NBER WORKING PAPER SERIES

UNIT ROOTS IN MACROECONOMIC TIME SERIES: SOME CRITICAL ISSUES

Bennett T. McCallum

Working Paper No. 4368

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 1993

This paper has been prepared for the Research Department of the Federal Reserve Bank of Richmond. The author is indebted to John Campbell, David DeJong, Marvin Goodfriend, Robert King, Peter Schmidt, and Mark Watson for suggestions and useful criticism. This paper is part of NBER's research programs in Economic Fluctuations and Monetary Economics. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4368
May 1993

UNIT ROOTS IN MACROECONOMIC TIME SERIES: SOME CRITICAL ISSUES

ABSTRACT

This paper suggests that the relevant question concerning "unit roots" in the U.S. real GNP time series pertains to the relative importance of difference-stationary and trend-stationary components. Various analytical approaches indicate that an accurate answer is not obtainable with existing data. The paper next considers whether trending series should be differenced prior to use in regression analysis and suggests that it may not matter greatly if autocorrelated residuals are avoided. Finally, the paper argues that the absence of cointegration among variables does not imply the absence of any practically useful long-run relationship.

Bennett T. McCallum
Graduate School of Industrial Administration
Carnegie Mellon University
Pittsburgh, PA 15213-3890
(412) 268-2347
and NBER

I. Introduction

An enormous amount of analytical literature has recently appeared on the topic of "unit roots" in macroeconomic time series. Indeed, tests for the presence of unit roots and techniques for dealing with them have together comprised one of the most active areas, over the past decade, in the entire field of macroeconomics. The issues at hand have involved substantive questions about the nature of macroeconomic growth and fluctuations in developed economies and technical questions about model formulation and estimation in systems that include unit-root variables. The present paper attempts to describe several of the main issues and to evaluate alternative positions. It does not pretend to be a comprehensive survey of the literature or to provide an "even-handed" treatment of issues, however.[1] Instead, it attempts to develop a convincing perspective on the topic, one that is consistent with the views of many active researchers in the area but that may nevertheless be somewhat idiosyncratic.

The exposition that is presented below is designed to be predominantly non-technical in nature. Indeed, it takes a rather old-fashioned approach to econometric issues and uses recently developed concepts only sparingly. It does, however, rely extensively on notational conventions involving the time series "lag operator," L. Under these conventions the symbol L may be manipulated as if it were an algebraic symbol while its effect, when applied to a time-series variable, is to shift the variable's date back in time by one period. Thus $Lx_t = x_{t-1}$ while $bLLx_t = bL^2x_t = bx_{t-2}$, etc. In addition, the notation $\alpha(L)$ will denote a polynomial expression in the lag operator as follows: $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \alpha_3 L^3 + \ldots$. Therefore, $\alpha(L)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots$. Using this notation, then, a distributed-lag regression relation of y_t on current and lagged values of x_t could be written as $y_t = \alpha(L)x_t + \varepsilon_t$, with ε_t a stochastic disturbance term. Furthermore,

polynomials in L, which are often restricted to have only a finite number of terms, may be "multiplied" as in the following example:[2] if $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2$ and $\beta(L) = \beta_0 + \beta_1 L$, then $\alpha(L)\beta(L) = \beta_0\alpha_0 + \beta_0\alpha_1 L + \beta_0\alpha_2 L^2 + \alpha_0\beta_1 L + \alpha_1\beta_1 L^2 + \alpha_2\beta_1 L^3$. Finally, "division" by a lag polynomial means that the implied inverse, $\alpha^{-1}(L)$, is a polynomial such that $\alpha^{-1}(L)\ \alpha(L) = 1$. Thus $\alpha(L)\beta^{-1}(L)$ yields a polynomial $\gamma(L)$ such that $\beta(L)\gamma(L) = \alpha(L)$. It should be mentioned that the first coefficient of a lag polynomial, such as α_0, is often normalized so as to equal 1.0.

A brief outline of our discussion is as follows. In Section II, the distinction between trend-stationarity and difference-stationarity of time series is introduced. That distinction is then related to the "unit root" concept in Section III, which is primarily devoted to a description of attempts by researchers to determine whether the time series of real GNP values for the United States is difference or trend-stationary (i.e., does or does not have an autoregressive unit root). Two approaches, involving different strategies for the specification of maintained and tested hypotheses, are discussed. Then in Section IV a third approach, which presumes that the real GNP series is a sum of trend-stationary and difference-stationary components, is considered. From the discussion in Sections III and IV together, it is concluded that the relevant question is not whether the GNP series is difference-stationary, but what is the relative contribution of the two components. It is also concluded that an accurate answer is not obtainable with the amount of data available.

In Section V the topic changes to the question of how to process trending data before conducting regression studies relating two or more variables. The answer that is developed is that the choice between differencing and deterministic trend removal is normally of secondary importance, the principal relevant consideration being the serial correlation

properties of the regression residuals. This regression context is continued in Section VI, which concerns the topic of cointegration. It is argued that strict cointegration is probably rather rare, since relationship disturbances will usually be--like shocks to univariate series--a sum of stationary and difference-stationary processes. Examples relating to money demand and purchasing-power-parity studies are provided. In Section VII, finally, some conclusions are tentatively put forth.

