The Uncertain Trend in U.S. GDP

Charles R. Nelson and Christian J. Murray
University of Washington

January 14, 1997
(Corrections January 21, 1997)

Abstract: Several recent papers conclude that U.S. real GDP is trend stationary, implying that all shocks are transitory and the long run path is deterministic. These inferences fail to take into account two problems: the distortion of test size in finite samples due to data-based model selection, and the fragility of unit root tests in the face of plausible departures from the maintained hypothesis of temporal homogeneity. Indeed, additive outliers that alter the level of output for only one period reliably trigger false rejections of the unit root hypothesis when it is true and signal the presence of permanent shifts in trend that did not occur. Trend stationarity is not supported by the more homogeneous post-war data and if imposed would imply business cycles of implausible duration and pattern - the economy was 8% below the trend line in 1994.

Key words: business cycle, unit root tests, largest AR root, stochastic trend, deternimistic trend.

We acknowledge with thanks the helpful comments of Tim Cogley, Charles Engel, Neil Ericsson, Jon Faust, John Rogers, Glenn Rudebusch, Richard Startz, David Wilcox, and Eric Zivot, but responsibility for errors and the views expressed is entirely the authors'. This research was supported by the Ford and Louisa Van Voorhis endowment at the University of Washington.

Address correspondence to:
Charles R. Nelson
Box 353330, Dept. of Economics
University of Washington; Seattle, WA 98195
cnelson@u.washington.edu
(206) 685-1382

1. Introduction

Recent papers bring the literature full circle on the issue of whether the trend in U.S. real GDP is deterministic or stochastic. The modeling of aggregate output as transitory fluctuation around a deterministic trend line was routine in empirical work until Nelson and Plosser (1982) showed that data for 1909-70 were consistent with the hypothesis that the trend is instead a non-stationary stochastic process akin to a random walk. Such processes contain a unit root in their autoregressive representation and require first differencing for stationarity. The model suggested by Nelson and Plosser implies that the stochastic trend contributes more to variation in output than does the transitory component. They argued that an economic implication of this finding is that real shocks are much more important than previously thought, since they will have a permanent impact on output, while monetary and fiscal shocks, presumably transitory in their effects, play a less important role.

Perron (1989), however, found that the same data reject the stochastic trend hypothesis in favor of the deterministic alternative if a break in the level of the trend is allowed to occur at 1929. His model implies that there has been one permanent shock to output during the 1909-70 period, that being a negative one, and that all other shocks have been transitory. Zivot and Andrews (1992) showed that this finding still holds after critical values are adjusted to reflect estimation of the break date.

Most recently, Ben-David and Papell (1995), Cheung and Chinn (1996), and Diebold and Senhadji (1996) have conducted tests with longer time series, extending U.S. output data back to 1870 and forward to the more recent past. All find that the longer time series strongly reject the stochastic trend hypothesis in favor of a deterministic trend without breaks. Diebold and Senhadji conclude that this result is robust to variation in sample period. These papers would suggest that as the more data has become available, the evidence has become sharper, moving strongly in the direction of long-run determinism, leaving no role for permanent shocks.

Whether the trend in aggregate output is deterministic or stochastic has far-reaching implications for modeling the economy and for judging the success of macro-stabilization policy. The deterministic trend view implies that it is only because of transitory shocks, presumably primarily monetary and fiscal in origin, that the economy deviates from a smooth, constant-growth-rate path. The performance of monetary policy should then be measured by its success in achieving small departures from that path. If, on the other hand, shocks to the trend component are an important source of macro-economic fluctuations, then the modeling and identification of real shocks becomes critical for the conduct and evaluation of monetary policy.

The two views of trend also have strikingly different implications for long run uncertainty: under the deterministic view, long run uncertainty is limited by the stationarity of the cycle, while under the stochastic trend view, uncertainty about future output grows without bound.

This paper examines the robustness of recent findings with respect to two issues: the finite sample implications of data-based model specification and the effect on test size of plausible departures from the maintained hypothesis that the data are generated by a homogeneous process. Section 2 of the paper reviews standard unit root test on U.S. real GDP 1870-1994 and examines the data for homogeneity across sub-periods. Section 3 presents Monte Carlo experiments designed to study the two issues of size and robustness to departures from homogeneity. Section 4 focuses on the evidence from the post-war period which we regard as more likely to represent a homogeneous sample. Section 5 summarizes our results and presents our conclusions.

2. Trends and Non-homogeneity in U.S. real GDP

The evidence against the unit root hypothesis reported in recent papers is reflected in the test results shown in Table 1 for U.S. real GDP 1870-1994. The data is from Maddison (1995), incorporating estimates of early data due to Balke and Gordon (1989). Before interpreting these results, we briefly review the tests and their assumptions.

The test developed in Dickey (1976), Fuller (1976), and Dickey and Fuller (1979) takes as its null hypothesis a unit root in the AR representation of the time series. The first difference then follows an AR process of known and finite order, say k, with i.i.d. innovations. The alternative hypothesis is that the largest AR root is smaller than one so the series is a stationary AR of order k+1 added to a deterministic trend. The Dickey-Fuller test runs the following regression in which y_t denotes the natural log of GDP:

$$y_t = \rho y_{t-1} + \alpha + \beta t + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t$$

Dickey and Fuller showed that the t-statistic for the unit root null hypothesis ($\rho=1$) has a non-standard distribution, and they tabulated critical values for various sample sizes. Under the alternative hypothesis, $\rho<1$ and y is said to be “trend stationary” in the terminology of Nelson and Plosser.

In practice, the value of k, the number of lagged first differences that "augment" the model, is unknown and is chosen by a data-dependent procedure. Nelson and Plosser argued that the choice of k is one between size and power. If k is too large, then the test will have the correct size, but sacrifice power to the estimation of extraneous lag coefficients. If k is too small, then the test will not have the correct size since relevant lags have been omitted. Following the principle of parsimony advocated by Box and Jenkins (1970), they started with k=0, adding lags until the t-statistic on the next lag was not significant at the .05 level. The prior implicit in this procedure is that dynamics are of low order.

Building on work later published in Hall (1994), Campbell and Perron (1991) suggested starting with a maximum value of k chosen *a priori*, deleting lags until encountering a t-statistic greater than 1.6 in absolute value. This is the procedure followed by Perron, Zivot and Andrews, and others. Theoretical support for this has been provided by Hall for the pure AR case and by Ng and Perron (1995) for the ARMA case. They show that if the maximum k is greater than the true value of k, then the limiting distribution of the unit root test statistic is that derived by Dickey and Fuller. We refer to this is the backward selection strategy, and the Nelson-Plosser procedure as the forward selection strategy. Which performs better in terms of size and power for a given sample size has not been established.

