Cowles Foundation Paper 811

Journal of Business & Economic Statistics, July 1992, Vol. 10, No. 3

Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis

Eric Zivot and Donald W. K. Andrews
Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520

Recently, Perron has carried out tests of the unit-root hypothesis against the alternative hypothesis of trend stationarity with a break in the trend occurring at the Great Crash of 1929 or at the 1973 oil-price shock. His analysis covers the Nelson–Plosser macroeconomic data series as well as a postwar quarterly real gross national product (GNP) series. His tests reject the unit-root null hypothesis for most of the series. This article takes issue with the assumption used by Perron that the Great Crash and the oil-price shock can be treated as exogenous events. A variation of Perron's test is considered in which the breakpoint is estimated rather than fixed. We argue that this test is more appropriate than Perron's because it circumvents the problem of data-mining. The asymptotic distribution of the *estimated breakpoint* test statistic is determined. The data series considered by Perron are reanalyzed using this test statistic. The empirical results make use of the asymptotics developed for the test statistic as well as extensive finite-sample corrections obtained by simulation. The effect on the empirical results of fat-tailed and temporally dependent innovations is investigated. In brief, by treating the breakpoint as endogenous, we find that there is less evidence against the unit-root hypothesis than Perron finds for many of the data series but stronger evidence against it for several of the series, including the Nelson–Plosser industrial-production, nominal-GNP, and real-GNP series.

KEY WORDS: Breakpoint; Gaussian process; Macroeconomic time series; Structural change; Trend stationary; Weak convergence.

1. INTRODUCTION

A major debate concerning the dynamic properties of macroeconomic and financial time series has been going on since Nelson and Plosser (1982) published their stimulating article in the *Journal of Monetary Economics* a decade ago. The primary issue involves the long-run response of a trending data series to a current shock to the series. The traditional view holds that current shocks only have a temporary effect and that the long-run movement in the series is unaltered by such shocks. Nelson and Plosser challenged this view and argued, using statistical techniques developed by Dickey and Fuller (1979, 1981), that current shocks have a permanent effect on the long-run level of most macroeconomic and financial aggregates. Others, including Campbell and Mankiw (1987, 1988), Clark (1987), Cochrane (1988), Shapiro and Watson (1988), and Christiano and Eichenbaum (1989), have argued that current shocks are a combination of temporary and permanent shocks and that the long-run response of a series to a current shock depends on the relative importance or "size" of the two types of shocks.

Recent research has cast some doubt on Nelson and Plosser's conclusions. In particular, Perron (1988, 1989) argued that if the years of the Great Depression are treated as points of structural change in the economy and the observations corresponding to these years are removed from the noise functions of the Nelson and Plosser data, then a "flexible" trend-stationary representation is favored by 11 of the 14 series. Similarly, Perron showed that if the first oil crisis in 1973 is treated as a point of structural change in the economy, then one can reject the unit-root hypothesis in favor of a trend-stationary hypothesis for postwar quarterly real gross national product (GNP). These results imply that the only observations (shocks) that have had a permanent effect on the long-run level of most macroeconomic aggregates are those associated with the Great Depression and the first oil-price crisis.

We enter this debate by taking issue with the unit-root testing procedure used by Perron (1989) (hereafter referred to as Perron). In particular, we examine the sensitivity of Perron's results to his exogeneity assumption concerning the Great Depression and the 1973 oil crisis. A skeptic of Perron's approach would argue that his choices of breakpoints are based on prior observation of the data and hence problems associated with "pre-testing" are applicable to his methodology. Simple visual inspection of the Nelson and Plosser data shows that there is an obvious jump down for most of the series occurring in 1929. Due to the sudden change in the data at 1929, Perron chooses to treat the drop in the Nelson and Plosser series as an exogenous event. This jump, however, could be interpreted as a realization from the tail of the distribution of the underlying data-generating process. This interpretation views the Great Depression as a shock or a combination of shocks from the underlying errors.

Similarly, an examination of the postwar quarterly GNP data shows a slowdown in GNP growth after the oil crisis in 1973. Analogous to his treatment of the Nelson and Plosser data, Perron's statistical model handles the slowdown in growth after the 1973 oil crisis as an event external to the domestic economy. Although it seems reasonable to regard the formation of the Organization of Petroleum Exporting Countries as an exogenous event, there are other big events such as the 1964 tax cut, the Vietnam War, and the financial deregulation in the 1980s that could also be viewed ex ante as possible exogenous structural breakpoints. Perron's preference for the 1973 oil-price crisis is undoubtedly influenced by his prior examination of the data.

If one takes the view that these events are endogenous, then the correct unit-root testing procedure would have to account for the fact that the breakpoints in Perron's regressions are data dependent. The null hypothesis of interest in these cases is a unit-root process with drift that excludes any structural change. The relevant alternative hypothesis is still a trend-stationary process that allows for a one-time break in the trend function. Under the alternative, however, we assume that we do not know exactly when the breakpoint occurs. Instead, a data-dependent algorithm is used to proxy Perron's subjective procedure to determine the breakpoints. Such a procedure transforms Perron's unit-root test, which is conditional on a known breakpoint, into an unconditional unit-root test.

We develop a unit-root testing procedure that allows for an estimated break in the trend function under the alternative hypothesis. Using our procedure on the data series analyzed by Perron, we find less conclusive evidence against the unit-root hypothesis than he found. In particular, using our asymptotic critical values, we cannot reject the unit-root hypothesis at the 5% level for 4 of the 10 Nelson and Plosser series for which Perron rejected the hypothesis—namely, real per capita GNP, GNP deflator, money stock, and real wages. We still reject the unit-root hypothesis, however, for 6 of the series. Furthermore, contrary to Perron, we cannot reject the unit-root null at the 5% or 10% level for the postwar quarterly real GNP series.

We also investigate the accuracy of our asymptotic approximations by computing the exact finite-sample distributions of our test statistics for the two data sets by Monte Carlo methods, assuming normal autoregressive moving average (ARMA) innovations. Here we find that our asymptotic critical values are more liberal than the finite-sample critical values. Using the finite-sample critical values, we cannot reject the unit-root hypothesis at the 5% level for three more of the series that Perron rejected—namely, employment, nominal wages, and common-stock prices (although the latter two are very close to being rejected at the 5% level). We can, however, still reject the unit-root null at the 5% level for the real-GNP and nominal-GNP series, and we can reject the unit-root null at the 1% level for the industrial-production series.

For the series for which we reject the unit-root null using our finite-sample critical values, we investigate the possibility that the distributions of the innovations driving these series have tails thicker than the normal distribution. Our estimates of the kurtosis of these series lead us to believe that Student-t innovations may be more appropriate than normal innovations for some of these series. We recompute the finite-sample distributions using Student-t ARMA innovations, with degrees of freedom determined by equating sample kurtosis values to theoretical kurtosis values. Although the percentage points of the finite-sample distributions using the t innovations are uniformly larger (in absolute value) than the corresponding percentage points assuming normality, our unit-root testing conclusions remain the same as in the normal case. Thus our finite-sample results for these series are robust to some relaxations of the normality assumption.

Last, we consider the effects of relaxing the assumption of finite variance by computing the finite-sample distributions of our test statistics using stable ARMA innovations. Our conclusion is that it would take only slightly more than infinite variance for us not to reject the unit-root hypothesis for all of the series. On the other hand, the estimates of kurtosis do not indicate that the series have infinite-variance innovations.

The approach of this article is similar to that taken by Christiano (1992). His results, however, were based solely on bootstrap methods. The latter have questionable reliability in regression models with dependent errors and small sample sizes. Christiano also limited his analysis to the postwar quarterly real-GNP series.

The asymptotic distribution theory developed here is quite similar to that of Banerjee, Lumsdaine, and Stock (1992), although our empirical applications are substantially different. Our asymptotic theory was developed simultaneously and independently of the theory presented by Banerjee et al.

The empirical results of this article are quite similar in many respects to some of those given by Perron (1990) subsequent to the article under discussion. (In particular, Perron's [1990] results given in his column labeled "p-value (t sig)" in tables VI and VIII correspond closely to ours except for the wage and real-wage series. The differences between his results and ours that arise for the latter series are due to his inclusion of a one-time dummy at the time of structural change in the estimated regression model, which we do not include.) Perron (1990) looked at several different data-dependent methods of determining k, whereas we consider only the method used by Perron (1989).

Other works in the literature that are related to this one include those of Rappoport and Reichlin (1989), Rappoport (1990), and Banerjee, Dolado, and Galbraith (1990).

The outline of this article is as follows. Section 2 reviews Perron's unit-root testing methodology and presents our testing strategy. Section 3 contains the requisite asymptotic distribution theory for our unit-root test in time series models with estimated structural breaks. We derive the asymptotic distributions for the test statistics, tabulate their critical values, and compare the latter to the critical values used by Perron. Section 4 applies our results to the Nelson and Plosser data and the postwar quarterly real GNP data. Section 5 investigates the finite-sample distributions of the test statistics by Monte Carlo methods. This section determines the difference in test size between the finite-sample distributions and the asymptotic distributions and determines the effect of fat-tailed innovations on the finite-sample distributions of our test statistics. Section 6 contains our concluding remarks.

2. MODELS AND METHODOLOGY

Perron developed a procedure for testing the null hypothesis that a given series $\{y_t\}_1^T$ has a unit root with drift and that an exogenous structural break occurs at time $1 < T_B < T$ versus the alternative hypothesis that the series is stationary about a deterministic time trend with an exogenous change in the trend function at time T_B. He considered three parameterizations of the structural break under the null and the alternative. Following the notation of Perron, the unit-root null hypotheses are

$$\text{Model (A):}\quad y_t = \mu + dD(T_B)_t + y_{t-1} + e_t,$$

$$\text{Model (B):}\quad y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t,$$

and

$$\text{Model (C):}\quad y_t = \mu_1 + y_{t-1} + dD(T_B)_t + (\mu_2 - \mu_1)DU_t + e_t,$$

where $D(T_B)_t = 1$ if $t = T_B + 1$, 0 otherwise; $DU_t = 1$ if $t > T_B$, 0 otherwise; $A(L)e_t = B(L)v_t$, $v_t \equiv \text{iid}(0, \sigma^2)$, with $A(L)$ and $B(L)$ pth and qth order polynomials in the lag operator. Model (A) permits an exogenous change in the level of the series, Model (B) allows an exogenous change in the rate of growth, and Model (C) admits both changes.

The trend-stationary alternative hypotheses considered are

$$\text{Model (A):}\quad y_t = \mu_1 + \beta t + (\mu_2 - \mu_1)DU_t + e_t,$$

$$\text{Model (B):}\quad y_t = \mu + \beta_1 t + (\beta_2 - \beta_1)DT_t^* + e_t,$$

and

$$\text{Model (C):}\quad y_t = \mu + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t^* + e_t,$$

where $DT_t^* = t - T_B$ if $t > T_B$ and 0 otherwise. As with the unit-root hypotheses, Model (A) allows for a one-time change in the level of the series, and, appropriately, Perron called this the "crash" model. The difference $\mu_2 - \mu_1$ represents the magnitude of the change in the intercept of the trend function occurring at time T_B. Perron labeled Model (B) the "changing growth" model, and the difference $\beta_2 - \beta_1$ represents the magnitude of the change in the slope of the trend function occurring at time T_B. Model (C) combines changes in the level and the slope of the trend function of the series.

Perron proposed Model (A) (the crash model) for all of the Nelson and Plosser series except the real-wage and common-stock-price series, for which he suggested Model (C). He submitted Model (B) as the representation for the postwar quarterly real-GNP series. His arguments for these representations were based primarily on visual inspection of the data.

Perron employed an adjusted Dickey–Fuller (ADF) type unit-root testing strategy (see Dickey and Fuller 1981; Said and Dickey 1984). His test for a unit root in Models (A), (B), and (C) involve the following augmented regression equations:

$$y_t = \hat{\mu}^A + \hat{\theta}^A DU_t + \hat{\beta}^A t + \hat{d}^A D(T_B)_t + \hat{\alpha}^A y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^A \Delta y_{t-j} + \hat{e}_t, \quad (1)$$

$$y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT_t^* + \hat{\alpha}^B y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^B \Delta y_{t-j} + \hat{e}_t, \quad (2)$$

and

$$y_t = \hat{\mu}^C + \hat{\theta}^C DU_t + \hat{\beta}^C t + \hat{\gamma}^C DT_t^* + \hat{d}^C D(T_B)_t + \hat{\alpha}^C y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^C \Delta y_{t-j} + \hat{e}_t. \quad (3)$$

The k extra regressors in the preceding regressions are added to eliminate possible nuisance-parameter dependencies in the limit distributions of the test statistics caused by temporal dependence in the disturbances. The number k of extra regressors is determined by a test of the significance of the estimated coefficients $\hat{c}_j^i$ (i = A, B, C) (as will be described).

For Model (B), we use the *intervention outlier model* instead of the two-step *additive outlier model* used by Perron. This latter regression is of the form

$$\tilde{y}_t^B = \hat{\alpha}^B \tilde{y}_{t-1}^B + \sum_{j=1}^{k} \hat{c}_j^B \Delta \tilde{y}_{t-j} + \hat{e}_t, \quad (2a)$$

where $\{\tilde{y}_t^B\}$ are the residuals from a regression of y_t on a constant, time trend, and DT_t^*. Perron and Vogelsang (1991) showed that the asymptotic distribution of $\hat{\alpha}^B$ from (2) (when T_B is known) is given in theorem 2 of

Perron and that this distribution is different from the asymptotic distribution of $\hat{\alpha}^B$ from (2a). We note, however, that the two models give essentially the same empirical results for the postwar quarterly real-GNP series and that the finite-sample distributions of $\hat{\alpha}^B$ from the two models for this series are very close.

