Macroeconomic Dynamics, **6**, 2002, 614–632. Printed in the United States of America.
DOI: 10.1017.S1365100501010094

UNIT ROOTS, TREND BREAKS, AND TRANSITORY DYNAMICS: A MACROECONOMIC PERSPECTIVE

LUTZ KILIAN
University of Michigan, Ann Arbor
and
CEPR

LEE E. OHANIAN
University of California, Los Angeles
and
Federal Reserve Bank of Minneapolis

It is common to interpret rejections of the unit-root null hypothesis in favor of a trend stationary process with possible trend breaks as evidence that the data are better characterized as stationary about a broken trend. This interpretation is valid only if the model postulated under the alternative hypothesis is the only plausible alternative to the model postulated under the null. We argue that there are economically plausible models that are not well captured under either the null hypothesis or the alternative hypothesis of these tests. We show that applied researchers who ignore this possibility are likely to reject the unit-root null with high probability in favor of a trend stationary process with possible breaks. Our evidence shows that this potential pitfall is both economically relevant and quantitatively important. We explore the extent to which applied users may mitigate inferential errors by using finite-sample and bootstrap critical values.

Keywords: Unit Root, Trend Break, Transitory Dynamics, Bootstrap

1. INTRODUCTION

It is common practice in time-series analysis of economic data to test the null hypothesis of a unit root against the alternative of trend stationarity with one or more possible breaks in the trend function. (Table 1 lists several examples).

When applied to long-run annual time series, such as the data used by Nelson and Plosser (1982), the null hypothesis is frequently rejected [e.g., Perron (1989, 1992, 1997); Zivot and Andrews (1992), Lumsdaine and Papell (1997)]. Users

We thank Bob Barsky, Valentina Corradi, Frank Diebold, Atsushi Inoue, Serena Ng, Bob Stine, and, in particular, an anonymous referee for helpful comments and suggestions. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The second author acknowledges financial support from the National Science Foundation and the Sloan Foundation. Address correspondence to: Lee E. Ohanian, Bunche Hall 8391, Department of Economics, UCLA, Box 951477, Los Angeles, CA 90095-1477, USA; e-mail: ohanian@econ.ucla.edu.

 1365-1005/02 $9.50

Table 1. Applications of unit-root tests against trend-break alternatives

Application	Examples
Nelson–Plosser data	Perron (1989, 1991), Rappoport and Reichlin (1989), Zivot and Andrews (1992), Lumsdaine and Papell (1997), Nuñes et al. (1997), Perron (1997)
Tariff rates	Sadorsky (1994)
Real exchange rates	Edison and Fisher (1991), Perron and Vogelsang (1992a), Culver and Papell (1995), Jorion and Sweeney (1996)
Net exports	Husted (1992)
Income inequality	Carlino and Mills (1993), Raj and Slottje (1994), Loewy and Papell (1996)
Commodity prices	Perron (1990), Trivedi (1995)
Interest rates	Perron (1990), Duck (1992), Perron and Vogelsang (1992b), Evans and Lewis (1995), Clemente et al. (1998)
Unemployment rates	Perron (1990), Perron and Vogelsang (1992b), Papell et al. (2000)
Price level	Balke and Fomby (1991), Duck (1992)
Inflation	Evans and Lewis (1995), Culver and Papell (1997)
Money	Duck (1992)
Velocity of money	Serletis (1995)
U.S. postwar output	Perron (1989), Balke and Fomby (1991), Pischke (1991), Christiano (1992), Zivot and Andrews (1992)
U.S. long-run output	Banerjee et al. (1990), Balke and Fomby (1991)
International output	Banerjee et al. (1992), Perron (1992), Raj (1992), De Haan and Zelhorst (1993), Alba and Papell (1995), Ben-David and Papell (1995,1998), Bradley and Jansen (1995), Zelhorst and De Haan (1995),Cheung and Chinn (1996), Perron (1997), Ben-David et al. (1999).

typically interpret these rejections as evidence that the data are well characterized as stationary about a broken trend. For example, Perron (1989, p. 1362) concludes from his study of twentieth century U.S. data "that most macroeconomic time series are not characterized by the presence of a unit root and that fluctuations are indeed transitory" once allowance for a possible one-time trend break is made.[1]

The implicit assumption underlying this interpretation is that the model postulated under the alternative hypothesis is the only plausible alternative to the model postulated under the unit-root null.[2] In this paper, we argue that this assumption is often questionable. For example, samples of U.S. data covering the 1930's and 1940's include the Great Depression and World War II. Economists widely agree that both events had very large *transitory* effects on aggregate economic activity. However, neither the null hypothesis nor the alternative hypothesis of unit-root tests allows for these types of very large transitory movements in the data. We conjecture that researchers following the conventional interpretation of these tests will therefore tend to incorrectly associate the transitory dynamics from one or both of these events with a *permanent* change in the deterministic trend function. This possibility is consistent with the fact that many estimated trend break dates in the data coincide with either the Great Depression or World War II.

The goal of this paper is to analyze formally how likely applied researchers are to mistake transitory dynamics from the Great Depression and World War II as trend breaks if they follow the tradition established by Perron (1989, 1997). We conduct a Monte Carlo study of the rejection rates of the widely used Zivot–Andrews (1992) test, which allows for trend-break alternatives. In our Monte Carlo experiment, we generate data from a simple statistical model of U.S. real output that includes transitory fluctuations from the Depression and World War II as well as a random-walk component. We test the null hypothesis of a unit root against the alternative of trend stationarity with a possible one-time break in the trend function.

Our main finding is that large transitory fluctuations, such as those arising from wars or depressions, can lead to high rejection rates of the unit-root null, despite the presence of a random-walk component. For commonly used sample sizes, we observe rejection rates as high as 67% for the nominal 10% test based on asymptotic critical values. Because of the poor performance of the asymptotic test, we also analyze the performance of finite-sample critical values of the type recently proposed by Perron (1997). These tests are less likely to reject the null, but still may have rejection rates as high as 50% for a 10% test. Finally, we study the performance of bootstrap critical values. Although the bootstrap test is much less likely than the other tests to reject the unit-root null, it still is biased in favor of rejecting the unit-root null in the presence of large transitory fluctuations with rejection rates of up to 22%.