Perron (1989) generalized the Dickey-Fuller test to allow for the possibility of structural change taking the form of a one-time break in level or slope or both. In the case of a break in level only, which he considered appropriate for U.S. real GNP, the Perron test adds step and impulse dummy variables to the Dickey-Fuller regression:

$$y_t = \rho y_{t-1} + \alpha + \beta t + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \delta S(b)_t + \gamma I(b)_t + \varepsilon_t .$$

where S is a "step" dummy that is zero through year b and one thereafter, and I is an "impulse" dummy that is one in year (b+1) only and zero otherwise. Under the null hypothesis that $\rho=1$, the impulse dummy accounts for a break in level, while under trend stationarity alternative, the step dummy does so. He provided critical values for testing the unit root hypothesis under the assumption that the break date is known and the innovations are i.i.d.

The test of Zivot and Andrews (1992) differs from the Perron test in two regards. First, the null hypothesis is that the series has a unit root but does not contain a break; accordingly, their test regression does not include an impulse dummy. Second, Zivot and Andrews recognize that if the break date is not known *a priori* it is estimated from the data. They consider the estimated break date to be that

which maximizes the absolute value of the unit root statistic. Innovations are assumed to be i.i.d. The test regression is

$$y_t = \rho y_{t-1} + \alpha + \beta t + \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \delta S(\hat{b})_t + \varepsilon_t$$

where $\hat{b}$ is the estimated break date. Critical values given by Zivot and Andrews are larger in absolute value than Perron's. They confirmed Perron's 1929 break date and his rejection of the unit root hypothesis for the Nelson-Plosser real GNP series.

Results of these tests are reported in Table 1 for the full Maddison sample and for the 1909-1970 sub-period studied by Nelson and Plosser. The number of lagged first differences, k, is chosen alternatively by the backward and forward selection methods discussed above. The Zivot-Andrews procedure identifies 1929 as the break date for both time spans. Several features of the results seem worthy of note:

1. The unit root hypothesis is strongly rejected by all three tests for the full period, and by the Perron and Zivot-Andrews tests for the Nelson-Plosser sub-period. The p-value for the augmented Dickey-Fuller (ADF) test for the Nelson-Plosser period is in the range .05 to .10, which is less favorable to the unit root hypothesis than the result originally reported by Nelson-Plosser using the data available prior to the revisions by Balke and Gordon.

2. Rejection of the unit root is stronger when break dummies are included, and is as strong or stronger when the break date is assumed to be known rather than estimated.

3. Evidence against the unit root is stronger for the full time period.

4. The choice of k differs greatly between backward and forward selection procedures. The backward often chooses the maximum while the forward in every case chooses only one.

5. For the Perron and Zivot-Andrews tests, backward selection of k results in smaller p-values.

6. The Step dummy is highly significant by conventional standards in every case. Its t-statistic is larger in absolute value for the Zivot-Andrews regression, which excludes the impulse dummy, and when k has been chosen by backward selection.

These unit root tests have as a maintained hypothesis that the series is homogeneous, in particular that the first differences are generated by an AR(k) process with fixed parameters and innovations that are i.i.d. It is unclear how deviations from these maintained

hypotheses might affect size or power. This is a concern since GDP exhibits substantial non-homogeneity over sub-periods, the dating of which is strongly suggested both by historical events and by changes in data collection.

The most recent estimates of data to 1929 are those by Balke and Gordon (1989) used by Maddison, and by Romer (1989). Both build on the pioneering methodology of Kuznets (1941, 1946) and the extensions made by Kendrick (1961) and Gallman (1966). Briefly, the Kuznets methodology relies on trends extrapolated between benchmark years thought to be similar in their stage of the business cycle. Deviations of GNP from trend are based on indicator variables such as commodity output. It would be surprising if construction of the series as deviations from trends did not affect its time series representation and, particularly, unit root tests. Indeed, the ADF test applied to the Balke-Gordon data rejects the unit root at the 10% level. Contemporaneous measurement of GDP began in 1929 with the work of the Dept. of Commerce.

The period immediately following 1929 is one of unprecedented upheaval in the U.S. economy: banking panics on a scale never before seen and repeated failures by the relatively new Federal Reserve System, established only in 1916, to restore order to the banking system; see Friedman and Schwartz (1963). Before the economy had recovered from the Great Depression it was jolted by the commencement of World War II. The delineation of the years 1930-1945 as possibly having different statistical properties does not rely on observing the GDP data itself. The abnormality of the Depression years is manifest in many economic variables and in extensive institutional change, while World War II is certainly an external shock of extraordinary magnitude.

Table 2 displays sample statistics for both growth rates and linearly detrended log GDP for the full 1870-1994 period and the sub-periods 1870-1929, 1930-1945, and 1946-1994 (post-war growth rates start in 1947 since 1946 is the first post war GDP level). Coefficients of AR(3) models fitted to each series are reported as a way of summarizing dynamics; order 3 being large enough to capture a wide range of behavior while not relying on the data to determine the lag. The Jarque-Bera test of normality is based on the residuals from these models. In addition to the Balke-Gordon data for 1870-1929, we also include growth rates for the Romer data.

We note that the 1930-45 period differs from the earlier and later periods, not so much in mean growth rate or mean deviation from fitted trend, but in dispersion. This is also evident in plots of the levels with the fitted trend line in Figure 1 and growth rates in Figure 2. Standard deviation and range are about four times larger during 1930-45 than in the post-war period. This difference persists in the

standard errors of AR(3) models fitted to the sub-periods and can be compared with the three fold increase in the volatility of stock returns reported by Schwert (1989b) for the Depression period. Normality is strongly rejected by the Jarque-Bera test for the full period, while each sub-period is consistent with normality. This suggests that the 1870-1994 period is a sequence of three sub-periods which are very different from each other but are internally homogeneous. If true, it would be inappropriate to apply unit root tests, which assume homogeneity, to the entire period.

3. The Sensitivity of Unit Root Tests to Departures From Homogeneity and the Effect of Data-based Model Selection on Size

If the historical data are not homogeneous, then we need to know how unit root tests are affected by plausible departures from homogeneity. Conceivably, the sharp decline in output following 1929 and the surge during World War II were unusually large but only temporary departures of output from its natural level. In this section we study the effect of additive outliers in the *level* of the series on tests of the unit root hypothesis and on inferences of a permanent break in trend using Monte Carlo methods. We also obtain estimates of the actual size of these tests under the null hypothesis when the order of the AR model used in the test regression is based on the data.