To formally test for the presence of a unit root, Perron considered the following statistics computed from (1)–(3):

$$t_{\hat{\alpha}^i}(\lambda), \qquad i = \text{A, B, C}, \tag{4}$$

which represents the standard t statistic for testing $\alpha^i = 1$. These statistics depend on the location of the break fraction (or breakpoint) $\lambda = T_B/T$, and we exhibit this dependence explicitly because this notation will be useful for the analysis that follows. Perron's test for a unit root using (4) can be viewed as follows: Reject the null hypothesis of a unit root if

$$t_{\hat{\alpha}^i}(\lambda) < \kappa_\alpha(\lambda), \tag{5}$$

where $\kappa_\alpha(\lambda)$ denotes the size α critical value from the asymptotic distribution of (4) for a fixed $\lambda = T_B/T$. Perron derived the asymptotic distributions for these statistics under the preceding null hypotheses and tabulated their critical values for a selected grid of λ values in the unit interval. Based on the critical values for (4), he rejected the unit-root hypothesis at the 5% level of significance for all of the Nelson and Plosser data series except consumer prices, velocity, and interest rates. He also rejected the unit-root hypothesis at the 5% level for the postwar quarterly real GNP series.

We construe Perron's test statistic (4) in a different manner. Perron's null hypotheses take the break fraction λ to be exogenous. We question this exogeneity assumption and instead treat the structural break as an endogenous occurrence; that is, we do not remove the Great Crash and the 1973 oil-price shock from the noise functions of the appropriate series. Our null hypothesis for the three models is

$$y_t = \mu + y_{t-1} + e_t, \tag{6}$$

Since we consider the null that the series $\{y_t\}$ is integrated without an exogenous structural break, we view the selection of the breakpoint, λ, for the dummy variables in Perron's regressions (1)–(3) as the outcome of an estimation procedure designed to fit $\{y_t\}$ to a certain trend-stationary representation; that is, we assume that the alternative hypothesis stipulates that $\{y_t\}$ can be represented by a trend-stationary process with a one-time break in the trend occurring at an unknown point in time. The goal is to estimate the breakpoint that gives the most weight to the trend-stationary alternative. Our hope is that an explicit algorithm for selecting the breakpoints for the series will be consistent with Perron's (subjective) selection procedure. (Note that several recent works in the econometric literature consider the problem of testing for *structural change* with unknown changepoint; see Ploberger, Krämer, and Kontrus [1989], Andrews [1989], Chu [1989], and Hansen [1992]. The problem considered here differs from that considered in the aforementioned works. It is one of testing for a *unit root* against the alternative of stationarity with structural change at some unknown point.)

One plausible estimation scheme, consistent with the preceding view, is to choose the breakpoint that gives the least favorable result for the null hypothesis (6) using the test statistic (4); that is, λ is chosen to minimize the one-sided t statistic for testing $\alpha^i = 1$ (i = A, B, C), when small values of the statistic lead to rejection of the null. Let $\hat{\lambda}^i_{\text{inf}}$ denote such a minimizing value for model i. Then, by definition,

$$t_{\hat{\alpha}^i}[\hat{\lambda}^i_{\text{inf}}] = \inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda), \qquad i = \text{A, B, C}, \tag{7}$$

where Λ is a specified closed subset of (0, 1).

With the null model defined by (6) we no longer need the dummy variable $D(T_B)_t$ in (1) and (3). Therefore, following Perron's ADF testing strategy, the regression equations we use to test for a unit root are

$$y_t = \hat{\mu}^A + \hat{\theta}^A DU_t(\hat{\lambda}) + \hat{\beta}^A t + \hat{\alpha}^A y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^A \Delta y_{t-j} + \hat{e}_t, \tag{1$'$}$$

$$y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT_t^*(\hat{\lambda}) + \hat{\alpha}^B y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^B \Delta y_{t-j} + \hat{e}_t, \tag{2$'$}$$

and

$$y_t = \hat{\mu}^C + \hat{\theta}^C DU_t(\hat{\lambda}) + \hat{\beta}^C t + \hat{\gamma}^C DT_t^*(\hat{\lambda}) + \hat{\alpha}^C y_{t-1} + \sum_{j=1}^{k} \hat{c}_j^C \Delta y_{t-j} + \hat{e}_t, \tag{3$'$}$$

where $DU_t(\lambda) = 1$ if $t > T\lambda$, 0 otherwise; $DT_t^*(\lambda) = t - T\lambda$ if $t > T\lambda$, 0 otherwise. We put "hats" on the λ parameters in (1$'$)–(3$'$) to emphasize that they correspond to estimated values of the break fraction. [Except for one series, the effect of excluding $D(T_B)_t$ from (1$'$) and (3$'$) is to increase in absolute value the magnitude of the t statistic for testing $\alpha^i = 1$. The actual changes in the t statistics for the series with estimated breakpoints equal to Perron's breakpoints are −.55 (real GNP), −.40 (nominal GNP), −.52 (real per capita GNP), −.48 (industrial production), −.44 (employment), −.08 (GNP deflator), .11 (nominal wages), and −.05 (money stock). For the series with estimated breakpoints different from Perron's choices, the changes in the t statistics are −1.48 (consumer prices), −1.73 (velocity), −.53 (interest rates), −.01 (quarterly real GNP), −.74 (common-stock prices), and −.46 (real wages).

Table 1 reports the values of $\hat{T}_B(= [T\hat{\lambda}])$ that correspond to $\hat{\lambda}^i_{\text{inf}}$ and the minimum values of $t_{\hat{\alpha}^i}(\lambda)$ obtained from the procedure defined in (7) for the data series analyzed by Perron. The breakpoints and minimum t statistics were determined as follows. For each

Table 1. Minimum t Statistics

Series^i (i = A, B, C)	Rank 1 t stat	Year	2 t stat	Year	3 t stat	Year
Real GNP^A	−5.58***	1929	−4.34	1928	−3.89	1927
Nominal GNP^A	−5.82***	1929	−4.36	1927	−4.23	1928
Real per capita GNP^A	−4.61***	1929	−4.29	1928	−4.09	1927
Industrial production^A	−5.95***	1929	−5.40	1928	−5.09	1927
Employment^A	−4.95***	1929	−4.71	1928	−4.38	1927
GNP deflator^A	−4.12**	1929	−3.90	1928	−3.82	1930
Consumer prices^A	−2.76	1873	−2.69	1872	−2.64	1864
Nominal wages^A	−5.30***	1929	−5.10	1930	−4.61	1928
Money stock^A	−4.34***	1929	−4.34	1928	−4.32	1930
Velocity^A	−3.39	1949	−3.35	1947	−3.21	1946
Interest rate^A	−.98	1932	−.96	1965	−.91	1967
Quarterly real GNP^B	−4.08**	1972:II	−4.07	1972:III	−4.07	1972:I
Common-stock prices^C	−5.61***	1936	−5.60	1937	−5.52	1939
Real wages^C	−4.74***	1940	−4.67	1941	−4.59	1931

NOTE: The minimum t statistics were determined as follows. For each series, Equation (1′), (2′), or (3′) was estimated with the breakpoint, T_B, ranging from t = 2 to t = T-1. For each regression, k was determined as described in the paragraph following Equation (3′), and the t statistic for testing $\alpha^i = 1$ was computed. The minimum t statistic reported is the minimum over all T-2 regressions. The symbols *, **, and *** indicate that the unit-root hypothesis is rejected at the 10%, 5%, and 1% levels, respectively, using Perron's critical values.

series, (1′), (2′), or (3′) was estimated by ordinary least squares with the break fraction, $\lambda = T_B/T$, ranging from $j = 2/T$ to $j = (T - 1)/T$. (This range corresponds to our choice of $\Lambda = [.001, .999]$. In fact, the results are not sensitive to this particular choice of Λ.) For each value of λ, the number of extra regressors, k, was determined using the same procedure as that of Perron, and the t statistic for testing $\alpha^i = 1$ was computed. The minimum t statistics reported are the minimums over all $T - 2$ regressions, and the break years are the years corresponding to the minimum t statistics. (It is important to note that the number of extra regressors, k, required for the ADF regressions was allowed to vary *for each* tentative choice of λ. We determined k using the same selection procedure as that used by Perron; that is, working backward from $k = \bar{k}$, we chose the first value of k such that the t statistic on $\hat{c}_k$ was greater than 1.6 in absolute value and the t statistic on $\hat{c}_\ell$ for $\ell > k$ was less than 1.6. For the Nelson and Plosser series, we set $\bar{k} = 8$, and for the postwar quarterly real GNP series, we set $\bar{k} = 12$. These are the same values of $\bar{k}$ used by Perron, although a typographical error in his article erroneously indicates that he used $\bar{k} = 12$ for the Nelson and Plosser series.)

From Table 1 we see that the break year that minimizes the one-sided t statistic for testing $\alpha^A = 1$ does, in fact, correspond to the year of the Great Depression, 1929, for the eight series that Perron rejected the unit-root hypothesis. The three series with estimated breakpoints not consistent with Perron's choice are consumer prices, velocity, and the interest rate. These are also the series for which Perron does not reject the unit-root hypothesis. The break years for these series are 1873, 1949, and 1932, respectively. The estimated break date for the velocity series corresponds to the widely noted leveling off of the series in the mid to late 1940s.

For the postwar quarterly real GNP series, the minimizing breakpoint occurs in the second quarter of 1972. Perron's choice of 1973:I produces the fifth smallest t statistic. The numerical difference between the t statistics for these two dates, however, is very small. The break years corresponding to the minimum t statistics for the Model (C) series do not coincide with the year of the Depression. The estimated break year for the common-stock price series is 1936 and the break year for the real-wage series is 1940. As these results show, our breakpoint algorithm is generally, though not completely, consistent with the subjective selection procedure used by Perron for the Nelson and Plosser series and the postwar quarterly real GNP series.

When we treat the selection of λ as the outcome of an estimation procedure, we can no longer use Perron's critical values to test the unit-root hypothesis. To see this, consider the minimum t statistic breakpoint estimation procedure. With this definition of the break fraction, our interpretation of Perron's unit-root test becomes the following: Reject the null of a unit root if

$$\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda) < \kappa^i_{\inf,\alpha}, \qquad i = \text{A, B, C}, \tag{8}$$

where $\kappa^i_{\inf,\alpha}$ denotes the size α left-tail critical value from the asymptotic distribution of $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda)$. By definition, the left-tail critical values in (8) are at least as large in absolute value as those computed for an arbitrary fixed λ. If one takes this unconditional perspective, then Perron's unit-root tests are biased toward rejecting the unit-root null hypothesis because he uses critical values that are too small (in absolute value). The extent of this size distortion depends on the magnitude of the difference between the critical values defined in (8) and those defined in (5). To determine this difference, the asymptotic distributions of the test statistics $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda)$ (i = A, B, C) are required. These distributions are derived in Section 3.

Table 2. Percentage Points of the Asymptotic Distribution of $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}A}(\lambda)$ *and* $t_{\hat{\alpha}A}(\lambda)$ *for a Fixed* λ

λ	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
					A. $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}A}(\lambda)$				
	−5.34	−5.02	−4.80	−4.58	−3.75	−2.99	−2.77	−2.56	−2.32
					B. $t_{\hat{\alpha}A}(\lambda)$ for a fixed λ				
.1	−4.30	−3.93	−3.68	−3.40	−2.35	−1.38	−1.09	−.78	−.46
.2	−4.39	−4.08	−3.77	−3.47	−2.45	−1.45	−1.14	−.90	−.54
.3	−4.39	−4.03	−3.76	−3.46	−2.42	−1.43	−1.13	−.83	−.51
.4	−4.34	−4.01	−3.72	−3.44	−2.40	−1.26	−0.88	−.55	−.21
.5	−4.32	−4.01	−3.76	−3.46	−2.37	−1.17	−0.79	−.49	−.15
.6	−4.45	−4.09	−3.76	−3.47	−2.38	−1.28	−0.92	−.60	−.26
.7	−4.42	−4.07	−3.80	−3.51	−2.45	−1.42	−1.10	−.82	−.50
.8	−4.33	−3.99	−3.75	−3.46	−2.43	−1.46	−1.13	−.89	−.57
.9	−4.27	−3.97	−3.69	−3.38	−2.39	−1.37	−1.04	−.74	−.47

NOTE: λ = time of break relative to total sample size. Percentage points are based on 5,000 repetitions.

3. ASYMPTOTIC DISTRIBUTION THEORY

The asymptotic distributions of the minimum t statistics may be compactly expressed in terms of standardized Brownian motions. Following Phillips (1988), Park and Phillips (1988), and Ouliaris, Park, and Phillips (1989), define $W^i(\lambda, r)$ to be the stochastic process on $[0, 1]$ that is the projection residual in $L_2[0, 1]$ of a Brownian motion projected onto the subspace generated by the following: (a) i = A: $1, r, du(\lambda, r)$; (b) i = B: $1, r, dt^*(\lambda, r)$; (c) i = C: $1, r, du(\lambda, r), dt^*(\lambda, r)$; where $du(\lambda, r) = 1$ if $r > \lambda$ and 0 otherwise and $dt^*(\lambda, r) = r - \lambda$ if $r > \lambda$ and 0 otherwise. Here, $L_2[0, 1]$ denotes the Hilbert space of square integrable functions on $[0, 1]$ with inner product $(f, g) = \int_0^1 fg$ for $f, g \in L_2[0, 1]$. For example, in Model (A), $W^A(\lambda, r)$ is the L_2 projection residual from the continuous time regression

$$W(r) = \hat{\alpha}_0 + \hat{\alpha}_1 r + \hat{\alpha}_2 du(\lambda, r) + W^A(\lambda, r); \quad (9)$$

that is, $\hat{\alpha}_0$, $\hat{\alpha}_1$, and $\hat{\alpha}_2$ solve

$$\min_{\alpha_0, \alpha_1, \alpha_2} \int_0^1 |W(r) - \alpha_0 - \alpha_1 r - \alpha_2 du(\lambda, r)|^2 dr. \quad (10)$$

Notice that if we allow that $\lambda = 0$ or 1, the preceding minimization problem, and the minimization problems for Models (B) and (C), do not have unique solutions due to the singularity of the matrix defining the normal equations.