Our results also raise questions about the conventional interpretation that estimated break dates "yield interesting conclusions about the identification of major economic events that had a permanent effect on the levels of economic activity" [Perron (1992, p. 144)]. The major economic events in our Monte Carlo study by construction have no permanent effects on output, yet casual interpretations of the test results routinely seem to suggest otherwise.

Our analysis is related to recent work by Murray and Nelson (2000), who explore the sensitivity of unit-root test results to a variety of departures from conventional assumptions. Whereas their work focuses on the economic plausibility of the cyclical component estimated from the trend-stationary model, we focus on the economic plausibility of a permanent break in the trend function. Another important distinction is that Murray and Nelson propose to reduce the possibility of spurious rejections of the unit-root null by discarding the long-run annual output data altogether and working with quarterly postwar data instead. In contrast, we focus on the use of bootstrap critical values to reduce the probability of spurious rejections of the unit-root null.

The remainder of the paper is organized as follows: Section 2 provides the motivation for our choice of data generating process (DGP) and places our analysis into the context of the literature. Section 3 reviews the methodology of unit-root tests against trend-break alternatives. Section 4 describes our statistical model. Section 5 describes the simulation design. Section 6 presents the findings of the Monte Carlo analysis. Section 7 concludes.

2. CHOICE OF DGP

Economists widely agree that there were large transitory changes in output during the Great Depression and World War II. This is because the large changes in output during these episodes theoretically are easy to reconcile with temporary shocks, but are difficult to reconcile with permanent shocks. Figure 1 plots the log of real per-capita output over the 1869–1993 interval. The substantial fluctuations in output during the Great Depression and World War II are clearly seen in the figure. We describe these two episodes, and the shocks assumed to be important for these periods, below.

During the Depression, real per-capita output fell about 30% between 1929 and 1933. Many economists, including Friedman and Schwartz (1963) and Lucas and Rapping (1969), view the Depression as a temporary response to a very large, temporary decline in the money stock. The stock of currency and demand deposits (M1) fell over 30% between 1929 and 1933. After 1933, both the money stock and

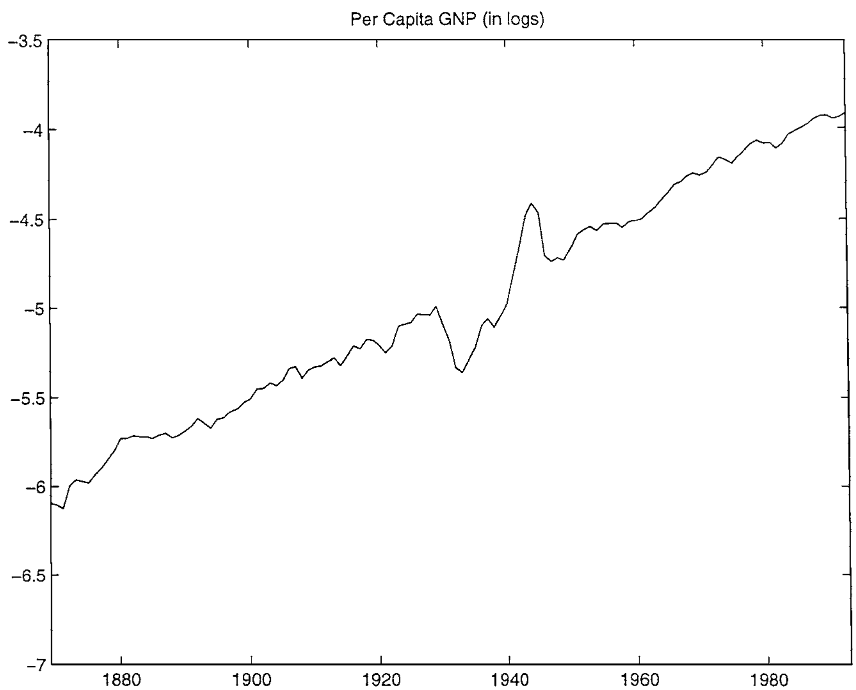

Figure 1. U.S. per-capita GNP series (1869–1993). *Source*: Diebold and Senhadji (1996).

https://doi.org/10.1017/S1365100501010094 Published online by Cambridge University Press

output began to recover. This pattern is consistent with the predictions of monetary business-cycle theory: Temporary changes in the money stock lead to *temporary* changes in economic activity. The main mechanism through which money shocks can have large temporary effects on economic activity is through *intertemporal substitution*. This is true whether the nonneutrality of money is due to imperfect price or wage adjustment or imperfect information [see Lucas (1972), Bordo et al. (2000), Chari et al. (2000)].

Intertemporal substitution also plays an important role in accounting for changes in U.S. output during the 1940's. During World War II, real per-capita GNP grew 41%. Most economists agree that this large increase was due to the enormous rise in temporary government purchases brought about by the war effort. Between 1940 and 1944, government purchases rose 232%. After the war, government purchases fell by 49%. Standard neoclassical theory predicts that temporary changes in government purchases can lead to a large, but temporary, increase in economic activity through intertemporal substitution. For example, Ohanian (1997) shows that, with temporary increases in government purchases, a neoclassical growth model can account for the large increase in output during World War II. The same channel of intertemporal substitution has been emphasized by Barro (1981).

Since intertemporal substitution is a key factor in the quantitative analyses of these two episodes, it is interesting to note that this mechanism is not operative for shocks that are perceived to have permanent effects. Thus, one would not expect the effects of permanent shocks on output to be as large or as rapid as for a transitory shock. For these reasons, economists have concluded that a substantial fraction of the movements in output during these two episodes must have been transitory.

In our Monte Carlo analysis, we develop a simple reduced-form model of annual U.S. per-capita GNP that captures the view that there were large transitory fluctuations in output during World War II and the Great Depression. We develop the model for the 1909–1970 period considered by Nelson and Plosser (1982) and later extend it to a longer sample. Our model generates time series as the sum of a latent random walk with drift and occasional large transitory movements driven by a regime-switching process. By construction, this process has a unit root but does not include any trend breaks. For expository purposes, we abstract from any permanent effects from the Great Depression and World War II, as well as other wars (World War I, the Korean War, and the Vietnam War). This simplification reflects our focus on the implications of large transitory dynamics for interpreting the results of unit-root tests against trend-break alternatives.