II. Stochastic vs. Deterministic Trends

As most readers are well aware, many macroeconomic data series display upward tendencies or "trends" when observations are plotted against time. For many purposes it is useful and/or conventional to work with detrended values of these variables--i.e., versions from which trend components have been removed. Traditionally, most researchers would effect this detrending step by subtracting from the raw numbers (or their logs) a deterministic[3] trend expression such as $\alpha_0 + \alpha_1 t$, where t is a time index. For various reasons it is often useful to express the basic series in terms of logarithms of the raw data, in which case α_1 becomes a measure of the per-period growth rate of the variable in question. Thus if y_t is the basic variable, the traditional detrending procedure implicitly splits y_t into two components, one representing the trend and the other a cyclical or non-trend component.[4] With y_t the basic variable and ε_t a white-noise[5] disturbance, we have

$$(1) \qquad y_t = \alpha_0 + \alpha_1 t + \gamma(L)\,\varepsilon_t$$

where $\alpha_0 + \alpha_1 t$ is the trend component and $\gamma(L)\varepsilon_t$ is the non-trend component (or the detrended series). In this traditional decomposition, it is assumed that the detrended series $\gamma(L)\varepsilon_t$ is a stationary stochastic process, which requires (among other things) that the population means $E[\gamma(L)\varepsilon_t]$, variances $E[\gamma(L)\varepsilon_t]^2$, and autocovariances $E[\gamma(L)\varepsilon_t\gamma(L)\varepsilon_{t-j}]$ are the same for all t. (Here the variance and covariance expressions are written under the presumption that the means equal zero.) Accordingly, y_t is said to be a <u>trend-stationary</u> variable; it may have a trend component but its deviations from a deterministic trend are stationary. A variable's status with regard to stationarity is of importance in its own right, as we shall see in a moment, and also because there is a large body of statistical techniques

whose validity depends upon stationarity of the variables being analyzed.

At least since 1982,[6] however, many researchers have preferred an alternative model of the trend vs. non-trend decomposition. Instead of (1), they use a formulation such as (2),

$$(2) \qquad \Delta y_t = \beta + A(L)\varepsilon_t,$$

where $A(L)\varepsilon_t$ is stationary and β represents the average per-period change (or growth rate) of the variable y_t (or the variable whose log is y_t). In this formulation y_t is said to be a <u>difference-stationary</u> variable, i.e., one that is generated by a difference-stationary time series. Such a variable cannot in general be made stationary by the removal of a deterministic trend; instead, the series needs to be first-differenced prior to processing.

The basic distinction between trend-stationary (TS) and difference-stationary (DS) variables is that the former do, and the latter do not, tend to return to a fixed deterministic trend function. Since the non-trend component $\gamma(L)\varepsilon_t$ in (1) is stationary with mean zero, the process is such that y_t tends to fluctuate about the fixed trend function $\alpha_0 + \alpha_1 t$. In formulation (2), by contrast, the tendency is for y_t to grow at the rate β from its <u>current</u> position, whatever that might be. There is, except in a special limiting case, no tendency for y_t to return to any fixed trend path.

The distinction between TS and DS series was emphasized in a highly influential paper by Nelson and Plosser (1982). In this paper, the authors clearly described the TS vs. DS distinction and also discussed the undesirable statistical consequences of detrending by the traditional technique of removing a deterministic time function when in fact the series is generated by a DS process. In addition, Nelson and Plosser (1982)

presented evidence suggesting that many important U.S. time series are of the DS class and went on to argue that evidence indicates that U.S. business cycles are largely real as opposed to monetary in nature, i.e., that real shocks have been the principal sources of cyclical variability with the contribution of monetary shocks being entirely of secondary importance. The last of these arguments was not found convincing by the macroeconomics profession--see, e.g., McCallum (1986) and West (1989)--but the hypothesis that many important series (including real GNP) reflect DS processes became quite widely accepted. More recently, opinion has partially reversed itself--as we shall see below--but for the past 8 to 10 years the idea that real GNP is not trend stationary has been viewed by a large number of researchers as true (and important). It will be useful, consequently, to devote some attention to the logic of the statistical tests that led researchers to that position. In the process of presenting this logic, several relevant points of importance will be brought out--including the meaning of the term "unit root."

III. A Unit Root in U.S. GNP?

Consider now the TS representation (1) with the lag polynomial $\gamma(L)$ written as the ratio of two other polynomials $\theta(L)$ and $\phi(L)$, both assumed with little loss of generality to be finite[7] of order q and p, respectively. Thus we have

$$(3) \qquad y_t = \alpha_0 + \alpha_1 t + \theta(L)\phi^{-1}(L)\varepsilon_t.$$

Now suppose that $1/\rho$ is the smallest root of the polynomial $\phi(L)$, i.e., is the smallest number[8] that satisfies the equation $1 + \phi_1 z + \ldots + \phi_p z^p = 0$.[9] Then $\phi(L)$ could be written as $(1-\rho L)\tilde{\phi}(L)$ and multiplication of (3) by $(1-\rho L)$ would give[10]

$$(4) \qquad (1-\rho L)y_t = \alpha_0(1-\rho) + \rho\alpha_1 + \alpha_1(1-\rho)t + \theta(L)\tilde{\phi}^{-1}(L)\varepsilon_t.$$

And in the special case in which $\rho = 1$, the latter collapses to

$$(5) \qquad (1-L)y_t = \alpha_1 + \theta(L)\tilde{\phi}^{-1}(L)\varepsilon_t.$$

Since $(1-L)y_t$ equals Δy_t, then, the latter is of the same form as (2). Consequently, when there is a unit root to $\phi(L)$ -- when $1/\rho = 1.0$ -- representation (1) yields, as a special case, the DS formulation (2).[11]

