INTERNATIONAL ECONOMIC REVIEW
Vol. 33, No. 3, August 1992

TRENDS AND RANDOM WALKS IN MACROECONOMIC TIME SERIES: A RE-EXAMINATION*

BY GLENN D. RUDEBUSCH[1]

In their 1982 article, Nelson and Plosser provided evidence supporting the existence of an autoregressive unit root in a variety of macroeconomic time series. I re-examine their evidence using small-sample distributions for various unit root test statistics. These distributions are calculated from specific null and alternative models (including median-unbiased models that correct for OLS coefficient bias) estimated from the data. Contrary to earlier assertions, the null and alternative models of many macroeconomic series provide very different characterizations of persistence but cannot be distinguished with unit root tests.

1. INTRODUCTION

One of the most influential papers in macroeconomics during the last decade, in terms of its effect on research agendas and methodology, was "Trends and Random Walks in Macroeconomic Time Series" by Charles Nelson and Charles Plosser (1982). Nelson and Plosser were unable to reject the hypothesis of a single unit root in the autoregressive representations of a wide variety of macroeconomic time series, including employment, GNP, prices, interest rates, and stock prices. Following Nelson and Plosser's lead, macroeconomics has often been preoccupied, for better or worse, by unit roots.

The ramifications of unit roots have been widely felt. For example, application of unit root tests resulted in re-evaluations of the permanent income hypothesis of consumption (e.g., Mankiw and Shapiro 1985, Deaton 1987, and Diebold and Rudebusch 1991), the sustainability of government deficits (e.g., Hamilton and Flavin 1986), and the efficiency of stock markets and foreign exchange markets (e.g., Poterba and Summers 1987 and Meese and Singleton 1982). In addition, the general inability to reject a unit root was considered by many to be proof of the *existence* of a permanent component whose fluctuations are not eventually eliminated through reversion to trend. This naturally spurred attempts to determine the importance of that permanent component. Thus, Nelson and Plosser's work can be seen as a precursor to research on the closely related questions of the persistence of macroeconomic shocks (e.g., Campbell and Mankiw 1987, Cochrane 1988, and Diebold and Rudebusch 1989 in univariate contexts, and Shapiro and Watson 1988

* Manuscript received February 1991; revised October 1991.

[1] David Wilcox, Danny Quah, Doug Steigerwald, Gene Savin, Spencer Krane, Steve Durlauf, Frank Diebold, and two anonymous referees provided helpful comments. Michelle Phillips and Roberto Sella provided research assistance. The views expressed are my own and are not necessarily shared by the Board of Governors of the Federal Reserve System or its staff.

and Blanchard and Quah 1989 in multivariate ones) and the decomposition of macroeconomic dynamics into trend and cycle (e.g., Harvey 1985).

Nelson and Plosser's results have also influenced the direction of business cycle theorizing. Under the condition that monetary shocks have only temporary real effects, Nelson and Plosser asserted that their evidence of a permanent component in real output suggested that the source of business fluctuations is nonmonetary. This argument was one factor in the early popularity of real business cycle models (e.g., King and Plosser 1984).[2]

Finally, Nelson and Plosser's work helped to stimulate a general interest in econometrics about issues related to stationarity. The associated research led to several advances that have been widely applied, notably cointegration (e.g., Engle and Granger 1987) and new tests for unit roots (e.g., Phillips 1987).

In light of the influence of Nelson and Plosser (1982), this paper re-examines the data set used in their analysis to delineate the evidence regarding the existence and economic significance of a unit root. To examine existence, I estimate a difference-stationary (DS) null model and a trend-stationary (TS) alternative model for each variable and use resampling techniques to obtain the distributions of unit root test statistics under each model. The distributions obtained from the estimated DS models provide the appropriate small-sample size of the tests, while the distributions from the estimated TS models provide, for specific and plausible alternatives, the power of the tests, that is, their ability to reject a false null. By simulating specific DS and TS models estimated from the data, this study uses the most relevant models, initial conditions, sample sizes, and error distributions available for calculation of size and power.

However, an examination merely of test power is not enough. Although unit root tests may have low power against particular TS alternatives, that inadequacy does not necessarily compromise Nelson and Plosser's results. As Nelson and Plosser (1982, p. 152) note, no unit root test " ... can have power against a TS alternative with an [autoregressive] root arbitrarily close to unity. However, if we are observing stationary deviations from linear trends in these series then the tendency to return to the trend line must be so weak as to avoid detection even in samples as long as sixty years to over a century." Thus, Nelson and Plosser argue that the evidence they provide against the DS models is so weak that plausible alternative TS models most likely exhibit persistence properties that are similar to those of the DS models over economically relevant horizons. I examine the persistence properties of the estimated TS and DS models at a variety of horizons in order to assess the validity of this argument.

Section 2 describes the two unit root tests that I will employ and introduces a measure of persistence. The simulation technique used to obtain the small-sample distributions of the test statistics is outlined in Section 3. Section 4 contains empirical results from the Nelson and Plosser data set; for each series, I compare likelihoods of the estimated DS and TS models and contrast their persistence properties. Section 5 examines the consequences of small-sample bias in the

[2] Recently, others have noted that substantial persistence in output is compatible with a wide range of theoretical models, including monetary ones (e.g., West 1988).

estimates of the parameters of the TS models; examination of median-unbiased TS models qualifies some of the earlier results, notably for the nominal series. Section 6 offers some concluding comments and discusses the relationship of my results to recent related work.

2. THE PERMANENT COMPONENT: EXISTENCE AND IMPORTANCE

Consider the k-th order autoregressive representation ($AR(k)$),

$$x_t = \mu + \gamma t + \sum_{i=1}^{k} \rho_i x_{t-i} + \varepsilon_t, \tag{1}$$

where ε_t is a white noise innovation (and $k \geq 1$). There are two special cases of interest: (1) the *trend-stationary* (TS) model, where stationarity is assumed around a linear deterministic trend, so the roots of the lag operator polynomial $\rho(L) = 1 - \rho_1 L - \cdots - \rho_k L^k$ lie outside the unit circle, which implies that the sum of the autoregressive coefficients is less than one ($\Sigma_{i=1}^{k} \rho_i < 1$); and (2) the *difference-stationary* (DS) model, where $\gamma = 0$ and $\rho(L)$ contains one (positive, real) unit root, which implies $\Sigma_{i=1}^{k} \rho_i = 1$.[3] Note that equation (1) can be rearranged as[4]

$$x_t = \mu + \gamma t + \left(\sum_{i=1}^{k} \rho_i \right) x_{t-1} + \sum_{i=1}^{k-1} \left(- \sum_{j=i+1}^{k} \rho_j \right) \Delta x_{t-i} + \varepsilon_t. \tag{2}$$

Thus, the DS model has the $AR(k-1)$ representation in first differences,

$$\Delta x_t = \mu + \sum_{i=1}^{k-1} \phi_i \Delta x_{t-i} + \varepsilon_t, \tag{3}$$

where $\phi_i = -\Sigma_{j=i+1}^{k} \rho_j$.

A test to distinguish the TS model from the DS model, assuming the order k is known, is easily constructed.[5] Estimate equation (2), the "augmented Dickey-Fuller" regression, as

$$x_t = \hat{\mu} + \hat{\gamma} t + \hat{\delta} x_{t-1} + \sum_{i=1}^{k-1} \hat{\phi}_i \Delta x_{t-i} + \hat{\varepsilon}_t, \tag{4}$$

and test the unit root (or DS model) null hypothesis, H_0: $\delta = 1$.[6] A natural test

[3] More precisely, these models are integrated of orders zero and one, respectively.