Our data generating process under the null hypothesis of a unit root is based on the first differences of the post-war Maddison log real GDP. The SIC criterion chooses ARMA(0,0), while the AIC chooses ARMA(1,0) but the estimated AR coefficient has a p-value of .24. We adopt the AR(1) model as the data generating process for the growth rate to allow some dynamics. The process is run for 20 periods before recording a realization corresponding to the years 1870-1994. After integrating the realization of first differences to obtain the levels of the underlying series, we add one of a number of types of outliers to see its effect on the test statistics. Thus, the underlying process is:

$$\Delta y_t = .17\Delta y_{t-1} + .027 + \varepsilon_t$$
$$\varepsilon_t \sim \text{i.i.d. } N(0,\ 0.025)$$

and the observed data, say y*, is then

$$y_t^* = y_t + O_t$$

where the outlier sequence $\{O_t\}$ varies across experiments.

A One-Period Additive Outlier

In the first experiment, reported in Table 3, we have subtracted a fixed quantity from the level of simulated log real GDP in 1930 only. The value of O_{1930} ranges in successive experiments between 0 and -.4, the latter representing a one third reduction in output. The dip lasts for only one year, so the observed series resumes its underlying path in 1931 with no permanent change in level. For comparison, we also include the non-parametric unit root test of Phillips and Perron (1988) which does not rely on a finite order AR representation of the series but instead employs a correction for serial correlation based in part on the spectral representation of the innovation sequence at frequency zero. The quadratic spectral kernel is used to estimate the spectrum, and Andrews' (1991) selection procedure is used to determine the number of autocovariance terms included in forming the estimate of the spectrum.

We report actual frequencies of rejection of the unit root hypothesis at the nominal 5% significance level based on critical values reported in Fuller (1976), Zivot and Andrews, and Perron (with corrections to the Perron critical values by Zivot and Andrews), respectively. For the Perron and Zivot-Andrews tests the critical values are asymptotic, but for the ADF test the finite sample critical values are exact if k is known to be zero and the innovations are Gaussian. The Phillips-Perron test has the same limiting distribution as the ADF test and so uses the same critical values. The number of lagged first differences included in a regression, denoted k, is selected alternatively by the backward and forward procedures described above. Note that in searching for the break date in the Zivot-Andrews test, the selection of k is repeated at every potential break date. In Table 3 we report the frequency with which the selected k is zero or one. Finally, we report the frequency with which the t-statistic for the step dummy is significant at a nominal 5% level. Two panels give results for sample lengths of 62 and 125 corresponding to the Nelson-Plosser 1909-1970 sub-period and the full 1870-1994 period, respectively, based on 1,000 replications.

In the experiment reported in the first line of each panel, no outlier has been added. In this case the series does in fact have a unit root and i.i.d. innovations, so the frequency reported for each unit root test is its size, the probability of rejecting the null hypothesis when it is true. It is apparent that the method for selecting k in the ADF, Perron, and Zivot-Andrews tests has a large impact on the value of k and, more important, on the sizes of the tests. Forward selection almost invariably selects k=0 or 1, while backward selection usually chooses a large k. In the case of the ADF and Perron tests, forward selection produces roughly the correct size, while in backward selection results in a size that is too large, about 10% when T=62 and only slightly less when T=125. The Zivot-Andrews test suffers from size distortion under both selection strategies, but the distortion is entirely due to selecting k from the data. When the correct value k=1 is imposed, the actual size is correct at 5%.

The forward selection strategy works well in this situation because the AR process is of low order and the single AR coefficient is small. The penalty for underestimating k in this case is slight since the process is well approximated by white noise. The prior that underlies the principle of parsimony is correct for our data generating process. While Hall showed that the backward strategy is valid asymptotically, his Monte Carlo results demonstrated that there may be substantial size distortions in finite samples such as we see here. It is clear that lag selection is not a simple trade-off between size and power, with strategies favoring large k offering more correct size. The analogy to including extraneous variables in a regression which use up degrees of freedom but do not create a bias is misleading because the particular value of k is based on pretesting. If k were set *a priori* and was larger than the true value of k in any particular case, then the test *would* have correct size but lower power; see Ng and Perron (1995). An appropriate analogy is to the problem of data-based selection of instruments in 2 stage least squares, where Hall, Rudebusch, and Wilcox (1994) have shown that searching for the best instruments severely distorts the size of tests on structural coefficients.

The Phillips-Perron test relies on the data for selecting lag length for truncation of the autocovariance function used in estimation of the spectrum at frequency zero, rather than for selecting AR order. It is the only one of the tests considered here that has too small a size. The size distortion of this test is evidently dependent on the form of the autocorrelation function, since Schwert (1989a) found that the size of this test was too large in the MA(1) case.

Finally, the empirical sizes of the t-tests for the step dummy are much too large; about 20% in the Perron regressions and over 90% in the Zivot-Andrews regressions, sample length seeming to have little influence. The size of the t-test for the impulse dummy in the Perron regression is also excessive (but not shown in Table 3). It is evident that standard distribution theory does not hold for these coefficients, since there is not in fact any break in level. The lesson for practitioners is clear: a “significant” t-statistic for step or impulse dummies does not constitute evidence of a permanent shift in level.

Successive experiments reported in Table 3 add outliers of -.1, -.2, -.3, and -.4 to the level of simulated log real GDP in 1930 only, after which it returns to the underlying process. This range of magnitudes was motivated by the range of deviations reported in Table 2 for the 1930-1945 period. The observed process still contains a unit root, but it is no longer homogeneous.

What we see in Table 3 is that this type of non-homogeneity triggers rejection of the unit root hypothesis with high probability. As the size of the outlier increases, rejection rates for all tests rise and

the forward k selection is associated with higher rejection rates. Nominal significance of the step dummy becomes greater as well. Doubling sample length reduces the probability of spurious rejection of the unit root but does not mitigate the apparent significance of the step dummy. Rejection frequencies for the unit root are plotted in Figures 3-5 for the ADF, Perron, and Zivot-Andrews tests, respectively. These results confirm and extend those of Frances and Haldrup (1994) for the ADF test who showed that additive outliers bias that test toward rejecting the unit root null when k is known.