The following theorem gives the asymptotic distributions for the minimum t statistics in terms of $W^i(\lambda, r)$.

Theorem 1. Let $\{y_t\}$ be generated under the null hypothesis (6) and let the errors $\{e_t\}$ be iid, mean 0, variance σ^2 random variables with $0 < \sigma^2 < \infty$. Let $t_{\hat{\alpha}^i}(\lambda)$ denote the t statistic for testing $\alpha^i = 1$ computed from either (1′), (2′), or (3′) with $k = 0$ for Models i = A, B and C, respectively. Let Λ be a closed subset of (0, 1). Then,

$$\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda) \xrightarrow{d} \inf_{\lambda \in \Lambda} \left(\int_0^1 W^i(\lambda, r)^2 dr \right)^{-1/2} \times \left(\int_0^1 W^i(\lambda, r) dW(r) \right) \text{ as } T \to \infty$$

for i = A, B, and C, where $\xrightarrow{d}$ denotes convergence in distribution.

The proof is given in Appendix A.

Table 3. Percentage Points of the Asymptotic Distribution of $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}B}(\lambda)$ *and* $t_{\hat{\alpha}B}(\lambda)$ *for a Fixed* λ

λ	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
					A. $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}B}(\lambda)$				
	−4.93	−4.67	−4.42	−4.11	−3.23	−2.48	−2.31	−2.17	−1.97
					B. $t_{\hat{\alpha}B}(\lambda)$ for a fixed λ				
.1	−4.27	−3.94	−3.65	−3.36	−2.34	−1.35	−1.04	−.78	−.40
.2	−4.41	−4.08	−3.80	−3.49	−2.50	−1.48	−1.18	−.87	−.52
.3	−4.51	−4.17	−3.87	−3.58	−2.54	−1.59	−1.27	−.97	−.69
.4	−4.55	−4.20	−3.94	−3.66	−2.61	−1.69	−1.37	−1.11	−.75
.5	−4.55	−4.20	−3.96	−3.68	−2.70	−1.74	−1.40	−1.18	−.82
.6	−4.57	−4.20	−3.95	−3.66	−2.61	−1.71	−1.36	−1.11	−.78
.7	−4.51	−4.13	−3.85	−3.57	−2.55	−1.61	−1.28	−.97	−.67
.8	−4.38	−4.07	−3.82	−3.50	−2.47	−1.49	−1.16	−.87	−.54
.9	−4.26	−3.96	−3.68	−3.35	−2.33	−1.34	−1.04	−.77	−.43

NOTE: λ = time of break relative to total sample size. Percentage points are based on 5,000 repetitions.

Table 4. Percentage Points of the Asymptotic Distribution of $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}C}(\lambda)$ and $t_{\hat{\alpha}C}(\lambda)$ for a Fixed λ

λ	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
				A. $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}C}(\lambda)$					
	−5.57	−5.30	−5.08	−4.82	−3.98	−3.25	−3.06	−2.91	−2.72
				B. $t_{\hat{\alpha}C}(\lambda)$ for a fixed λ					
.1	−4.38	−4.01	−3.75	−3.45	−2.38	−1.44	−1.11	−.82	−.45
.2	−4.65	−4.32	−3.99	−3.66	−2.67	−1.60	−1.27	−.98	−.67
.3	−4.78	−4.46	−4.17	−3.87	−2.75	−1.78	−1.46	−1.15	−.81
.4	−4.81	−4.48	−4.22	−3.95	−2.88	−1.91	−1.62	−1.35	−1.04
.5	−4.90	−4.53	−4.24	−3.96	−2.91	−1.96	−1.69	−1.43	−1.07
.6	−4.88	−4.49	−4.24	−3.95	−2.87	−1.93	−1.63	−1.37	−1.08
.7	−4.75	−4.44	−4.18	−3.86	−2.77	−1.81	−1.47	−1.17	−.79
.8	−4.70	−4.31	−4.04	−3.69	−2.67	−1.63	−1.29	−1.04	−.64
.9	−4.41	−4.10	−3.80	−3.46	−2.41	−1.44	−1.12	−.80	−.50

NOTE: λ = time of break relative to total sample size. Percentage points are based on 5,000 repetitions.

The limiting distributions presented in Theorem 1 are for the case in which the disturbances are independent and there are no extra lag terms in the regression equations (1′)–(3′). If we allow the disturbances to be correlated and heterogeneously distributed, then the asymptotic distributions in the theorem become nonstandard in that they depend on the nuisance parameters $\sigma^2 = \lim_{T\to\infty} ET^{-1}(\Sigma_1^T e_t)^2$ and $\sigma_e^2 = \lim_{T\to\infty} ET^{-1}\Sigma_1^T e_t^2$.

Two approaches have been employed in the time series literature to eliminate this nuisance-parameter dependency. One approach is due to Phillips (1987). His technique is based on the result that if consistent estimators of σ^2 and σ_e^2 are available then one can derive a nonparametric transformation of the test statistics whose limiting distributions are independent of the population parameters σ^2 and σ_e^2. It should not be too difficult to extend Theorem 1 to incorporate Phillips's technique of handling serial correlation. The other approach is the ADF approach referred to previously. It is based on the addition of extra lags of first differences of the data as regressors. The number of extra regressors must increase with the sample size at a controlled rate. With the ADF procedure, the errors are restricted to the class of ARMA (p, q) processes. Since we follow Perron and use the ADF approach, we consider the following assumption.

Assumption 1. (a) $A(L)e_t = B(L)v_t$; $A(L)$ and $B(L)$ are pth and qth order polynomials in the lag operator L and satisfy the standard stationarity and invertibility conditions. (b) $\{v_t\}$ is a sequence of iid$(0, \sigma^2)$ random variables with $E|v_t|^{4+\delta} < \infty$ for some $\delta > 0$. (c) $k \xrightarrow{p} \infty$ and $T^{-1}k^3 \xrightarrow{p} 0$ as $T \to \infty$.

When the error sequence $\{e_t\}$ satisfies Assumption 1, we conjecture, based on arguments outlined by Said and Dickey (1984), that the limiting distributions of the test statistics computed from the ADF regression equations (1′)–(3′) are free of nuisance parameter dependencies and have the limiting distributions presented in the theorem. In fact, the proof of this conjecture is apt to be tricky and rather involved. For the case in which $q = 0$ and k $(\geq p)$ is fixed and independent of T in Assumption 1, the proof is clearly easier than when $k \xrightarrow{p} \infty$; see Banerjee et al. (1992) for a treatment of this case. We do not give a proof of the efficacy of the ADF procedure, but we use it in the following empirical applications.

Critical values for the limiting distributions in the theorem are obtained by simulation methods; that is, the integral functions in the theorem are approximated by functions of sums of partial sums of independent normal random variables. The method used is described in Appendix B.

The critical values for the limiting distributions of the minimum t statistics and for the t statistics used by Perron are presented in Tables 2–4. Estimates of their densities are plotted in Figure 1. As expected, for a given size of a left-tailed test, the critical values for $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda)$ are larger in absolute value (more negative) than the critical values obtained by Perron for any fixed value of the break fraction λ. The biggest difference occurs for the Model (A) densities. At the 5% level, the critical value for $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^A}(\lambda)$ is -4.80 and the average value, over λ, of Perron's critical values is -3.74. Thus, at the 5% level, our critical value is roughly 24% larger (in absolute value) than Perron's and at the 1% level our critical value is about 23% larger. For the Model (B) densities, our 5% critical value is -4.42 and Perron's average critical value is -3.84. For the Model (C) densities, our 5% critical value is -5.08 and Perron's average value is -4.07.

(The critical values for Perron's test statistics presented in the B panels of our Tables 2, 3, and 4 were generated from projection-residual approximations instead of the approximations used in Perron's theorem 2 to give more accurate comparisons with the critical values for our test statistics. The two techniques give approximately the same results and any difference can be attributed to simulation error.)

We can now address the magnitude of the size dis-

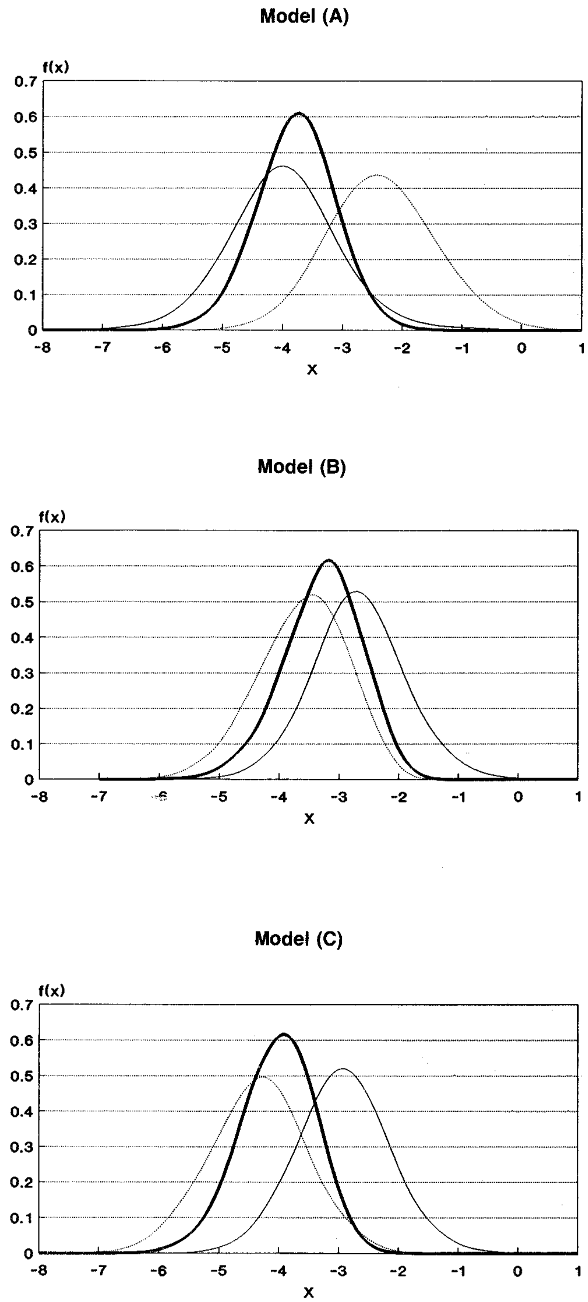

Figure 1. Density Plots: Model (A): ——, Finite Sample; ▬▬, Asymptotic; · · · , Perron; Models (B) and (C): · · · , Finite Sample; ▬▬, Asymptotic; ——, Perron.

tortion of Perron's test statistics incurred by ignoring the pretest information concerning the location of the trend break. Table 5 gives the actual asymptotic sizes of tests based on the statistic $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^i}(\lambda)$ that uses Perron's 5% critical values. We see that the size distortion is quite dramatic for Models (A) and (C), where the

Table 5. Size Distortions

Model	Critical value	Size
A	−3.68	.551
B	−3.96	.142
C	−4.24	.345

actual sizes of Perron's 5% tests are 55.1% and 34.5%, respectively. The size distortion for Model (B) is more moderate with an actual size of 14.2%. The density plots in Figure 1 clearly illustrate this distortion. For all models, the asymptotic densities of the minimum t statistics are shifted to the left of the Perron densities. The densities for the minimum t statistics also have thinner tails than the Perron densities.

4. EMPIRICAL APPLICATIONS

We now apply the unit-root test developed in the previous sections to the data series analyzed by Perron. (The Nelson and Plosser data were generously provided by Charles Nelson. The postwar quarterly real GNP series [GNP82] was extracted from the Citibase data bank.) We analyze the natural logarithm of all the data except for the interest-rate series, which is analyzed in levels form. Table 6, A—C, presents the estimated regressions for all of the series using the regression equations (1′)–(3′), t statistics are in parentheses. The t statistic for $\hat{\alpha}^i$ is for testing the hypothesis that $\alpha^i = 1$ (i = A, B, C).

These results are somewhat different from the results in Perron's table 7 for two reasons. First, the break years defining the dummy variables are estimated according to (7) instead of being fixed at 1929 or 1973:I. This has relevance only for the series whose estimated break years are different from the ones used by Perron. Second, we do not impose a structural break under our null hypothesis, and hence, the variable $D(T_B)_t$ is not included in the regressions. This affects only the Model (A) and Model (C) regressions. For this effect, the most notable change in the regression results is that the estimated t statistics for testing $\alpha^i = 1$ (i = A and C) increased (in absolute value) for a majority of the series, see the preceding paragraph that contains Equation (3′). Often this increase was substantial. For example, the absolute change in the t statistic for the real per capita GNP series is .52 ($-4.61 - (-4.09)$), which is roughly 13%. For most of the affected series, this change favors the trend stationary alternative.