[4] If $k = 1$, summations over $i = 1, \ldots, k-1$ are ignored; the DS model is simply a random walk with drift.

[5] Like Nelson and Plosser (1982), I only consider unit root tests that allow for a trend under the alternative. Such tests appear to be appropriate for most macroeconomic time series (note the time trend coefficients in the estimated TS models below); furthermore, West (1987) has shown that the tests without a trend have no power asymptotically against a trend alternative.

[6] The analysis below was also conducted with the Dickey-Fuller normalized bias and likelihood ratio tests, and qualitatively similar results were obtained.

statistic for this hypothesis is the t-test, defined as $\hat{\tau} = (\hat{\delta} - 1)/SE(\hat{\delta})$, where $SE(\hat{\delta})$ is the standard error of the estimated coefficient. However, Dickey (1976) and Fuller (1976) show that this statistic does not have the usual Student-t distribution but has a distribution that is skewed toward negative values. For $k = 1$, Dickey and Fuller report critical values for this test at conventional significance levels for a variety of sample sizes. For the case of k greater than one, Dickey and Fuller show that $\hat{\tau}$ will have an *asymptotic* distribution that is the same as when $k = 1$; thus, their reported asymptotic critical values for $\hat{\tau}$ with $k = 1$ can be applied in the case of arbitrary k.

In addition to the augmented Dickey-Fuller $\hat{\tau}$ test for stationarity, which was employed by Nelson and Plosser (1982), I also consider a second, more recent, test of stationarity that has been developed by Phillips (1987) and Phillips and Perron (1988). To test for a unit root in the $AR(k)$ process (1),[7] they estimate the first-order regression,

$$x_t = \bar{\mu} + \bar{\gamma} t + \bar{\delta} x_{t-1} + \bar{\eta}_t, \tag{5}$$

but modify—in a nonparametric way—the t-statistic $\bar{\tau} = (\bar{\delta} - 1)/SE(\bar{\delta})$ for likely serial correlation in residuals $\bar{\eta}_t$. The Phillips test statistic (for sample size T) takes the form[8]

$$\bar{Z} = \bar{\tau}(s_\eta / s_q) - (s_q^2 - s_\eta^2) T^3 [4 s_q (3 D_{XX})^{1/2}]^{-1},$$

where D_{XX} is the determinant of the regressor cross-product matrix from (5) and

$$s_\eta^2 = T^{-1} \sum_{t=1}^{T} \bar{\eta}_t^2,$$

$$s_q^2 = s_\eta^2 + 2T^{-1} \sum_{j=1}^{q} [1 - j/(1+q)] \sum_{t=j+1}^{T} \bar{\eta}_t \bar{\eta}_{t-j},$$

where q is a parameter that determines the number of autocorrelations used in the estimate of s_q^2. Under the unit root null that $\delta = 1$, the Phillips t-test, $\bar{Z}$, converges in the limit to the distribution of $\hat{\tau}$; thus, the asymptotic critical values tabulated in Dickey and Fuller are also appropriate for $\bar{Z}$.

It should be emphasized that in the general case of the k-th order autoregression in equation (1) only asymptotic critical values are available for either the augmented Dickey-Fuller or the Phillips unit root test. In finite samples, the distributions of $\hat{\tau}$ and $\bar{Z}$ under the null will depend on the sample size, the parameter values, and the distribution of the disturbances (see Evans and Savin 1984 and Schmidt 1990). The empirical section below reports critical values for $\hat{\tau}$ and $\bar{Z}$ that account for these

[7] Following Nelson and Plosser, I limit consideration in this paper to pure autoregressive models. Both the augmented Dickey-Fuller test (as noted by Said and Dickey 1984) and the Phillips test can be applied to more general mixed ARMA processes. I consider the issue of model specification in more detail below.

[8] More precisely, this is the form that I use; see Schwert (1989) for a discussion of various estimators of s_η^2 and s_q^2.

factors in the Nelson and Plosser data samples. In addition, in these same samples, I examine the power of the tests to reject the unit root null for plausible TS alternatives.

While $\hat{\tau}$ and $\tilde{Z}$ are able to test for the existence of a unit root, and hence test the DS model restrictions, these statistics will, of course, have low power against TS alternatives with an autoregressive root very close to unity. However, here, the assertion of Nelson and Plosser that was quoted in the introduction about the *economic importance* of trend reversion comes to the fore: the near-unit root in the TS model that makes it difficult to reject the unit root null may also imply that trend reversion in the TS model occurs only with a very long lag. Thus, it is necessary to supplement the unit root tests with a measure of the persistence of shocks. A unit root simply implies a nonzero permanent response of the time series to an innovation, while an examination of the amount of trend reversion at various horizons allows one to distinguish the DS model from the TS model in terms of economic significance.

The measure of persistence that I use is the size (at a given horizon) of the impulse response of the time series to a unit innovation.[9] Consider the moving-average representation of the first difference of the (TS or DS) time series x_t,

$$\Delta x_t = k + A(L)\varepsilon_t = k + (1 + a_1 L + a_2 L^2 + \cdots)\varepsilon_t, \tag{6}$$

where k is some constant. The measure of interest is the sum of the coefficients of the lag operator $A(L)$. A unit shock in period t affects the *change* in x at time $t + h$ by a_h and affects the *level* of x at time $t + h$ by $c_h = 1 + a_1 + \cdots + a_h$. For various horizons, the cumulative response c_h answers the question of interest: "How does a shock today affect x in the short, medium, and long run?" For example, with annual data, c_{10} measures the impact of a shock today on the level of x ten years hence. In the limit, the effect of a unit shock today on the level of x infinitely far in the future is c_∞. For any TS series, $c_\infty = 0$, because the effect of any shock is transitory as reversion to the deterministic trend eventually dominates. For a DS series, $c_\infty \neq 0$; that is, each shock has a permanent component. At shorter horizons, the dynamic responses of particular TS and DS representations may be quite similar or quite different, and below, I shall investigate these responses for TS and DS models that are estimated from a common data sample.

3. MONTE CARLO METHODOLOGY

My methodology is distinguished by careful construction of a sampling experiment that closely mimics the actual unit root inference problem for a typical macroeconomic time series. For each of the 14 macroeconomic variables used by Nelson and Plosser (1982), I formulate DS and TS data generating processes with parameters and innovations that are obtained from regressions estimated from the

[9] This measure is described in further detail in Diebold and Rudebusch (1989) and Diebold and Nerlove (1990).

data.[10] Doing so allows me to ascertain the DS model and TS model small-sample distributions of $\hat{\tau}$ and $\tilde{Z}$ for the Nelson and Plosser data samples and to describe the exact amount of support from the data for these models. In addition, and as importantly, I am able to clearly delineate the persistence properties, and thus the economic characteristics, of the relevant DS and TS models. In this section, I will illustrate the procedures using Nelson and Plosser's sample of real GNP.