How do we understand the rejection of the null hypothesis of a unit root when the data do in fact contain a unit root? The null hypothesis is rejected because the maintained hypothesis of homogeneity is violated when we introduce an additive outlier; the first difference is no longer an AR(1) with i.i.d. innovations. It is completely general that rejection of a null hypothesis does not imply that the alternative hypothesis depicted by the test regression is true. Faced with the choice between the unit root null hypothesis and the trend stationary alternative when neither is true, these tests reject the null. If misinterpreted, these tests spuriously signal trend stationarity, with a level shift if allowed, when in fact the series has a unit root and the outlier event affect only one observation.

It is interesting that frequencies of rejection of the unit root diminish when the sample size is doubled, seemingly at odds with the idea that the power of a test should increase with sample size. However, the present case is one where the null hypothesis also becomes less wrong as sample size grows, since the departure from homogeneity becomes relatively less severe the longer the time series.

Additive Outliers That Persist

The next set of experiments, reported in Table 4, is designed to capture various dimensions of the turmoil that characterized the period 1930 through 1946. Starting from the benchmark case of no outlier (as in Table 3), we go to fixed outliers of -.2 added, successively, to 1930 only (as in Table 3), to the five years 1930-34, to the ten years 1930-39, and to every year from 1930 on. In every case the series contains a unit root, and in the last also contains a permanent break in level. Further experiments add a stochastic outlier generated by an AR(2) process modeled on departures from a local trend connecting 1929 to 1946, in one case with purely random start-up and in the other with actual declines in 1930 and 1931 establishing initial conditions for each realization of the AR(2). Finally, we impose the fixed pattern of actual de-meaned cumulative changes from 1930 through 1945. This last case is in the spirit of the experiments reported by Kilian and Ohanian (1996).

Persistence in itself reduces the frequency of rejection of the unit root in the ADF, Phillips-Perron, and Perron tests. When the -.2 outlier persists indefinitely, the size of the Perron test is close to 5%. Indeed, the Perron regression is correctly specified in that case with inclusion of the impulse dummy at the correct date, allowing for a permanent change in level, while the step dummy is irrelevant. Indeed, the impulse dummy is almost always significant (not shown in Table 4) in the Perron regressions, however, the size of the t-test for the step dummy is still too large. If those t-statistics were taken at face value, they would imply a permanent break in the growth rate which did not in fact occur.

Rejection of the unit root by the Zivot-Andrews test is less sensitive to persistence in the outlier. When the outlier persists indefinitely, the Zivot-Andrews test rejects the unit root null much more frequently than does the Perron test, and the step dummy is almost always significant. Recall that the null hypothesis for the Zivot-Andrews test is a unit root without any break in level or slope, so in this case their null hypothesis is untrue. Recall also that the regressions run for the Zivot-Andrews and Perron tests differ in two respects. First, the Zivot-Andrews regression does not include the impulse dummy, only the step dummy, so under the unit root the shift in level of -.2 represents a very large residual. When the Zivot-Andrews test was rerun with the impulse dummy included in the regression, the results were essentially the same. Second, the break date is estimated from the data. The break date is correctly estimated to be in the range 1920-29 in 76% of the replications for T=62 and in only 52% for T=125. Apparently, then, it is the search for the break date that accounts for the difference in results between the Perron and Zivot-Andrews tests. This suggests the importance of information about the timing of structural change.

In the last three experiments reported in Table 4, the underlying process is inundated with a high amplitude wave which distorts the level of the series for 16 years but then is gone. In all three, the unit root and no-step-dummy null hypotheses are rejected often, except by the Phillips-Perron test. The AR(2) used in the first two of these experiments is designed to mimic the amplitude and persistence of the 1930-1945 period, but does not seem to account for three striking characteristics of the historical results of Table 1: the sharp contrast between the extreme values of k chosen in forward verses backward selection, the contrast in p-values for the unit root based on that choice, and the stronger rejection of the unit root in the longer time series. Judging from the results in the last line, it is the *particular* pattern of real GDP during the period 1930-45 that accounts for much of these effects. In particular, when we use the actual historical pattern in repeated realizations, the value 0 or 1 is almost always chosen for k by forward selection and much less frequently by backward selection, a greater contrast emerges in

rejections of the unit root based on that choice (at least for T=62), and rejection of the unit root is more frequent in the longer series.

4. Evidence from the Post-war Annual Data

While more data is usually preferred to less, we find two compelling reasons to focus on the post-war GDP data, 1946-94, for testing the unit root hypothesis. Recall that the Balke-Gordon data used by Maddison up to 1929 are constructed by linear interpolation between benchmark years, so unit root tests are presumably biased toward rejection. During the next 16 years the economy was subject to the large shocks associated with the Great Depression and World War II. The experiments reported above imply that even if these events were only transitory, they could account for rejection of the unit root hypothesis and be misconstrued as evidence of structural change. By focusing on post-war data we hope to minimize the chance of spurious rejection of the unit root hypothesis due to violation of the homogeneity assumption.

However, as first pointed out by Perron and extended with Monte Carlo evidence by Hendry and Neale (1991) and Campos, Ericsson, and Hendry (1996), a permanent level shift biases tests toward acceptance of the unit root if a break is not allowed under the alternative. Indeed, the existence and causes of such a structural change, in the form of a productivity slowdown, have been discussed since the 1970s and are the subject of a large and continuing literature; see Baily and Gordon (1988) and Gordon (1993). Perron reported evidence of a break in the slope of the trend function at 1973 in quarterly post-war data, although Zivot and Andrews were not able to confirm this finding when they searched for the break. Further, Gordon (1993) has pointed out that a productivity slow-down will be obscured in a series based on fixed weight price deflators, such as the real GDP data used by Maddison. The recently released chained-dollar real GDP measure is free of this distortion, so we shall report unit root tests based on this data as well. Before taking up the issue of structural change, we report the results of ADF tests and confidence intervals for the largest AR root.

ADF Tests and Confidence Intervals for the Largest AR Root.

Results of the ADF test are strongly consistent with a unit root in the post-war segment of the Maddison data. The backward selection procedure chooses one lag and gives an ADF test statistic of -2.18 which is the median of the Dickey-Fuller distribution. The forward procedure chooses zeros lags and the ADF test statistic is -1.44 which is in the right tail of the sampling distribution; simulation gives a p-value of 0.84. However, a common criticism of ADF tests is that they

have low power against the alternative hypothesis that the largest AR root is close to unity. To address this issue, we perform the unit root test proposed by Elliot, Rothenberg, and Stock (1996), which they refer to as DF-GLS. The DF-GLS test employs a local-to-unity detrending procedure designed to maximize power against local alternatives. As in the conventional ADF test, the DF-GLS test chooses one and zero lags in the backward and forward procedures respectively. The test statistic for one lag is -2.33 with a p-value of 0.29 obtained by simulation, and when no lags are included a statistic of -1.59 with a p-value = 0.73.

