The Uncertain Trend in U.S. GDP

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

First draft: November 15, 1996
This draft: March 4, 1998

Abstract

Several recent papers conclude that GDP is trend stationary, implying that all shocks are transitory. We re-examine the evidence in light of test size distortion due to data-based lag selection and departures from the maintained hypothesis of temporal homogeneity, and find both effects 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, but if imposed implies cycles of implausible duration and pattern - 1997 output was 8% below trend.

Key words: business cycle, unit root, stochastic trend, trend.

JEL classifications: C22, E32, N1, O47.

We acknowledge with thanks comments from Tim Cogley, Charles Engel, Neil Ericsson, Jon Faust, Chang-Jin Kim, John Rogers, Glenn Rudebusch, Richard Startz, Chris Weber, David Wilcox, and Eric Zivot, but responsibility for errors and the views expressed is entirely the authors'. Support from the National Science Foundation under grant SBR-9711301 and the Ford and Louisa Van Voorhis endowment at the University of Washington is gratefully acknowledged.

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

1. Introduction

Several recent papers have brought 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 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 estimated by Nelson and Plosser implies that the stochastic trend contributes more to the 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 it is presumably real shocks that impact the trend while monetary and fiscal shocks have only transitory effects.

Perron (1989) argued that by failing to allow for structural change, Nelson and Plosser vastly overstated the frequency of permanent shocks. He 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 data-based selection of the break date.

More recently, Ben-David and Papell (1995), Cheung and Chinn (1997), 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. The implicit argument in these papers is that rejection of the unit root hypothesis can be attributed to an increase in power derived from a longer sample. These papers would thus suggest that as more data has become available, the evidence has become sharper, pointing now in the direction of 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 tests 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 stochastic trend view is reflected in the test statistics shown in Table 1 for the annual U.S. real GDP series, 1870-1994, assembled by Maddison (1995). Before interpreting these results, we briefly review the tests and their maintained hypotheses.

Dickey (1976), Fuller (1976), and Dickey and Fuller (1979) developed a test of the null hypothesis that a unit root in the AR representation, rather than a deterministic trend, accounts for the non-stationarity of a trending time series. The Dickey-Fuller test runs the regression

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

Under the unit root null, $\rho=1$, the first difference is a stationary AR process and the series is said to be "difference stationary" in the parlance of Nelson and Plosser. Under the alternative hypothesis $\rho<1$, the series is "trend stationary," a stationary AR process added to a deterministic linear trend. Dickey and Fuller showed that the t-statistic for testing $\rho=1$ has a non-standard distribution, and they tabulated Monte Carlo critical values for various sample sizes for a random walk with i.i.d. Normal shocks. They show that the limiting distribution remains the same when k lagged first differences "augment" the model to account for serial correlation (see also Hamilton, 1994, for further discussion).

In practice the lag length k, is unknown and is chosen by a data-dependent procedure. 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 indicating significance at the .10 level (greater than 1.645 in absolute value). This general-to-specific (GS) procedure has been followed by Perron (1989), Zivot and Andrews (1992), and others. Theoretical support for GS, as well as for various information criteria, was provided by Hall for the pure AR case and by Ng and Perron (1995) for the ARMA case. They showed that if the maximum is at least as large as the true lag, then asymptotically inference is unaffected by the data-based lag selection.

Since we will be interested in departures from the maintained hypothesis of i.i.d. shocks in Dickey-Fuller, we also include the heteroskedastic-consistent unit root test of Phillips and Perron (1988). This test does not rely on a finite order AR representation, 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. The Phillips-Perron test has the same limiting distribution as the Dickey-Fuller test.

Perron (1989) provided a generalization of 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 zero through year b and one thereafter, and I is one in year b+1 only and zero otherwise. Under the unit root hypothesis, the impulse dummy accounts for a break in level, while under trend stationarity alternative, the step dummy does. Perron provided critical values under the maintained hypotheses that the break date is known, the innovations are i.i.d. Normal, and lag k is known.

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 *and* does not contain a break; accordingly, their test regression does not include an impulse dummy. Second, Zivot and Andrews recognize that the break date is unknown *a priori* and estimate it to be that which maximizes the absolute value of the unit root test statistic. 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. Zivot and Andrews tabulate Monte Carlo critical values for $t(\rho=1)$ in the case of a random walk with i.i.d. Normal innovations and where k is assumed known to be zero. They confirmed Perron's choice of 1929 as the break date and his rejection of the unit root hypothesis for the Nelson-Plosser real GNP series. In practice, k is unknown and Zivot and Andrews did a GS search at each potential break date.

Leybourne and McCabe (1994) have developed a test of the null hypothesis that a series is trend stationary, with difference stationarity (a unit AR root) being the alternative hypothesis. They assume that the series has an unobserved components representation where the trend is a random walk, the stationary component is AR(k), and the innovations are independent across components and are i.i.d. This implies that the univariate representation of the first differences is ARMA(k,1) and the MA part will have a unit root if the trend is deterministic (zero variance in the random walk). Critical values for the Leybourne-McCabe test are tabulated in Kwiatowski et al. (1992). As in the Dickey-Fuller test, the Leybourne-McCabe test necessitates the preliminary step of selecting the lag length k to account for serial correlation. However, under the null hypothesis the series follows a non-invertible ARMA(k,1) process. The distribution of the AR terms is thus unknown. Therefore, in contrast to the Dickey-Fuller test, there does not exist a set of results which guarantees that inference based on the Leybourne-McCabe test is asymptotically unaffected by data-based lag selection.

Results of these tests are reported in Table 1 for the full Maddison sample and the sub-period 1909-1970 studied by Nelson and Plosser. The lag length is chosen alternatively by GS and Schwarz' (1978) information criteria (SIC). Following Perron and Zivot and Andrews, the maximum lag we consider for annual data is 8. As noted before, if the true lag is less than or equal to 8, the results of Hall (1994) and Ng and Perron (1995) state that in the limit, both GS and SIC will choose the correct lag with probability one. The Zivot-Andrews procedure identifies 1929 as the break date for both time spans. P-values are obtained by simulation as in the original articles; the DGP under the null hypothesis is a random walk for the unit root tests and a trend plus random error for Leybourne-McCabe.

Several features of the results seem worthy of note:

1. The null hypothesis of a unit root is strongly rejected by all three tests for the Maddison data. The Dickey-Fuller test for the sub-period studied by Nelson and Plosser is less favorable to the unit root hypothesis than they reported using data available prior to the work of Balke and Gordon.
2. Evidence against a unit root is stronger for the full time period.
3. Rejection of the unit root null is stronger when a break is allowed if the break date is assumed known.
4. The choice of lag k differs greatly between GS and SIC selection, the former often choosing the maximum allowed while the latter in every case chooses only one lag.
5. Lag selection matters for inference. For the Perron and Zivot-Andrews tests, GS lag selection leads to stronger rejection of the unit root. For the Leybourne-McCabe test, lag selection determines the outcome. SIC does not lead to rejection of trend stationarity (as reported by Cheung and Chinn, 1997), but GS does.
6. The step dummy is highly significant by conventional standards in every case However, Banerjee, Lumsdaine, and Stock (1992) demonstrate that the distributions of break-dummy coefficients are non-standard.