The first step is to estimate both a TS and a DS model from the sample of the log of U.S. real GNP (Y_t), which consists of 62 annual observations (1909 through 1970). This requires specifying k, the order of the model, and I follow the choice of Nelson and Plosser (1982).[11] For real GNP, Nelson and Plosser set $k = 2$; thus, I estimate the TS model with a linear deterministic trend and second-order dependence for GNP as:

TS Model, $(k = 2)$

$$Y_t = \underset{(.270)}{.819} + \underset{(.0019)}{.0056}\ t + \underset{(.121)}{1.24}\ Y_{t-1} - \underset{(.121)}{.419}\ Y_{t-2} + \hat{u}_t,\ \hat{\sigma}_u = .0583. \tag{7}$$

The first two years of the sample are used up by initial conditions; the standard errors of the coefficients appear in parentheses. Estimating the DS model for this sample yields:

DS Model,

$$\Delta Y_t = \underset{(.009)}{.019} + \underset{(.124)}{.341}\ \Delta Y_{t-1} + \hat{v}_t,\ \hat{\sigma}_v = .0618. \tag{8}$$

As noted above, the DS model imposes two restrictions: (1) that the time trend coefficient is zero and (2) that the sum of the autoregressive coefficients is one (i.e., the unit root hypothesis).

Both models appear to fit the data quite well; the standard deviations of the fitted residuals, $\hat{u}_t$ and $\hat{v}_t$, are very close, and plots of the residuals are basically indistinguishable and suggest no obvious outliers. In addition, Q-statistics computed from the fitted residuals are similar at a variety of lags and provide little evidence against the null of no serial correlation. However, the two models have very different implications about the dynamics of GNP. I transform the estimated DS and TS models into moving average form (6) and calculate the cumulative impulse responses (c_h) for each model at a variety of horizons. These responses are shown in Table 1, with standard errors for the responses in parentheses.[12]

[10] Section 5 will examine the implications of the small-sample bias in the coefficients of these estimated autoregressive models.

[11] Nelson and Plosser implicitly specify k by selecting the number of Dickey-Fuller regression augmentation lags (which equals $k - 1$). Their selection is based on examination of the sample autocorrelation and partial autocorrelation functions; as noted below, uncertainty about the true value of k would only add to the uncertainty involved in the unit root inference.

[12] The standard errors are calculated as follows. Let the cumulative impulse response at horizon h be given by $c_h = F(\rho_1, \ldots, \rho_k)$, and let f denote the vector of partials of F with respect to the parameters. Then, the standard error equals $\sqrt{f'\Sigma f}$, where Σ is the estimated variance-covariance matrix of the autoregressive parameters.

TABLE 1
CUMULATIVE IMPULSE RESPONSE OF REAL GNP

	Horizon (years)						
	1	2	3	4	5	10	15
DS Model	1.34	1.46	1.50	1.51	1.52	1.52	1.52
	(.12)	(.21)	(.25)	(.27)	(.28)	(.29)	(.29)
TS Model	1.24	1.12	.87	.61	.39	.00	.00
	(.12)	(.20)	(.24)	(.26)	(.27)	(.13)	(.01)

The impulse response of the DS model implies not only shock persistence but shock magnification, because an innovation is not reversed but eventually changes the level of real GNP by more than one and a half times the size of the innovation ($c_{10} = 1.52$). In contrast, the TS model exhibits fairly rapid reversion to trend, with the effect of a shock essentially completely disappearing after ten years ($c_{10} = 0$). The cumulative impulse responses for the two models, each estimated from a common data sample, are clearly very different in both economic and statistical terms.

Ideally, the unit root tests could distinguish between these two models. For this sample of real GNP, Nelson and Plosser (1982) report a $\hat{\tau}$ test statistic equal to -2.99, which is insignificant at even the 10 percent level, according to the asymptotic critical value of -3.12 given in Fuller (1976); thus, the data offer no evidence to reject the DS model. Nelson and Plosser (p. 160) go even further and argue that this result would be consistent with a TS model "only if the fluctuations around a deterministic trend are so highly autocorrelated as to be indistinguishable from nonstationary series themselves in realizations as long as one hundred years." Clearly, this statement would seem to rule out the TS model estimated for real GNP above in light of that model's rapid trend reversion displayed in Table 1.

I will, in effect, re-examine Nelson and Plosser's evidence and their assertion. By repeatedly simulating models (7) and (8), I calculate distributions of the unit root test statistics, under the DS model and under the TS model, that are precisely tailored to Nelson and Plosser's GNP data sample. For example, from the estimated DS model, I generate 10,000 data sets. Each data set includes the historical levels of the log of GNP in 1909 and 1910 as initial conditions and evolves according to the estimated equation in (8) with disturbances obtained by random draws (with replacement) from the set of fitted residuals $\{\hat{v}_1, \ldots, \hat{v}_{60}\}$.[13] For each data set, $\hat{\tau}$ is computed, and the resulting 10,000 realizations of this statistic provide the appropriate DS model distribution. Similarly, using the parameters from the estimated TS model (7), I generate 10,000 data sets using disturbances obtained by random draws from $\{\hat{u}_1, \ldots, \hat{u}_{60}\}$. The resulting 10,000 realizations of $\hat{\tau}$ provide

[13] I also obtained virtually identical results below from simulations using normal errors (with the variance of the residuals) instead of the actual redrawn (bootstrapped) residuals. Both simulation procedures, of course, assume that the disturbances are identically and independently distributed (iid). This last assumption was also maintained in Nelson and Plosser's application of the Dickey-Fuller test, and they selected k for each series to obtain iid residuals. Formal tests, such as the Ljung-Box Q-statistic and the Breusch-Pagan heteroskedasticity test generally support this assumption for each of the time series.

the sampling distribution of the augmented Dickey-Fuller statistic from a plausible TS model. Of interest for examining the existence of a unit root in this sample of real GNP are the probabilities of obtaining a value of $\hat{\tau}$ as extreme as -2.99 from these TS model and DS model test statistic distributions. The next section provides these probabilities and similar ones for the Phillips $\tilde{Z}$ test. Application of this same methodology also provides DS and TS model probabilities for all of the other macroeconomic time series examined by Nelson and Plosser.

4. EMPIRICAL EVIDENCE FOR THE DS AND TS MODELS

This section implements the resampling procedure described above for each of the fourteen data samples considered in Nelson and Plosser (1982).[14] The first step is to estimate TS and DS models, of form (1) and (3) respectively, for each series. The TS model estimates are shown in Table 2, with the real GNP estimates described above repeated in the first row. For each series, the selection of k, the number of autoregressive lags used in the TS model, follows the choice of Nelson and Plosser. Note that for all but one of the estimated models the sum of the autoregressive parameters is close to, but less than, unity. The exception is the model of the bond yield, where $\Sigma_{i=1}^{k} \rho_i > 1$, so the estimated TS model is not stationary; thus, I shall not include this series in the simulations below.

The DS models for the fourteen variables, estimated in first differences, are shown in Table 3. As was the case for real GNP, the standard errors of the TS and DS models for each of the series are very close in size.

Table 4 provides cumulative impulse responses for the estimated TS and DS models shown in Tables 2 and 3, as well as, in the final column, the standard error (conditional on each model) for the 10-year response. The rapid divergence between the DS and TS model responses that was evident for real GNP is typical of the other series as well. There are some exceptions: after 30 years, the TS model for consumer prices displays substantial persistence, while the DS model for industrial production displays some shock dampening. Overall, however, the estimated TS and DS models provide very different descriptions of the dynamics of each of the series, with the TS models rapidly reverting to trend after a unit innovation and the DS models exhibiting shock magnification.

Before considering the ability of the unit root tests to discriminate between the TS and DS models, I first examine the tests under the DS model in order to indicate how appropriate the widely-used asymptotic Dickey-Fuller critical values are for statistical inference in the Nelson and Plosser data set. For a given critical value (of nominal size, say, 5 percent) the proportion of rejections among the 10,000 simulated DS model data sets, given the truth of the null in the data generation process, is the empirical size of a test. This empirical size may differ from nominal size because it accounts for several factors: (1) the finite size of the sample, (2) the

[14] The fourteen series are real GNP (sample size, $T = 62$), nominal GNP (62), real per capita GNP (62), industrial production (111), employment (81), unemployment rate (81), GNP deflator (82), consumer prices (111), nominal wages (71), real wages (71), money stock (82), velocity (102), bond yield (71), and an index of common stock prices (100). The series, which are annual and in logs (except for the bond yield), are described further in Nelson and Plosser (1982).