Although these results are consistent with a unit root process, we know that they are also consistent with a range of trend stationary alternatives. It is simply not possible to distinguish in a finite sample between the realization of a unit root process and a trend stationary one with a root close enough to unity. This is the observational equivalence problem identified by Nelson and Plosser and emphasized by Christiano and Eichenbaum (1990) among others. Cochrane (1991) has identified a corresponding observational equivalence between a trend stationary process and one with a stochastic trend where the variance of the innovations is small. Thus, the range of models that cannot be rejected by any finite data set must always include both unit root and trend stationary alternatives. We would like to know how wide that range is in any given case.

To see the range of the largest AR root, ρ, that is consistent the post-war Maddison data, we computed one-sided confidence intervals (excluding explosive values *a priori*) using the procedure developed by Stock (1991). These are based on inverting the ADF test statistic to determine the values of ρ consistent with it; the results are in Table 5.

Since the Stock confidence interval is asymptotic in nature, we also constructed confidence intervals for ρ based on finite sample realizations of trend stationary representations of the data. Data was generated according to:

$$y_t = d_t + c_t \text{ where } d_t = \alpha + \beta t \text{ and } c_t = (\rho + \phi)c_{t-1} - \phi c_{t-2} + \varepsilon_t .$$

where ϕ is non-zero in the case k=1 and zero if k=0 and parameters are estimated from the post-war data. The only variable parameter in the experiment is ρ. To determine the probability of obtaining the observed test statistic under various trend stationary alternatives, we let ρ vary between 0.65 and 1, by increments of 0.01, generating 10,000 time series under each specification, and computing the frequency with which the generated unit root test statistics exceed the value in the historical data. In Table 5 we report empirical 95% and 90% confidence intervals based on both ADF and DF-GLS statistics.

We note that the analytical confidence intervals of Stock do not differ greatly from the corresponding empirical intervals that are also based on the ADF statistic. The empirical interval based on the DF-GLS statistic is wider in spite of the greater power of that test because the p-value of the actual statistic is less favorable to the unit root than was the ADF statistic. Our results here are similar to those of Rudebusch (1992) who showed that the ADF statistic (k=1) from the Nelson-Plosser real GNP series is consistent with an AR root of 0.82. We also note that all the intervals are considerably wider for k=1 than for k=0. Again, inferences are substantially affected by this nuisance parameter.

To give an economically meaningful calibration to ρ, we also report the corresponding average length of the cycle, defined here as the time between crossings of the trend line from below. The mean distance between these upcrosses serves as a measure of cycle length in cases where the characteristic equation does not contain complex roots and was obtained here by simulation. Under trend stationarity, these cycles would be generated by monetary and fiscal shocks and must be entirely transitory. However, the cycles lengths implied by these confidence intervals seem implausibly long if they are to be interpreted as *minimum* estimates of the persistence of such shocks. Indeed, the average length of NBER business cycles, about 6 years, is well below the low end of the range of these minimum estimates.

Two estimates of the cyclical component implied by the trend stationary hypothesis are plotted in Figure 6 for the post-war period. In one case we take deviations from the trend line fitted to post-war observations only, in the other deviations from the trend line fitted over the entire 1870-1994 period, and the two are indistinguishable. Neither corresponds to what most observers would regard as a reasonable description of economic activity during the post-war period. First, there are only five upcrosses, with a mean cycle length of 9.25 years. Second, the sign of the usual cycle is reversed: the economy was below “trend” during most of the post-war period until 1963, above trend for the next 20 years, and has been below trend since 1988, finishing the sample period in 1994 about 8% below trend. This picture is contradicted by employment-based measures as well as the Fed’s index of capacity utilization which suggest that the economy was strong in the 1960s, that the 1970s was a period of poor performance, and that the economy was close to full employment in 1994. Indeed, the danger of renewed inflation in an economy close to full employment has been a preoccupation of the Fed in recent years. A similar picture emerges if we allow for a break in the level of the trend line at 1929, but the mid-1990s slump is even more severe, the economy falling about 14% below trend in 1994.

The Variance Ratio Statistic

Cochrane (1988) criticized the use of unit root tests to determine the long run dynamic properties of a time series. Since unit root tests rely on parsimonious representations of the short run dynamics, they only use the first few terms of the autocorrelation function and may fail to capture the long run behavior of a time series. Cochrane advocated a non-parametric measure of long run persistence, the ratio of the variance of the j^{th} difference to the variance of the first difference, normalized by the factor 1/j. If a series is trend stationary, the variance ratio approaches zero as $j \to \infty$. If a series is integrated, then it can always be decomposed into a random walk plus a stationary component (Beveridge and Nelson, 1981) and the variance ratio then approaches the ratio of the variance of the random walk to the variance of the first difference, so it is unity for a pure random walk. Thus, the variance ratio provides an estimate of the contribution of the stochastic trend to the long run dynamics of a time series.

Figure 7 plots the sample variance ratios for j=1,..,20 for post-war annual GDP based on Cochrane's unbiased estimate of the variance of the j^{th} difference of a random walk (his equation A3). It is evident that the variance ratio for post-war GDP shows no tendency to decay toward zero. At the 20 year horizon, the variance ratio is 1.15, suggesting that variation in post-war GDP is dominated by its random walk component, in contrast to the finding for the longer time series reported by Cochrane.

We have constructed confidence intervals for ρ based on the sampling distribution of the variance ratio, similar in spirit to our confidence intervals for ρ based on simulation of unit root statistics. Using the same data generating processes for k= 0 and 1, we determined the range of values of ρ that is consistent with the observed variance ratio at a given lag. We arbitrarily focus on values of j equal to 10 and 20 years. The results, reported in Table 5, are qualitatively similar to those obtained in the unit root testing framework. The confidence intervals are considerably tighter for the larger value of j. Two factors drive this result. One is that the sample statistic is more favorable to the unit root for j=20 than for j=10. The other is that for trend stationary alternatives, the decay in the variance ratio will be more apparent for larger values of j. The cycle lengths implied by the variance ratio for j = 20 are as implausible as those implied by the unit root statistics.