In light of the foregoing result, a natural test procedure is suggested for determining whether "y_t has a unit root"--i.e., whether the AR polynomial $\phi(L)$ has a unit root so that y_t is DS. What is required is that the researcher maintains the hypothesis that (1) is true, represents it as in equation (4), and then tests the ("null") hypothesis that ρ in (4) is equal to 1.0. If the latter hypothesis is rejected, then one concludes that y_t is

not a DS series. But if the hypothesis $\rho = 1.0$ is not rejected, then one can in a sense conclude that y_t is a DS variable--or that the behavior of y_t is not significantly different from that of a DS variable. Because ordinary asymptotic distribution theory breaks down in the case in which ρ is precisely equal to 1.0, a consistent test requires that the relevant "t-statistic" on the coefficient of y_{t-1} be compared with a critical value taken from a non-standard distribution. But this can readily be done, since Dickey and Fuller (1976) have provided the profession with the pertinent tables.

The foregoing procedure was in fact employed by Nelson and Plosser (1982) to test for unit roots in over a dozen important U.S. time series. In only one of these could the tested hypothesis that $\rho = 1.0$ be rejected at a conventional significance level (i.e., 0.01 or 0.05), so the authors' tentative conclusion was that most U.S. macroeconomic data series are of the DS class, i.e., are unit-root variables.

There was, however, one rather obvious difficulty with this tentative conclusion,[12] as follows: while it was not possible to reject the hypothesis that the series' roots like ρ were equal to 1.0, it would also have been impossible to reject hypotheses asserting that these roots equaled 0.99, for example, or even 0.95. But with ρ equal to 0.99 or 0.95, the model would be one of the TS class. Continuing with this perspective, it might be argued that it is highly implausible that the tested hypothesis of ρ equal to unity would hold <u>precisely</u>, as opposed to approximately. The data, that is, could do no more than show that the value of ρ is <u>close to</u> 1.0 Consequently, this entire testing approach, which begins with a TS representation and posits the DS model as a special case, seemed highly unconvincing to a number of analysts.[13]

An alternative approach would be to begin with a maintained hypothesis

implying difference stationarity and then express trend stationarity--the absence of an AR unit root--as a special case. Thus the time series process for y_t could be written as in (2) but with $A(L) = \theta(L)\phi^{-1}(L)$:

$$(6) \qquad \Delta y_t = \beta + \theta(L)\phi^{-1}(L)\varepsilon_t.$$

Then if the moving-average lag polynomial $\theta(L)$ were to have a unit root so that $\theta(L) = (1-L)\tilde{\theta}(L)$, expression (6) could be operated on by $(1-L)^{-1}$ to yield

$$(7) \qquad y_t = \beta_0 + \beta t + \tilde{\theta}(L)\phi^{-1}(L)\varepsilon_t.$$

[That $(1-L)^{-1}\beta$ equals $\beta_0 + \beta t$ can be justified by multiplying each by $(1-L)$.] Consequently, it would be possible to express (6) as

$$(8) \qquad \phi(L)\Delta y_t = \beta\phi(L) + (1-\gamma L)\,\tilde{\theta}(L)\varepsilon_t,$$

estimate the latter, and test the hypothesis that $\gamma = 1$. If it were possible to <u>reject</u> the latter, then the outcome might be viewed as providing more convincing evidence in favor of the DS view.[14]

In fact, the influential paper by Campbell and Mankiw (1987) proceeded in precisely this fashion, using quarterly postwar data for the United States, 1947-1985. So what did these authors find? As it happens, the answer is not straightforward because it is unclear how many terms should be included in estimation of the $\phi(L)$ and $\theta(L)$ polynomials in (8). In their paper, Campbell and Mankiw reported results for 16 different cases representing all possible combinations of zero to three AR parameters and zero to three MA parameters. Of these, it is arguable that only those with

Table 1

Test Statistics from Campbell and Mankiw (1987, Table I)

Number of AR Parameters	Number of MA Parameters 1	2	3
1	22.96*	11.73*	0.00
2	2.06	4.02*	0.00
3	0.95	1.31	0.00

Notes: Tabulated entries are values of 2 log (SSE°/SSE), where SSE denotes the sum of squared residuals and SSE° indicates imposition of the constraint that makes A(1) = 0. The ARMA models are estimated for Δy_t where y_t is the log of U.S. real GNP, seasonally adjusted, quarterly for 1947.1 - 1985.4. Asterisks indicate values that are significantly different from zero (0.05 significance level) under the usual test, but this test is inappropriate (as discussed in the text).

at least one AR and one MA term should be seriously entertained. The usual test statistics for those 9 cases are given in Table 1. For each case, the reported number is the likelihood ratio statistic for a test of the hypothesis that $\theta(L)$ has a unit root--i.e., that the TS hypothesis is true. In most cases this statistic has asymptotically,[15] under the null hypothesis, a chi-square distribution with one degree of freedom, so that the critical value is 3.84 for a test with significance level 0.05 (or 6.63 for a 0.01 level). Based on these values, the table indicates that in three of the nine cases--i.e., for 3 of the 9 specifications--the null TS hypothesis can be rejected at the 0.05 level. Described in this fashion, then, the Campbell and Mankiw (1987) results did not provide strong evidence against the TS hypothesis (or, in favor of the unit root hypothesis). But under the particular hypothesis of concern in this case, that $\theta(L)$ has a unit root, the usual asymptotic distribution theory breaks down--as it does when testing for a unit root in the AR polynominal $\phi(L)$. This breakdown tends to reduce the critical level for the likelihood ratio statistics and to produce an excessive number of extreme values such as those in the final column of Table 1. Thus the figures in that table are actually more unfavorable for the TS hypothesis than they appear to be at first glance.