These tests have as maintained hypothesis that the series is homogeneous, generated by an AR process of known order with constant parameters and i.i.d. Normal innovations. It is not clear how deviations from these maintained hypotheses might affect size or power, although recent contributions to the theoretical literature, discussed below, suggest that they will. This is a concern in the context of U.S. GDP since 1929, when methods of data collection change, the Great Depression, and World War II are points at which the GDP data process might be expected to exhibit changes in both volatility and serial correlation.

For the period to 1929, Maddison used estimates by Balke and Gordon (1989); an alternative series is by Romer (1989). Both build on the pioneering methodology of Kuznets (1941,1946) and extensions by Kendrick (1961) and Gallman (1966). Briefly, the Kuznets methodology relies on trends extrapolated between benchmark years, then deviations from trend are based on indicator variables such as commodity output. It would be surprising if this method did not affect serial correlation and unit root tests. Indeed, the Dickey-Fuller test applied to the Balke-Gordon data rejects the unit root at the 10% level.

The period immediately following 1929 was one of banking failures on an unprecedented scale and repeated failure of the new Federal Reserve System to stabilize the system (Friedman and Schwartz, 1963). By the time the economy had recovered from the Great Depression it was jolted by World War II. The magnitude of these shocks is apparent in Figures 1 and 2 for levels and growth rates, and in the summary statistics in Table 2 for growth rates and deviations from the least squares trend line. The largest observations during the Depression-WWII period are three to four times the sample standard deviation measured over the full period. Fitted AR(3) models reveal large changes in serial correlation and much higher residual variance during the 1930-45 sub-period. The extremely small probability of this occurring in a homogeneous Normal sample is reflected in the asymptotic Jarque-Bera p-values of zero for the full sample. The three fold increase in the standard deviation of shocks is comparable to that for stock returns reported by Schwert (1989b). Separating pre-1929, 1930-45, and post-war periods, however, one obtains apparently homogeneous Normal samples.

More formal evidence on non-homogeneity comes from an extension of the Clark (1987) model in which stochastic trend and cycle components are augmented by an additive irregular component that switches on and off according to a Markov process. Details are given in the Appendix and in Murray (1997). As seen in Figure 3, the irregular component switches on in 1893, but shows only small fluctuations until 1930 when it reflects the huge swings in output of the Depression and WWII. Then it switches off in 1947. This pattern is consistent with larger measurement errors in the pre-1929 data and the 1930-46 period characterized by a sequence of additive outliers that do not occur elsewhere. Further, the cyclical component is no longer significant once this irregular component is included in the model. Our results are consistent with those of Balke and Fomby (1991) who also identified additive outliers associated with the Depression and World War II, but failed to detect permanent breaks in level.

3. The Sensitivity of Unit Root and Trend Stationarity Tests to Lag Selection and Additive Outliers

3.1 Design of the Experiments

This section presents a series of experiments designed to investigate how tests for a unit root or trend stationarity are affected by data-based lag selection and departure from the i.i.d. Normal assumption in the form of additive outliers. Our strategy is to specify a data generating process for 1870-1994 that contains a unit root and replicates the main statistical features of post-war GDP and use it to study size or power under GS and SIC lag selection, then introduce various types of additive outliers to see what effect they have on the tests. Our choice of the post-war data as a guide to the DGP is based on the results of the Markov-switching state-space model discussed above. We do not attempt to model the measurement errors in the pre-1929 data.

For the post-war Maddison data in first differences of logs, SIC chooses ARMA(0,0) and AIC chooses ARMA(1,0). We adopt the AR(1) model to include some dynamics. The estimated model, and our underlying DGP, is:

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

This DGP is run for 20 periods before recording a realization corresponding to 1870-1994. After integrating the first differences to obtain levels, we add one of a number of types of outliers to see its effect on the test statistics. The observed data, say y*, is then

$$y_t^* = y_t + O_t$$

where the outlier sequence $\{O_t\}$ varies across experiments. Two panels give results for sample lengths of 125 and 62 years corresponding to the full 1870-1994 period and the Nelson-Plosser 1909-1970 sub-period, respectively, based on 1,000 replications. The upper bound for the standard error of the rejection frequencies we report is .016 (see Davidson and MacKinnon (1993)).

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. In the second experiment, reported in Table 4, we consider outlier events that begin in 1930 and then follow either fixed or stochastic paths.

Reported in each table are actual frequencies of rejection of the unit root hypothesis at a nominal .05 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 Dickey-Fuller test the finite sample critical values are exact if k is known to be zero and the innovations are Gaussian. We also report the frequency with which the t-statistic for the step dummy in the Perron and Zivot-Andrews regressions is significant at the conventional nominal .05 level.

In the case of the LM test, the null hypothesis is trend stationarity, so the frequency of rejection is the power of the test against the alternative represented by our DGP. Since these frequencies are meaningless unless the size of the test is correct, we follow Cheung and Chinn in setting critical values by simulation of the trend stationary AR model suggested by the historical data.

The number of lagged first differences included in any of these regressions, denoted k, is selected alternatively by GS and SIC 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.

3.2 The Effect of Data-based Lag Selection on Test Size or Power

In the experiment reported in the first line of each panel of Table 3, no outlier has been added. In this case the series does in fact have a unit root with i.i.d. Normal innovations, so the frequency reported for each unit root test is its actual size, the probability of rejecting the null hypothesis at a nominal .05 level when it is true. As reported previously by Hall (1994) for the Dickey-Fuller test, size depends importantly on the method of lag selection. In the case of the Dickey-Fuller and Perron tests, SIC produces roughly the correct size, while GS results in a size of about .10. The Zivot-Andrews test suffers from greater size distortion under both lag selection strategies, and the distortion is entirely due to selecting k from the data; when the correct value k=1 is imposed, we find that the actual size is correct.

While Hall showed that both GS and SIC are 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 simply a trade-off between size and power, with strategies favoring large k offering more correct size but lower power. 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 to select 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.

As noted above, Banerjee, Lumsdaine, and Stock (1992) demonstrate that the asymptotic distributions of break dummy coefficients are non-standard, although they maximized the F-stat of the dummy variable, rather than the unit root statistic, across break dates. Using the t-distribution for inference would create the false impression that a break occurred. As seen here, the sizes of the t-tests for the step dummy are much too large; about .2 in the Perron regressions and about .95 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 the case of the Leybourne-McCabe test, the frequency of rejection reflects power against the alternative of the ARIMA(1,1,0) process generating the data. We note that power is higher under SIC lag selection. For T=125, the test correctly rejects trend stationarity about 80% of the time using SIC and about 60% for GS. For T=62, the power is significantly decreased; 44% for SIC and 24% for GS.

