and is therefore omitted. It uses the limiting distribution for fixed λ derived in Perron and Vogelsang (1993a, b) (see also Vogelsang, 1993).

References

Ash, R.B., 1972. Real analysis and probability. Academic Press, New York.

Banerjee, A., R.L. Lumsdaine, J.H. Stock, 1992. Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence. Journal of Business and Economic Statistics 10, 271–287.

Billingsley, P., 1968. Convergence of probability measures. Wiley, New York.

Brockwell, P.J., R.A. Davies, 1991. Time series: theory and methods, 2nd ed. Springer, New York.

Campbell, J.Y., N.G. Mankiw, 1989. International evidence on the persistence of economic fluctuations. Journal of Monetary Economics 23, 319–333.

Christiano, L.J., 1992. Searching for breaks in GNP. Journal of Business and Economic Statistics 10, 237–250.

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

Friedman, M., A.J. Schwartz, 1982. Monetary trends in the United States and the United Kingdom: their relation to income, prices and interest rates. 1867–1975, The University of Chicago Press, Chicago.

Gregory, A.W., B.E. Hansen, 1996. Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics 70.

Hall, A., 1994. Testing for a unit root in time series with pretest data based model selection. Journal of Business and Economic Statistics 12, 461–470.

Nelson, C.R., C.I. Plosser, 1982. Trends and random walks in macroeconomic time series. Journal of Monetary Economics 10, 139–162.

Ng, S., 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.

Parthasarathy, K.R., 1978. Introduction to probability and measure. Springer, New York.

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

Perron, P., 1991. Test consistency with varying sampling frequency. Econometric Theory 7, 341–368.

Perron, P., 1994. Trend, unit root and structural change in macroeconomic time series. In B.B. Rao (ed.), Cointegration for the applied economists. MacMillan Press, Basingstoke, pp. 113–146.
Perron, P., T.J. Vogelsang, 1992. Nonstationarity and level shifts with an application to purchasing power parity. Journal of Business and Economic Statistics 10, 301–320.
Perron, P., T.J. Vogelsang, 1993a. A note on the asymptotic distributions of unit root tests in the additive outlier model with breaks. Revista de Econometria 8, 181–202.
Perron, P., T.J. Vogelsang, 1993b. The Great Crash, the oil price shock and the unit root hypothesis: Erratum. Econometrica 61, 248–249.
Press, W.H., B.P. Flanery, S.A. Teukolsky, W.T. Vetterling, 1986. Numerical recipes: the art of scientific computing, Cambridge University Press, Cambridge.
Said, S.E., D.A. Dickey, 1984. Testing for unit roots in autoregressive – moving average models of unknown order. Biometrika 71, 599–608.
Vogelsang, T.J., 1993. Essays on testing for nonstationarities and structural change in time series models. Unpublished Ph.d. Dissertation. Department of Economics, Princeton University, Princeton, NJ.
Vogelsang, T.J. and P. Perron, 1994. Additional tests for a unit root allowing for a break in the trend function at an unknown time. Manuscript. Department of Economics, Cornell University, Ithaca, NY.
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 Economic Statistics 10, 251–270.