Other authors have argued that tests for unit roots in general may be sensitive to the presence of transitory components in the DGP. Most prominently, Schwert (1987, 1989) shows that standard ADF tests reject the unit-root null too often if the true model is an ARIMA(0,1,1) model with an MA coefficient close to -1. In practice, this warning has led researchers to investigate the possible presence of a large negative MA coefficient in growth rates. Since the growth rates of many macroeconomic series do not have MA(1) coefficients near −1, however, many researchers have dismissed the notion that transitory dynamics in their data can interfere with statistical inference.[3] We demonstrate that this dismissal can be a serious mistake. The unobserved-components model that we postulate as our DGP is substantially different from Schwert's linear ARIMA(0,1,1) model and can result in severe inference problems not foreseen by Schwert. The problem we study cannot be identified with the diagnostics that work for Schwert's problem, and procedures that result in correct inference in his model do not in ours [see Kilian and Ohanian (1998) for details]. Indeed, our results indicate that fundamental changes are needed in the application and interpretation of unit-root tests against trend-break alternatives.

Our approach is also related to, but distinct from, the work of Franses and Haldrup (1994), who show that adjacent outliers (or temporary-change outliers) may bias unit-root tests in favor of the trend-stationary alternative (without breaks). There are four main differences: First, we view the transitory fluctuations as an endogenous response of the underlying time-series process to economic shocks rather than an exogenous and extraneous influence to be removed from the analysis (such as the effect of a strike or a change in the definition of a price series). Second, Franses and Haldrup establish the existence of a bias arising from transitory dynamics, but they do not investigate its quantitative importance for the rejection rates of the widely used ADF test. Third, we show that this bias has important implications for the interpretation of unit-root tests against trend-break alternatives that few applied researchers appear to be aware of. Fourth, we provide new evidence about the extent to which finite-sample and bootstrap critical values may mitigate size distortions of the Zivot–Andrews test in the standard random-walk model and about the extent to which they may protect against spurious rejections due to transitory dynamics.

3. UNIT-ROOT TESTS AGAINST TREND-BREAK ALTERNATIVES

Many methodologies exist for unit-root tests against trend-break alternatives.[4] The early literature, for example, Perron (1989), assumed that the breakpoints are exogenously determined on the basis of economic theory before inspecting the data. Today, most researchers favor the assumption that the breakpoints must be estimated endogenously [see Perron (1997) for further discussion]. In this paper, we follow that convention and focus on the sequential breakpoint selection tests developed by Zivot and Andrews (1992). These tests are popular in applied work.[5] They are also similar to the methodology used by Banerjee et al. (1992) and Perron (1997). We study three versions of the Zivot–Andrews test that consider the same null hypothesis of a unit root, but differ in the type of structural break considered under the alternative. All tests assume a one-time break under the alternative at date TB. We let $DU_t = 1(t > TB)$ and $DT_t = (t - TB)\,1(t > TB)$, where $t = 1, \ldots, T$ and $1(\cdot)$ is the indicator function. Following Zivot and Andrews (1992), the regression models are

https://doi.org/10.1017/S1365100501010094 Published online by Cambridge University Press

$$Model\ A: \quad \Delta y_t = \mu + \beta t + \theta DU_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t,$$

$$Model\ B: \quad \Delta y_t = \mu + \beta t + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t,$$

$$Model\ C: \quad \Delta y_t = \mu + \beta t + \theta DU_t + \gamma DT_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + \varepsilon_t,$$

where $\{y_t\}_{t=1}^{T}$ denotes the time series of interest. We follow Zivot and Andrews (1992) in determining the number of augmented lags k by Perron's (1989) sequential t-value procedure, starting with $k^{\max} = 8$ [also see Ng and Perron (1995)]. We test H_0: $\alpha = 0$ against the one-sided alternative. Since the breakpoint is assumed to be unknown, the *Dickey–Fuller statistic* is defined as the infimum of the sequence of the *Dickey–Fuller statistics* over all possible breakpoints, not including the endpoints of the sample. The asymptotic critical values for t_α are taken from Zivot and Andrews (1992). They also propose a small-sample extension of their test that involves resampling the ADF statistic under the null by fitting an ARIMA(p, 1, q) model to the data. The bootstrap critical values of that test can be read from the empirical distribution of t_α. Perron (1997) proposes simulating finite-sample critical values that allow for lag-order uncertainty under the null hypothesis of a random-walk model. We will consider all three types of critical values in Section 6.

4. AN UNOBSERVED-COMPONENTS MODEL WITH OCCASIONAL TRANSITORY FLUCTUATIONS

Here we describe the statistical model that underlies our simulation study. Large transitory responses in output to temporary government spending shocks or to monetary shocks arise naturally in dynamic equilibrium models [see Ohanian (1997)]. In principle, we could base the DGP on a dynamic optimizing macroeconomic model with regime switching in the decision-rule coefficients. However, for tractability and simplicity, we focus on a univariate reduced-form time-series model that captures these same features. We focus on the annual U.S. per-capita output series of Nelson and Plosser (1982) for the 1909–1970 period because it has been analyzed by a number of authors and because output is central to a number of macroeconomic questions.

We treat World War II and the Great Depression as random events that occur with positive probability along the sample path. In the model, wars and depressions follow two independent Markov chains modeled after World War II and the Great Depression. We specify the Markov chains in detail later in this section. We assume that wars and depressions have only transitory effects. Let s_{1t} denote the state underlying the war variable w and s_{2t} the state underlying the depression variable d. The variables w_{s1t} and d_{s2t} measure the transitory effects on output of World War II

https://doi.org/10.1017/S1365100501010094 Published online by Cambridge University Press

and the Great Depression, respectively. These variables take on nonzero values if the respective underlying state is activated, and zero values otherwise.

We assume that real per-capita output (y_t) follows a process that is the sum of a latent random walk with drift (z_t) and the two state-dependent transitory components w_{s1t} and d_{s2t}. The specification of our model differs in important ways from standard outlier models in that the effects of wars and depressions, by construction, are not permanent and in that the random walk is not directly observable:

$$\begin{aligned} y_t &= z_t + w_{s1t} + d_{s2t} \\ z_t &= \mu + z_{t-1} + \varepsilon_t \\ s_{1t} &= \Pi^w s_{1t-1} + v_t \\ s_{2t} &= \Pi^d s_{2t-1} + u_t \end{aligned} \tag{1}$$

where $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, v_t, and u_t are martingale difference sequences, and Π^w and Π^d are the transition probability matrices associated with s_{it}, $i = 1, 2$ [see Hamilton (1994)]. Lowercase letters denote natural logs.