3.3 The Effect of Additive Outliers

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

As the magnitude of the outlier increases, rejection rates for all of the unit root tests rise sharply. Frances and Haldrup (1994) studied the effect of a stochastic additive outlier that occurs with some probability each time period in an I(1) process, and showed that even asymptotically the distribution of the Dickey-Fuller t-statistic (k fixed) is shifted to the left, increasing the frequency of rejection of the unit root null. A stochastic outlier introduces an MA(1) component into the process, so no finite AR representation exists and this is the situation in which Schwert (1989a) had shown that unit root tests have poor finite sample properties. Whether the outlier is stochastic or fixed, the maintained hypothesis of a finite order AR with i.i.d. Normal errors is violated, and rejection is triggered.

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 and the trend stationary alternative when neither is true, these tests reject the unit root. If misinterpreted, these tests spuriously signal trend stationarity, with a level break if allowed, when in fact the series has a unit root and the outlier event affects 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.

For the Leybourne-McCabe test, the issue is how power is altered by introduction of the outlier. For the one period outliers considered in this section, the power is marginally reduce for T=125, while T=62 results in a more severe power reduction.

3.4 Additive Outliers That Persist

The experiments reported in Table 4 are designed to explore the impact of outlier duration and pattern. Starting from the benchmark case of no outlier, we add -.2 in 1930 only (as in Table 3), in each of the ten years 1930-39, and to every year from 1930 on. The last case is motivated by Perron's (1989) finding of a permanent break in the level of output in 1930. Further experiments adds a stochastic additive outlier sequence generated by an AR(2) modeled on departures from a local trend connecting 1929 to 1946 or, alternatively, the fixed, 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).

Note that persistence in itself reduces the frequency of rejection of the unit root in the Dickey-Fuller, Phillips-Perron, and Perron tests. When the outlier is permanent, the size of the Perron test is close to the correct .05. 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. In contrast, the Zivot-Andrews test rejects the unit root null much more frequently if the level of the series shifts permanently, and the step dummy is almost always significant. It turns out that this is not due to the absence of an impulse dummy in the Zivot-Andrews regression; when it was rerun with the impulse dummy the results were essentially the same. Rather, it is having to search for the break date (which is identified quite poorly) that accounts for excessively frequent rejection of the unit root. This suggests a high value for information about the timing of structural change.

In the last two experiments in Table 4, the underlying process is inundated with a high amplitude wave which distorts the level of the series for 16 years and then is gone. In both, the unit root and no-step-dummy null hypotheses are rejected often, except by Phillips-Perron. Rejection rates differ sharply depending on which lag selection method is used. GS generally chooses a much larger value of k than does SIC, reflecting the contrast we saw in Table 1. It also appears that the *particular* pattern of real GDP during the period 1930-45 as opposed to the random outcomes of the AR(2) process do matter; the tendency to stronger rejection of the unit root in the longer sample being more apparent for the fixed pattern.

Finally, the Leybourne-McCabe test has substantially lower power in the last two experiments, failing to reject trend stationarity much more frequently if the underlying unit root process is overlaid by a transitory component of large amplitude. This is not surprising, since the stochastic trend will appear relatively smooth compared to the transitory component. As pointed out by Cochrane (1991), there is an observational equivalence between a trend stationary process and one with a stochastic trend where the variance of the innovations is small enough relative to the transitory component.

4. Evidence from the Post-war Data

The argument that a longer span of data yields a test statistic with greater power is valid only if a time series is temporarlly homogeneous. Thus while more data is usually preferred to less, we find two compelling reasons to focus on post-war GDP data 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 may be 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 entirely transitory, they could account for rejection of the unit root hypothesis and be misconstrued as evidence of trend stationarity with or without structural change. Thus by focusing on the post-war data, we hope to minimize the chance of spuriously rejecting the unit root hypothesis due to violation of the homogeneity assumption.

To study the post-war period we shift from the annual Maddison data to the recently available quarterly series of real GDP in chained (1992) dollars. One advantage of the chained data for our purposes is that it addresses the concern of Gordon (1993) that a productivity slow-down would be obscured in a series based on fixed weight price deflators, such as the real GDP data used by Maddison. Indeed, the existence and causes of such a structural break have been discussed since the 1970s and are the subject of a large and continuing literature; see also Baily and Gordon (1988). Indeed, Perron reported evidence of a break in 1973 in the *slope* of the trend function for post-war quarterly real GNP, indicating a slow down in long-term growth, although Zivot and Andrews were not able to confirm this finding when they searched for the break date. Results of unit root, trend break, and trend stationarity tests for chained post-war quarterly real GDP are presented in Table 5. Nominal p-values are based on tabulated asymptotic distributions, while exact p-values will depend on the method of lag selection and are obtained by simulation under the null hypothesis where the data generating process is either an AR(1) model fitted to first differences of the log of the post-war data, or an AR(2) around a deterministic time trend, the orders chosen both by GS and SIC.

4.1 Unit Root Tests and Confidence Intervals for the Largest AR Root.

The Dickey-Fuller tests reported in Table 5 are fully consistent with a unit root in post-war real GDP. Following Perron and Zivot and Andrews, we start with a maximum lag of 12 for quarterly data. For both methods of lag selection the nominal and exact p-values are greater than 0.50; the expected value of the test statistic being about -2.2. For completeness, we include the Phillips-Perron test, although its nominal size is too small, and it too is fully consistent with a unit root. A common criticism of Dickey-Fuller tests is that they have low power against local alternatives, an AR root close to unity. A modified test by Elliot, Rothenberg, and Stock (1996), which they call DF-GLS$^{\tau}$, employs a local-to-unity detrending procedure designed to maximize power against local alternatives. Although lag selection differs sharply between GS and SIC, the test results do not, both being entirely consistent with a unit root.

Although these results are consistent with a unit root process, they are also consistent with a range of trend stationary alternatives since it is not possible to distinguish in a finite sample between the realization of a unit root process and a trend stationary process with an AR 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. As noted above, 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 enough, and Engel (1997) shows that an economically significant random walk component can be missed. 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 chained GDP data, we computed two-sided confidence intervals using the procedure developed by Stock (1991). These are based on inverting the augmented Dickey-Fuller test statistic to determine the values of ρ consistent with it. The 95% interval based on GS selection of 12 lags is (0.961, 1.026) and based on SIC selection of 1 lag it is (0.931, 1.022). We note that both include trend stationary as well as explosive alternatives. The former possibility has received considerable attention in the recent literature (Rudebusch 1992, 1993). Rudebusch (1993) demonstrates that the augmented Dickey-Fuller test applied to post-war quarterly GNP lacks power against a specific non-local alternative. He fits an AR(2) to deviations from the linear trend and shows that for this parameterization a unit root statistic greater than or equal to the observed statistic occurs 22% of the time, suggesting that the distribution of the statistic is not radically different under the unit root and this stationary alternative. Applying Rudebusch's technique to the chained data, we find that the trend stationary representation yields statistics greater than or equal to the observed unit root statistics 2.47% and 14.55% of the time for GS and SIC respectively. Thus, the disparity between the unit root and trend stationary alternative is greater in the chained data.