TABLE 2
ESTIMATED COEFFICIENTS OF TS MODELS[a]

Series	Const.	t	Y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	Standard Error
Real GNP	.819	.0056	1.24	−.419			.0583
	(.270)	(.0019)	(.121)	(.121)			
Nominal GNP	1.06	.0056	1.39	−.489			.0871
	(.448)	(.0024)	(.117)	(.117)			
Real GNP, p.c.	1.28	.0035	1.23	−.410			.0590
	(.419)	(.0012)	(.122)	(.121)			
Industrial Prod.[b]	.103	.0067	.931	−.134	.084	−.091	.0973
	(.024)	(.0027)	(.099)	(.136)	(.136)	(.135)	
Employment	1.42	.0021	1.27	−.479	.072		.0353
	(.529)	(.0008)	(.117)	(.180)	(.116)		
Unemployment	.514	−.0005	1.09	−.585	.445	−.243	.4068
	(.183)	(.0021)	(.107)	(.148)	(.146)	(.105)	
GNP Deflator	.260	.0021	1.37	−.454			.0460
	(.102)	(.0008)	(.102)	(.101)			
Consumer Prices	.090	.0006	1.66	−.953	.358	−.091	.0419
	(.051)	(.0002)	(.094)	(.182)	(.180)	(.092)	
Wages	.566	.0038	1.45	−.604	.068		.0599
	(.246)	(.0017)	(.126)	(.208)	(.126)		
Real Wages	.488	.0035	1.08	−.252			.0346
	(.157)	(.0011)	(.118)	(.116)			
Money Stock	.133	.0049	1.58	−.663			.0468
	(.038)	(.0016)	(.086)	(.086)			
Velocity	.052	−.0003	.941				.0671
	(.052)	(.0018)	(.035)				
Bond Yield	−.187	.0032	1.13	.204	−.305		.2836
	(.197)	(.0018)	(.127)	(.189)	(.139)		
Stock Prices	.096	.0032	1.20	−.425	.134		.1543
	(.056)	(.0013)	(.104)	(.157)	(.105)		

[a] These models are estimated on the level of the series. All data, except bond yields, are in log form. Standard errors are given in parentheses.

[b] The coefficients on the fifth and sixth lags of the TS model for industrial production are −.175 and .222, respectively, with standard errors equal to .135 and .097.

estimated values of the parameters, and (3) the estimated error distribution. Using the asymptotic critical values reported in Dickey and Fuller, Table 5 provides, for each series, the actual percentage of rejections of the unit root null out of the 10,000 DS model replications. Critical values at the 1 percent and 5 percent level are used for both the $\hat{\tau}$ and $\tilde{Z}$ tests. Below these empirical sizes, Table 5 also provides the actual 1 percent and 5 percent cutoff values. The $\hat{\tau}$ test exhibits an empirical size that is very close to its nominal size for all variables and at both significance levels; at the nominal 5 percent level, for example, empirical size ranges from 4.29 to 6.51 percent. In contrast, the Phillips $\tilde{Z}$ test appears to be mis-sized in these small samples; at the nominal 5 percent level, the empirical size of $\tilde{Z}$ ranges from 0.58 to 27.62 percent.[15] Thus, Table 5 indicates that for classical hypothesis tests of the unit root null hypothesis in the Nelson and Plosser samples, the asymptotic critical

[15] This is consistent with the results of Schwert (1989) that suggest the Phillips test is incorrectly sized in samples of ARIMA (0, 1, 1) processes. In the Phillips test, I set q equal to 6; however, the results were qualitatively identical with q equal to 10.

TABLE 3
ESTIMATED COEFFICIENTS OF DS MODELS[a]

Series	Const.	Δy_{t-1}	Δy_{t-2}	Δy_{t-3}	Δy_{t-4}	Standard Error
Real GNP	.019 (.009)	.341 (.124)				.0618
Nominal GNP	.031 (.013)	.439 (.118)				.0897
Real GNP, p.c.	.011 (.008)	.332 (.124)				.0627
Industrial Prod.[b]	.069 (.014)	.006 (.096)	−.126 (.096)	−.030 (.096)	−.116 (.095)	.0997
Employment	.012 (.005)	.375 (.113)	−.171 (.113)			.0365
Unemployment	−.014 (.050)	.217 (.108)	−.352 (.102)	.122 (.106)		.4357
GNP Deflator	.012 (.006)	.434 (.102)				.0475
Consumer Prices	.005 (.004)	.708 (.094)	−.291 (.113)	.078 (.093)		.0432
Wages	.025 (.009)	.533 (.123)	−.150 (.123)			.0614
Real Wages	.014 (.005)	.191 (.120)				.0366
Money Stock	.022 (.008)	.622 (.089)				.0490
Velocity	−.012 (.007)					.0684
Bond Yield	.040 (.036)	.178 (.126)	.369 (.130)			.2862
Stock Prices	.027 (.016)	.266 (.102)	−.187 (.102)			.1574

[a] These models are estimated on the first difference of the series. All data, except bond yields, are in log form. Standard errors are given in parentheses.

[b] The coefficient on the fifth lag of the DS model for industrial production is −.286 with standard error equal to .096.

values appear to be appropriate approximations for the $\hat{\tau}$ test but not for the $\tilde{Z}$ test.[16]

Of more interest, however, than just the appropriateness of the asymptotic critical values is the actual likelihood of obtaining the sample values of the test statistics, denoted $\hat{\tau}_s$ and $\tilde{Z}_s$, from their estimated DS model and TS model small-sample distributions. For real GNP, Figure 1 displays the estimated DS model density function for $\hat{\tau}$, denoted $f_{DS}(\hat{\tau})$, and the estimated TS model density function for $\hat{\tau}$, denoted $f_{TS}(\hat{\tau})$; these are empirical densities formed from the realizations of the test statistic in the 10,000 samples from each model. The sample value of the augmented Dickey-Fuller t-test for real GNP, which is equal to −2.99, is shown as a vertical dotted line. There are two areas in Figure 1 of special interest. The hatched area under $f_{DS}(\hat{\tau})$ and to the left of $\hat{\tau}_s$ represents the probability of

[16] It should be stressed that the $\hat{\tau}$ test, unlike the Phillips test, incorporates the order of the data-generating process, which is assumed to be known. If the incorrect order were used, it is likely that the divergence between the empirical and nominal size would be greater.