The results of our unit-root tests are also presented graphically in Figure 2, which contains time plots of the natural logarithm of the 14 data series. Superimposed on the time plot of each series are the estimated t statistics (in absolute value) for testing $\alpha^i = 1$ for each possible break date $T_B = [T\lambda]$, a line indicating the

Table 6. Tests for a Unit Root

Series	T	$\hat{T}_B$	k		$\hat{\mu}^A$	$\hat{\theta}^A$	$\hat{\beta}^A$	$\hat{\alpha}^A$	$S(\hat{e})$
Model A: Regression: $y_t = \hat{\mu}^A + \hat{\theta}^A DU(\hat{\lambda})_t + \hat{\beta}^A t + \hat{\alpha}^A y_{t-1} + \Sigma_1^k \hat{c}_j^A \Delta y_{t-j} + \hat{e}_t$									
Real GNP	62	1929	8		3.514	−.195	.027	.267	.05
					(5.62)	(−4.92)	(5.71)	(−5.58)***	
Nominal GNP	62	1929	8		5.040	−.311	.032	.532	.07
					(5.85)	(−5.12)	(5.97)	(−5.82)***	
Real per capita GNP	62	1929	7		3.584	−.117	.012	.494	.056
					(4.62)	(−3.41)	(4.69)	(−4.61)*	
Industrial production	111	1929	8		.122	−.317	.034	.290	.088
					(4.46)	(−5.12)	(5.91)	(−5.95)***	
Employment	81	1929	7		3.564	−.051	.006	.651	.029
					(4.97)	(−3.14)	(4.79)	(−4.95)**	
GNP Deflator	82	1929	5		.641	−.091	.007	.786	.044
					(4.17)	(−3.23)	(4.14)	(−4.12)	
Consumer prices	111	1873	2		.217	−.055	.001	.941	.043
					(2.79)	(−2.51)	(3.27)	(−2.76)	
Nominal wages	71	1929	7		2.126	−.161	.017	.660	.054
					(5.35)	(−4.16)	(5.32)	(−5.30)**	
Money stock	82	1929	6		.288	−.064	.011	.823	.044
					(4.76)	(−2.54)	(4.25)	(−4.34)	
Velocity	102	1949	0		.224	.095	−.002	.840	.064
					(2.99)	(3.09)	(−2.95)	(−3.39)	
Interest rate	71	1932	2		.065	−.444	.013	.945	.272
					(.31)	(−2.55)	(3.09)	(−.98)	
Model B: Regression: $y_t = \hat{\mu}^B + \hat{\beta}^B t + \hat{\gamma}^B DT^*(\hat{\lambda})_t + \hat{\alpha}^B y_{t-1} + \Sigma_1^k \hat{c}_j^B \Delta y_{t-j} + \hat{e}_t$									
Quarterly real GNP	159	72:II	10		1.044	.001	−.0004	.851	.010
					(4.12)	(3.93)	(−2.86)	(−4.08)	
Model C: Regression: $y_t = \hat{\mu}^C + \hat{\theta}^C DU(\hat{\lambda})_t + \hat{\beta}^C t + \hat{\gamma}^C DT^*(\hat{\lambda})_t + \hat{\alpha}^C y_{t-1} + \Sigma_1^k \hat{c}_j^C \Delta y_{t-j} + \hat{e}_t$									
Common-stock prices	100	1936	1	.471	−.226	.007	.021	.642	.139
				(5.12)	(−3.25)	(4.83)	(4.80)	(−5.61)***	
Real wages	71	1940	8	2.678	.085	.012	.008	.115	.030
				(4.81)	(4.33)	(4.49)	(3.68)	(−4.74)	

NOTE: t statistics are in parentheses. The t statistic for $\hat{\alpha}^i$ is for testing $\alpha^i = 1$. k is determined as described in the paragraph following Equation (3′). The symbols *, **, and *** denote significance of the test of $\alpha^i = 1$ at the 10%, 5%, and 1% levels, respectively, using the critical values from Table 2A, 3A, or 4A.

appropriate asymptotic 5% critical value (in absolute value) for the minimum t statistic, and a line depicting the appropriate 5% critical value from Perron's asymptotic distributions for a fixed break date. A line labeled "Finite Sample 5% C.V.," which will be explained later, is also superimposed on the time plots.

Consider first the results for the Model (A) series, presented in Table 6, panel A. From Table 1, we know that Perron's break fraction for 8 of the 11 series corresponds to the break fraction associated with the minimum t statistic for testing $\alpha^A = 1$. This can also be seen graphically from Figure 2, where, clearly, the largest t statistic (in absolute value) for these series occurs at $T_B = 1929$. These 8 series are also the ones for which Perron rejects the unit-root null hypothesis at a 5% significance level using his critical values for a fixed breakpoint. Now, treating the break fraction as the outcome of the estimation procedure defined by (7) and using the critical values from Table 2, panel A, we can reject the unit-root null at the 1% level for the real-GNP, nominal-GNP, and industrial-production series. We can reject the unit-root null at the 2.5% level for the nominal-wage series, at the 5% level for the employment series, and at the 10% level for the real per capita GNP series. We *cannot* reject the unit-root null at the 5% or 10% level, however, for the GNP-deflator, consumer-prices, money-stock, velocity, and interest-rate series. In fact, the p values for these series, computed from the asymptotic distribution of $\inf_{\lambda \in \Lambda} t_{\hat{\alpha}^A}(\lambda)$, are .278, .951, .174, .737, and .999, respectively. Thus, by endogenizing the breakpoint selection procedure, we reverse Perron's test conclusions for the GNP deflator and nominal-money-stock series and weaken the evidence against the unit-root hypothesis for the remaining series.

Next, consider the results for the Model (B) series presented in Table 6, panel B. The estimated break date for the postwar quarterly real GNP series occurs three quarters prior to Perron's choice of 1973:I, so it seems reasonable to apply our methodology to this series. Using the critical values from Table 3, panel A, we find, contrary to Perron, that we cannot reject the unit-root null at the 5% level. The asymptotic p value for the t statistic is .105.

Finally, the results for the Model (C) series are given in Table 6, panel C. For these series the estimated break years do not coincide with Perron's choices. Nevertheless, using our estimated breakpoints for these series

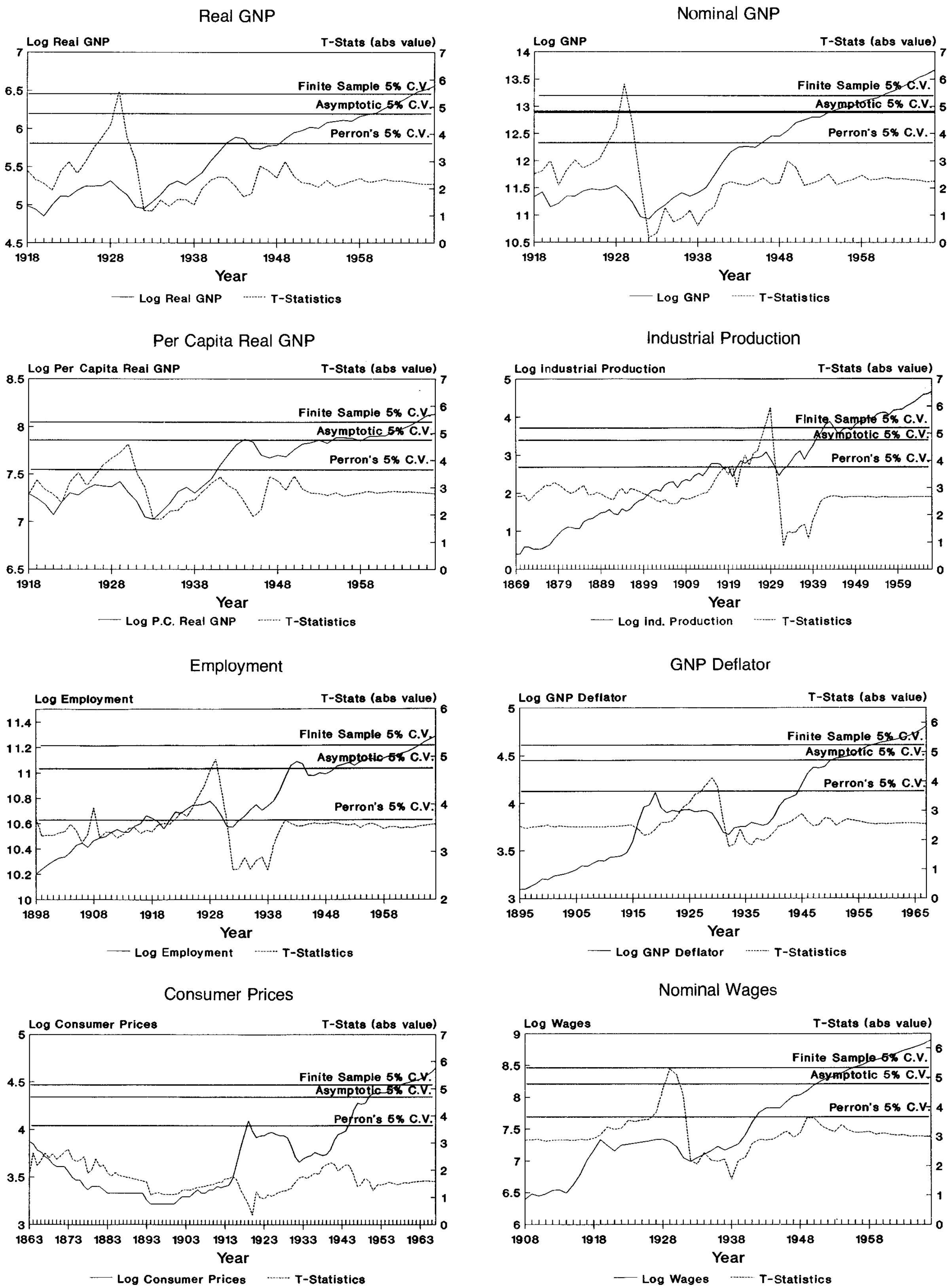

Figure 2. Time Plots of t Statistics.

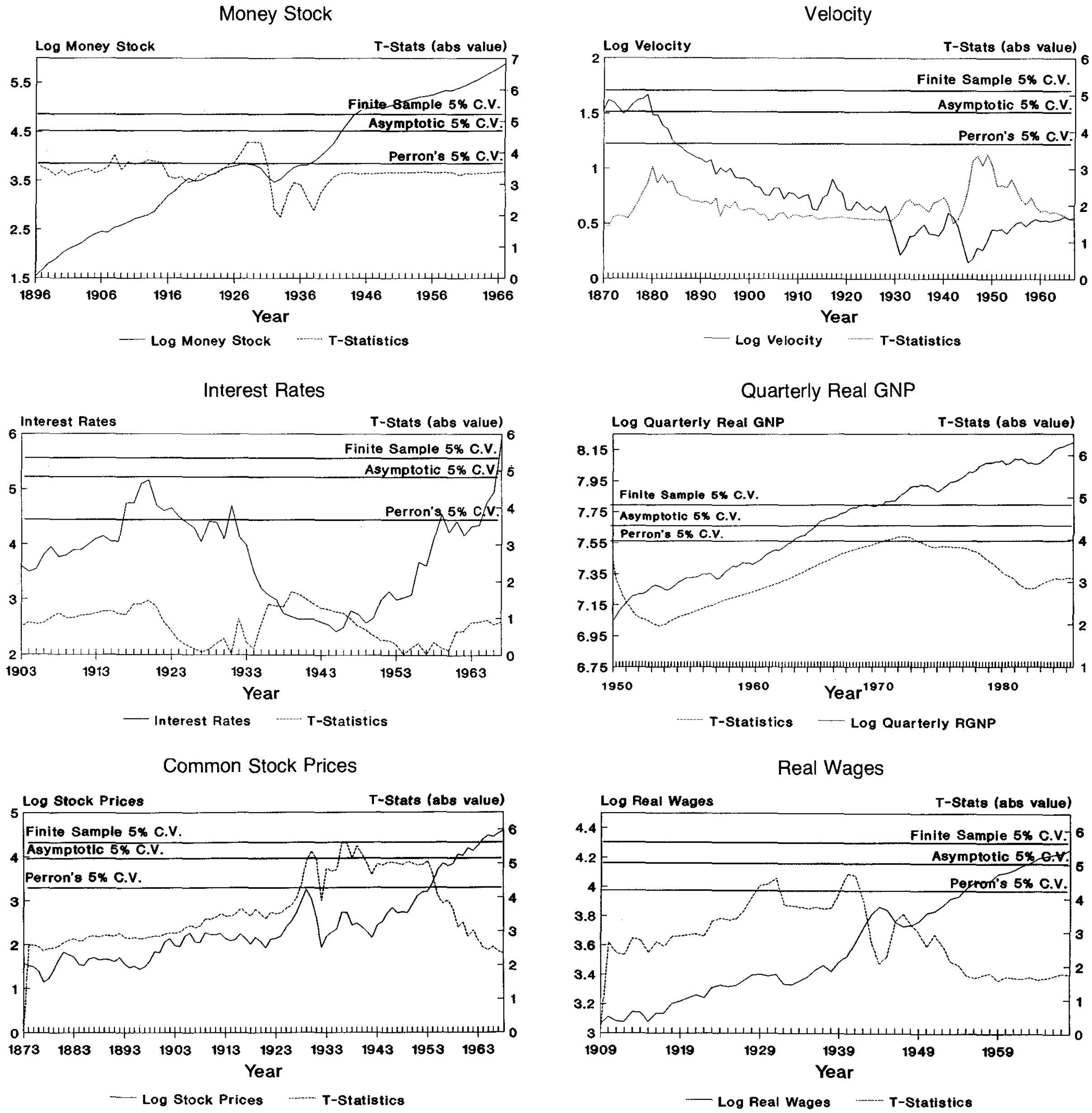

Figure 2. (continued).

and the critical values from Table 6, we reject the unit-root null for the common-stock price series at the 1% level, but, contrary to Perron, we cannot reject the unit-root null at the 1%, 5%, or 10% level for the real wage series. The asymptotic p value for the t statistic is .119.

Table 7 compares the p values computed from Perron's fixed-λ distributions to the p values computed from our asymptotic distributions, as well as p values from distributions that will be explained later. The table clearly shows the effects of incorporating the pretest trend-break information on the asymptotic distributions of the unit-root tests. In sum, by endogenizing Perron's breakpoint selection procedure, we reverse his conclusions for 5 of the 11 series for which he rejects the unit-root null hypothesis at 5% and for 4 of the 11 series for which he rejects at 10%. (Of course, our inability to reject the unit-root null hypothesis for these series should not be interpreted as an acceptance of the unit-root hypothesis.) On the other hand, even after adjusting for pretest examination of the data, we reject the unit-root null for 6 series using our 5% asymptotic *estimated breakpoint* critical values and for 7 series using 10% critical values.