4.2 Implications of Trend Stationarity for the Cyclical Behavior of GDP

While Jones (1995) and Diebold and Senhadji (1996) argue for the efficacy of trend stationarity in forecasting the long run path of output, little attention has been given to the particular realization of the transitory component that is implied for the post-war U.S. and whether it corresponds to an economically meaningful deviation from a long run growth path. Figure 4 plots the deviation of the log of chained GDP from the fitted trend line, with NBER reference cycles shaded. While the deviation from trend does dip in concert with NBER recessions, its variation is dominated by a very low frequency wave that says that the economy was well below trend most of the period from 1947 through the early 1960s, consistently above trend until 1981, finally falling sharply below trend during the last recession and continuing downward through 1997. Conventional measures of economic performance would suggest a very different pattern, unemployment having been very high during the

1974-75 and 1981-82 recessions and very low in 1997. The implied deviations from trend are also of large amplitude, starting at -7% in 1947, peaking at +10% in 1973, and ending at -8% in 1997. We are not aware of any estimates that an additional 8% of output was available to the U.S. economy in 1997. A forecast based on trend stationarity would imply growth rates about one percentage point above average for the next several years.

The impression of a long wave in Figure 4 is reflected in the long wave observed in the correlogram in Figure 5 and the low frequency peak in sample spectrum plotted in Figure 6. These features are reminiscent of the spurious periodicity, identified by Nelson and Kang (1981), that characterizes residuals from the regression of a random walk, and I(1) processes in general, on time. They show that the spurious cycle typically has a period equal to about 0.83 of the length of the series, here about 42 years. Indeed, the peak of the sample spectrum occurs at a frequency of 0.035, which implies a period of 45 years, slightly above that predicted by Nelson and Kang. These low frequency dynamics, as well as the economic implausibility of the implied cycle, suggest to us that the trend component of output is much more flexible than a straight line, probably accounting for much of the long wave that the trend stationarity would attribute to the transitory component we see in Figure 4.

4.3 How Big is the Random Walk in GDP?

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∅ . If a series is integrated, it can 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. The sample variance ratio using Cochrane's unbiased estimate (his equation A3) for the post-war chained GDP is plotted in Figure 7 and, unlike the shorter series used by Cochrane, shows no tendency to decline at longer lags, suggesting that the variation in GDP is dominated by the variation in the stochastic trend.

4.4 Is There a Productivity Slow-Down in 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 annual growth rate over the period 1947.1-1973.1 was 3.9% while in 1973.2-1997.3 it fell to 2.5%. 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 and impulse dummies 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 GS, choosing k = 10 quarters.

For the post-war chained data, both GS and SIC provide no evidence against the unit root hypothesis. As seen in Table 5, 1972.2 is chosen as the break date as in Zivot and Andrews, and the nominal p-values are 0.54 and 0.14 respectively. This corroborates the finding of Zivot and Andrews that post-war GDP is not well characterized as stationary fluctuations around a kinked time trend. Exact p-values reflect the finite sample size distortion induced by lag selection.

4.5 The Leybourne-McCabe Test for Trend Stationarity

Finally, Table 5 also reports the results of the Leybourne-McCabe (1994) test for trend stationary applied to the post-war chained data. As noted by Cheung and Chinn (1997), the asymptotic p-values provided by Kwiatowski et. al. (1992) are not useful guides for inference when the sample is finite. Indeed, the Leybourne-McCabe statistics based on GS and SIC lead to rejection of the trend stationary null at any significance level. We also computed exact p-values for the observed statistics based on the trend stationary AR(2) parameterization discussed above. The likelihood of observing the Leybourne-McCabe statistics under GS and SIC is 9.3% and 2.1% respectively, offering little evidence in favor of pure trend stationarity for the post-war data.

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, for parameterizations suggested by the data, the actual probability of rejecting the unit root hypothesis when it is true is substantially greater under data-based lag selection than is indicated by the nominal significance levels upon which rejections of the unit root have been based in the recent empirical literature.

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.

To reduce the possibility of spurious rejection of the unit root null due to heterogeneity in the data, we focus on post-war chained GDP. The unit root statistics in all cases not only fail to reject, but lie in the upper half of the distribution under the null hypothesis. While we also cannot reject a range of trend stationarity alternatives, we find that the implied cycle component contains a low frequency peak in the sample spectrum with a period of 45 years, much longer than the 6.5 year average peak-to-peak length in the NBER chronology. This is reminiscent of the spurious periodicity phenomenon analyzed by Nelson and Kang (1981) for a detrended unit root process. Furthermore, the cycle implied 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 1997. These results cast serious doubt on the trend stationary model as an economically credible representation of real GDP.

In our view, a constructive direction for modeling aggregate output will be one that moves beyond 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 Thomas B. Fomby, 1991, "Shifting trends, segmented trends, and infrequent permanent shocks," *J. of Monetary Economics*, 28, 61-86.

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.

Banerjee, Anindya, Robin L. Lumsdaine, and James H. Stock, 1992, "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *J. of Business and Economic Statistics*, 10, 271-287.

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.

Beveridge, Stephen and Charles R. Nelson, 1981, "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle," *J. of Monetary Economics*, 7, 151-174.

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.

Clark, Peter K., 1987, "The cyclical component of U.S. economic activity," *The Quarterly Journal of Economics*, 102, 797-814.

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.

Davidson, Russell, and James G. McKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

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.

Engel, Charles, 1997, "Long-run PPP may not hold after all," University of Washington discussion paper.

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, 461-70.

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

Hamilton, James D., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Jones, Charles I., 1995, "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics*, 110, 495-525.

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.

Kim, Chang-Jin, 1994, "Dynamic Linear Models with Markov Switching," *Journal of Econometrics*, 60, 1-22.