Furthermore, Campbell and Mankiw did not describe the results as in the previous paragraph. Instead, they suggested that the ARMA (2,2) model[16]--the case with two autoregressive and two moving average parameters--commands special attention because it is not significantly worse than the (2,3) or (3,2) models and is significantly better than the (2,1) model (and somewhat better than the (1,2) specification).[17] But in this case, the results (see Table 1) call for rejection of the TS null hypothesis, even given the test's bias toward acceptance. The suggestion of Campbell and Mankiw, consequently, was that postwar quarterly evidence supports the

notion that real GNP for the U.S. is not trend stationary, but instead is generated by a DS (or unit root) process. We shall return to the persuasiveness of this suggestion below. But first it will be useful to discuss a different aspect of the Campbell and Mankiw analysis, which their discussion emphasized.

In particular, a notable feature of the Campbell-Mankiw (1987) paper is its presentation of an attractive measure of the ultimate or "long run" response of y_t to a unit shock, i.e., a 1.0 realization of the disturbance ε_t. To define this measure, consider again the DS formulation (2), $\Delta y_t = \beta + A(L)\varepsilon_t$, and write it out as

$$(9) \qquad y_t = y_{t-1} + \beta + \varepsilon_t + A_1\varepsilon_{t-1} + A_2\varepsilon_{t-2} + \ldots .$$

From the latter expression, it can be seen that the per unit effect of ε_t on y_t is 1.0 (in the sense that if ε_t were to equal some positive value instead of its mean zero, then y_t would be higher by the same amount.) But then the per unit effect of ε_t on y_{t+1} would be $1 + A_1$, with the part A_1 occurring "directly" and the remainder through its effect on y_t. Continuing with this line of reasoning, it is found that the (per unit)[18] effect on y_{t+k} would be $1 + A_1 + A_2 + \ldots + A_k$. In the limit as $k \to \infty$, then, we would have $1 + A_1 + A_2 + \ldots$, which may be denoted A(1). (That expression arises from writing $A(L) = 1 + A_1L + A_2L^2 + \ldots$ and inserting 1 wherever L appears.) Thus the measure A(1) reflects the ultimate or long-run effect of ε_t on y_t when the process generating y_t is of form (2).

An important property of the measure A(1) is that its value is zero for any TS process. To see that, write $A(L) = \theta(L)\,\phi^{-1}(L)$ and recall that for a TS variable the MA polynomial can be written as $\theta(L) = (1-L)\tilde{\theta}(L)$. Thus we have

$$(10) \qquad A(L) = (1-L)\tilde{\theta}(L)\phi^{-1}(L) \equiv (1-L)\,a(L) = a(L) - La(L)$$

where $a(L) \equiv \tilde{\theta}(L)\phi^{-1}(L)$. But then we obtain

(11) $A(1) = a(1) - La(1) = a(1) - a(1) = 0$

since $La(1) = L(1+a_1 + a_2 + \ldots) = 1 + a_1 + a_2 + \ldots$. Thus if $\theta(L)$ can be written as $(1-L)\tilde{\theta}(L)$, as it can when the process at hand is TS, it is true that $A(1) = 0$.

What about the values of A(1) implied by various DS processes? For each of these A(1) will be non-zero, but will take on various values depending on the response pattern. In particular, A(1) will exceed 1.0 if the ultimate impact on y_t of a shock is greater than the first-period impact. An important special case is provided by the random walk process in which $\Delta y_t = \beta + \varepsilon_t$. In this case $A(L) = 1 + 0L + 0L^2 + \ldots = 1$ so $A(1) = 1$. Next, the first-order MA case has $\Delta y_t = \beta + \varepsilon_t + \theta\varepsilon_{t-1}$ so $A(L) = 1 + \theta L$ and $A(1) = 1 + \theta$. Then A(1) is greater than or smaller than 1.0 depending on whether θ is positive or negative.

A somewhat more general process is the ARMA (1,1) model for Δy_t, namely,

(12) $(1-\phi L)\Delta y_t = (1+\theta L)\varepsilon_t$.

In this case $A(L) = (1 + \theta L)(1 - \phi L)^{-1}$ so $A(1) = (1 + \theta)/(1 - \phi)$. An example application is provided by the Campbell-Mankiw (1987) estimates with the U.S. GNP series. Their point estimates of ϕ and θ are 0.522 and -0.179, respectively, so that $A(1) = (1 - .179)/(1 - .522) = 0.821/0.478 = 1.717$. Thus the ARMA (1,1) model for Δy_t suggests that the long-run response of y_t (log of GNP) to a shock will be about 1.7 times as large as the immediate

(within one quarter) response.