TABLE 4
CUMULATIVE IMPULSE RESPONSES OF ESTIMATED TS AND DS MODELS[a]

Series	Model	Horizon (years) 1	2	3	4	5	10	15	30	$se(c_{10})$
Real GNP	DS	1.34	1.46	1.50	1.51	1.52	1.52	1.52	1.52	.29
	TS	1.24	1.12	.87	.61	.39	.00	.00	.00	.13
Nominal GNP	DS	1.44	1.63	1.72	1.75	1.77	1.78	1.78	1.78	.37
	TS	1.39	1.44	1.33	1.14	.93	.24	.04	.00	.35
Real GNP, p.c.	DS	1.33	1.44	1.48	1.49	1.50	1.50	1.50	1.50	.28
	TS	1.23	1.10	.85	.60	.38	.00	.00	.00	.12
Industrial Prod.	DS	1.01	.88	.85	.75	.75	.79	.79	.79	.06
	TS	.93	.73	.64	.49	.17	.22	.07	.01	.14
Employment	DS	1.38	1.35	1.27	1.25	1.25	1.26	1.26	1.26	.21
	TS	1.27	1.13	.90	.70	.53	.14	.04	.00	.20
Unemployment	DS	1.22	.91	.89	1.02	1.02	.98	.99	.99	.18
	TS	1.09	.60	.47	.40	.16	−.07	.00	.00	.08
GNP Deflator	DS	1.43	1.62	1.70	1.74	1.76	1.77	1.77	1.77	.32
	TS	1.37	1.42	1.33	1.17	1.00	.38	.13	.01	.31
Consumer Prices	DS	1.71	1.92	1.94	1.95	1.96	1.98	1.98	1.98	.38
	TS	1.66	1.80	1.77	1.72	1.67	1.28	.96	.41	.35
Wages	DS	1.53	1.67	1.66	1.63	1.62	1.62	1.62	1.62	.33
	TS	1.45	1.50	1.37	1.17	.98	.34	.12	.00	.36
Real Wages	DS	1.19	1.23	1.23	1.24	1.24	1.24	1.24	1.24	.18
	TS	1.08	.91	.72	.54	.41	.09	.02	.00	.12
Money Stock	DS	1.62	2.01	2.25	2.40	2.49	2.63	2.64	2.65	.59
	TS	1.58	1.83	1.85	1.71	1.47	.23	−.13	.01	.45
Velocity	DS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.00
	TS	.94	.89	.83	.78	.74	.54	.40	.16	.21
Stock Prices	DS	1.27	1.15	1.07	1.07	1.09	1.09	1.09	1.09	.15
	TS	1.20	1.02	.84	.74	.67	.38	.21	.04	.22

[a] The standard error of the estimated cumulative impulse reponse at a horizon of 10 years is given in the last column as $se(c_{10})$.

obtaining a value of the t-test equal to or smaller than -2.99, conditional on the DS model. This probability, which equals 0.150 for real GNP, is denoted as

$$\text{DS } p\text{-value} = \text{prob } (\hat{\tau} \leq \hat{\tau}_s | \text{DS model}).$$

This p-value is the marginal significance level for rejection of the DS model null hypothesis; that is, in a classical hypothesis testing framework, given the sample test statistic $\hat{\tau}_s$, we could not reject the DS model at anything less than the 15 percent level. This is consistent with the inability of Nelson and Plosser to reject the DS model.

The other area of interest is the shaded region under $f_{\text{TS}}(\hat{\tau})$ and to the right of $\hat{\tau}_s$, which represents the probability of obtaining a value of the t-test equal to or greater than -2.99, conditional on the TS model. This probability is denoted

$$\text{TS } p\text{-value} = \text{prob } (\hat{\tau} \geq \hat{\tau}_s | \text{TS model}).$$

For real GNP, the TS p-value is 0.216; in other words, it is not very unlikely that a sample statistic as large as -2.99 could have been generated from the estimated

TABLE 5
EMPIRICAL SIZE OF UNIT ROOT TESTS[a]

	$\hat{\tau}$		$\tilde{Z}$	
Series (k)	1%	5%	1%	5%
Real GNP (2)	1.66	6.40	.24	1.17
	−4.13	−3.53	−3.49	−2.93
Nominal GNP (2)	1.72	6.38	.43	1.19
	−4.23	−3.52	−3.54	−2.82
Real GNP, p.c. (2)	1.56	6.26	.22	1.25
	−4.12	−3.51	−3.50	−2.95
Industrial Prod. (6)	1.41	5.42	8.01	27.62
	−4.08	−3.45	−4.67	−4.12
Employment (3)	1.02	4.88	.11	.84
	−3.97	−3.40	−3.37	−2.97
Unemployment (4)	1.55	6.08	2.74	10.30
	−4.12	−3.50	−4.29	−3.73
GNP Deflator (2)	1.25	5.12	.50	1.50
	−4.05	−3.42	−3.58	−2.93
Consumer Prices (4)	1.46	5.93	.86	3.60
	−4.10	−3.47	−3.87	−3.27
Wages (3)	1.35	5.40	.11	.80
	−4.10	−3.45	−3.33	−2.85
Real Wages (2)	.99	4.29	.08	.58
	−3.96	−3.34	−3.30	−2.82
Money Stock (2)	1.78	6.51	.21	.87
	−4.14	−3.55	−3.38	−2.82
Velocity (1)	1.27	5.49	1.39	6.47
	−4.06	−3.45	−4.07	−3.51
Stock Prices (3)	1.14	5.02	.61	3.83
	−3.99	−3.42	−3.80	−3.32

[a] The upper number in each cell is the percentage of rejections using the asymptotic Dickey-Fuller critical values, namely, −3.96 at the 1% level and −3.41 at the 5% level. The actual cutoff value for the 1% or the 5% fractile is given as the lower number. The number of augmentation lags in the Dickey-Fuller test equals k-1, while q equals 6 in each Phillips test.

TS model.[17] In short, at conventional significance levels, there is very little evidence against either the DS model or the TS model for real GNP. The evidence in Nelson and Plosser (1982) on the DS model probabilities thus provides only one side of the story for inference regarding the unit root hypothesis. The other side is that the TS model is at least as consistent with the sample test statistic.

The other variables provide similar results. For each series, the estimated densities, f_{DS} and f_{TS}, have a substantial region of overlap, and the sample value, $\hat{\tau}_s$, falls into the range of this overlap. The overlap is obvious in Table 6, which gives $\hat{\tau}_s$ for each series with its associated DS p-value and TS p-value. Among all of the variables, there is only one model that can be rejected at the 5 percent level: the DS model for the unemployment rate.

Table 7 provides, for each series, the sample Phillips test statistics, $\tilde{Z}_s$, along with their p-values conditional on the DS or TS models. The qualitative results are similar to those from the augmented Dickey-Fuller test: there is a substantial region

[17] An equivalent statement of this result is that the $\hat{\tau}$ test of the DS null at the 15 percent significance level has a power against the TS alternative of only 78.4 percent.

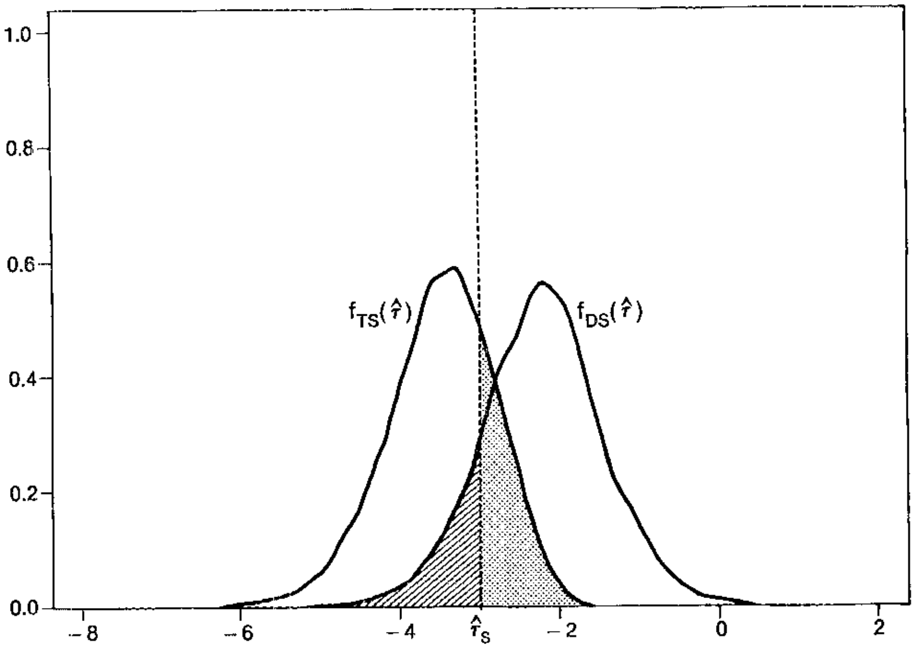

FIGURE 1

EMPIRICAL DENSITIES OF $\hat{\tau}$ FROM DS AND TS MODELS OF REAL GNP

of overlap between the estimated DS model and TS model densities of $\tilde{Z}$, and the sample test statistic falls into this region of overlap. Only two of the 26 estimated models have p-values less than 5 percent, namely, the TS models for velocity and stock prices.