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

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

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992, “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root,” *J. of Econometrics*, 54, 159-178.
Leybourne S. J., and McCabe, B. P. M., 1994, “A Consistent Test for a Unit Root,” *J. of Business and Economic Statistics*, 12, 157-66.
Maddison, Angus, 1995, *Monitoring the World Economy 1820-1992*, Paris: OECD.
Murray, Christian J., 1997, “Trend Stationarity in Annual U.S. RGDP? A Reeximination of Recent Evidence,” Univ. of Washington.
Nelson, Charles R. and Heejoon Kang, 1981, “Spurious Periodicity in Inappropriately Detrended Time Series,” *Econometrica*, 49, 741-51.
Nelson, Charles R. and Charles I. Plosser, 1982, "Trends and random walks in macroeconomic time series:Some Evidence and Implications," *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., 1992, “Trends and random walks in macroeconomic time series: A re-examination,” *International Economic Review*, 33, 661-80.
Rudebusch, Glenn D., 1993, “The uncertain unit root in real GNP,” *American Economic Review*, 83, 264-72.
Schwartz, G., 1978, “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461-64.
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.

This print May 22, 1998

Table 1
Tests for a Unit Root or Trend Stationarity
in U.S. Real GDP
Annual data; Maddison (1995)

Test: GS/SIC	AR Lag (4)	Test Statistic	Nominal p-value (5)	Step Dummy t-statistic
1870-1994:				
Unit root:				
Dickey-Fuller	6	-3.74	0.03	-
	1	-4.14	< .01	-
Phillips-Perron	-	-3.52	0.04	-
Perron (1)	8	-5.58	< .01	-3.74
(break in level)	1	-4.72	< .01	-2.31
Zivot&Andrews (2)	8	-6.10	< .01	-4.33
	1	-5.10	0.02	-2.82
Trend stationary:				
Leybourne	5	0.42	<.01	-
& McCabe (3)	2	0.05	0.40	-
1909-1970:				
Unit root:				
Dickey-Fuller	1 & 1	-3.43	0.06	-
Phillips-Perron	-	-2.63	0.27	-
Perron (1)	8	-4.89	< 0.01	-3.87
(break in level)	1	-4.26	0.02	-2.40
Zivot&Andrews (2)	8	-5.61	< 0.01	-4.63
	1	-4.72	0.07	-3.03
Trend stationary:				
Leybourne	4	0.47	< 0.01	-
& McCabe (3)	2	0.10	> .85	-

(1) Break in level assumed at 1929 as in Perron (1989).
(2) Break date maximizes unit root t-statistic;
choose 1929 under GS & SIC.
(3) Null hypothesis is trend stationarity.
(4) GS starts with 8 lags, reducing lags until t>1.645,
SIC maximizes criterion of Schwarz (1978) over lags 0 to 8.
(5) Nominal p-values obtained by simulation under the null,
following the original papers where these tests are described.

Table 2
Summary Statistics for U.S. Real GDP

Growth Rates	Mean	Std Dev	Range	AR Coefficient Estimates			S E	J-B p
1870-1994	0.033	0.056	.18/-.23	0.27 *	0.00	-0.12	0.06	0.00
1870-1929	0.037	0.048	.13/-.08	-0.28 *	-0.20	-0.04	0.05	0.91
1930-1946	0.026	0.118	.18/-.23	0.90 *	-0.16	-0.49	0.08	0.93
1947-1994	0.031	0.026	.09/-.02	0.20	-0.11	-0.19	0.03	0.71

Detrended	Mean	Std Dev	Range	AR Coefficient Estimates			S E	J-B p
1870-1994	0.000	0.11	.31/-.37	1.13 *	-0.24	-0.10	0.05	0.00
1870-1929	0.029	0.082	.18/-.14	0.64 *	0.05	0.13	0.05	0.66
1930-1945	-0.100	0.222	.31/-.37	1.48 *	-0.57	-0.23	0.07	0.60
1946-1994	-0.003	0.048	.08/-.10	1.09 *	-0.24	-0.05	0.02	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; DGP is ARIMA(1,1,0) with Additive Outlier at 1930

Frequencies of Rejection of Null and Step Dummy Significant at Nominal .05 Level

Lag k Selection Alternatively General-to-Specific and SIC

Series Length: T=125

	Dickey-Fuller		Phillips-	Perron; 1929 Break Date				Zivot-Andrews; Search for Da				Leybourne & McCabe	
1930	GS	SIC	Perron	GS		SIC		GS		SIC		GS	SIC
Outlier	Unit Root		UR	UR	Step	UR	Step	UR	Step	UR	Step	Trend	Stationary
0	0.084	0.052	0.036	0.092	0.189	0.063	0.171	0.108	0.965	0.073	0.961	0.57	0.81
-0.1	0.076	0.045	0.054	0.147	0.256	0.115	0.239	0.115	0.959	0.066	0.961	0.55	0.79
-0.2	0.125	0.135	0.139	0.301	0.321	0.316	0.312	0.180	0.961	0.187	0.965	0.5	0.76
-0.3	0.142	0.215	0.270	0.374	0.393	0.414	0.379	0.261	0.959	0.378	0.964	0.48	0.74
-0.4	0.198	0.222	0.491	0.453	0.458	0.513	0.445	0.277	0.959	0.450	0.969	0.49	0.71

Series Length; T=62

	Dickey-Fuller		Phillips-	Perron; 1929 Break Date				Zivot-Andrews; Search for Da				Leybourne & McCabe	
1930	GS	SIC	Perron	GS		SIC		GS		SIC		GS	SIC
Outlier	Unit Root		UR	UR	Step	UR	Step	UR	Step	UR	Step	Trend	Stationary
0	0.097	0.058	0.031	0.100	0.246	0.081	0.212	0.155	0.930	0.128	0.947	0.24	0.44
-0.1	0.083	0.061	0.079	0.260	0.309	0.216	0.276	0.171	0.940	0.102	0.918	0.22	0.41
-0.2	0.228	0.286	0.299	0.448	0.425	0.508	0.402	0.364	0.923	0.366	0.931	0.18	0.35
-0.3	0.377	0.542	0.616	0.529	0.471	0.620	0.470	0.692	0.919	0.725	0.932	0.13	0.24
-0.4	0.488	0.711	0.837	0.634	0.506	0.731	0.518	0.864	0.928	0.912	0.943	0.07	0.16

Table 4

Monte Carlo Study of Unit Root Tests; DGP is ARIMA(1,1,0) Additive Outlier Process Starting at 1930

Frequencies of Rejection of Null and Step Dummy Significant at Nominal .05 Level

Lag k Selection Alternatively General-to-Specific and SIC

Series Length: T=125

	Dickey-Fuller		Phillips-	Perron; 1929 Break Date				Zivot-Andrews; Search for Dat				Leybourne-McCabe	
	GS	SIC	Perron	GS		SIC		GS		SIC		GS	SIC
Outlier	Unit Root		UR	UR	Step	UR	Step	UR	Step	UR	Step	Trend	Stationary
None	0.084	0.052	0.036	0.092	0.189	0.063	0.171	0.108	0.965	0.073	0.961	0.57	0.81
(-.2) @ 1930	0.125	0.135	0.139	0.301	0.321	0.316	0.312	0.180	0.961	0.187	0.965	0.549	0.762
(-.2) '30-'39	0.112	0.042	0.042	0.169	0.160	0.056	0.134	0.233	0.974	0.128	0.966	0.470	0.705
(-.2) 1930 on	0.050	0.033	0.033	0.056	0.272	0.040	0.256	0.227	0.975	0.187	0.974	0.577	0.813
AR(2) '30-'45	0.416	0.764	0.599	0.498	0.275	0.772	0.276	0.628	0.955	0.817	0.970	0.298	0.286
Fixed pattern	0.499	0.673	0.083	0.506	0.256	0.642	0.143	0.643	0.979	0.651	0.988	0.170	0.278