Table 7. One Sided p Values for the Minimum t Statistics

Series	t stat	Perron's p value	Asymptotic p value	F.S.N. p value	F.S.T. p value
Real GNP	−5.58	.000***	.003***	.029**	.035** (9)
Nominal GNP	−5.82	.000***	.001***	.017**	.050* (4)
Real per capita GNP	−4.61	.003***	.091*	.216	
Industrial production	−5.95	.000***	.000***	.005***	.009*** (9)
Employment	−4.95	.001***	.031**	.101	.124 (10)
GNP deflator	−4.12	.017**	.278	.392	
Consumer prices	−2.76	.340	.951	.939	
Nominal wages	−5.30	.000***	.012**	.053*	.117 (5)
Money stock	−4.34	.008***	.174	.293	
Velocity	−3.39	.104	.737	.774	
Interest rate	−.98	.939	.999	.999	
Quarterly real GNP	−4.08	.027**	.105	.251	
Common stock prices	−5.61	.000***	.009***	.055*	.075* (6)
Real wages	−4.74	.005***	.119	.298	

NOTE: The symbols *, **, and *** denote rejection at the 10%, 5%, and 1% levels, respectively. The column labeled Perron's p value gives the p values computed from Perron's fixed λ distributions for the appropriate λ value, the column labeled F.S.N. gives the p values computed from the finite-sample distributions using normal innovations, and the column labeled F.S.T. gives the p values computed from the finite-sample distributions using Student-t innovations. The degrees of freedom for the t-distribution p values are in parentheses.

5. FINITE-SAMPLE RESULTS

The sample sizes for the series under consideration range from $T = 62$ to $T = 111$. In addition, there appears to be considerable temporal dependence in the data. In consequence, our asymptotic critical values may differ from the appropriate finite-sample critical values. In this section, we investigate this possibility by computing the finite-sample distributions of our test statistics, under specific distributional assumptions, by Monte Carlo methods.

To compute the finite-sample distributions of the minimum t statistics, one has to make specific assumptions concerning the underlying error sequence $\{e_t\}$ for each series. First, we suppose the errors driving the data series are normal ARMA(p, q) processes. In this case, the first differences of the series are normal ARMA(p, q) processes, possibly with nonzero mean, under the null hypothesis. To determine p and q, we fit ARMA(p, q) models to the first differences of each series, and we use the model-selection criteria of Akaike (1974) and Schwarz (1978) to choose the optimal ARMA(p, q) model with $p, q \leq 5$. The Akaike criterion minimizes $2 \ln L + 2(p + q)$, where L denotes the likelihood function. The Schwarz criterion minimizes $2 \ln L + (p + q) \ln T$, where T is the sample size. The Schwarz criterion penalizes extra parameters more heavily than does the Akaike criterion. We then treat the optimal estimated ARMA(p, q) models as the true data-generating processes for the errors of each of the series.

Table 8, panels A and B, presents the chosen models for each of the data series. In most cases the Akaike and Schwarz criteria select the same model. ARMA(1, 0) models are selected by both criteria for the real-GNP, nominal-GNP, real per capita GNP, GNP-deflator, and money-stock series, whereas ARMA(0, 1) models are selected by both criteria for the employment, nominal-wages, velocity, and real-wages series. In addition, an ARMA(0, 5) model is selected by both criteria for the industrial-production series. The Akaike criterion favors an ARMA(5, 0) model for stock prices, an ARMA(3, 0) model for interest rates, an ARMA(1, 1) model for consumer prices, and an ARMA(0, 3) model for postwar quarterly real GNP; the Schwarz criterion chooses ARMA(0, 1), ARMA(2, 0), ARMA(0, 1), and ARMA(1, 0) models for these series, respectively. In those cases in which the two criteria choose different models, we select the most parsimonious model.

To determine the finite-sample distributions of our test statistics under the null hypothesis with the preceding error distributions, we perform the following Monte Carlo experiment. For each series, we construct a pseudo sample of size equal to the actual size of the series using the optimal ARMA(p, q) models described previously with iid $N(0, \sigma^2)$ innovations, where σ^2 is the estimated innovation variance of the optimal ARMA(p, q) model. Then, for each $j = 2, \ldots, T - 1$, we set $\lambda = j/T$, determine k as described in the paragraph following Equation (3′), and compute $t_{\hat{\alpha}^i}(\lambda)$ using either (1′), (2′), or (3′). Our test statistic is then determined to be the minimum t statistic over all $T - 2$ regressions. We repeat this process 5,000 times, and the critical values for the finite-sample distributions are obtained from the sorted vector of replicated statistics.

Table 9, panels A—C, display the percentage points of the finite-sample distributions of the minimum t statistics for all of the data series under the assumption of normal ARMA(p, q) errors. The salient feature of these critical values is that they are all uniformly larger (in absolute value) than the corresponding average asymptotic critical values. At the 5% level, the Model (A) finite-sample critical values range from −5.12 to −5.38, the average of which is 9.2% larger (in absolute value) than the corresponding asymptotic critical value. At the same level, the Model (B) finite-sample critical value is −4.84 and the Model (C) finite-sample critical values are −5.63 and −5.68. These finite-sample critical values are 9.0% and 11.8% larger (in absolute value), respectively, than their corresponding asymptotic values. For the Model (A) series, the difference between the finite-sample and asymptotic critical values abates for the series with larger sample sizes. For the two Model (C) series, however, the critical values are nearly identical even though the sample size for the real-wage series is 71 and the sample size for the common-stock-price series is 100. Furthermore, for the series with comparable sample sizes, the finite-sample distributions generated from different ARMA models are fairly similar. The latter result suggests that the ADF methodology works fairly well in finite samples for our data set.

(We note that the finite-sample distributions of our test statistics are sensitive to the procedure used to de-

Table 8. Selected ARMA Models

Series	Model	$\hat{\theta}$	$\tilde{\psi}$	$\hat{\mu}$	$\hat{\sigma}$	AIC	SBIC	Q(x)
A. Model: $\Delta y_t = \hat{\mu} + \hat{\theta}\Delta y_{t-1} + e_t + \hat{\psi}e_{t-1}$								
Real GNP	(1,0)	.341	—	.029	.061	−165	−161	Q(22)
		(2.78)	(—)	(2.50)				18.06
Nominal GNP	(1,0)	.440	—	.055	.089	−120	−116	Q(22)
		(3.76)	(—)	(1.36)				23.57
Real per capita GNP	(1,0)	.331	—	.016	.062	−164	−160	Q(22)
		(2.69)	(—)	(1.36)				17.60
Employment	(0,1)	—	.388	.016	.036	−302	−278	Q(22)
		(—)	(3.72)	(2.38)				18.53
GNP deflator	(1,0)	.434	—	.20	.047	−262	−257	Q(22)
		(4.27)	(—)	(2.19)				21.55
Consumer prices	(0,1)	—	.655	.012	.046	−365	−360	Q(22)
		(—)	(9.21)	(1.73)				25.78
Nominal wages	(0,1)	—	.474	.040	.061	−192	−188	Q(22)
		(—)	(4.44)	(3.81)				34.90
Money stock	(1,0)	.622	—	.059	.048	−257	−253	Q(22)
		(7.06)	(—)	(4.28)				22.19
Velocity	(0,1)	—	.116	−.012	.068	−254	−248	Q(22)
		(—)	(1.16)	(−1.55)				21.53
Quarterly real GNP	(1,0)	.368	—	.005	.010	−992	−986	Q(28)
		(4.94)	(—)	(6.08)				19.80
Stock prices	(0,1)	—	.313	.029	.156	−84.6	−79.5	Q(22)
		(—)	(3.22)	(1.40)				21.41
Real wages	(0,1)	—	.205	.018	.036	−263	−258	Q(22)
		(—)	(1.72)	(3.50)				10.80
B. Model: $\Delta y_t = \hat{\mu} + \Sigma_1^p\hat{\theta}_i\Delta y_{t-i} + e_t + \Sigma_1^q\tilde{\psi}_i e_{t-i}$								
Industrial production	(0,5)	—	.033	.043	.095	−198	−182	Q(18)
		(—)	(.36)					13.42
		—	−.087					
		(—)	(−.98)					
		—	−.022					
		(—)	(−.24)					
		—	−.199					
		(—)	(−2.27)					
		—	−.402					
		(—)	(−4.46)					
Interest rates	(2,0)	.177	—	.079	.282	24.5	31.3	Q(21)
		(1.43)	(—)	(1.10)				14.98
		.367	—					
		(2.86)	(—)					

NOTE: All models were estimated using PROC ARIMA in SAS. t statistics are in parentheses. AIC denotes the Akaike information criterion, SBIC denotes the Schwarz criterion, and $Q(x)$ denotes the Box–Pierce statistic.

termine k, the number of lags of first differences of the data used in the regressions (1′)–(3′). In particular, when k is fixed at some value, say k^*, for each tentative choice of the break fraction λ instead of being allowed to vary, the fixed-k finite-sample distributions of the minimum t statistics are much closer to the appropriate asymptotic distributions than the random-k finite-sample distributions. Furthermore, this result obtains regardless of the value of k^* chosen for $k^* \le 8$ for the Nelson and Plosser data and $k^* \le 12$ for the postwar quarterly real-GNP data. For example, the 1%, 2.5%, 5%, and 10% points based on 5,000 repetitions of the fixed-k distributions for the nominal-GNP series are (a) $k^* = 2$: −5.55, −5.21, −4.89, −4.62; (b) $k^* = 4$: −5.52, −5.18, −4.94, −4.63; (c) $k^* = 6$: −5.56, −5.19, −4.91, −4.60; and (d) $k^* = 8$: −5.61, −5.17, −4.88, −4.60. These percentage points are, on average over k^*, 10% smaller in absolute value than the random-k percentage points reported in Table 9, panel A. The p values for nominal GNP computed from the preceding four fixed-k distributions are .005, .003, .005, and .005, whereas the asymptotic p value is .003 and the p value computed from the random-k distribution is .017.)

Using the finite-sample distributions of the Model (A) t statistics, the actual sizes of the asymptotic 5% tests range from 10.7% to 16.0%, producing an average size distortion of 8.2%. The size of the Model (B) asymptotic 5% test is 13.6%, and the average size for the model (C) t statistics is 16.0%. These size distortions are presented graphically in Figure 1, where we see that the finite-sample densities of the minimum t statistics are shifted to the left of the asymptotic densities of the minimum t statistics.

Assuming that the fitted ARMA models of the Nelson and Plosser series and the postwar quarterly real GNP series are correct, we can use the Monte Carlo generated finite-sample distributions of our test statistics to test these series for a unit root. From the preceding discussion, we know that the asymptotic tests are too liberal, allowing us to reject the unit-root null too often. This effect can be seen graphically in Figure 2,

Table 9. Percentage Points of the Finite-Sample Distribution of $inf_{\lambda\in\Lambda}t_{\hat{\alpha}A}(\lambda)$, $inf_{\lambda\in\Lambda}t_{\hat{\alpha}B}(\lambda)$, and $inf_{\lambda\in\Lambda}t_{\hat{\alpha}C}(\lambda)$ Assuming Normal ARMA Innovations

Series/model	T	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
				A. $inf_{\lambda\in\Lambda}t_{\hat{\alpha}A}(\lambda)$						
Asymptotic	∞	−5.34	−5.02	−4.80	−4.58	−3.75	−2.99	−2.77	−2.56	−2.32
Real GNP ARMA(1,0)	62	−6.03	−5.65	−5.35	−4.99	−3.96	−2.90	−2.47	−2.07	−1.51
Nominal GNP ARMA(1,0)	62	−6.12	−5.67	−5.38	−5.05	−4.00	−2.90	−2.53	−2.14	−1.52
Real per capita GNP ARMA(1,0)	62	−6.03	−5.63	−5.32	−5.01	−3.99	−2.92	−2.52	−2.23	−1.62
Ind. prod. ARMA(0,5)	111	−5.73	−5.41	−5.14	−4.86	−3.88	−3.01	−2.74	−2.52	−2.15
Employment ARMA(0,1)	81	−5.92	−5.55	−5.26	−4.95	−3.97	−3.00	−2.66	−2.26	−1.71
GNP deflator	82	−5.85	−5.50	−5.21	−4.87	−3.91	−2.99	−2.62	−2.33	−1.82
CPI ARMA(0,1)	111	−5.76	−5.46	−5.14	−4.85	−3.88	−2.97	−2.68	−2.34	−1.95
Nominal wages ARMA(0,1)	71	−5.93	−5.69	−5.33	−5.02	−4.01	−2.96	−2.55	−2.16	−1.90
Money stock ARMA(1,0)	82	−5.91	−5.49	−5.19	−4.90	−3.96	−2.94	−2.60	−2.22	−1.71
Velocity ARMA(0,1)	102	−5.67	−5.37	−5.12	−4.85	−3.88	−3.02	−2.75	−2.47	−2.16
Interest rate ARMA(3,0)	71	−5.90	−5.64	−5.30	−5.00	−3.99	−2.95	−2.60	−2.31	−1.96
				B. $inf_{\lambda\in\Lambda}t_{\hat{\alpha}B}(\lambda)$						
Asymptotic	∞	−4.93	−4.67	−4.42	−4.11	−3.23	−2.48	−2.31	−2.17	−1.97
Quarterly real GNP ARMA(1,0)	159	−5.40	−5.14	−4.84	−4.57	−3.53	−2.70	−2.49	−2.32	−2.20
				C. $inf_{\lambda\in\Lambda}t_{\hat{\alpha}C}(\lambda)$						
Asymptotic	∞	−5.57	−5.30	−5.08	−4.82	−3.98	−3.25	−3.06	−2.91	−2.72
Stock prices ARMA(0,1)	100	−6.30	−5.93	−5.63	−5.31	−4.30	−3.30	−3.09	−2.85	−2.64
Real wages ARMA(0,1)	71	−6.25	−5.92	−5.68	−5.38	−4.32	−3.36	−3.04	−2.81	−2.57

NOTE: Percentage points are based on 5,000 repetitions.

which shows the finite-sample 5% critical values lying above the corresponding asymptotic critical values. Using the finite-sample distributions, we can no longer reject the unit-root null at the 5% level for the employment, nominal-wage, and common-stock-price series. On the other hand, we can reject the unit root null at the 1% level for the industrial-production series, we can reject the null at the 2.5% level for the nominal-GNP series, and we can reject the null at the 5% level for the real-GNP series. The p values computed from the preceding finite-sample distributions are presented, for comparison with the previous asymptotic results, in column 4 of Table 7. Thus, after endogenizing the breakpoint selection procedure and correcting for small-sample biases, we do not reject the unit-root hypothesis for 8 of the 11 series for which Perron rejects the hypothesis. In accordance with Perron, however, we do reject the unit-root null for the real-GNP, nominal-GNP, and industrial-production series.