The conclusion to be drawn from Tables 6 and 7 is that there is very little

TABLE 6

SAMPLE $\hat{\tau}$ TEST STATISTICS AND PROBABILITIES[a]

Series	$\hat{\tau}_s$	DS p-value	TS p-value
Real GNP	−2.99	.150	.216
Nominal GNP	−2.32	.429	.148
Real GNP, p.c.	−3.05	.137	.223
Industrial Prod.	−2.53	.301	.177
Employment	−2.66	.229	.196
Unemployment	−3.55	.045	.268
GNP Deflator	−2.52	.280	.228
Consumer Prices	−1.97	.613	.128
Wages	−2.24	.425	.165
Rela Wages	−3.05	.094	.403
Money Stock	−3.08	.134	.261
Velocity	−1.66	.761	.075
Stock Prices	−2.12	.522	.104

[a] The sample augmented Dickey-Fuller test statistic, $\hat{\tau}_s$, is given with its p-values from the TS and DS models. For each series, k equals the value given in Table 5.

TABLE 7
SAMPLE $\tilde{Z}$ TEST STATISTICS AND PROBABILITIES[a]

Series	$\tilde{Z}_s$	DS p-value	TS p-value
Real GNP	−2.25	.307	.131
Nominal GNP	−1.77	.592	.095
Real GNP, p.c.	−2.34	.255	.161
Industrial Prod.	−3.31	.321	.201
Employment	−2.39	.271	.156
Unemployment	−3.08	.198	.235
GNP Deflator	−2.34	.234	.263
Consumer Prices	−1.45	.814	.117
Wages	−2.05	.427	.212
Real Wages	−2.65	.094	.383
Money Stock	−2.33	.193	.341
Velocity	−1.47	.864	.047
Stock Prices	−1.96	.665	.044

[a] The sample Phillips test statistic, $\tilde{Z}_s$, is given with its p-values from the TS and DS models. For each series, q equals six.

evidence to support the rejection of the DS model or the TS model in *any* of the Nelson and Plosser time series. Given these data and test statistics, little can be said about the choice between the DS and TS models. The proper conclusion from application of the unit root $\hat{\tau}$ test to these samples is that we cannot answer the question of the existence of a unit root.

5. CORRECTION OF BIAS IN TS MODEL ESTIMATES

There is one qualification to the preceding analysis that must be considered. For each series, the specific alternative examined is the TS model estimated from the available data sample with OLS. Is this model the *most* plausible alternative? Perhaps not. Although the OLS estimates of the coefficients of the TS model are consistent and asymptotically normal, they are biased in small samples because of the presence of lagged dependent variables.[18] An arguably more plausible TS alternative would correct the coefficient estimates for this bias, and this section constructs and analyzes such an alternative.[19]

The small-sample bias of the OLS estimates of autoregressive model coefficients is most easily documented for an AR(1) process. The middle column of Table 8 provides the median value of the OLS estimate $\hat{\rho}_1$, based on repeated samples from the first-order process $y_t = \mu + \gamma t + \rho_1 y_{t-1} + \varepsilon_t$, for a variety of values of ρ_1 (with $\mu = \gamma = 0$). This estimate is downwardly biased over a wide range of ρ_1, with the deviation of the median estimate from the true value being particularly

[18] This bias is most severe in the presence of a unit root, the so-called "Dickey-Fuller bias," but it is present for all autoregressions. For further discussion of this bias, see Evans and Savin (1984) and Stine and Shaman (1989).

[19] The coefficient estimates of the DS model, which is an autoregressive model in differences, are biased as well. However, because the roots of the DS model are distant from the unit circle, the biases in the DS model coefficients were small and inconsequential for the analysis.

TABLE 8
SMALL-SAMPLE BIAS OF OLS ESTIMATE OF AR(1) MODEL[a]

ρ_1	median ($\hat{\rho}_1$)	prob ($\hat{\rho}_1 \geq \rho_1$)
0.70	0.676	0.333
0.72	0.695	0.325
0.74	0.715	0.315
0.76	0.735	0.306
0.78	0.754	0.296
0.80	0.773	0.286
0.82	0.792	0.273
0.84	0.810	0.257
0.86	0.829	0.240
0.88	0.847	0.221
0.90	0.864	0.194
0.92	0.880	0.158
0.94	0.893	0.117
0.96	0.902	0.066
0.98	0.906	0.022
1.00	0.910	0.005
1.02	0.991	0.088

[a] Column 2 provides the median OLS estimate of ρ_1 in the regression $y_t = \hat{\mu} + \hat{\gamma}t + \hat{\rho}_1 y_{t-1} + \varepsilon_1$, while column 3 gives the proportion of estimates that are greater than or equal to ρ_1. These are based on 10,000 samples of size 100 generated from $y_t = \rho_1 y_{t-1} + \varepsilon_t$ with normal disturbances and $\sigma_\varepsilon = 0.1$ and an initial condition of zero.

pronounced for values of ρ_1 that are just less than one.[20] The third column of Table 8 gives the probability of obtaining an OLS estimate equal to or greater than ρ_1; these probabilities also indicate that a given estimate is more likely to be below rather than above the true value of the autoregressive parameter.

Based on Table 8, it is likely that the methodology of Sections 4 and 5 would employ a TS model (when $k = 1$) that incorporated an estimate of ρ_1 that is lower than the true value of ρ_1. A more plausible TS model of the data-generating process would correct for this downward small-sample bias.[21] I define the "median-unbiased" TS model as the one that, across repeated simulations, has a median OLS estimate of each parameter that is equivalent to the actual sample estimate of that parameter. Formally, let the vector of median-unbiased TS model coefficients be denoted as $\Phi_{\text{MUE}} = (\mu, \gamma, \rho_1, \dots, \rho_k)$; across repeated samples the median OLS estimates of these coefficients is median ($\hat{\Phi}_{\text{MUE}}$). Let $\hat{\Phi}_s$ be the vector of OLS parameters estimated from the data sample under consideration (which formed the parameters of the TS data-generating processes in Table 2). The vector Φ_{MUE} is

[20] The distribution of $\hat{\rho}_1$ is negatively skewed, so the median is a better measure of central tendency than the mean: I use the term "bias" in this paper to denote the deviation of the *median* from the true value.

[21] Several bias corrections have been proposed for the OLS estimates of the autoregressive model. For example, Orcutt and Winokur (1969) and Rudebusch (1993) recommend mean-unbiased estimators, and Andrews (1990), independently of the present paper, proposes a median-unbiased one. The simulation strategy pursued below can approximate these estimators to any desired degree of accuracy. One advantage to *median*-unbiased estimates of the parameters is that they imply a median-unbiased estimate of the cumulative impulse response (which is a nonlinear function of the parameters). This property does not hold for *mean*-unbiased estimates.