Series Length; T=62

	Dickey-Fuller		Phillips-	Perron; 1929 Break Date				Zivot-Andrews; Search for Dat				Leybourne-McCabe	
	GS	SIC	Perron	GS		SIC		GS		SIC		GS	SIC
Outlier	Unit Root		UR	UR	Step	UR	Step	UR	Step	UR	Step	Trend	Stationary
None	0.097	0.058	0.031	0.100	0.246	0.081	0.212	0.155	0.930	0.128	0.947	0.238	0.439
(-.2) @ 1930	0.228	0.286	0.299	0.448	0.425	0.508	0.402	0.364	0.923	0.366	0.931	0.182	0.346
(-.2) '30-'39	0.048	0.007	0.016	0.157	0.278	0.018	0.212	0.503	0.978	0.235	0.989	0.172	0.374
(-.2) 1930 on	0.060	0.018	0.033	0.084	0.412	0.053	0.347	0.483	0.953	0.411	0.963	0.319	0.501
AR(2) '30-'45	0.641	0.740	0.163	0.730	0.317	0.823	0.278	0.798	0.898	0.789	0.920	0.042	0.045
Fixed pattern	0.185	0.233	0.000	0.553	0.518	0.332	0.216	0.855	0.999	0.597	0.996	0.005	0.068

Table 5

Tests for a Unit Root or Trend Stationarity in Post-war U.S. Real GDP

Quarterly Chained Real GDP; 1947.1 - 1997.3

Test: GS/SIC	AR Lag (4)	Test Statistic	Nominal p-value (5)	Exact p-value (6)
Unit root				
Dickey-Fuller	12	-1.52	0.82	0.816
	1	-2.05	0.58	0.552
Phillips-Perron	n/a	-1.81	0.69	0.560
Elliott, Rothenber	12	-0.84	0.90	0.621
& Stock	1	-1.50	0.59	0.660
Perron (1)	12	-3.3	0.19	0.269
(break in slope)	1	-3.96	0.05	0.056
Zivot-Andrews (2)	12	-3.31	0.54	0.628
	1	-3.99	0.14	0.173
Trend stationarity				
Leybourne	3	2.25	0.00	0.103
& McCabe (3)	1	3.17	0.00	0.021

Notes:

(1) Break in slope assumed to occur at 1973:1 as in Perron (1989).

(2) Break date maximizes unit root t-statistic; choose1972:2 under GS

(3) Null hypothesis is trend stationarity.

(4) GS starts with 12 lags, reducing lag until t>1.645 in absolute value,

SIC maximizes criterion of Schwarz (1978) over lags 0 to 12.

(5) Nominal p-values obtained by simulation under null hypothesis,

as in the original articles in which these tests are described.