We define the states s_{1t} underlying the transitory component w_{s1t} so that wars in the model will have the same duration as World War II in the U.S. data. For the purpose of this model, the World War II period is assumed to last six years, corresponding to the years 1941–1946 in the U.S. data. We include 1946 in the war period because we are interested in the transitory effects of the war on the level of economic activity. In 1946, immediately after the war, government expenditures and output fell dramatically. For expositional purposes, we discount the possibility of longer-lasting effects. We define the *primitive states* $s_{1t}^* \in \{W, P\}$, corresponding to whether a particular year t is a war year or a peace year. By dividing the 62 observations of the Nelson–Plosser output series into overlapping blocks of six consecutive annual observations, we can completely characterize the Nelson–Plosser data by 12 states s_{1t}, each consisting of six consecutive annual observations. For example, the six-year period from 1938 to 1943 is given by $\{P, P, P, W, W, W\}$, and the period for 1939 through 1944 is given by $\{P, P, W, W, W, W\}$.

The states s_{2t} underlying d_{s2t} are constructed in a similar fashion so that depressions in the model will have the same duration as the Great Depression in the U.S. data. For the purpose of the model, we assume that the Great Depression lasts five years, corresponding to the years 1929–1933 in the U.S. data. We define $s_{2t}^* \in \{D, N\}$, corresponding to whether a particular year t is a depression year or a normal year. We then define 10 states s_{2t}, each consisting of five consecutive annual observations, such that these states completely characterize the observed U.S. data. Based on the Nelson–Plosser data, we determine the empirical transition probabilities as shown in Table 2. The resulting transition probability matrix Π^w in terms of the states s_{1t} and s_{1t+1} is shown in Figure 2.

Similarly, we summarize the empirical transition probabilities from s_{2t} to s_{2t+1} in the matrix Π^d. Our model is the simplest DGP that generates wars of the

TABLE 2. Empirical transition probabilities from s_{1t} to s_{1t+1}, conditional on the last six states s^*_{1t}

s_{1t}	s^*_{1t}	s^*_{1t-1}	s^*_{1t-2}	s^*_{1t-3}	s^*_{1t-4}	s^*_{1t-5}	s_{2t}	s^*_{2t}	s^*_{2t-1}	s^*_{2t-2}	s^*_{2t-3}	s^*_{2t-4}
1	W	W	W	W	W	W	1	D	D	D	D	D
2	P	W	W	W	W	W	2	N	D	D	D	D
3	P	P	W	W	W	W	3	N	N	D	D	D
4	P	P	P	W	W	W	4	N	N	N	D	D
5	P	P	P	P	W	W	5	N	N	N	N	D
6	P	P	P	P	P	W	6	N	N	N	N	N
7	P	P	P	P	P	P	7	D	N	N	N	N
8	W	P	P	P	P	P	8	D	D	N	N	N
9	W	W	P	P	P	P	9	D	D	D	N	N
10	W	W	W	P	P	P	10	D	D	D	D	N
11	W	W	W	W	P	P						
12	W	W	W	W	W	P						

same duration as World War II and depressions of the same duration as the Great Depression.

5. SIMULATION DESIGN

We consider three alternative DGPs for the Monte Carlo analysis. Our benchmark model is a simple random walk with drift (DGP1):

$$y_t = \mu + y_{t-1} + \varepsilon_t. \tag{2}$$

This model contains no transitory components, and all shocks (ε_t) have permanent effects. Without loss of generality, we set the drift μ equal to $\hat{\mu}$ where $\hat{\mu} = 0.0164$ is obtained by regressing the first-differenced logged Nelson–Plosser data on a constant. The corresponding estimated residual standard error is 0.0653. However, the presence of measurement error, in particular for the pre-1929 period, and the possible presence of large transitory fluctuations during World War II and the Great Depression suggest that this estimate is considerably inflated.[6] We therefore base σ_ε on the residual standard error of the same regression for annual postwar data. That estimate is $\hat{\sigma}_\varepsilon = 0.024$.[7]

DGP2 is based on model (1). This reflects our concern that, in small samples, large transitory movements, w_{s1t} and d_{s2t}, may resemble trend breaks, leading conventional trend-break tests to overreject the null of a unit root. For expository purposes, we approximate the transitory variation in output over the Great Depression (1929–1933) and World War II (1941–1946) by the excess growth relative to the drift over the periods 1929–1933 (for the Great Depression) and 1941–1946 (for World War II). That is, $d_{s2t} = 1(s^*_{2t} = D)(\Delta y_t - \hat{\mu})$ and $w_{s1t} = 1(s^*_{1t} = W)(\Delta y_t - \hat{\mu})$. We set $\mu = 0.0164$, $\sigma_\varepsilon = 0.024$, $d_{s2t} = \{0.0370, -0.1311, -0.1048, -0.1833, -0.0410\}$ for $s^*_{2t} = D$, and $w_{s1t} = \{0.1228, 0.0941, 0.0937,$

https://doi.org/10.1017/S1365100501010094 Published online by Cambridge University Press

(A) **Transition Probability Matrix Π^w:**

							s_{1t}						
		1	2	3	4	5	6	7	8	9	10	11	12
	1	0	0	0	0	0	0	0	0	0	0	0	1
	2	1	0	0	0	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	0	0	0	0	0	0	0
	4	0	0	1	0	0	0	0	0	0	0	0	0
	5	0	0	0	1	0	0	0	0	0	0	0	0
s_{1t+1}	6	0	0	0	0	1	0	0	0	0	0	0	0
	7	0	0	0	0	0	1	45/46	0	0	0	0	0
	8	0	0	0	0	0	0	1/46	0	0	0	0	0
	9	0	0	0	0	0	0	0	1	0	0	0	0
	10	0	0	0	0	0	0	0	0	1	0	0	0
	11	0	0	0	0	0	0	0	0	0	1	0	0
	12	0	0	0	0	0	0	0	0	0	0	1	0

(B) **Transition Probability Matrix Π^d:**

						s_{2t}					
		1	2	3	4	5	6	7	8	9	10
	1	0	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	0	0	0	0	0
	4	0	0	1	0	0	0	0	0	0	0
s_{2t+1}	5	0	0	0	1	0	0	0	0	0	0
	6	0	0	0	0	1	48/49	0	0	0	0
	7	0	0	0	0	0	1/49	0	0	0	0
	8	0	0	0	0	0	0	1	0	0	0
	9	0	0	0	0	0	0	0	1	0	0
	10	0	0	0	0	0	0	0	0	1	0

Figure 2. Markov-switching models for wars and depressions.