For the preceding series for which we do reject the unit-root hypothesis, we investigate the effect of relaxing the normality assumption on the finite-sample distributions of our test statistics. In particular, the large changes in the series at the estimated breakpoints suggest that the distributions of the innovations underlying the series may have fatter tails than the normal distribution. In repeated samples under a distribution with a higher probability of generating tail events than the normal, our breakpoint selection procedure will tend to produce larger (in absolute value) t statistics than in the normal case. Therefore, with fat-failed innovations, we expect the finite-sample distributions of our test statistics to shift further to the left.

To assess the normality assumption, Table 10 gives the estimated skewness and kurtosis values for the residuals from the optimal ARMA models for the first differences of the logarithms of the data series. Most of the series exhibit mild negative skewness. The estimated kurtosis values for real GNP, industrial production, and employment are only slightly larger than 3 (the kurtosis for a standard normal random variable), whereas the values for nominal GNP, nominal wages, and common-stock prices are considerably larger than 3. [For a sequence of iid normal random variables, the sample kurtosis has standard error equal to $(24/T)^{1/2}$. Using this formula, the estimated kurtosis values for the nominal GNP, nominal wages, and common-stock prices are 4.3, 2.3, and 5.7 standard deviations larger than the kurtosis values of a normal random variable.] Hence there is some evidence of leptokurtosis for some of the series.

A plausible family of distributions close to the normal but with thicker tails is the Student-t family with η df. To determine the appropriate degrees of freedom, we use a modified method-of-moments approach. In par-

Table 10. Skewness and Kurtosis Values for Residuals From the ARMA Model for Δy_t

Series	Sample skewness	Sample kurtosis	Finite-sample kurtosis	df
Real GNP	−.317	3.400	3.426	9
Nominal GNP	−1.146	5.868	4.804	4
Real per capita GNP	−.235	3.257		
Industrial production	−.737	3.722	3.815	9
Employment	−.424	3.469	3.420	10
GNP deflator	−.907	9.739		
Consumer prices	1.023	6.852		
Nominal wages	−.304	4.658	4.283	5
Money stock	−.270	4.636		
Velocity	−.329	2.927		
Interest rate	.751	4.413		
Quarterly real GNP	−.061	3.828		
Common stock prices	−.390	4.324	4.351	6
Real wages	−.040	3.161		

NOTE: The column labeled "Finite sample kurtosis" gives the mean kurtosis value obtained from the finite-sample distribution of the sample kurtosis using Student-t ARMA innovation with degrees of freedom given in the adjacent column.

Table 11. Percentage Points of the Finite-Sample Distribution of $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}i}(\lambda)$ *Assuming Student-t ARMA Innovations*

Series	df	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
Nominal GNP	4	−7.56	−6.40	−5.86	−5.31	−4.05	−3.01	−2.63	−2.20	−1.84
Real GNP	9	−6.16	−5.75	−5.39	−5.04	−3.98	−2.92	−2.57	−2.13	−1.56
Ind. prod.	9	−5.94	−5.60	−5.29	−4.91	−3.95	−3.03	−2.73	−2.48	−1.93
Employment	10	−5.98	−5.67	−5.27	−5.01	−3.98	−3.06	−2.66	−2.32	−1.90
Nominal wages	5	−7.26	−6.31	−5.81	−5.39	−4.11	−3.12	−2.82	−2.40	−1.86
Stock prices	6	−6.66	−6.09	−5.84	−5.46	−4.29	−3.37	−3.12	−2.91	−2.71

NOTE: The column labeled "df" gives the degrees of freedom of a Student-*t* random variable that gives the closest match between the observed sample kurtosis and the finite-sample mean kurtosis value. Percentage points are based on 1,000 repetitions.

ticular, for each series under consideration we compute by Monte Carlo the means of the sample kurtosis statistic using the appropriate ARMA(p, q) model with iid Student-t innovations for various values of η. We then determine the t distribution for each series by finding the closest match between the observed sample kurtosis and the finite-sample mean kurtosis values. The finite-sample mean values of the sample kurtosis and the degrees of freedom of the selected t distributions are given in the third and fourth columns of Table 10. For nominal GNP, 4 df are chosen; for nominal wages, 5 are chosen; for common-stock prices, 6; for both industrial production and real GNP, 9; and for employment, 10.

[Note that we also tried the traditional method-of-moments approach to estimate η by using the fact that the kurtosis of a Student-t random variable with η df is $3 + 6/(\eta - 4)$. Using this method, 6 df were determined for nominal GNP and common-stock prices, 8 were chosen for nominal wages, 12 for both industrial production and employment, and 16 for real GNP. Our test conclusions based on these t distributions are the same as in the normal case.]

Table 11 gives the percentage points of the finite-sample distributions of our test statistics for the preceding series. Table 7 (col. 5) reports the p values computed from these distributions. The percentage points obtained using the Student-t innovations are uniformly larger than the corresponding percentage points determined from normal innovations. Our test conclusions based on the Student-t distributions, however, remain essentially the same as in the normal case; that is, we reject the unit-root null at the 1% level for the industrial-production series, we reject at the 5% level for the real-GNP and nominal-GNP series, and we reject at the 10% level for the common-stock-price series. We no longer reject at the 10% level for the nominal-wage series, but we are still close to rejecting because the p value is only 11.7%. Thus our rejections of the unit-root hypothesis for these series are not very sensitive to the relaxation of the normality assumption.

Last, we briefly investigate the effect that infinite-variance innovations would have on our test results, although the sample kurtosis estimates are not indicative of innovations whose tails are that fat. We compute the finite-sample distribution of our test statistic, using the parameters of the nominal-GNP series, under the assumption of ARMA errors with innovations that follow a symmetric stable distribution with characteristic exponent α. (Note that the asymptotic distribution of our test statistic given in Theorem 1 does not apply in this case because the partial sum process of the errors does not satisfy a Gaussian functional central limit theorem.) Table 12 reports the percentage points of this distribution for various values of α. From these results, we see that if the innovations have only slightly less than two moments finite (e.g., $\alpha = 1.9$) one cannot reject the unit-root hypothesis at the 5% level for any of the series.

6. CONCLUDING REMARKS

In this article, we transform Perron's unit-root test that is conditional on structural change at a known point in time into an unconditional unit-root test. We also take into consideration the effects of fat-tailed innovations on the performance of the tests. Our analysis

Table 12. Percentage Points of the Finite-Sample Distribution of $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}A}(\lambda)$ *Assuming Stable ARMA Innovations*

α	1.0%	2.5%	5.0%	10.0%	50.0%	90.0%	95.0%	97.5%	99.0%
2.0	−6.12	−5.65	−5.37	−5.05	−4.00	−2.90	−2.53	−2.15	−1.51
1.9	−9.99	−6.69	−6.00	−5.33	−4.08	−2.92	−2.43	−1.95	−1.52
1.8	−12.0	−9.14	7.05	−5.87	−4.11	−2.94	−2.51	−1.96	−1.54
1.5	−49.3	−18.2	−12.2	−9.32	−4.57	−3.00	−2.57	−1.70	−1.50
1.0	−221	−93.0	−42.9	−20.8	−5.72	−3.02	−2.45	−1.52	−1.01
.5	−15,622	−4,518	−893	−205	−7.18	−3.42	−2.32	−1.41	−.73

NOTE: α is the characteristic exponent of a standard stable random variable. $\alpha = 2$ corresponds to a normal variate, and $\alpha = 1$ corresponds to a Cauchy variate. The ARMA innovations use the nominal GNP parameters. Percentage points are based on 5,000 repetitions.

is motivated by the fact that the breakpoints used by Perron are data dependent and plots of drifting unit-root processes often are very similar to plots of processes that are stationary about a broken trend for some breakpoint. The null hypothesis that we believe is of most interest is a unit-root process without any exogenous structural breaks, and the relevant alternative hypothesis is a trend-stationary process with possible structural change occurring at an unknown point in time.

We systematically address the effects of endogenizing the breakpoint selection procedure on the asymptotic distributions and finite-sample distributions of Perron's test statistics for a unit root. Using our *estimated breakpoint* asymptotic distributions, we find less conclusive evidence against the unit-root hypothesis than Perron found for many of the data series. We reverse his conclusions for 5 of the 11 Nelson and Plosser series for which he rejected the unit-root hypothesis at the 5% level, and we reverse his unit-root rejection for the postwar quarterly real-GNP series. When we take into consideration small-sample biases and the effects of fat-tailed (but not infinite-variance) innovations, we reverse his conclusions for 3 more of the Nelson and Plosser series.

The reversals of some of Perron's results should not be construed as providing evidence for the unit-root null hypothesis, because the power of our test against Perron's trend-stationary alternatives is probably low for small to moderate changes in the trend functions. Rather, the reversals should be viewed as establishing that there is less evidence against the unit-root hypothesis for many of the series than the results of Perron indicate. On the other hand, for some of the series (industrial production, nominal GNP, and real GNP), we reject the unit-root hypothesis even after endogenizing the breakpoint selection procedure and accounting for moderately fat-tailed errors. For these series, our results provide stronger evidence against the unit-root hypothesis than that given by Perron.

ACKNOWLEDGMENTS

We thank Peter Phillips, Pierre Perron, Jim Stock, two referees, and an associate editor for helpful comments. Donald W. K. Andrews acknowledges the research support of the National Science Foundation via Grant SES-8821021.

APPENDIX A: PROOF OF THEOREM

One way to establish the convergence result in the theorem is to first show that the finite dimensional distributions of $t_{\hat{\alpha}^i}(\lambda)$, indexed by λ, converge; that is, for any finite number J of λ values one must show that $(t_{\hat{\alpha}^i}(\lambda_1), \ldots, t_{\hat{\alpha}^i}(\lambda_J))'$ converges weakly to $(L(\lambda_1), \ldots, L(\lambda_J))'$. Next one must show that the sequence of probability measures associated with $t_{\hat{\alpha}^i}(\lambda)$ is tight. If the preceding two conditions hold, then we have the weak convergence result $t_{\hat{\alpha}^i}(\cdot) \Rightarrow L(\cdot)$. Then, provided $\inf_{\lambda\in\Lambda} L(\lambda)$ is a continuous functional of $L(\cdot)$ a.s. $[L(\cdot)]$, we get the desired result: $\inf_{\lambda\in\Lambda} t_{\hat{\alpha}^i}(\lambda) \Rightarrow \inf_{\lambda\in\Lambda} L(\lambda)$.

Establishing the finite-dimensional convergence of $t_{\hat{\alpha}^i}(\lambda)$ is trivial given Perron's results. Showing tightness, however, is a difficult task. We avoid the problem of establishing tightness by using a different method of proof from the "fidi plus tightness" method. The method we use appeals directly to the continuous mapping theorem (CMT). The idea is to express the test statistic in the theorem as a functional, say $g(\cdot, \cdot, \cdot, \cdot, \cdot)$, of the partial sum process $X_T(\cdot)$, a rescaled version of the deterministic regressors $Z_T(\cdot, \cdot)$, the process $T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t$, the average squared innovations σ_T^2, and an estimate $s^2(\cdot)$ of the error variance. If we have joint weak convergence of the process $(X_T(\cdot), Z_T(\cdot, \cdot), T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t, \sigma_T^2, s^2(\cdot))'$ to a process $(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))'$ and if g is continuous with respect to $(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))'$ on a set C with $P\{(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))' \in C\} = 1$, then $g(X_T(\cdot), Z_T(\cdot, \cdot), T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t, \sigma_T^2, s^2(\cdot)) \Rightarrow g(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))$ by the CMT.

Let $S_t = \Sigma_1^t e_j$ $(S_0 = 0)$ and $X_T(r) = \sigma^{-1}T^{-1/2}S_{[Tr]}$, $(j - 1)/T \le r < j/T$ for $j = 1, \ldots, T$, where $\sigma^2 = \lim_{T\to\infty} T^{-1}ES_T^2$ and $[Tr]$ denotes the integer part of Tr. Here, as in the theorem, we assume that the disturbances are iid so that $\sigma^2 = Ee_1^2 = \sigma_e^2 \in (0, \infty)$. Under these assumptions, the disturbances $\{e_t\}$ satisfy an invariance principle. Specifically, as processes indexed by $r \in [0, 1]$, we have $X_T(\cdot) \Rightarrow W(\cdot)$ as $T \to \infty$ (using the uniform metric on the space of functions on $[0, 1]$), where $W(r)$ denotes a standard Brownian motion or Wiener process on $[0, 1]$. In addition, $\sigma_T^2 = T^{-1}\Sigma_1^T e_t^2 \xrightarrow{p} \sigma^2$. As in the theorem, we take Λ to be a closed subset of $(0, 1)$.

Throughout what follows, "$\Rightarrow$" denotes weak convergence (with respect to a specified metric) and "$\equiv$" denotes equivalence in distribution. For notational convenience, we shall often denote $W(r)$ by W. Similarly, we shall often write integrals with respect to Lebesgue measure such as $\int_0^1 W(r)dr$ as $\int_0^1 W$.

We consider least squares regressions of the form $y_t = \hat{\beta}^i(\lambda)'z_{tT}^i(\lambda) + \hat{\alpha}^i(\lambda)y_{t-1} + \hat{e}_t(\lambda)$ $(t = 1, \ldots, T)$ for Models i = A, B, and C. The vector $z_{tT}^i(\lambda)$ encompasses the deterministic components of the regression equation, and it depends explicitly on the location of the break fraction and the sample size. For example, in Model (A) we have $z_{tT}^A(\lambda)' = (1, t, DU_t(\lambda))$, where $DU_t(\lambda) = 1$ if $t > T\lambda$ and 0 otherwise.