(6) Exact p-values obtained by simulation with lag selection,

under unit root null for unit root tests, under trend stationarity for L-M

DGP is the AR process selected by SIC for the actual data under the nu

Figure 1 - Log of Real GDP; Maddison (1995) Data

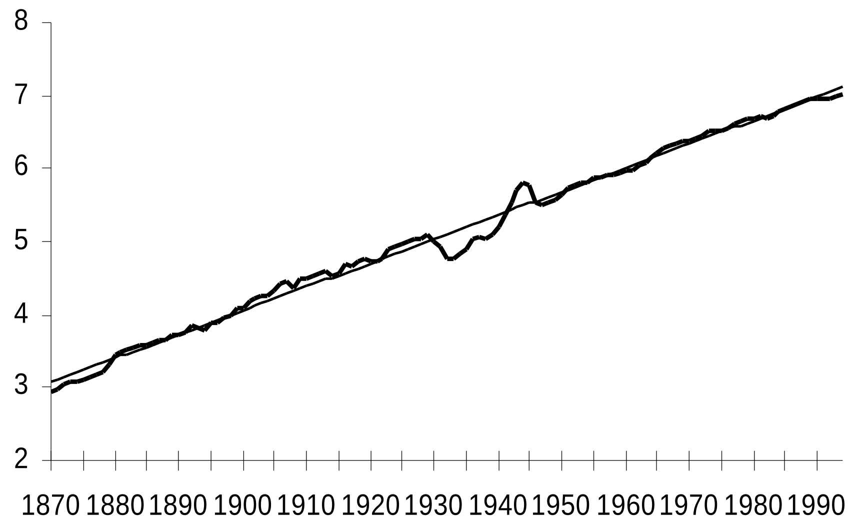

Figure 2 - Growth Rate of U.S. Real GDP

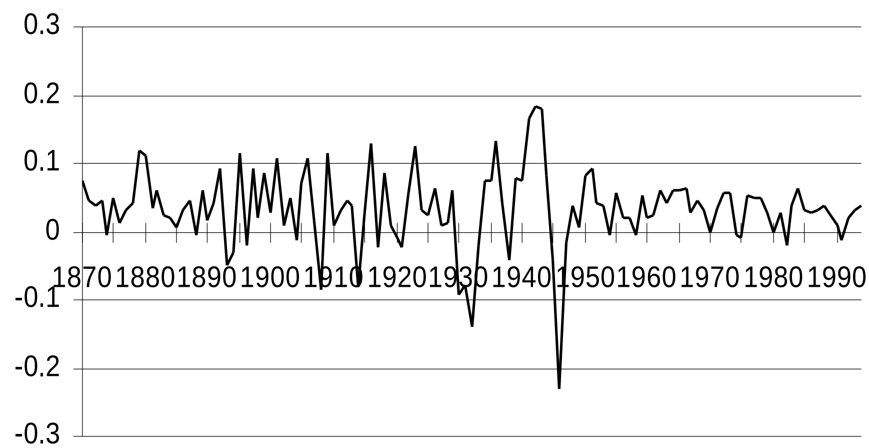

Figure 3 - Irregular Component, U.S. real GDP

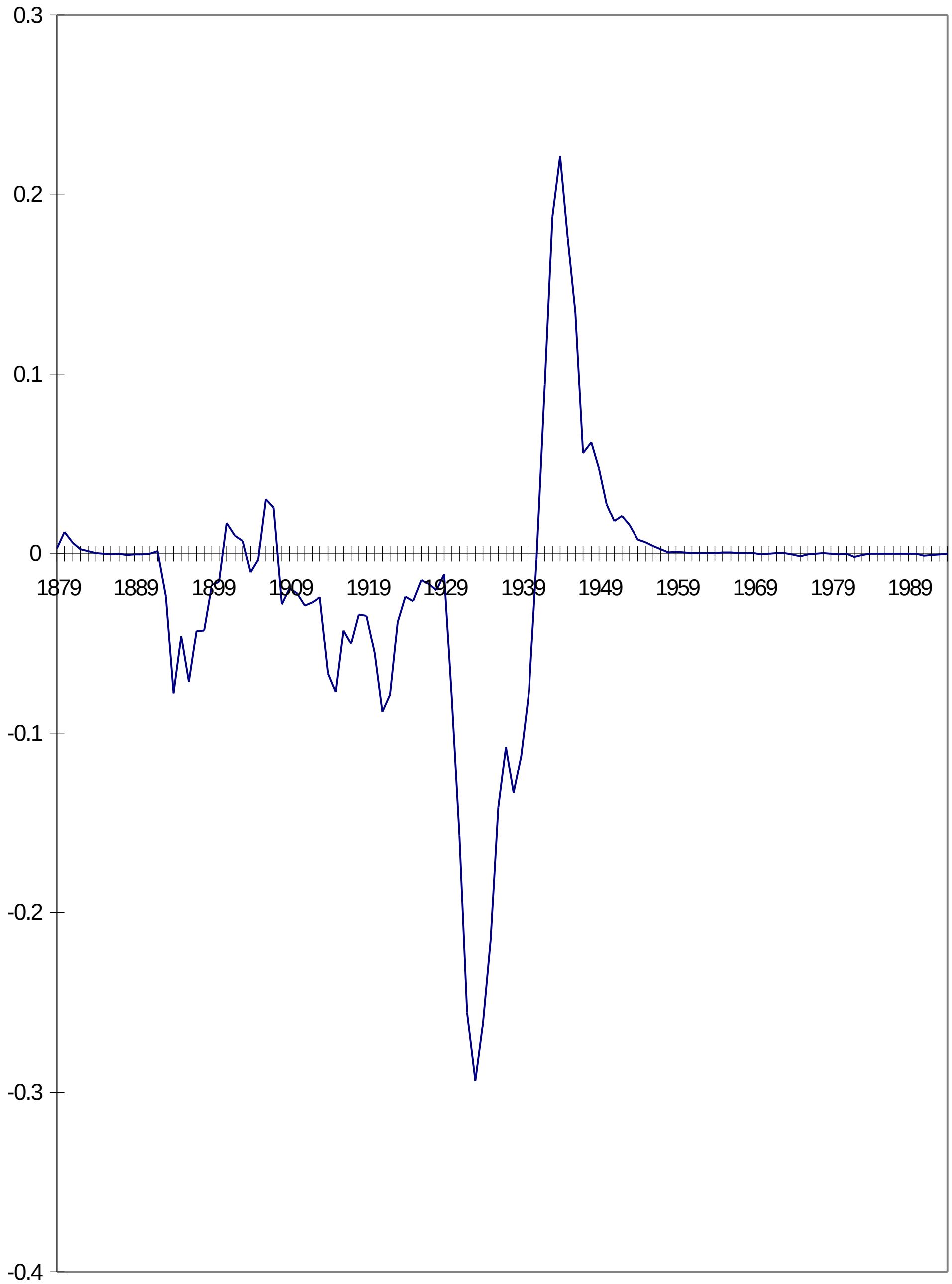