0.0411, −0.0448, −0.1544} for $s_{1t}^* = W$, with Π^w and Π^d as given in Section 4. Note that the number, sequencing, and timing of wars and depressions in DGP2 is completely unrestricted.

By allowing for multiple wars and depressions in the Markov chain, the data generated by DGP2 are likely to exhibit fewer apparent one-time trend breaks than the U.S. data. We therefore also considered a third DGP that resembles DGP2 except that we exogenously impose the number of Great Depressions and World Wars and their timing. There are exactly one war and one depression in each Monte Carlo trial of DGP3, and the depression and the war occur on the same dates as in the U.S. data. Only the random-walk component varies across trials.

6. SIMULATION RESULTS FOR THE ZIVOT–ANDREWS TEST

Our Monte Carlo analysis examines two sample sizes. The first set of results in Table 3 is based on a sample size of 62 observations (the same sample size as the

Table 3. Rejection rates of nominal 10% ADF tests for $T = 62^a$

H_1	DGP[b]	Critical values: Asymptotic	Finite-sample	Bootstrap
No-break model	1	18.7	9.8	9.7
	2	42.4	32.5	17.8
	3	54.3	35.1	18.6
Model A	1	23.4	10.9	9.3
	2	53.7	35.8	19.0
	3	54.2	33.2	17.6
Model B	1	34.9	9.3	8.2
	2	59.1	28.1	15.4
	3	53.8	16.8	8.2
Model C	1	28.9	10.6	9.7
	2	60.0	35.0	19.6
	3	66.5	34.6	22.4

[a] Based on 1,000 Monte Carlo trials.
[b] DGP1: Random-walk with drift. DGP2: Sum of random walk with drift and two transitory components driven by independent Markov chains for world wars and for depressions. DGP3: Like DGP2 with the additional restrictions that there are exactly one world war and one depression in the sample period, and the depression and the war occur on the same dates as in the U.S. data.
Sources: The asymptotic critical values for the no-break model are from Hamilton (1994). The asymptotic critical values for the other models are from Zivot and Andrews (1992). The finite-sample critical values for all four models are based on ADF regressions [see Perron (1997)]. All bootstrap critical values were calculated on the basis of ARIMA approximation suggested by Zivot and Andrews (1992).

original Nelson–Plosser data). Although this sample size may appear low, some studies use the asymptotic Zivot–Andrews test for samples as small as 44 annual observations [Raj and Slottje (1994)] or 27 annual observations [Alba and Papell (1995)]. However, since other studies use larger sample sizes, we also repeated the analysis with 124 observations. This sample size is larger than most used in the literature—among all the studies we survey, only three use more annual observations.[8] The results for $T = 124$ are in Table 4. All results are based on common random numbers and 1,000 Monte Carlo trials. We present rejection rates based on three types of critical values:

- asymptotic critical values provided by Zivot and Andrews (1992),
- finite-sample critical values of the type provided by Perron (1997),
- bootstrap critical values based on the ARIMA specification suggested by Zivot and Andrews (1992).

6.1. Test Performance Based on Asymptotic and Finite-Sample Critical Values

Before analyzing ADF tests against alternative hypotheses with possible breaks in the trend function, we first consider the test against the simple no-break alternative. We first evaluate the behavior of the test based on asymptotic critical values.

https://doi.org/10.1017/S1365100501010094 Published online by Cambridge University Press

TABLE 4. Rejection rates of nominal 10% ADF tests for $T = 124$

H_1	DGP	Critical values: Asymptotic	Finite-sample	Bootstrap
No-break model	1	12.9	8.8	8.1
	2	29.4	25.3	10.8
	3	30.1	24.0	15.7
Model A	1	15.4	10.1	9.7
	2	49.2	42.7	17.3
	3	43.3	31.9	20.2
Model B	1	23.4	11.1	10.1
	2	52.4	40.8	17.5
	3	46.2	27.8	16.5
Model C	1	18.7	10.9	9.1
	2	58.7	49.6	20.0
	3	47.7	33.6	19.8

Source: See Table 3.

Clearly, the asymptotic test suffers from severe size distortions, as evidenced by the rejection rate of about 19% for DGP1 (pure random-walk DGP) in Table 3. In addition, we find substantial bias for DGP2 and DGP3 (rejection rates of between 42 and 54%, respectively) for the asymptotic test.

The existence of size distortions raises the question of whether a test based on finite-sample critical values can perform better. To address this question, we present results based on finite-sample critical values of the type proposed by Perron (1997). While the finite-sample critical values control successfully for size (see the results for DGP1), we find rejection rates of between 33% and 35% for DGP2 and DGP3. This finding illustrates that the ability of the ADF test to reject the unit root in long-run per-capita GNP data may not provide conclusive evidence against the unit root. A similar point has recently been made by Murray and Nelson (2000). These results for the simple no-break alternative provide a quantitative complement to recent theoretical work by Franses and Haldrup (1994) who show that additive outliers in a random walk with drift can lead to the wrong impression that the series is trend stationary.[9] Although the DGPs that we consider differ from the one studied by Franses and Haldrup, our results show that this type of problem can be quantitatively important for ADF tests against the no-break alternative.

We now turn to the main goal of our paper and study the performance of ADF tests that allow for a one-time trend break under the alternative. In short, we find that substantial biases also exist for these tests. Column 1 of Table 3 shows that the test based on the widely used asymptotic critical values of Zivot and Andrews (1992) reject the unit-root null as often as 54–67% of the time. For the finite-sample statistic, we follow the general procedure of Perron (1997) to generate 10% finite-sample critical values based on 5,000 Monte Carlo trials.[10] Column 2 of Table 3 shows that the ADF test based on finite-sample critical values successfully controls size for all three trend-break models. However, for DGP2 and DGP3, rejection rates range from 17% to 36%.