In sum we see that the measure A(1) provides an attractive way of expressing the magnitude of the "long run response" of a variable (y_t) to a unit shock (ε_t).[19] And in their study of postwar U.S. GNP, Campbell and Mankiw (1987) find that for all but three of their nine cases[20] A(1) is substantially larger than 1.0 implying that the impact of shocks is to cumulate, rather than dissipate, as time passes. Their values are reported in Table 2, where it may be noted that for the ARMA (2,1), (3,1) and (3,2) cases (marked with asterisks), the point estimates of A(1) are large even though the A(L) polynomials are not significantly different from ones with A(1) = 0 according to the usual test (compare Table 1). Consequently, although they are guarded in their statements, Campbell and Mankiw seem to conclude that there are grounds for being reasonably confident that the long run response of a unit shock to U.S. GNP is substantially greater than 1.0. In this sense, shocks to GNP have no trend-reversion tendency at all.[21]

This conclusion has not, however, held up to subsequent criticism. One major reason for skepticism was provided in a study by Christiano and Eichenbaum (1989). Basically, their study emphasized that the different ARMA specifications considered by Campbell and Mankiw give rise to quite different values of A(1)--as is evident in Table 2--and that there is very little reason to conclude that any one of them is appropriate--or even that one of those with A(1)>1 is. The Christiano-Eichenbaum argument is that relevant inferences are sensitive to the choice of ARMA specification employed, even within the set of those that provide approximately equally good fits to the data.

One of the experiments conducted by Christiano and Eichenbaum will illustrate their results. In this experiment they conducted simulations with a model with parameters matching those estimated by Campbell and Mankiw in

Table 2

Estimates of A(1) from Campbell and Mankiw (1987, Table IV)

Number of AR Parameters	Number of MA Parameters 1	2	3
1	1.72	1.73	0.03
2	1.77*	1.52	0.00
3	1.36*	1.60*	0.00

Notes: See Table 1

the ARMA (3,3) case. In other words, they pretended that this case--which implies A(1) = 0--is true, and then considered what would happen if it were studied under the assumption that the ARMA (2,2) specification were correct. For each simulation they would generate 150 "data" points using the (3,3) parameters, then estimate a (2,2) model and test the hypothesis that A(1) = 0. They conducted 2000 such simulations and found that the hypothesis A(1) = 0, which was true in the process studied, was nevertheless rejected in 74 percent of the simulations.[22] Similarly, in 2000 more simulations, based on the ARMA (1,3) parameter estimates from Campbell and Mankiw, it was found that the true hypothesis A(1) = 0 was rejected in 38 percent of the simulations.

The conclusion reached by Christiano and Eichenbaum was as follows: on the basis of 150 observations, about the number of quarterly postwar data periods, it is not possible to make accurate inferences about the long-run response measure A(1). Equivalently, it is not possible to determine with high reliability whether the stochastic process generating real GNP observations is of the TS or DS class.

In the last couple of years, since the publication of the Christiano-Eichenbaum study, numerous additional papers on the topic have appeared. Sims (1988) has suggested that Bayesian techniques of statistical inference are more appropriate than classical in this particular context and DeJong and Whiteman (1989) (1991) have presented Baysian results that provide support for the view that the U.S. GNP process is actually of the TS class. That conclusion has been strongly challenged by Phillips (1991), in a paper that provided the basis for a symposium occupying an entire issue of the Journal of Applied Econometrics. The symposium includes rejoinders by Sims and DeJong-Whiteman; it would be difficult to identify any clear outcome. Others, including Stock (1991), Sowell (1992) and Rudebusch (1992), have

reached the Christiano-Eichenbaum (1991) conclusion--i.e., that it is not possible with existing data to settle the issue--by alternative means. To the present writer, this last conclusion seems generally appropriate, but there is another way of approaching the issue that is conceptually rather simple and perhaps illuminating.

IV. The Unobserved Components Approach

In the previous section two approaches were mentioned, based on equations (3) and (6). In the first of these, the maintained hypothesis is trend stationarity with difference stationarity viewed as a zero-measure[23] special case, whereas in (6) the DS hypothesis is maintained and TS is treated as the (zero-measure) special case. Let us now consider an alternative approach that proceeds within a framework in which both TS and DS components are presumed to play a role, the implied statistical problem being to determine how much weight to give to each. Aspects of this "unobserved components" approach have been developed by Harvey (1985), Watson (1986), Clark (1987), and Cochrane (1988).

The analysis presented by Clark (1987) provides a useful introduction and perspective. It begins by writing the observable variable under study, y_t, as the sum of a DS "trend" term z_t and a stationary "cycle" term x_t:

$$(13) \quad y_t = z_t + x_t.$$

Although a more general specification would be possible, Clark assumes that the cyclical component is a pure AR process so that $\phi(L)x_t = v_t$, with v_t white noise. Indeed, in his empirical implementation with U.S. GNP data Clark makes $\phi(L)$ a second order polynomial, so that x_t is an ARMA (2,0). The trend component is assumed to obey

$$(14) \quad z_t = z_{t-1} + d + w_t$$

where w_t is white noise, independent of v_t. Actually Clark takes the drift term d to be itself a random walk: $d_t = d_{t-1} + u_t$ with u_t white. But empirically he finds the variability of u_t to be very small, so we shall for

simplicity view d_t as a constant, as in (14). The model at hand for y_t is therefore

$$(15) \quad y_t = (1-\phi_1 L - \phi_2 L^2)^{-1} v_t + (1-L)^{-1}(d+w_t)$$

Let us consider, then, how (15) fits the U.S. quarterly postwar GNP data.