The Productivity Slow-Down and Chained Real GDP

We now turn to the issue of a productivity slow-down in the U.S. economy and any implications it might have for tests of the unit root hypothesis. It is a fact that growth has been slower since 1973: the

growth rate over the period 1947-1973 was 3.9% while in 1974-1994 it fell to 2.4%. Whether this difference is statistically significant and, if so, whether it represents an abrupt structural change or a gradual evolution toward slower growth is unclear. Model B of Perron allows for a break in the growth rate under the trend stationary alternative, though not under the null. It differs from Model A in replacing the step dummy with a "ramp" dummy that is zero through the break date then increasing arithmetically, so the trend function is allowed to bend but not shift. Perron applied this test to post-war quarterly real GNP, 1947-86, setting the break date at 1973:1, and rejected the null hypothesis. Zivot and Andrews estimated the break date at 1972:2 but did not reject the unit root. Both used the backward lag selection procedure, choosing k = 10 quarters. Since the null hypothesis in this test encompasses *both* the unit root and constant growth, evidence against the null can be due to trend stationarity, structural change, or both, or to some other departure from the null.

For the post-war segment of Maddison's data, backward lag selection chooses seven lags, 1977 is the estimated break date, and the null hypothesis is rejected at the 10% level (critical values for Model B are tabled in Zivot and Andrews). Forward selection chooses one lag, picks 1973 as the break date, but the test statistic is not significant at the 10% level. However, just as lag selection distorts significance levels in Model A it does also in Model B. Using the same data generating process that we have in prior simulations, we find that a nominal significance of level of 10% from the Zivot-Andrews' Table 3 corresponds to a rejection rate of 40% under backward selection and 18% under forward selection when the null is true. The p-values obtained by simulation for the sample statistics are 36% under backward selection and 27% under forward selection. Evidently, there is no evidence against the compound null hypothesis of a unit root and a constant expected growth rate when we recognize size distortions induced by selecting k.

Evidence of a slow-down may be obscured in real GDP based on fixed weight price indices, as Gordon (1993) has pointed out. Growth before the base year is understated and growth after it is overstated. The recently released chained real GDP data are free of this bias but are only available from 1959; see *Survey of Current Business* (1996). Model B rejects the null hypothesis at the 5% level in the chained data with both selection procedures choosing 1 lag. The estimated break date is 1969, the first break date possible given that eight lags are reserved. To obtain p-values that reflect data-based selection of k, we simulated the chained data as an AR(1) estimated in first differences in the spirit of prior experiments. The p-value obtained under backward selection is 42% and under forward selection 15%. Thus, the chained GDP strengthens our conclusion that there is a lack of evidence against the null hypothesis.

5. Summary and Conclusions

Recent research has demonstrated that standard tests reject the null hypothesis of a unit root in U.S. real GDP over the period 1870-1994 in favor of the alternative of stationarity around a log-linear trend. If valid, these findings would imply that all shocks are temporary and that the long run path of the economy is deterministic. This paper calls that inference into question on two grounds.

First, the size of these tests is distorted in finite samples by the necessary preliminary step of selecting the number of lagged first differences to be included in the regression. We show that the actual probability of rejecting the unit root hypothesis when it is true is substantially greater under data-based lag selection than the asymptotic significance levels upon which rejections of the unit root have been based.

Second, the long historical time series used in the literature violate the maintained hypothesis that the data generating process is temporally homogeneous. The period 1930-45 was one of unusually large disturbances that may have been largely temporary in their effect on the level of output. However, we find that outliers added to the level of a unit root process for only one period are sufficient to trigger rejections of the unit root hypothesis with high probability. Given the choice between two wrong models, the unit root tests lean towards trend stationarity although it is false.

These pitfalls remain if a potential break in the level of output is included in the model, but the problems of inference in finite samples become more severe. Conventional t-tests for break dummies are spurious, indicating a break when none has occurred. As Kilian and Ohanian (1996) also report, purely transitory disturbances trigger rejection of the unit root hypothesis and spurious inference that there has been a permanent break in level.

To avoid spurious rejection of the unit root due to heterogeneity in the data, we focus on the post-war segment of GDP. The data are consistent with the hypothesis of a unit root, as well as with a range of trend stationary alternatives. To give an economic meaning to this range, we calculate the mean cycle length implied by the *lower bound* of confidence intervals for the largest AR root and find that it is much longer than the average of NBER reference cycles. Thus, although unit root tests have low power against trend stationary alternatives close enough to unity, in this case they rule out trend stationary alternatives that would have economically plausible dynamics. Furthermore, the cycle produced by detrending post-war GDP contradicts employment based measures of economic activity; it implies below-trend performance during the 1960s, above-trend performance in the

1970s, and then a decline that puts real GDP 8% below trend in 1994. These results cast serious doubt on the trend stationary model as a credible representation of real GDP.

When the possibility that post-war GDP has undergone a productivity slowdown is considered, what little evidence there is against the unit root hypothesis is eliminated once the finite sample size distortion induced by lag selection is taken into account. The choice of models permitted by existing unit root tests is an unattractive one in this setting. Under the null hypothesis there is a unit root in output and no change in the underlying mean rate of growth; under the alternative, a one-time change in growth rate is permitted. Thus, a gradual but sufficiently large productivity slow-down would trigger a rejection of the null hypothesis, but not because the series is trend stationary or because there has been a discrete change in the mean rate of growth.

In our view, a constructive direction for modeling aggregate output will be one that moves away from the unit root issue and the use of dummy variables to represent shifts in level or growth rate. Determinism is not an hypothesis that is supported either in economic theory or in history. Dummy variables restrict the frequency of permanent shocks, and give no guidance as to the likelihood or size of future shocks. A statistical model implies a conditional distribution of future observations given the data, not simply an accounting of past events.

References

Andrews, Donald W. K., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

Baily, Martin Neil, and Robert J. Gordon, 1988, "The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power, *Brookings Papers on Economic Activity*, 2, 347-420.

Balke, Nathan S., and Robert J. Gordon, 1989, "The prewar business cycle reconsidered: new estimates of gross national product 1869-1908," *J. of Political Economy*, 97, 38-92.

Ben-David, Dan, and David H. Papell, 1995, "The great wars, the great crash, and steady state growth: Some new evidence about an old stylized fact," *Journal of Monetary Economics*, 36, no. 3, 453-75.

Box, George E. P. and Gwilym, M. Jenkins, 1970, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Campbell, John Y. and Pierre Perron, 1991, "Pitfalls and opportunities: What macroeconomists should know about unit roots," in *NBER Macroeconomic Annual*, eds. O. J. Blanchard and S. Fischer, MIT Press, Cambridge, 141-201.

Cheung, Yin-Wong, and Menzie D. Chinn, 1996, "Further investigation of the uncertain unit root in GNP," *J. of Business and Economic Statistics*, in press.