The important implication of these findings is that applied users following the tradition of Perron (1989, 1997) would routinely conclude that the data from DGP2 and DGP3 are better characterized as a trend-stationary process with a one-time break, rather than an integrated process. Consequently, these users would tend to use the wrong statistical model for testing other hypotheses about these data, and may mistake the importance of permanent shocks and transitory shocks in accounting for fluctuations in these data.

It can be shown that these problems are not just an artifact of the small sample size. Table 4 presents the results for $T = 124$. We find that the rejection rates of the asymptotic test decline with increasing sample size, but only slowly. Moreover, in many cases, the rejection rates based on finite-sample critical values *increase* substantially, reflecting the convergence of asymptotic and finite-sample critical values. This result is to be expected because the limit distribution under the null for DGP1 differs from those for DGP2 and DGP3, and we know the standard critical values to be incorrect in the limit [see Franses and Haldrup (1994)].

6.2. Can Bootstrapping Eliminate the Test Bias?

The fundamental problem with using the asymptotic and finite-sample critical values is that critical values that are compiled under the pure random-walk hypothesis are inappropriate in the presence of the additional transitory dynamics in DGP2 and DGP3. A natural conjecture is that a bootstrap approach that explicitly models the nuisance parameters may result in more accurate inference. Bootstrap critical values have rarely been used in practice because they can be computationally expensive, but are likely to play a more important role, as computing power becomes cheaper.[11] Our bootstrap procedure closely follows that of Zivot and Andrews (1992).[12] For each of the Monte Carlo trials, we fit an ARIMA model under the null and calculate the bootstrap critical values based on 1,000 bootstrap replications of the Zivot–Andrews ADF test statistic.[13] The last column in Tables 3 and 4 shows that bootstrapping reduces the test bias but does not eliminate it. Like finite-sample critical values, bootstrap critical values all but eliminate the problem of size distortions in the pure random-walk case (DGP1). However, unlike finite-sample critical values, they also greatly reduce the rejection rates under DGP2 and DGP3. For example, for $T = 62$, the bootstrap test lowers rejection rates to, at most, 22% (down from 36% for the finite-sample critical values and 67% for the asymptotic critical values). For $T = 124$, it reduces rejection rates to, at most, 20% (down from 50% for the finite-sample critical values and 59% for the asymptotic critical values). However, even the bootstrap test fails to eliminate the test bias completely. In fact, there is not much evidence that the rejection rates fall as the sample size is doubled, and in some cases they increase. The failure of theboot-strap to completely eliminate the size distortions indicates that these dynamics are difficult to capture by conventional ARIMA models.[14]

7. SUMMARY AND CONCLUSION

Using a quantitatively plausible model, we show that occasional, large changes in economic activity, such as those that occurred during the Great Depression and World War II, can lead researchers to reject the null hypothesis of difference stationarity with high probability, despite the presence of a unit root. Our evidence suggests that rejections of the unit-root hypothesis for samples covering the 1930's and 1940's in the United States [e.g., Perron (1989), (1992), (1997), Zivot and Andrews (1992), Lumsdaine and Papell (1997)] may not be very informative and raises questions about whether the U.S. data are well characterized as trend stationary, once allowance for a possible trend break is made.

These results do not reflect a shortcoming of the Zivot–Andrews test (or of related tests), but rather a shortcoming in the *interpretation* of the test results. Unless the model under the null and under the alternative reasonably comprise all relevant states of the world, a rejection of the null hypothesis does not imply the acceptance of the alternative hypothesis. We showed that the common interpretation of the Zivot–Andrews test as providing *evidence* of trend breaks fails precisely when the DGP departs from the assumptions of Zivot and Andrews (1992).[15]

We conjecture that excessive rejection rates resulting from transitory shocks are not just confined to samples covering the 1930's and 1940's in the United States, but may also occur in other samples where there are occasional, large fluctuations in economic activity. For example, the rebuilding of the capital stock after World War II in many European countries and in Japan is likely to have induced large transitory dynamics in the postwar period.

The main goal of this paper was (1) to make applied users aware of this pitfall in interpreting unit-root tests; (2) to show that large, occasional fluctuations can have quantitatively important consequences for statistical inference and model selection for economic time series; and (3) to provide recommendations to applied users for dealing with this potential problem.

How can applied users guard against spurious rejections of the unit-root null due to these types of large transitory shocks? The fundamental problem is that critical values compiled under the random-walk null hypothesis are invalid in this case, even asymptotically. Franses and Haldrup (1994) have argued that, in principle, the asymptotic and finite-sample critical values can be adjusted to account for the presence of temporary outliers. Unfortunately, implementing this approach requires that the researcher know the underlying DGP. In the absence of this knowledge, this proposal is not operational. When the underlying DGP is unknown, an alternative to the Franses–Haldrup approach is the bootstrap. We explored the extent to which bootstrapping based on ARIMA models may capture transitory components added to the random-walk model. Our results suggest that the use of bootstrap critical values is essential in reducing spurious rejections of the unit-root hypothesis. Although bootstrap critical values tend to be considerably more accurate than asymptotic and finite-sample critical values, they do not completely eliminate the problem.

We conclude that a purely statistical analysis of the trend properties of economic time series is not sufficient. Instead, we recommend that applied researchers should first assess whether there are important events in the sample that could have led to large, transitory fluctuations, such as the Great Depression or World War II. Moreover, we recommend that bootstrap, rather than asymptotic or finite-sample, critical values be used to test the null of difference stationarity. These two changes in the application of unit-root tests can be important in preventing spurious rejections of the null hypothesis.

The current paper used bootstrap inference to overcome the limitations of asymptotic inference. An alternative approach would have been to compare the models of interest from a Bayesian perspective. Given the potential sensitivity of the Bayesian approach to the choice of priors, a systematic study along these lines clearly is beyond the scope of the current paper. Future work in this area would be an interesting complement to our paper.

NOTES

1. Similar interpretations for this and other series can be found in Banerjee et al. (1992, pp. 279, 282), Christiano (1992, p. 249), Lumsdaine and Papell (1997, pp. 212, 214–215, 218), Perron (1997, pp. 355–356), Rappoport and Reichlin (1989, p. 176), Zivot and Andrews (1992, pp. 251, 258, 267). Some authors, notably Banerjee et al. 1992, (p. 283), have cautioned against overly literal interpretations of rejections of the unit-root null in favor of trend stationarity with possible breaks, whereas others are less reserved in their judgment.