As a preliminary, it will be instructive to consider a comparison that begins by expressing (15) as

$$(16) \quad (1 - \phi_1 L - \phi_2 L^2)\, \Delta y_t = (1-L)\, v_t + (1-\phi_1 L-\phi_2 L^2)(d+w_t).$$

Here the right-hand side is the sum of two independent MA processes, with the higher order one being an ARMA (0,2). Using Granger's Lemma,[24] then, we can write (16) as

$$(17) \quad (1 - \phi_1 L - \phi_2 L^2)\, \Delta y_t = \delta + (1 + \theta_1 L + \theta_2 L^2)\, \varepsilon_t$$

where ε_t is an implied, constructed white-noise disturbance and where $\delta = d(1 - \phi_1 - \phi_2)$. But the representation in (17) has 6 parameters (ϕ_1, ϕ_2, θ_1, θ_2, σ_ε^2, and δ) whereas the basic model (15) has only five (ϕ_1, ϕ_2, σ_v^2, σ_w^2, and d). So the particular components model at hand, which sums an AR (2) component and a random walk component, can be viewed as a <u>constrained</u> version of an ARMA (2,2) model for Δy_t.

It is of course true that the unconstrained model (17) must fit the data at least slightly better than the constrained version (15). But Clark's estimates, reported in Table 3, indicate that in the case at hand there is almost no difference, i.e., almost no deterioration in fit, from imposing the

Table 3

Estimates of ARMA (2,2) Models from Clark (1987)

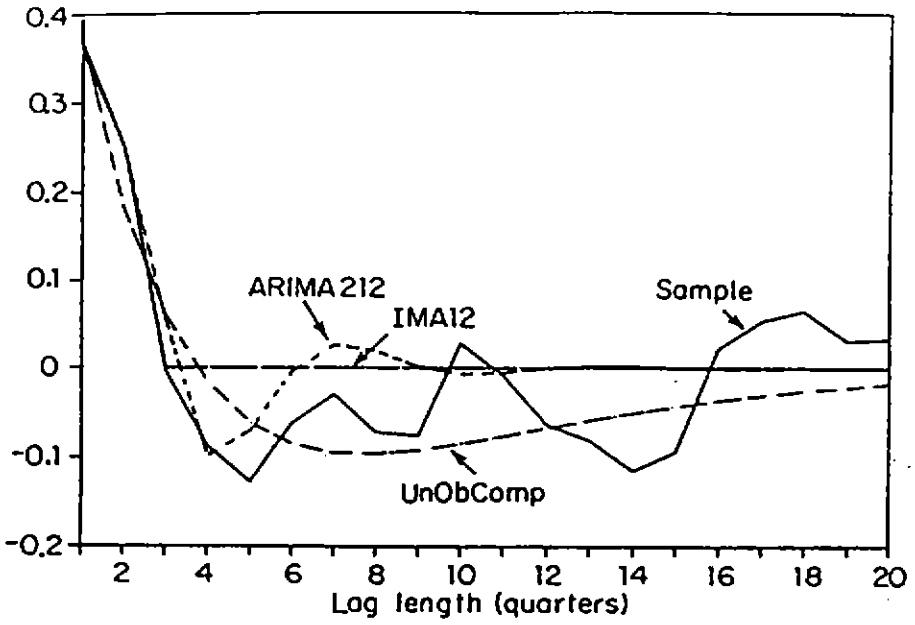
Parameters and Statistics	Constrained	Unconstrained
ϕ_1	1.548	0.658
ϕ_2	-0.601	-0.420
θ_1	-1.214	-0.335
θ_2	0.248	0.529
SE	0.0103	0.0103
Q(10)	7.9	4.8
A(1)	0.64	1.57

Notes: ARMA models estimated for Δy_t (See Table 1) for 1948.1 - 1985.4. Q(10) denotes the Box-Pierce Q statistic for 10 autocorrelations of the residuals; under the null hypothesis of white noise it has asymptotically a chi-square distribution with 10 degrees of freedom.

constraint. In particular, the estimated residual variance for (15) is essentially the same as with (17) and the Box-Pierce Q(10) statistic is not much worse. So the constrained version--the components model (15)--could as well be the preferred choice.[25] But although the constrained and unconstrained ARMA models fit the data about the same, they yield very different A(1) measures. Whereas the unconstrained version gives $\hat{A}(1) = 1.57$, virtually the same as estimated by Campbell and Mankiw, for the (unconstrained) components model the estimate is 0.64.

In two diagrams, Clark (1987) presented evidence apparently suggesting that for U.S. quarterly GNP the unobserved components model may provide a better estimate than the unconstrained ARMA of the long-run response statistic A(1). The first of these, presented here as Figure 1, plots the implied autocorrelations at various lags for the two models (plus one more, an ARMA (0,2)) and for the Δy_t sample. There it will be seen that the unconstrained ARMA (denoted ARIMA 212) matches the sample somewhat better at short lags (e.g., 1-5 quarters) but that the components model provides a better match at lags of 5-20 quarters. More striking are the related results shown in Figure 2, which plots the variance ratios V_k/V_1, where $V_k \equiv (1/k)\mathrm{Var}(y_t - y_{t-k})$, for lag lengths k up to 60 quarters.[26] In this case, the apparent superiority of the components model's match to the sample data is considerable. But, as Campbell and Mankiw (1987b) point out, sample values of V_k provide biased estimates of their population counterparts. Accordingly, Campbell and Mankiw suggest that the sample values should be multiplied by T/(T-k), where T is the sample size. Here T = 148, so the adjusted sample values of V_k are as indicated by the small circles in Figure 2. Clearly, this bias adjustment worsens substantially the match between sample and components-model values of V_k. The same point applies, but with less force, to Figure 1.