Figure 4 – Deviations from Trend Line
Post-war chained GDP

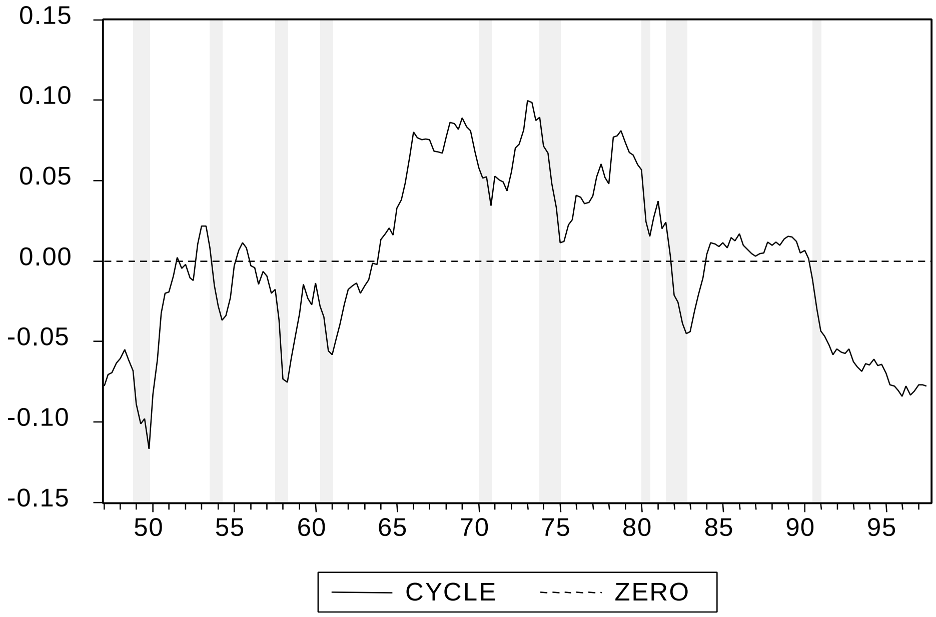

Figure 5 - ACF of Deviations from Trend Line
Post-war Chained GDP

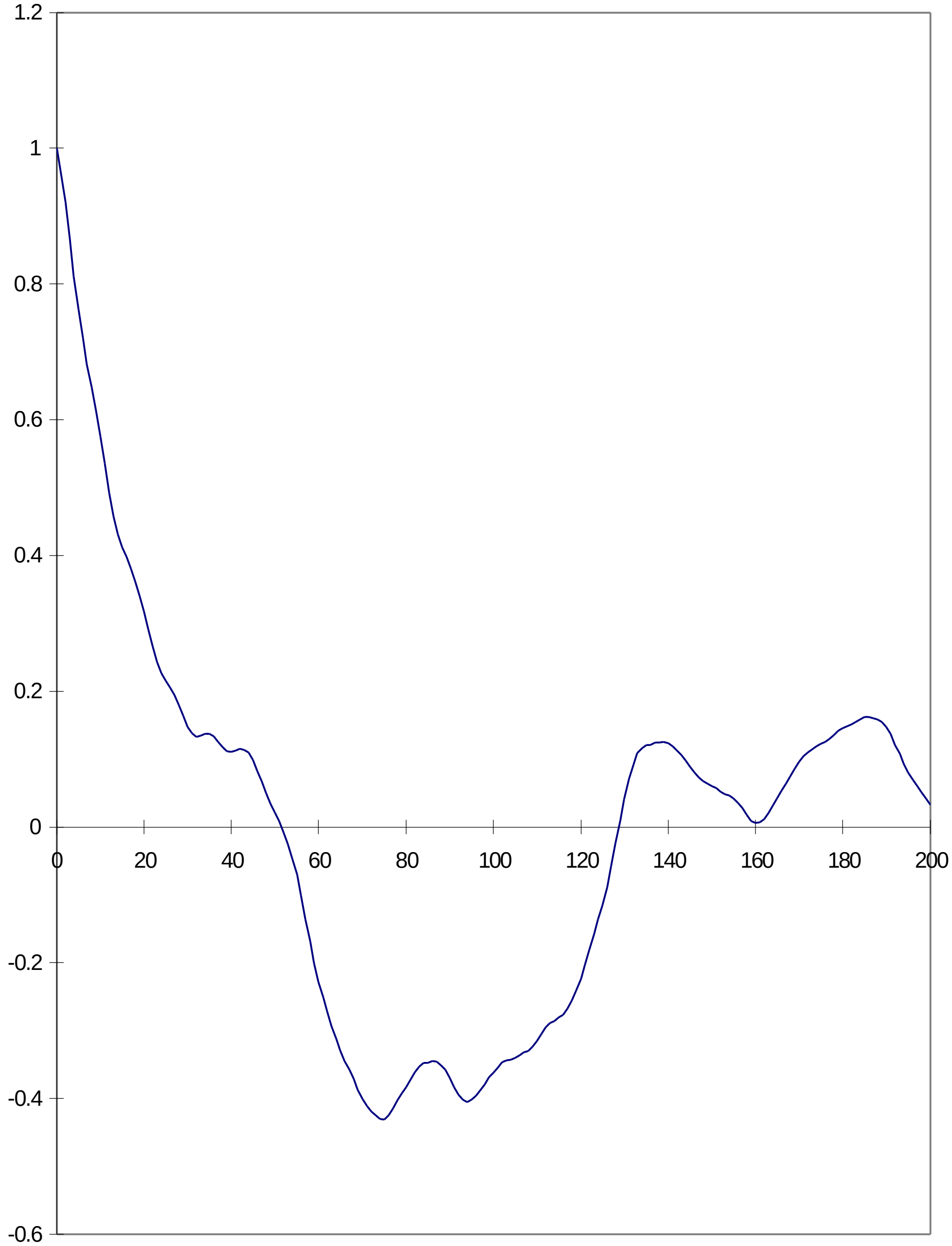

Fig. 6-Spectrum of Deviations from Trend Line
Post-war Chained GDP (lag window150)

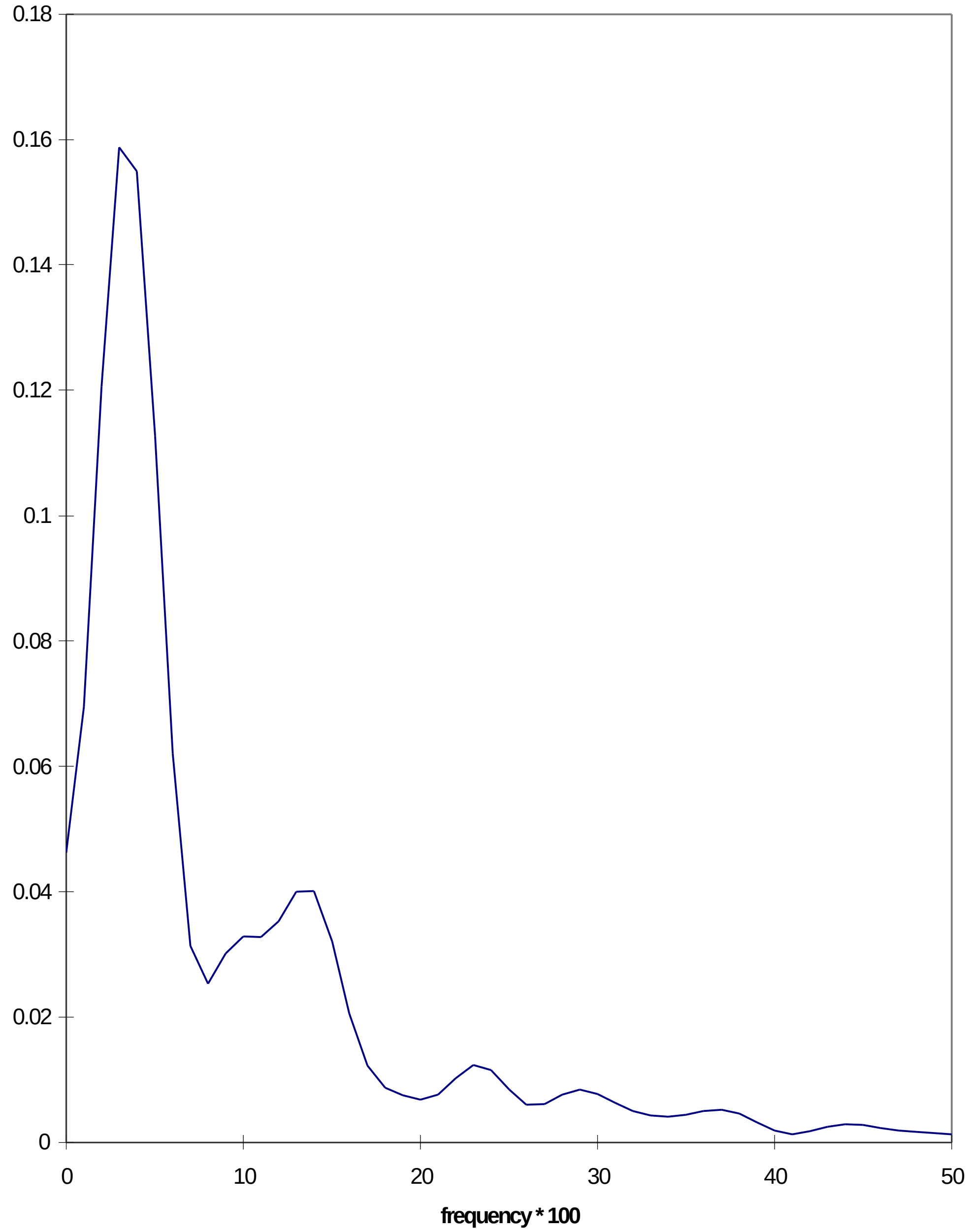

Figure 7
Variance Ratios for Log of Post-war Quarterly Chained GDP
Based on Cochrane's (1988) Bias Corrected Estimate