2. Without this assumption, a rejection of the unit-root null cannot be interpreted as evidence in favor of the alternative of trend stationarity with possible breaks.

3. A widespread view in applied work is that adding additional autoregressive lags effectively mitigates size distortions arising from MA roots [e.g., see Raj (1992) and De Haan and Zelhorst (1993)].

4. See Perron (1989, 1990, 1991), Rappoport and Reichlin (1989), Banerjee et al. (1990), Balke and Fomby (1991), Perron and Vogelsang (1992a,b; 1993a,b; 1995), Park and Sung (1994), Stock (1994), Bradley and Jansen (1995), Maddala and Kim (1996), Montañes (1997), Clemente et al. (1998), Leybourne et al. (1998), Montañes and Reyes (1998), Nuñes et al. (1997), Vogelsang and Perron (1998).

5. See, for example, Raj (1992), Raj and Slottje (1994), Sadorsky (1994), and Serletis (1995). The Zivot–Andrews tests also have recently been extended by Lumsdaine and Papell (1997) to allow for multiple trend breaks.

6. The effects of measurement error in the prewar data are discussed by Romer (1989), who argues that, after adjustment for data construction and collection methods, the variability of pre- and postwar fluctuations is fairly similar.

7. We also verify that the random-walk model with drift is an adequate representation of U.S. per-capita real GNP in the postwar period. For the first five coefficients of the autocorrelation function of the first-differenced data, we cannot reject the null that the autocorrelations were zero at conventional significance levels.

8. While our analysis is based on annual data, some studies analyze postwar data with as many as 160 quarterly observations or monthly data with close to 400 observations. However, in practice, power considerations dictate the use of the data with the longest time span, and postwar quarterly or monthly data are currently available for a span shorter than 50 years, even shorter than our sample of 62 annual observations. For a similar argument, see Perron (1992).

9. Franses and Haldrup (1994) propose to adjust the critical values to account for the presence of temporary change under the null. They derive asymptotic and finite-sample critical values. However,

their results depend critically on knowledge of the unknown DGP and are not operational in the absence of this knowledge.

10. Note that the critical values reported by Perron (1997) do not apply to Model C and, in general, are specific to his choice of sample size and lag order bounds. We therefore generated critical samples that are fully consistent with our assumptions.

11. Of all the studies in our survey that use endogenous breakpoint selection tests, only Christiano (1992), Zivot and Andrews (1992), Sadorsky (1994), and Lumsdaine and Papell (1997) bootstrapped the test statistic.

12. It has not been established whether this particular bootstrap algorithm is asymptotically valid for DGP2 and DGP3, but since this bootstrap algorithm has been used in the literature, examining its performance under our assumptions is obviously of interest.

13. Zivot and Andrews (1992) allow a maximum order of five for p and q in selecting the best-fitting ARIMA$(p, 1, q)$ model. In contrast, we impose a maximum order of three for both p and q because of computational considerations.

14. In addition to this basic analysis, we have examined the sensitivity of the results to a number of modifications, including modifications of our DGP to consider the effects of pretesting, changes in sample size, and alterations in the magnitude, timing, and sequencing of the transitory dynamics. The basic results are not sensitive to any of these effects. Details are not shown to conserve space.

15. It may be tempting to test for the presence of trend breaks directly. For example, Vogelsang (1997) proposed a test of the null of no break in the trend polynomial against the alternatives of a one-time break of unknown form that may be used even if the order of integration of the true process is unknown. However, this test does not help us in determining whether the data should be treated as I(0) or as I(1), conditional on rejecting or not rejecting a trend break, nor is the test designed to account for transitory effects of the type considered in this paper.

REFERENCES

Alba, J.D. & D.H. Papell (1995) Trend breaks and the unit-root hypothesis for newly industrializing and newly exporting countries. *Review of International Economics* 3, 264–274.

Balke, N.S. & T.B. Fomby (1991) Shifting trends, segmented trends, and infrequent permanent shocks. *Journal of Monetary Economics* 28, 61–85.

Banerjee, A., J. Dolado, & J.W. Galbraith (1990) Recursive Tests for Unit Roots and Structural Breaks in Long Annual GNP Series. Manuscript, University of Florida.

Banerjee, A., R.L. Lumsdaine, & J.H. Stock (1992) Recursive and sequential tests of the unit root and trend-break hypotheses: Theory and international evidence. *Journal of Business and Economic Statistics* 10, 271–287.

Barro, R. (1981) Output effects of government purchases. *Journal of Political Economy* 89, 1086–1121.

Ben-David, D. & D.H. Papell (1995) The Great Wars, the Great Crash, and steady state growth: Some new evidence about an old stylized fact. *Journal of Monetary Economics* 36, 453–475.

Ben-David, D. & D.H. Papell (1998) Slowdowns and meltdowns: Postwar growth evidence from 74 countries. *Review of Economics and Statistics* 80, 561–571.

Ben-David, D., R.L. Lumsdaine, & D.H. Papell (1999) Unit Roots, Postwar Slowdowns and Long-Run Growth: Evidence from Two Structural Breaks. Manuscript, University of Houston.

Bordo, M.D., C.J. Erceg, & C.L. Evans (2000) Money, sticky wages, and the Great Depression. *American Economic Review* 90, 1447–1463.

Bradley, M.D. & D.W. Jansen (1995) Unit roots and infrequent large shocks: New international evidence on output growth. *Journal of Money, Credit, and Banking* 27, 876–893.

Carlino, G.A. & L.O. Mills (1993) Are U.S. regional incomes converging? A time series analysis. *Journal of Monetary Economics* 32, 335–346.

Chari, V.V., P. Kehoe, & E. McGrattan (2000) Sticky price models of the business cycle: Can the contract multiplier solve the persistence problem? *Econometrica* 68, 1151–1179.

Cheung, Y.-W. & M.D. Chinn (1996) Deterministic, stochastic, and segmented trends in aggregate output: A cross-country analysis. *Oxford Economic Papers* 48, 134–162.

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

Clemente, J., A. Montañes, & M. Reyes (1998) Testing for a unit root in variables with a double change in the mean. *Economics Letters* 59, 175–182.

Culver, S.E. & D.H. Papell (1995) Real exchange rates under the gold standard: Can they be explained by the trend break model? *Journal of International Money and Finance* 14, 539–548.