TABLE 9
MEDIAN-UNBIASED TS MODEL COEFFICIENTS[a]

Series	Const.	t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
Real GNP	.507	.0035	1.296	−.398		
Nominal GNP	.657	.0031	1.449	−.503		
Real GNP, p.c.	.882	.0023	1.276	−.394		
Industrial Prod.[b]	.094	.0033	.976	−.127	.072	−.080
Employment	1.043	.0015	1.323	−.508	.085	
Unemployment	.402	−.0002	1.130	−.592	.456	−.221
GNP Deflator	.110	.0008	1.421	−.453		
Consumer Prices	.039	.0004	1.688	−.984	.360	−.079
Wages	.186	.0011	1.538	.672	.108	
Real Wages	.292	.0020	1.144	−.231		
Money Stock	.096	.0032	1.616	−.669		
Velocity	−.025	.0003	.995			
Stock Prices	.029	.0010	1.269	−.446	.161	

[a] When simulated 1000 times, these models have median OLS parameter estimates equal to those of the corresponding TS models in Table 2.

[b] The coefficients on the fifth and sixth lags are −.184 and .262, respectively.

defined by the equality of median $(\hat{\Phi}_{MUE})$ and $\hat{\Phi}_s$; that is, the median-unbiased model has median OLS parameter estimates equal to the sample estimates.

For multi-parameter models, finding Φ_{MUE} is complicated by the covariance among estimated parameters. This problem can be solved by repeated simulations of a baseline model where each baseline model coefficient is changed by a fraction of the difference between the resulting median estimate of that coefficient and the actual sample value. Convergence occurs when all of the median estimates match the sample ones. Table 9 presents the coefficients of these median-unbiased TS models for each of the series.[22]

What are the consequences of using the median-unbiased TS model coefficients for inference about unit roots? Examination of the AR(1) case in Table 8 is instructive; the median-unbiased estimate of ρ_1 will be closer to one than the OLS estimate (assuming $\rho_1 < 1$). Thus, the median-unbiased model will exhibit a greater persistence of shocks than the uncorrected model;[23] thus, in economic terms, the median-unbiased model will behave more like the DS model. On the other hand, the median-unbiased TS model will be even harder to distinguish from the DS model using unit root tests; that is, the power of the unit root tests will even be lower against the median-unbiased alternative.

These two results—higher persistence and lower power—usually generalize to the higher-order median-unbiased models as well. The cumulative impulse response of the median-unbiased model is shown in Table 10 for each series. The

[22] To be precise, when 1000 samples are generated from the models in Table 9 (of size equal to the sample data size of the series given in footnote 14) the difference between the median OLS estimate for each coefficient and the sample OLS estimate (given in Table 2) is less than .001 (except for the time coefficient when it was less than .0001). The disturbances for the simulations are generated from a normal distribution with standard error equal to the OLS estimate of the regression standard error (given for each series in Table 2). Letting this standard error be a free parameter along with the other coefficients led to qualitatively identical results.

[23] For an $AR(1)$, $c_n = \rho_1^n$, so the cumulative impulse response is higher at all horizons.

TABLE 10
CUMULATIVE IMPULSE RESPONSES OF MEDIAN-UNBIASED TS MODELS

Series	Horizon (years)							
	1	2	3	4	5	10	15	30
Real GNP	1.30	1.28	1.15	.98	.81	.28	.09	.00
Nominal GNP	1.45	1.60	1.59	1.49	1.37	.75	.38	.05
Real GNP, p.c.	1.28	1.24	1.07	.89	.71	.19	.05	.00
Industrial Prod.	.98	.83	.75	.62	.31	.42	.24	.12
Employment	1.32	1.24	1.06	.88	.73	.30	.13	.01
Unemployment	1.13	.69	.56	.52	.32	−.02	−.01	.00
GNP Deflator	1.42	1.57	1.58	1.54	1.47	1.10	.81	.32
Consumer Prices	1.69	1.87	1.85	1.81	1.78	1.54	1.31	.82
Wages	1.54	1.69	1.68	1.61	1.53	1.18	.92	.43
Real Wages	1.14	1.08	.97	.86	.76	.41	.22	.03
Money Stock	1.62	1.94	2.06	2.03	1.90	.86	.20	−.02
Velocity	1.00	.99	.99	.98	.98	.95	.93	.86
Stock Prices	1.27	1.16	1.07	1.04	1.04	.95	.86	.65

responses of these models display greater persistence for each series than the responses of the uncorrected OLS TS models shown in Table 4. For example, the median-unbiased TS model of real GNP has a cumulative impulse response at a horizon of ten years (c_{10}) of .28 as compared to the zero response for the uncorrected model. However, for about half of the series, the median-unbiased TS alternative still has much less persistence than the DS model. In particular, for the real series (real GNP, real GNP per capita, industrial production, employment, unemployment, and real wages), reversion to trend is fairly rapid (c_{10} is well below 0.5). The median-unbiased TS models for the remaining nominal series show considerably more persistence; indeed, for velocity and stock prices, the median-unbiased TS model is virtually indistinguishable from the DS model.

Finally, from repeated simulations of the median-unbiased TS models, Table 11

TABLE 11
PROBABILITIES OF SAMPLE TEST STATISTICS FROM MEDIAN-UNBIASED TS MODELS

Series	p-value ($\hat{\tau}_s$)	p-value ($\tilde{Z}_s$)
Real GNP	.587	.458
Nominal GNP	.300	.223
Real GNP, p.c.	.504	.409
Industrial Prod.	.547	.623
Employment	.227	.159
Unemployment	.505	.441
GNP Deflator	.704	.807
Consumer Prices	.397	.358
Wages	.457	.500
Real Wages	.692	.607
Money Stock	.536	.588
Velocity	.394	.310
Stock Prices	.468	.347

provides the p-values of the sample values of the unit root test statistics.[24] For each series, the p-values of $\hat{\tau}_s$ and $\tilde{Z}_s$ are quite large and provide no evidence against the TS alternative. For each nominal series, the lack of evidence is not surprising because the median-unbiased alternative is so similar in economic terms to the DS model. However, for the real series, Table 11 reinforces the conclusion of Section 4 that there are plausible TS alternatives that are different from DS models in economic terms but which cannot be identified by unit root tests.

6. CONCLUSION

Until about a decade ago, economists were in broad agreement that macroeconomic variables were trend stationary; as a prominent example, the business cycle fluctuations of real output were treated as stationary deviations from a steadily growing trend. This general agreement was shattered by Nelson and Plosser (1982), and a new consensus was formed that macroeconomic variables were best modeled as difference stationary. The evidence above indicates that, at least for *real* macroeconomic variables, this new consensus has no firmer statistical foundation than the one it replaced. The Nelson and Plosser sample of data does not support the proposition that unit roots are a pervasive element in real macroeconomic time series.[25] The unit root tests employed by Nelson and Plosser display low power, not against TS alternatives with a "root arbitrarily close to unity," but against plausible TS models estimated from the data. Furthermore, if the alternative TS model is true, the DS model does not provide a good approximation for even medium-term dynamic responses. For each nominal time series, the evidence is more ambiguous because the most plausible alternative TS model displays substantial persistence; however, in light of the impulse response of this TS model, an appropriate confidence interval around an estimate of medium-term persistence is much larger than the one suggested by conditioning on the DS model alone.[26]

The above analysis can be fruitfully contrasted with two other recent papers in this area. My results were obtained even though I limited the class of models under consideration to $AR(k)$ processes with specific and known k. Variation in the order of the model will likely lead to variation in the relative probabilities of the DS and TS models; thus, uncertainty about the underlying model would only add to the uncertainty about unit roots. This conjecture is supported by Christiano and Eichenbaum (1990) who examine a variety of ARMA representations for postwar U.S. real GNP. They argue that imposing the TS constraint that $c_\infty = 0$ leads to only small differences in likelihood and that errors in specifying the order of the

[24] The simulation methodology was the same as described in Sections 3 and 4, with initial conditions equal to the first k observations for each series. However, the disturbances were drawn from normal distributions instead of bootstrapping, but as noted in footnote 12, this change should have negligible effect.