Let $Z_T^i(\lambda, r) = \delta_T^i z_{[Tr]T}^i(\lambda)$ denote a rescaled version of the deterministic regressors, where δ_T^i is a diagonal matrix of weights. For each i = A, B, C, there is a nonrandom function $Z^i(\lambda, r)$ such that $Z_T^i(\cdot, \cdot) \to Z(\cdot, \cdot)$ as $T \to \infty$ with respect to a metric d^*. By definition, $d^*(Z_T^i(\cdot, \cdot), Z^i(\cdot, \cdot))$ is the maximum of the uniform metric distances between the elements of the vector-valued function $Z_T^i(\cdot, \cdot)$ and those of $Z^i(\cdot, \cdot)$, unless the elements are $DU_{[Tr]}(\lambda)$ and its limit $du(\lambda, r) = 1(r > \lambda)$. In the latter case, the metric distance between these two elements is taken to be a hybrid uniform/L^2 metric

$$d(DU_{[T\cdot]}(\cdot), du(\cdot,\cdot)) = \sup_{\lambda\in\Lambda}\left(\int_0^1 (DU_{[Tr]}(\lambda) - du(\lambda, r))^2 dr\right)^{1/2}.$$

Note that $|DU_{[Tr]}(\lambda) - du(\lambda, r)| = 1([Tr]/T \le \lambda < r) \le 1(\lambda < r \le \lambda + 1/T)$, so that $d(DU_{[T\cdot]}(\cdot), du(\cdot,\cdot)) \le 1/T \to 0$ as $T \to \infty$ as desired. We use the metric D for the functions $DU_{[Tr]}(\lambda)$ and $du(\lambda, r)$, because convergence under the uniform metric does not hold: $\sup_{\lambda,r}|DU_{[Tr]}(\lambda) - du(\lambda, r)| = 1 \not\to 0$ as $T \to \infty$. In fact, convergence does not hold even under the Skorohod metric for any fixed $\lambda \in \Lambda$. [The use of the metric d for $DU_{[Tr]}(\lambda)$ makes it somewhat more difficult to establish Lemma A.2 than if the uniform metric could be used but otherwise does not affect the proof of the theorem. Note that the metric d is just what is needed here, because it is weak enough so that $Z_T^i(\cdot,\cdot) \to Z^i(\cdot,\cdot)$ as $T\to\infty$ and strong enough so that continuity of H_1 and H_2 can be established in Lemma A.2.)

For example, in Model A we have

$$\delta_T^A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1}$$

and $Z_T^A(\lambda, r) \to Z^A(\lambda, r) = (1, r, du(\lambda, r))'$ under d^*.

Note that it is possible to circumvent the use of the hybrid metric d^* by writing the test statistic of interest as a continuous function g of the same terms as previously but excluding $Z_T(\cdot,\cdot)$. We do not do so, because the resulting expressions are excessively lengthy.

The coefficient $\hat\alpha^i(\lambda)$ and its t statistic are invariant with respect to the value of the drift μ in the null model (6). Therefore, without loss of generality, we set $\mu = 0$ in (6).

The normalized bias for testing the null hypothesis $\alpha^i = 1$ is given by

$$T(\hat\alpha^i(\lambda) - 1) = \left(T^{-2}\sum_1^T y_{t-1}^i(\lambda)^2\right)^{-1}\left(T^{-1}\sum_1^T y_{t-1}^i(\lambda)e_t\right)$$

and the t statistic for testing $\alpha^i = 1$ is given by

$$t_{\hat\alpha^i}(\lambda) = \left(T^{-2}\sum_1^T y_{t-1}^i(\lambda)^2\right)^{1/2} T(\hat\alpha^i(\lambda) - 1)/s(\lambda)$$
$$= \left(T^{-2}\sum_1^T y_{t-1}^i(\lambda)^2\right)^{-1/2} \times \left(T^{-1}\sum_1^T y_{t-1}^i(\lambda)e_t\right)\Big/ s(\lambda),$$

where $y_t^i(\lambda) = y_t - z_{tT}^i(\lambda)'(\Sigma_1^T z_{sT}^i(\lambda)z_{sT}^i(\lambda)')^{-1}\Sigma_1^T z_{sT}(\lambda)y_s$ and $s^2(\lambda) = T^{-1}\Sigma_1^T(y_t - \hat\beta^i(\lambda)'z_{tT}^i(\lambda) - \hat\alpha^i(\lambda)y_{t-1})^2$ for Models i = A, B, and C. For brevity we drop the superscript i and only consider Model (A) for which $z_{tT}^A(\lambda)' = (1, t, DU_t(\lambda))$. The proofs for Models (B) and (C) are analogous and are therefore omitted.

The test statistic of interest is

$$\inf_{\lambda\in\Lambda} t_{\hat\alpha}(\lambda) = \inf_{\lambda\in\Lambda}\left(T^{-2}\sum_1^T y_{t-1}(\lambda)^2\right)^{-1/2} \times \left(T^{-1}\sum_1^T y_{t-1}(\lambda)e_t\right)\Big/ s(\lambda),$$

which we may write as a function of X_T $(= X_T(\cdot))$, Z_T $(= Z_T(\cdot,\cdot))$, $T^{-1/2}\Sigma_1^T Z_T e_t$ $(= T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t)$, σ_T^2, and s^2 $(= s^2(\cdot))$ plus an asymptotically negligible term,

$$\inf_{\lambda\in\Lambda} t_{\hat\alpha}(\lambda) = g\left(X_T, Z_T, T^{-1/2}\sum_1^T Z_T e_t, \sigma_T^2, s^2\right) + o_{p\lambda}(1), \quad \text{(A.1)}$$

where g is as defined later. The symbol $o_{p\lambda}(1)$ denotes any random variable $\zeta(\lambda)$ such that $\sup_{\lambda\in\Lambda}|\zeta(\lambda)| \xrightarrow{p} 0$.

It will be useful to reexpress g as the following composite functional:

$$g\left(X_T, Z_T, T^{-1/2}\sum_1^T Z_T e_t, \sigma_T^2, s^2\right) = h^*\left(h\left[H_1[\sigma X_T, Z_T], H_2\left[\sigma X_T, Z_T, T^{-1/2}\sum_1^T Z_T e_t, \sigma_T^2\right], s^2\right]\right), \quad \text{(A.2)}$$

where H_1 maps a function on [0, 1] and a function on $\Lambda \times [0, 1]$ into a function on Λ, H_2 maps a function on [0, 1], a function on $\Lambda \times [0, 1]$, a function on Λ, and a positive real number into a function on Λ, h maps three functions on Λ into a function on Λ, and h^* maps a function on Λ into a real number. Specifically, for any real function $m = m(\cdot)$ on Λ,

$$h^*(m) = \inf_{\lambda\in\Lambda} m(\lambda), \quad \text{(A.3)}$$

and for any functions $m_1 = m_1(\cdot)$, $m_2 = m_2(\cdot)$, and $m_3 = m_3(\cdot)$ on Λ,

$$h[m_1, m_2, m_3](\cdot) = m_1(\cdot)^{-1/2}m_2(\cdot)/m_3(\cdot). \quad \text{(A.4)}$$

The functionals H_1 and H_2 are the functional analogs of the sample moments $T^{-2}\Sigma_1^T y_{t-1}(\lambda)^2$ and $T^{-1}\Sigma_1^T y_{t-1}(\lambda)e_t$ plus an $o_{p\lambda}(1)$ term. In particular,

$$T^{-2}\sum_1^T y_{t-1}(\lambda)^2$$
$$= T^{-2}\sum_1^T\left\{y_{t-1} - z_{tT}(\lambda)'\left(\sum_1^T z_{sT}(\lambda)z_{sT}(\lambda)'\right)^{-1}\sum_1^T z_{sT}(\lambda,s)y_s\right\}^2$$
$$= T^{-1}\sum_1^T\left\{T^{-1/2}S_{t-1} - z_{tT}(\lambda)'\delta_T\left(T^{-1}\sum_1^T \delta_T z_{sT}(\lambda)z_{sT}(\lambda)'\delta_T\right)^{-1} \times T^{-1}\sum_1^T \delta_T z_{sT}(\lambda)T^{-1/2}S_{s-1}\right\}^2 + o_{p\lambda}(1)$$
$$= \int_0^1\left\{\sigma X_T(r) - Z_T(\lambda, r)'\left(\int_0^1 Z_T(\lambda,s)Z_T(\lambda,s)'ds\right)^{-1} \times \int_0^1 Z_T(\lambda,s)\sigma X_T(s)ds\right\}^2 dr + o_{p\lambda}(1)$$
$$= H_1[\sigma X_T, Z_T](\lambda) + o_{p\lambda}(1) \quad \text{(A.5)}$$

and

$$T^{-1}\sum_1^T y_{t-1}(\lambda)e_t$$

$$= T^{-1}\sum_1^T\left\{y_{t-1} - z_{tT}(\lambda)'\left(\sum_1^T z_{sT}(\lambda)z_{sT}(\lambda)'\right)^{-1}\sum_1^T z_{sT}(\lambda)y_s\right\}e_t$$

$$= T^{-1}\sum_1^T S_{t-1}e_t - T^{-1/2}\sum_1^T e_t z_{tT}(\lambda)'\delta_T\left(T^{-1}\sum_1^T \delta_T z_{sT}(\lambda)z_{sT}(\lambda)'\delta_T\right)^{-1}$$

$$\times T^{-1}\sum_1^T \delta_T z_{sT}(\lambda)T^{-1/2}S_{s-1} + o_{p\lambda}(1)$$

$$= (1/2)(\sigma^2 X_T(1) - \sigma_T^2) - T^{-1/2}\sum_1^T e_t Z_T(\lambda, t/T)'$$

$$\times\left(\int_0^1 Z_T(\lambda,s)Z_T(\lambda,s)ds\right)^{-1}\int_0^1 Z_T(\lambda,s)\sigma X_T(s)ds + o_{p\lambda}(1)$$

$$= H_2\left[\sigma X_T, Z_T, T^{-1/2}\sum_1^T Z_T e_t, \sigma_T^2\right)(\lambda) + o_{p\lambda}(1). \qquad \text{(A.6)}$$

For the analysis that follows, we require the following lemma.

Lemma A.1. $T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t \Rightarrow \sigma\int_0^1 Z(\cdot, r)dW(r)$ (using the uniform metric).

Proof. The individual components of the vector $T^{-1/2}\Sigma_1^T Z_T(\lambda, t/T)e_t$ are $T^{-1/2}\Sigma_1^T e_t$, $T^{-3/2}\Sigma_1^T te_t$, and $T^{-1/2}\Sigma_{[T\lambda]+1}^T e_t$, respectively. By straightforward manipulations, we may express the preceding sums as functions of X_T; that is, $T^{-1/2}\Sigma_1^T e_t = \sigma X_T(1)$, $T^{-3/2}\Sigma_1^T te_t = \sigma(X_T(1) - \int_0^1 X_T(r)dr)$, and $T^{-1/2}\Sigma_{[T\lambda]+1}^T e_t = \sigma(X_T(1) - X_T(\lambda))$. By joint convergence and the CMT, we have that $(T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t) \Rightarrow (\sigma W(1), \sigma(W(1) - \int_0^1 W), \sigma(W(1) - W(\cdot)))' = \sigma\int_0^1 Z(\cdot, r)dW(r)$.

Note that the convergence result of Lemma A.1 holds jointly with $X_T(\cdot) \Rightarrow W(\cdot)$. Furthermore, using arguments similar to those used later it can be shown that $s^2(\lambda) = \sigma^2 + o_{p\lambda}(1)$. Since $Z_T(\cdot, \cdot)$ has the degenerate limiting distribution $Z(\cdot, \cdot)$, and σ_T^2 and $s^2(\cdot)$ have the degenerate limit distributions σ^2 and $\sigma^2 1(\cdot)$, where $1(\cdot)$ is the constant function equal to 1 for all $\lambda \in \Lambda$, it follows that $(X_T(\cdot), Z_T(\cdot, \cdot), T^{-1/2}\Sigma_1^T Z_T(\cdot, t/T)e_t, \sigma_T^2, s^2(\cdot))'$ converges weakly to $(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))'$ (using the uniform metric on the first, third, and fifth elements of this vector function, the d^* metric on the second element, and the Euclidean metric on the fourth). Hence the desired result follows from the CMT provided that (A.2) defines a continuous functional with probability 1 with respect to the limit process $(W(\cdot), Z(\cdot, \cdot), \int_0^1 Z(\cdot, r)dW(r), \sigma^2, \sigma^2 1(\cdot))'$. In what follows, continuity is defined using the uniform metric on every space of functions considered except that which contains $Z_T(\cdot, \cdot)$ and $Z(\cdot, \cdot)$, for which d^* is used.

We prove the continuity of g in a series of steps. The first step establishes continuity of H_1 at (W, Z) and H_2 at $(W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2)$ a.s. $[W]$. The second step establishes continuity of $h[m_1, m_2, m_3](\cdot)$ at $(m_1, m_2, m_3) = (H_1[\sigma W, Z], H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2], \sigma^2 1)$ a.s. $[W]$. The last step establishes the continuity of $h^*(m)$ at all real functions m on Λ. The continuity of g then follows from the continuity of a composition of continuous functions, and the result of the theorem follows from the CMT.

Lemma A.2. The functions H_1 and H_2 defined in (A.5) and (A.6) are continuous at (W, Z) and $(W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2)$, respectively, with W-probability 1.