Christiano, Lawrence J. and Martin Eichenbaum, 1990, "Unit Roots in Real GNP: Do We Know and Do We Care?," in *Unit Roots, Investment Measures, and Other Essays*, ed. by Allen H. Meltzer, Carnegie-Rochester Conference Series on Public Policy, 32, 7-61.

Cochrane, John H., 1988, ``How Big Is the Random Walk in GNP?," *J. of Political Economy*, 893-920.

Cochrane, John H., 1991, "A Critique of the Application of Unit Root Tests," *J. of Economic Dynamics and Control*," 15, 275-84.

Dickey, David A., 1976, Estimation and hypothesis testing in nonstationary time series, Unpublished doctoral dissertation, Iowa State University.

Dickey, David A. and Wayne A. Fuller, 1979, "Distribution of the estimators for autoregressive time series with a unit root," *J. of the American Statistical Assn*, 74, 427-431.

Diebold, Francis X., and Abdelhak S. Senhadji, 1996, "The Uncertain Unit Root in Real GNP: Comment," *American Economic Review*, 86, 1291-8.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock, 1996, "Efficient tests for an autoregressive unit root," *Econometrica*, 64, 813-36.

Frances, Philip H., and Niels Haldrup, 1994, "The Effects of Additive Outliers on Tests for Unit Roots and Cointegration," *J. of Business and Economic Statistics*, vol. 12, no. 4, 471-78.

Friedman, Milton and Anna J. Schwartz, 1963, *A Monetary History of the United States, 1867-1960*, New York: NBER.

Gallman, Robert E. 1966, "Gross national product in the United States 1834-1909," in *Studies in Income and Wealth*, vol. 30, New York: Columbia University Press (for NBER).

Gordon, Robert J., 1993, "The Jobless Recovery: Does It Signal a New Era of Productivity-led Growth?," *Brookings Papers on Economic Activity*, 1, 271-306.

Hall, Alastair, 1994, "Testing for a unit root in time series with pretest data-based model selection," *J. of Business and Economic Statistics*, vol. 12, no. 4, 449-70.

Hall, Alastair R., Glenn D. Rudebusch, and David W. Wilcox, 1994, "Judging Instrument Relevance in Instrumental Variables Estimation," unpublished paper, Federal Reserve Board.

Kendrick, John W., 1961, *Productivity Trends in the United States*, Princeton: Princeton University Press.

Kilian, Lutz and Lee E. Ohanian, 1996, "Is there a trend break in U.S. GNP? A macroeconomic perspective," unpublished manuscript.

Kuznets, Simon S., 1941, *National Income and Its Composition*, 191-1938, New York: NBER.

_____, 1946, *National Product Since 1869*, New York: NBER.

MacKinnon, James G., 1991, "Critical values for cointegration tests," Ch. 13 in *Long-run Economic Relationships, Readings in Cointegration*, eds. R. F. Engle and C. W. J. Granger, Oxford, Oxford University Press.

Maddison, Angus, 1995, *Monitoring the World Economy 1820-1992*, Paris: OECD.

Nelson, Charles R. and Charles I. Plosser, 1982, "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10, no. 2, 139-162.

Ng, Serena, and Pierre Perron, 1995, "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag," *J. of the American Statistical Assn*, 90, 268-81.

Perron, Pierre, 1989, "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361-1401.

Phillips, Peter C. B. and Pierre Perron, 1988, "Testing for a unit root in time series regression," *Biometrika*, 75, 335-46.

Romer, Christina D., 1989, "The prewar business cycle reconsidered, new estimates of gross national product 1869-1908," *J. of Political Economy*, 97, 1-37.

Rudebusch, Glenn D., 1993, "The uncertain unit root in real GNP," *American Economic Review*, 83, 264-72.

Rudebusch, Glenn D., 1992, "Trends and random walks in macroeconomic time series: A re-examination," *International Economic Review*, 33, 661-80.

Schwert, G. William, 1987, "Effects of model specification on tests for unit roots in macroeconomic data," *J. of Monetary Economics*, 20, 73-103.

Schwert, G. William, 1989a, "Tests for Unit Roots: A Monte Carlo Investigation," *J. of Business and Economic Statistics*, 7, 147-60.

Schwert, G. William, 1989b, "Why does stock market volatility change over time?," *Journal of Finance*, vol XLIV, No. 5.

Stock, James H., 1991, "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series," *J. of Monetary Economics*, 28, 435-459.

Survey of Current Business,1996, volume 76, number 1/2 (January/February), U.S. Department of Commerce.

Zivot, Eric and Donald W. K. Andrews, 1992, "Further evidence on the great crash, the oil-price shock, and the unit root hypothesis," *J. of Business and Economic Statistics*, 10, 251-270.

Table 1
Tests for a Unit Root in U.S. Real GDP

Time Period & Test	Lags (3)	Unit Root t-statistic	Asymptotic p-value (4)	Step Dummy t-statistic (5)
1909-1970:				
ADF	1 only	-3.43	.05 - .10	N/A
Perron (1)	8	-4.89	< 0.01	-3.87
	1	-4.26	.01 - .025	-2.40
Zivot & Andrews (2)	8	-5.61	< 0.01	-4.63
	1	-4.72	.05 - .10	-3.03
1870-1994:				
ADF	6	-3.74	.01 - .05	N/A
	1	-4.14	< .01	N/A
Perron (1)	8	-5.58	< .01	-3.74
	1	-4.72	< .01	-2.31
Zivot & Andrews (2)	8	-6.10	< .01	-4.33
	1	-5.10	.01 - .025	-2.82

Notes:
(1) Break date assumed to be 1929; both step and impulse dummies are included.
(2) Break date maximizes unit root t-statistic; only the step dummy is included.
(3) The number of lags of first difference to be included is chosen by:
a) from an initial 8 lags, reduction until t>1.6 in absolute value, and
b) by successively increasing lags until t<2 in absolute value.
(4) Asymptotic significance levels are for the unit root t-statistic based on:
ADF: MacKinnon (1991) as provided in EViews.
Perron and Zivot & Andrews: as provided in Zivot and Andrews (1992), Table 2.
(5) The break date was in every case 1929, by assumption in the Perron tests, and by maximization of the unit root t-statistic in the ZA tests.