[25] Further evidence on this issue for real GNP (where the Nelson and Plosser data are of dubious origin) is given in Rudebusch (1993) using a sample of postwar quarterly data.

[26] This suggests the importance of measuring the confidence intervals for estimates of persistence without conditioning on the TS or DS model. Diebold and Rudebusch (1989b) provide a first step in this direction using a model of fractional integration.

model can greatly affect these differences and hence affect the inference about unit roots.[27]

DeJong, Nankervis, Savin, and Whiteman (1989) provide a complementary investigation of the power of unit root tests against a comprehensive array of first-order autoregressive alternatives (at the nominal 5 percent significance level). Their conclusion also stresses the difficulty of detecting a TS process. My analysis differs from their work in that it focuses on specific, but arguably the most relevant, higher-order null and alternative models and provides an assessment of power at the most relevant significance level (the empirical marginal significance level of the DS model). More importantly, however, my analysis goes beyond a study of power and contrasts the persistence properties of the relevant TS and DS models.

In sum, the evidence in this paper and in other recent work suggests that a new consensus should be formed that stresses the difficulty of knowing anything about the existence of unit roots in macroeconomic time series. This recommends careful scrutiny of all macroeconometric results for possible sensitivity to the modeling of the trend component. An example of this sensitivity is given by Shapiro and Watson (1988), who provide two sets of estimates of their model under deterministic and stochastic detrending that have very different implications about the source of macroeconomic fluctuations.

Board of Governors of the Federal Reserve System, U.S.A.

REFERENCES

ANDREWS, D., "Exact Unbiased Estimation of First Order Autoregressive/Unit Root Models," manuscript, Yale University, 1990.

BLANCHARD, O., "What Is Left of the Multiplier Accelerator?" *American Economic Review, Papers and Proceedings* 71 (1981), 150–154.

——— AND D. QUAH, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* 79 (1989), 655–673.

CAMPBELL, J. Y. AND N. G. MANKIW, "Are Output Fluctuations Transitory?" *Quarterly Journal of Economics* 102 (1987), 857–880.

CHRISTIANO, L. J. AND M. EICHENBAUM, "Unit Roots in Real GNP: Do We Know, and Do We Care?" in A. H. Meltzer, ed., *Unit Roots, Investment Measures, and Other Essays*, Carnegie-Rochester Conference Series on Public Policy 32 (1990), 7–62.

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

DEATON, A., "Life Cycle Models of Consumption: Is the Evidence Consistent with the Theory?" in T. F. Bewley, ed., *Advances in Econometrics, Fifth World Congress*, Vol. 2 (Cambridge: Cambridge University Press, 1987).

DEJONG, D. N., J. C. NANKERVIS, N. E. SAVIN, AND C. H. WHITEMAN, "Integration Versus Trend Stationarity in Macroeconomic Time Series," Working Paper No. 89-31, Department of Economics, University of Iowa, 1989.

DIEBOLD, F. X. AND M. NERLOVE, "Unit Roots in Economic Time Series: A Selective Survey," in T. B. Fomby and G. F. Rhodes, eds., *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots* (Greenwich: JAI Press, 1990).

——— AND G. D. RUDEBUSCH, "Long Memory and Persistence in Aggregate Output," *Journal of*

[27] However, the analysis of Christiano and Eichenbaum (1990) is subject to the criticism of Section 5 because their estimated TS models do not take into account the small-sample estimation bias.

Monetary Economics 24 (1989), 189–209.

——— AND ———, "Is Consumption Too Smooth? Long Memory and the Deaton Paradox," *Review of Economics and Statistics* 73 (1991), 1–9.

DICKEY, D. A., "Estimation and Hypothesis Testing in Nonstationary Time Series," Ph.D. Dissertation, Iowa State University, 1976.

——— AND W. F. FULLER, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* 49 (1981), 1057–1072.

ENGLE, R. F. AND C. W. J. GRANGER, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55 (1987), 251–276.

EVANS, G. B. A. AND N. E. SAVIN, "Testing for Unit Roots II," *Econometrica* 52 (1984), 1241–1270.

FULLER, W. A., *Introduction to Statistical Time Series* (New York: Wiley, 1976).

HAMILTON, J. D. AND M. A. FLAVIN, "On the Limitations of Government Borrowing: A Framework for Empirical Testing," *American Economic Review* 76 (1986), 808–819.

HARVEY, A. C., "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics* 3 (1985), 216–227.

KING, R. G. AND C. I. PLOSSER, "Money, Credit, and Prices in a Real Business Cycle," *American Economic Review* 74 (1984), 363–380.

KYDLAND, F. AND E. C. PRESCOTT, "A Competitive Theory of Fluctuations and the Feasibility and Desirability of Stabilization Policy," in S. Fischer, ed., *Rational Expectations and Economic Policy* (Chicago: University of Chicago, 1980), 169–189.

MANKIW, N. G. AND M. D. SHAPIRO, "Trends, Random Walks, and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics* 16 (1985), 165–174.

MEESE, R. A. AND K. J. SINGLETON, "On Unit Roots and the Empirical Modeling of Exchange Rates," *Journal of Finance* 37 (1982), 1029–1035.

NELSON, C. R. AND C. I. PLOSSER, "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* 10 (1982), 139–162.

ORCUTT, G. H. AND H. S. WINOKUR, "First Order Autoregression: Inference, Estimation, and Prediction," *Econometrica* 37 (1969), 1–14.

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

——— AND P. PERRON," "Testing for a Unit Root in Time Series Regression," *Biometrika* 75 (1988), 335–346.

POTERBA, J. AND L. SUMMERS, "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review* 76 (1987), 1142–1151.

RUDEBUSCH, G. D., "The Uncertain Unit Root in Real GNP," *American Economic Review* (forthcoming 1993).

SAID, S. E. AND D. A. DICKEY, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika* 71 (1984), 599–607.

SCHWERT, G. W., "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics* 7 (1989), 147–159.

SCHMIDT, P., "Dickey-Fuller Tests with Drift," in T. B. Fomby and G. F. Rhodes, eds., *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots* (Greenwich: JAI Press, 1990).

SHAPIRO, M. AND M. WATSON, "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual* 3 (1988), 111–156.

STINE, R. A. AND P. SHAMAN, "A Fixed Point Characterization for Bias of Autoregressive Estimators," *Annals of Statistics* 17 (1989), 1275–1284.

STOCK, J. H. AND M. W. WATSON, "Interpreting the Evidence on Money-Income Causality," *Journal of Econometrics* 40 (1989), 161–181.

WEST, K., "A Note on the Power of Least Squares Tests for a Unit Root," *Economics Letters* 24 (1987), 249–252.

———, "On the Interpretation of Near Random-Walk Behavior in GNP," *American Economic Review* 78 (1988), 202–209.