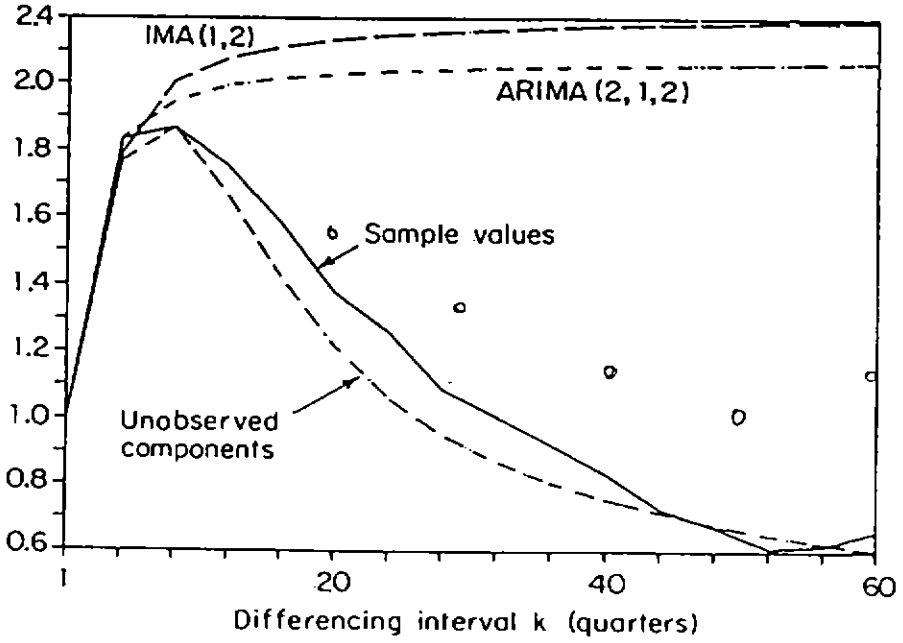
Figure 1



Sample and Theoretical Autocorrelations for Δlog GNP

—sample -----ARIMA 212 – – –UnObComp — — —IMA 12

Figure 2.



Variance Ratio of k-differences for Δlog GNP

More generally, the unobserved components approach to modeling the trend vs. cyclical decomposition seems conceptually attractive, in part because it does not treat either TS or DS processes as (zero-measure) special cases. The implied question is not whether one of these two possibilities can be rejected, but instead is "how important quantitatively is the z_t as opposed to the x_t component?" That question cannot be answered in precisely the stated form, since the variance of the DS component z_t depends on the horizon considered and goes to infinity in the limit. But one type of answer is provided by the A(1) measure itself and another by a comparison of the variances of v_t and w_t, i.e., the shocks to x_t and z_t. In the case at hand, Clark's estimates are $\hat{\sigma}_v = 0.0072$ and $\hat{\sigma}_w = 0.0066$.

An objection to the components approach as implemented by Clark (1987) and Watson (1986) was expressed by Campbell and Mankiw (1987b, p. 115). This objection is that with the DS component z_t modeled as a random walk, the estimated value of A(1) must lie between 0 and 1; thus values greater than 1.0 are ruled out a priori. But while this important objection is applicable to the Clark and Watson studies, it is not applicable to the approach in general, for the latter can accommodate other DS processes for z_t. Instead of a random walk, for example, the z_t process could be specified as a first-order MA: $\Delta z_t = d + w_t + \theta w_{t-1}$. For the z_t component alone, A(1) would then equal $1 + \theta$ so values in excess of 1.0 will be obtained if $\theta > 0$. And if the variability of z_t is large in relation to that for x_t, then the A(1) value for y_t could easily exceed 1.0.[27]

Another objection is that it is unreasonable to assume that x_t and z_t components are independent, as the approach presumes. There is undoubtedly some merit to this point, since technology shocks will presumably have both cyclical and long-lasting effects. But the Campbell-Mankiw ARMA approach amounts to use of a unobserved components model in which the shocks (like v_t

and w_t in (15)) are perfectly correlated,[28] which property seems even less desirable.

Perhaps the most important objection to the unobserved components modeling of trend vs. cycle is that it is computationally much more difficult than estimation of ARMA models, the necessary steps involving application of Kalman filter techniques. For a discussion of such techniques, the reader is referred to Harvey (1981).

On the basis of the foregoing discussion, it would seem reasonable to conclude that the postwar U.S. quarterly real GNP process is most likely of the DS class, since a sum of DS and TS components is itself a DS process.[29] But it is far from clear that the long-run impact of shocks exceeds that of the random-walk case in which A(1) = 1.0. Instead, a measure such as 0.6, which attributes a substantial share of GNP variability to a stationary component, is just as plausible. What does seem clear is that it is not possible, on the basis of currently available data, to estimate A(1) with much accuracy or reliability.

Conceptually, the basic components-approach idea, of viewing a time series as the sum of DS and TS processes, seems attractive as a framework for thinking about the properties of univariate time series. In many cases, both components would be presumed to be of non-negligible importance so many series will be of the DS class. That does not imply, however, that any particular method can be relied upon for trend vs. cyclical decomposition of time series data.