Table 2
Summary Statistics for U.S. Real GDP

Growth Rates	Mean	Std Dev	Range	AR(3): Coefficient Estimates			S E	J-B p
1870-1994	0.033	0.056	.18/-.23	0.27 *	0.00	-0.12	0.055	0.00
1870-1929 BG	0.037	0.048	.13/-.08	-0.28 *	-0.20	-0.04	0.048	0.91
1870-1929 R	0.038	0.035	.15/-.04	0.04	-0.25 *	0.04	0.031	0.85
1930-1946	0.026	0.118	.18/-.23	0.90 *	-0.16	-0.49	0.081	0.93
1947-1994	0.031	0.026	.09/-.02	0.20	-0.11	-0.19	0.025	0.71

Detrended	Mean	Std Dev	Range	AR(3): Coefficient Estimates			S E	J-B p
1870-1994	0.000	0.11	.31/-.37	1.13 *	-0.24	-0.1	0.052	0.00
1870-1929	0.029	0.082	.18/-.14	0.64 *	0.05	0.13	0.046	0.66
1930-1945	-0.100	0.222	.31/-.37	1.48 *	-0.57	-0.23	0.071	0.60
1946-1994	-.003	0.048	.08/-.10	1.09 *	-0.24	-0.05	0.024	0.35

Notes: * denotes asymptotic t-statistic significant at .05 level.
J-B p denotes significance level of Jarque-Bera test for Normality.

Table 3: Monte Carlo Study of Unit Root Tests and Lag Selection
when DGP is AR(1) in First Differences with Outlier Added to Level in 1930 Only.
Frequencies of Rejection of the Unit Root at 5% Level (UR), of Selecting k = 0 or 1,
and of Step Dummy Significance at 5% Level.
Lag k Selected Backward From 8 or Forward from Zero to Maximum of 8.

Nelson/Plosser Series Length; T=62

	ADF Test:					Phillips-	Perron Tests; 1929 Break Date Known					
	Backwardk			Forward k		Perron	Backward k			Forward k		
Outlier	U R	k=0,1		U R	k=0,1	U R	U R	k=0,1	Step	U R	k=0,1	Step
0	10	45		5	98	3	10	42	25	6	98	20
-0.1	8	57		5	100	8	26	43	31	19	99	26
-0.2	23	78		27	99	30	45	49	43	48	96	39
-0.3	38	89		50	99	62	53	42	47	61	85	45
-0.4	49	93		64	98	84	63	38	51	74	78	50

Maddison Series Length; T=125

Outlier	U R	k=0,1		U R	k=0,1	U R	U R	k=0,1	Step	U R	k=0,1	Step
0	8	46		7	97	4	9	45	19	6	97	17
-0.1	8	51		5	99	5	15	46	26	11	99	24
-0.2	13	68		13	99	14	30	51	32	31	97	31
-0.3	14	78		17	97	27	37	48	39	39	93	39
-0.4	20	72		21	89	49	45	31	46	50	63	45

Table 3 - Continued
Zivot-Andrews Test; Search for Break

Backward k			Forward k		
U R	k=0,1	Step	U R	k=0,1	Step
16	35	93	9	95	92
17	43	94	7	99	91
36	74	92	37	100	93
69	91	92	73	100	93
86	97	93	91	100	93

Backward k			Forward k		
U R	k=0,1	Step	U R	k=0,1	Step
11	41	97	7	95	96
12	43	96	7	98	96
18	64	96	18	99	97
26	75	96	33	98	97
28	80	96	34	95	97

Table 4: Monte Carlo Study of Unit Root Tests and Lag Selection
when DGP is AR(1) in First Differences with Outlier in Level Starting in 1930.
Frequencies of Rejection of the Unit Root at 5% Level (UR), of Selecting k = 0 or 1,
and of Step Dummy Significance at 5% Level.
Lag k Selected Backward From 8 or Forward from Zero to Maximum of 8.

	ADF Test:				Phillips-	Perron Test;Break Date Known					
T=62:	Backwardk		Forward k		Perron	Backward k			Forward k		
Outlier	U R	k=0,1	U R	k=0,1	U R	U R	k=0,1	Step	U R	k=0,1	Step
None	10	45	5	98	3	10	42	25	6	98	20
(-.2) @ 1930	23	78	27	99	30	45	49	43	48	96	39
(-.2) '30-'34	11	18	4	99	5	14	3	23	9	95	14
(-.2) '30-'39	5	68	1	100	2	16	40	28	2	99	14
(-.2) 1930 on	6	66	2	100	3	8	44	41	4	99	34
AR(2) '30-'45	64	42	75	79	16	73	22	32	83	66	28
Crash, AR(2)	61	42	77	85	9	64	35	20	79	81	18
Fixed pattern	19	32	23	100	0	55	9	52	33	100	22
T=125/Outlier:	U R	k=0,1	U R	k=0,1	U R	U R	k=0,1	Step	U R	k=0,1	Step
None	8	46	7	97	4	9	45	19	6	97	17
(-.2) @ 1930	13	68	13	99	14	30	51	32	31	97	31
(-.2) '30-'34	7	9	12	97	10	16	3	25	23	95	22
(-.2) '30-'39	11	59	5	99	4	17	46	16	5	98	14
(-.2) 1930 on	5	57	4	98	3	6	48	27	4	97	26
AR(2) '30-'45	42	18	80	79	60	50	14	28	83	70	29
Crash, AR(2)	38	16	78	79	50	42	15	20	79	77	21
Fixed pattern	50	18	68	98	8	51	12	26	64	98	15

Zivot -Aandrews Test; Search for Break Date

Backward k			Forward k		
U R	k=0,1	Step	U R	k=0,1	Step
16	35	93	9	95	92
36	74	92	37	100	93
36	14	94	22	93	92
50	34	98	23	99	99
48	43	95	38	98	97
80	28	90	79	59	91
81	36	89	80	70	92
86	27	100	59	96	100
U R	k=0,1	Step	U R	k=0,1	Step
11	41	97	7	95	96
18	64	96	18	99	97
14	10	95	15	93	96
23	41	97	13	97	97
23	47	98	19	96	98
63	26	96	83	64	97
58	25	96	85	73	98
64	21	98	65	95	98

TABLE 5
The Lower Bounds of One-Sided Confidence Intervals for the Largest AR Root in Post-war Real GDP

		Backward; k = 1		Forward; k = 0	
		ρ	cycle length	ρ	cycle length
Stock (1991)	95%	.72	9	.90	15
	90%	.77	10	.95	21
Simulation of:					
ADF	95%	.72	9	.88	13
	90%	.78	10	.93	18
DF-GLS	95%	.67	8	.86	12
	90%	.73	9	.90	15
Variance Ratios					
j = 10	95%	.65	8	.71	8
	90%	.70	9	.75	9
j = 20	95%	.80	11	.84	11
	90%	.84	12	.87	13

Figures are available from authors on request.