Culver, S.E. & D.H. Papell (1997) Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics* 12, 435–444.

De Haan, J. & D. Zelhorst (1993) Does output have a unit root? New international evidence. *Applied Economics* 25, 953–960.

Diebold, F.X. & A.S. Senhadji (1996) The uncertain unit root in real GNP: Comment. *American Economic Review* 86, 1291–1298.

Duck, N.W. (1992) U.K. evidence on breaking trend functions. *Oxford Economic Papers* 44, 426–439.

Edison, H.J. & E.O. Fisher (1991) A long-run view of the European Monetary System. *Journal of International Money and Finance* 10, 53–70.

Evans, M.D.D. & K.K. Lewis (1995) Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance* 50, 225–253.

Franses, P.H. & N. Haldrup (1994) The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics* 12, 471–478.

Friedman, M. & A.J. Schwartz (1963) *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Husted, S. (1992) The emerging U.S. current account deficit in the 1980s: A cointegration analysis. *Review of Economics and Statistics* 74, 159–166.

Jorion, P. & R.J. Sweeney (1996) Mean reversion in real exchange rates: Evidence and implications for forecasting. *Journal of International Money and Finance* 15, 535–550.

Kilian, L. & L.E. Ohanian (1998) Is There a Trend Break in U.S. GNP? A Macroeconomic Perspective. Federal Reserve Bank of Minneapolis Staff Report, 244.

Leybourne, S., P. Newbold, & D. Vougas (1998) Unit roots and smooth transitions. *Journal of Time Series Analysis* 19, 83–97.

Loewy, M.B. & D.H. Papell (1996) Are U.S. regional incomes converging? Some further evidence. *Journal of Monetary Economics* 38, 587–598.

Lucas, R.E., Jr. (1972) Expectations and the neutrality of money. *Journal of Economic Theory* 4, 103–123.

Lucas, R.E., Jr. & L.A. Rapping (1969) Real wages, employment and inflation. *Journal of Political Economy* 77, 721–754.

Lumsdaine, R.L. & D.H. Papell (1997) Multiple trend breaks and the unit root hypothesis. *Review of Economics and Statistics* 79, 212–218.

Maddala, G.S. & I.-M. Kim (1996) Structural change and unit roots. *Journal of Statistical Planning and Inference* 49, 73–103.

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

Montañes, A. & M. Reyes (1998) Effect of a shift in the trend function on Dickey–Fuller unit-root tests. *Econometric Theory* 14, 355–363.

Murray, C.J. & C.R. Nelson (2000) The uncertain trend in U.S. GDP. *Journal of Monetary Economics* 46, 79–95.

Nelson, C.R. & C.I. Plosser (1982) Trends and random walks in macroeconomic time series: Some evidence and implications. *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.

Nuñes, L.C., P. Newbold, & C.-M. Kuan (1997) Testing for unit roots with breaks: Evidence on the Great Crash and the unit root hypothesis reconsidered. *Oxford Bulletin of Economics and Statistics* 59, 435–448.

Ohanian, L.E. (1997) The macroeconomic effects of war finance in the United States: World War II and the Korean War. *American Economic Review* 87, 23–40.

Papell, D., C. Murray, & H. Ghiblawi (2000) The Structure of unemployment. *Review of Economics and Statistics* 82, 309–315.

Park, J.Y. & J. Sung (1994) Testing for unit roots in models with structural change. *Econometric Theory* 10, 917–936.

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

Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics* 8, 153–162.

Perron, P. (1991) A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series. Manuscript, Princeton University.

Perron, P. (1992) Trend, unit root, and structural change: A multi-country study with historical data. *Proceedings of the Business and Economics Statistics Section, American Statistical Association*, pp. 144–149.

Perron, P. (1997) Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355–386.

Perron, P. & T.J. Vogelsang (1992a) 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 (1992b) Testing for a unit root in a time series with a changing mean: Corrections and extensions. *Journal of Business and Economic Statistics* 10, 467–470.

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* 13, 181–207.

Perron, P. & T.J. Vogelsang (1993b) The Great Crash, the oil price shock, and the unit root hypothesis: Erratum. *Econometrica* 61, 248–249.

Perron, P. & T.J. Vogelsang (1995) Additional Tests for a Unit Root Allowing for a Break in the Trend Function at Unknown Time. Manuscript, Cornell University.

Pischke, J.-S. (1991) Measuring persistence in the presence of trend breaks. *Economics Letters* 36, 379–384.

Raj, B. (1992) International evidence on persistence in output in the presence of an episodic change. *Journal of Applied Econometrics* 7, 281–293.

Raj, B. & D.J. Slottje (1994) The trend behavior of alternative income inequality measures in the United States from 1947–1990 and the structural break. *Journal of Business and Economic Statistics* 12, 479–488.

Rappoport, P. & L. Reichlin (1989) Segmented trends and non-stationary time series. *Economic Journal* 99, 168–177.

Romer, C.D. (1989) The prewar business cycle reconsidered: New estimates of gross national product, 1869–1908. *Journal of Political Economy* 97, 1–37.

Sadorsky, P. (1994) The behavior of U.S. tariff rates: Comment. *American Economic Review* 84, 1097–1103.

Schwert, G.W. (1987) Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics* 20, 73–103.

Schwert, G.W. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–159.

Serletis, A. (1995) Random-walks, breaking trend functions, and the chaotic structure of the velocity of money. *Journal of Business and Economic Statistics* 13, 453–458.

Stock, J.H. (1994) Deciding between I(1) and I(0). *Journal of Econometrics* 63, 105–131.

Trivedi, P.K. (1995) Tests of some hypotheses about the time series behavior of commodity prices. In G.S Maddala list all (eds.), *Advances in Econometrics and Quantitative Economics*, pp. 382–412. Oxford, UK: Basil Blackwell.

Vogelsang, T.J. (1997) Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory* 13, 818–849.
Vogelsang, T.J. & P. Perron (1998) Additional tests for a unit root allowing for a break in the trend function at an unknown time. *International Economic Review* 39, 1073–1100.
Zelhorst, D. & J. De Haan (1995) Testing for a break in output: New international evidence. *Oxford Economic Papers* 47, 357–362.
Zivot, E. & D.W.K. Andrews (1992) Further evidence on the Great Crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics* 10, 251–270.