Proof. From (A.5) we see that the functional $H_1[\sigma W, Z](\lambda)$ is simply the sum of products of the functions $\int_0^1 W^2$, $\int_0^1 Z(\lambda, r)W(r)dr$, and $(\int_0^1 Z(\lambda, r)Z(\lambda, r)'dr)^{-1}$, each of which is being viewed as a map that maps $\sigma W(\cdot)$ on $[0, 1]$ and $Z(\cdot, \cdot)$ on $\Lambda \times [0, 1]$ to a function on Λ. From (A.6) we see that $H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2](\lambda)$ is similarly defined with the addition of the terms $W(1)$, $\int_0^1 Z(\lambda, r)dW(r)$, and σ^2. Continuity of H_1 with respect to (W, Z) and H_2 with respect to $(W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2)$ follows from continuity of each of the preceding functions, with respect to (W, Z) [$(W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2)$] with W-probability 1 provided each function is bounded over Λ with W-probability 1—that is, provided $\sup_{\lambda\in\Lambda}|\int_0^1 Z(\lambda, r)W(r)dr| < \infty$ with probability 1 and likewise for the other functions. Let $\int_0^1 ZZ' = \int_0^1 Z(\lambda, r)Z(\lambda, r)'dr$. With the possible exception of $(\int_0^1 ZZ')^{-1}$, it is easy to verify the boundedness over Λ with W-probability 1 of each of the preceding functions. The function $(\int_0^1 ZZ')^{-1}$ is bounded over Λ provided $\inf_{\lambda\in\Lambda}\det(\int_0^1 ZZ') > 0$. Since $\det(\int_0^1 ZZ') = (1/3)(1 - \lambda) - (1/4)(1 - \lambda^2)^2 - (1/4)(1 - \lambda)\lambda^2 + (1/4)(1 - \lambda)(1 - \lambda^2) - (1/3)(1 - \lambda)^2$, $\inf_{\lambda\in\Lambda}\det(\int_0^1 ZZ') > 0$ if Λ is a closed subset of $(0, 1)$, which we assume.

Next, we show continuity of the map $(W, Z) \to (\int_0^1 ZZ')^{-1}$ (using the metric on the domain given by the maximum of the d^* metric for Z and the uniform metric for W and using the uniform metric on the range). We have

$$\sup_{\lambda\in\Lambda}\left|\left[\int_0^1 Z_T Z_T'\right]_{33} - \left[\int_0^1 ZZ'\right]_{33}\right|^{1/2}$$

$$= \sup_{\lambda\in\Lambda}\left|\int_0^1 (DU_{[Tr]}^2(\lambda) - du^2(\lambda, r))dr\right|^{1/2}$$

$$= d(DU_{[T\cdot]}(\cdot), du(\cdot, \cdot)).$$

In consequence, the map $(W, Z) \to [\int_0^1 ZZ']_{33}$ is continuous. Similarly, one can show that the maps $(W, Z) \to [\int_0^1 ZZ']_{k\ell}$ are continuous for $k, \ell = 1, 2, 3$. Furthermore, the map $\int_0^1 ZZ' \to [\int_0^1 ZZ']^{-1}$ is continuous (using the uniform metric on its domain and range), because $\inf_{\lambda\in\Lambda}\det[\int_0^1 ZZ'] > 0$. Since the composition of continuous functions is continuous, we get the desired result that the map $(W, Z) \to [\int_0^1 ZZ']^{-1}$ is continuous.

We now consider the continuity of the maps involving the integral functions of W. For example, consider the map $(W, Z) \to \int_0^1 W^2$ (using the same metrics on the domain and range as previously). Let W and $\tilde{W}$ be two Wiener processes on $[0, 1]$ such that for some $\delta > 0$

$\sup_{r\in[0,1]}|W(r) - \tilde{W}(r)| < \delta$. Then

$$\left|\int_0^1 W^2 - \int_0^1 \tilde{W}^2\right| \le \sup_{r\in[0,1]} |W(r)^2 - \tilde{W}(r)^2|$$
$$\le \sup_{r\in[0,1]} |W(r) - \tilde{W}(r)| \cdot \sup_{r\in[0,1]} |W(r) + \tilde{W}(r)|$$
$$\le \sigma \cdot \sup_{r\in[0,1]} |W(r) + \tilde{W}(r)|.$$

Since W and $\tilde{W}$ are continuous functions with probability 1 on the compact set $[0, 1]$, $\sup_{r\in[0,1]}|W(r) + \tilde{W}(r)| < \infty$ with probability 1 and continuity follows on a set with W-probability 1.

Similar proofs to those just given establish the continuity of the maps $(W, Z) \to \int_0^1 Z(\lambda, r)W(r)dr$ and $(W, Z) \to \int_0^1 Z(\lambda, r)dW(r)$ with W-probability 1, using the fact that the latter function can be written as an explicit function of $W(\cdot)$ as in the proof of Lemma A.1.

Remark. The functions $H_1[\sigma W, Z](\lambda)$ and $H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2](\lambda)$ may be expressed as $\sigma^2\int_0^1 W(\lambda, r)^2dr$ and $\sigma^2\int_0^1 W(\lambda, r)dW(r)$, respectively, where $W(\lambda, r)$ is the limit expression of the projection residual $y_t(\lambda)$; that is,

$$W(\lambda, r) = W(r) - Z(\lambda, r)' \times \left(\int_0^1 Z(\lambda, s)Z(\lambda, s)'ds\right)^{-1} \int_0^1 Z(\lambda, s)W(s)ds. \quad \text{(A.7)}$$

Lemma A.3. The function h defined in (A.4) is continuous at $(m_1, m_2, m_3) = (H_1[\sigma W, Z], H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2], \sigma^2 1)$ with W-probability 1 (using the uniform metric on its domain and range).

Proof. Since $h[m_1, m_2, m_3](\cdot) = m_1(\cdot)^{-1/2}m_2(\cdot)/m_3(\cdot)$, h is continuous at $(m_1, m_2, m_3) = (H_1[\sigma W, Z], H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2], \sigma^2 1)$ with W-probability 1 provided $\sigma^2 > 0$ and $\inf_{\lambda\in\Lambda}|H_1[\sigma W, Z](\lambda)| > 0$ with probability 1. Suppose that $H_1[\sigma W, Z](\lambda) = 0$ with positive W-probability. Then, since $H_1[\sigma W, Z](\lambda)$ is continuous in λ with W-probability 1 and Λ is compact, there exists a $[0, 1]$-valued random variable λ_0 such that for those realizations of W for which $\inf_{\lambda\in\Lambda}|H_1[\sigma W, Z](\lambda)| = 0$ we have $H_1[\sigma W, Z](\lambda_0) = 0$ and $\lambda_0 \in \Lambda$, for other realizations of W we have $\lambda_0 = 0$, and $\lambda_0 > 0$ with positive probability. In consequence, on a set with positive probability $W(\lambda_0, r) = 0$, for all $r \in [0, 1]$. From the definition of $W(\lambda_0, r)$ given in (A.7), this implies that

$$W(r) = Z(\lambda_0, r)'\left(\int_0^1 Z(\lambda_0, s)Z(\lambda_0, s)'ds\right)^{-1} \times \int_0^1 Z(\lambda_0, s)W(s)ds$$
$$= (1, r, du(\lambda_0, r)) \cdot C(W, \lambda_0), \quad \text{(A.8)}$$

for all $r \in [0, 1]$, where $C(W, \lambda_0)$ is a (3×1) vector, independent of r, with elements $C_1(W, \lambda_0)$, $C_2(W, \lambda_0)$, and $C_3(W, \lambda_0)$. Now consider any $0 \le r_1 < r_2 < r_3 < \inf\{\lambda : \lambda \in \Lambda\}$. By definition of the Wiener process, the increments $W(r_3) - W(r_2)$ and $W(r_2) - W(r_1)$ are independent. On the other hand, by (A.8), $W(r_3) - W(r_2) = C_2(W, \lambda_0)(r_3 - r_2)$ and $W(r_2) - W(r_1) = C_2(W, \lambda_0)(r_2 - r_1)$ on a set with positive probability. This implies that these increments are not independent, which is a contradiction. Hence we conclude that $\lambda_0 = 0$ with probability 1 and the desired result follows.

Lemma A.4. The function h^* defined in (A.3) is continuous at all functions m on Λ (using the uniform metric on its domain and the Euclidean metric on its range).

Proof. Given that $\varepsilon > 0$, let m and $\tilde{m}$ be two functions on Λ such that $\sup_{\lambda\in\Lambda}|m(\lambda) - \tilde{m}(\lambda)| < \varepsilon$. Then the result follows from the inequality

$$|\inf_{\lambda\in\Lambda} m(\lambda) - \inf_{\lambda\in\Lambda} \tilde{m}(\lambda)| \le \sup_{\lambda\in\Lambda}|m(\lambda) - \tilde{m}(\lambda)| < \varepsilon.$$

The proof of the theorem follows from Lemmas A.1–A.4, the continuity of a composition of continuous functions, and the CMT. The expression for the limit distribution given in the theorem may be verified by using the integral representations of $H_1[\sigma W, Z](\cdot)$ and $H_2[\sigma W, Z, \int_0^1 Z(\cdot, r)dW(r), \sigma^2](\cdot)$ described in the preceding remark.

APPENDIX B: CONSTRUCTION OF TABLES OF CRITICAL VALUES

This appendix details the approach used to approximate the limiting distributions in Theorem 1. It is instructive to outline the steps of the approximation since our methodology for approximating the limiting distributions differs slightly from the procedure used by Perron. First, we generate $N = 1{,}000$ iid $N(0, 1)$ random variables, $\{e_t\}$, and form the $(N \times 1)$ vector of partial sums, S. Then for each value of $\lambda = j/N$, where j runs from 2 to 999, we create the data matrix $X^i(\lambda) = (Z^i(\lambda), S_{-1})$, where $Z^i(\lambda)$ contains the deterministic components of the regressions, and construct the projection residual vector $S^i(\lambda) = (I - X^i(\lambda)(X^i(\lambda)'X^i(\lambda))^{-1}X^i(\lambda)')S$ for each Model i = A, B, C. We then form sample moments that converge as $N \to \infty$ to the functions of the standardized Wiener processes that are involved in the expressions in the theorem; that is, we form $N^{-1}\Sigma_1^N S^i(\lambda)_{j-1}e_j$ ($\Rightarrow \int_0^1 W^i(\lambda; r)dW$ as $N \to \infty$) and $N^{-2}\Sigma_1^N S^i(\lambda)_{j-1}^2$ ($\Rightarrow \int_0^1 W^i(\lambda; r)^2dr$ as $N \to \infty$). Using these values, we form the approximate expressions for the limiting distributions of the statistics for a fixed value of λ—for example, $t_{\alpha^i}(\lambda) = (N^{-2}\Sigma_1^N S^i(\lambda)_{j-1}^2)^{-1/2} \times (N^{-1}\Sigma_1^N S^i(\lambda)_{j-1}e_j)$. We do this for each value of $\lambda = j/N$ for $1 < j < N$, and from these $N - 2$ expressions we define $\lambda_{\inf}^i$ to be the value of λ that minimizes the preceding expression. The test statistic approximations evaluated at these values of λ give the corresponding approximate limiting distributions of the test statistics. This process gives us one observation from the asymptotic distributions of the test statistics. We repeat this process 5,000 times and obtain the critical values for the limiting distributions from the sorted vector of replicated statistics.

[*Received December 1990. Revised February 1992.*]

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Andrews, D. W. K. (1989), "Tests for Parameter Instability and Structural Change With Unknown Change Point," Discussion Paper 943, Yale University, Cowles Foundation.

Banerjee, A., Dolado, J., and Galbraith, J. W. (1990), "Recursive Tests for Unit Roots and Structural Breaks in Long Annual GNP Series," unpublished manuscript, University of Florida, Dept. of Economics.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992), "Recursive and Sequential Tests for a Unit Root: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10, 271–287.

Campbell, J. Y., and Mankiw, N. G. (1987), "Permanent and Transitory Components in Macroeconomic Fluctuations," *American Economic Review: Papers and Proceedings*, 77, 111–117.

——— (1988), "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics*, 102, 875–880.

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

Christiano, L. J., and Eichenbaum, M. (1989), "Unit Roots in Real GNP: Do We Know, and Do We Care?" Discussion Paper 18, Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics.

Chu, C.-S. J. (1989), "New Tests for Parameter Constancy in Stationary and Nonstationary Regression Models," unpublished manuscript, University of California-San Diego, Dept. of Economics.

Clark, P. K. (1987), "The Cyclical Component of U.S. Economic Activity," *Quarterly Journal of Economics*, 102, 797–814.

Cochrane, J. H. (1988), "How Big Is the Random Walk in GNP?" *Journal of Political Economy*, 96, 893–920.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

——— (1981), "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49, 1057–1072.

Hansen, B. E. (1992), "Tests for Parameter Instability With I(1) Processes," *Journal of Business & Economic Statistics*, 10, 321–335.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Ouliaris, S., Park, J. Y., and Phillips, P. C. B. (1989), "Testing for a Unit Root in the Presence of a Maintained Trend," in *Advances in Econometrics and Modelling*, ed. B. Raj, Hingham, MA: Kluwer Academic.

Park, J. Y., and Phillips, P. C. B. (1988), "Statistical Inference in Regressions With Integrated Processes: Part 1," *Econometric Theory*, 4, 486–498.

Perron, P. (1988), "The Hump-shaped Behavior of Macroeconomic Fluctuations," unpublished manuscript, Université de Montréal, Dept. of Economics.

——— (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

——— (1990), "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," unpublished manuscript, Princeton University, Dept. of Economics.

Perron, P., and Vogelsang, T. J. (1991), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis: Corrections and Extensions of Some Asymptotic Results," unpublished manuscript, Princeton University, Dept. of Economics.

Phillips, P. C. B. (1987), "Time Series Regression With a Unit Root," *Econometrica*, 55, 277–301.

——— (1988), "Multiple Regression With Integrated Time Series," *Contemporary Mathematics*, 80, 79–105.

Ploberger, W., Krämer, W., and Kontrus, K. (1989), "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

Rappoport, P. (1990), "Testing for the Frequency of Permanent Shifts in Time Series," unpublished manuscript, Rutgers University, Dept. of Economics.

Rappoport, P., and Reichlin, L. (1989), "Segmented Trends and Non-stationary Time Series," *Economic Journal*, 99, 168–177.

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

Schwarz, G. (1978), "Estimating the Dimensions of a Model," *The Annals of Statistics*, 6, 461–464.

Shapiro, M., and Watson, M. (1988), "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual*, 3, 111–148.