V. Detrending Prior to Econometric Analysis

In this section we switch our attention away from trend estimation, conducted for the purpose of isolating trend and cyclical components of a series, and toward trend removal (or "detrending"), conducted for the purpose of obtaining series suitable for econometric analysis of relationships among variables. In this context, then, the issue is whether to process variables prior to (say) regression analysis by removal of an estimated deterministic trend[30] or by differencing of the series. A major reason for detrending is that the standard formulae for standard errors, test statistics, etc., are in most cases based on asymptotic distribution theory that assumes stationarity of the regressor variables.[31] Belief that some variable is generated by a process of the DS type--i.e., one with a unit root--would then lead to the presumption that data differencing would be preferable for that variable prior to its use in a regression study.

Other influential arguments for differencing of data prior to time series econometric analysis were put forth by Granger and Newbold (1974) and Nelson and Kang (1984). In the earlier of these papers it was shown that a regression relating y_t to x_t would spuriously tend to find a relationship when in fact y_t and x_t are generated independently but by random-walk processes. The Nelson-Kang piece emphasized a tendency for trendless random-walk variables to be spuriously related to time trends in estimated regressions.

As a result of these and other studies, considerable support developed during the mid-1980s for the position that differencing should routinely be carried out prior to regression analysis involving time series data.[32] The case for such a practice was succinctly summarized by Plosser and Schwert (1978, p. 653) as follows: "ignoring the effects of underdifferencing can be a far more costly error than ignoring the effects of overdifferencing." More

recently, there has been significant counter-movement based on phenomena related to the concept of "cointegration." A consideration of that position will be presented below, but it will be useful first to consider the merits of routine differencing, rather than detrending, of variables with an apparent trend component.

The issues at hand can be usefully introduced and illustrated in an example similar to that used by Plosser and Schwert (1978). Consider a linear regression relationship that is (by assumption) correctly specified in first differences, viz.,

$$(18) \quad \Delta y_t = \beta \Delta x_t + \varepsilon_t,$$

where ε_t is a white noise disturbance with variance σ_ε^2 and where x_t is exogenous, generated by a process independent of the process generating ε_t. Now, if instead of (18) the investigator estimates by ordinary least squares (OLS) the relationship between x_t and y_t in levels, he is in effect applying OLS to

$$(19) \quad y_t = \alpha + \beta x_t + \eta_t,$$

in which the disturbance term $\eta_t = \varepsilon_t + \varepsilon_{t-1} + \ldots$ is serially correlated and nonstationary. In this underdifferenced case, as Plosser and Schwert point out, the OLS estimator of β could be inconsistent, depending on the process generating x_t. In any event, whether or not the OLS estimator is consistent, its sampling distribution does not have finite moments. Inferences based on the usual OLS formulae are likely, accordingly, be highly inappropriate.

Next, suppose that instead the investigator applies OLS to the second differences of y_t and x_t, estimating

(20) $\Delta(\Delta y_t) = \beta\Delta(\Delta x_t) + \Delta c_t$.

In this case with overdifferencing the disturbance Δc_t is again serially correlated but now its distribution is stationary. The OLS estimator of β will be unbiased and consistent, but will be inefficient and its sampling variance will (except in special cases) not be consistently estimated by the usual formulae.

These foregoing considerations, discussed by Plosser and Schwert (1978), are of some interest but are actually relevant only under the presumption that the investigator is wrong about the appropriate degree of differencing and makes use of OLS estimators even though the implied disturbances are serially correlated. Of considerably greater interest, it would seem, are the consequences of estimating β with underdifferenced or overdifferenced data when the investigator recognizes the presence of serial correlation in the OLS residuals and responds by utilizing an estimator designed to take account of autocorrelated disturbances in the appropriate manner. In the overdifferenced case, for example, the true relation can be written as

(21) $\Delta(\Delta y_t) = \beta\Delta(\Delta x_t) + \varepsilon_t + \theta\varepsilon_{t-1}$

with $\theta = -1.0$. The interesting question, then, is whether the investigator will be led seriously astray if he regresses $\Delta(\Delta y_t)$ on $\Delta(\Delta x_t)$ using an estimation procedure designed for cases in which the disturbance process is a first-order MA.

Now precisely this last question has been investigated via Monte Carlo experimentation[33] by Plosser and Schwert (1977). They find that even though

the absolute value of θ tends to be somewhat underestimated--with a sample size of T = 100 the mean across 1000 replications of the estimates of θ is about -0.94 -- the estimates of β are not appreciably biased and the experimental sampling distribution is not such as to lead frequently to incorrect inference. Specifically, the frequency of rejection of a true hypothesis with a nominal significance level of 0.05 is 0.063 in one experiment and 0.081 in the other. Plosser and Schwert conclude, appropriately, that "the cost associated with overdifferencing may not be large when care is taken to analyze the properties of regression disturbances" (1978, p. 643).

The corresponding case of an investigation with underdifferencing arises if we write the true relation as

$$(22) \quad y_t = \alpha + \beta x_t + (1-\rho L)^{-1}\varepsilon_t \quad ,$$

with $\rho = 1.0$, and ask whether the investigator will be led seriously astray (regarding β) if he regresses y_t on x_t under the assumption that the disturbance process is a first-order AR. With respect to this possibility, Plosser and Schwert (1978, p. 643) recognize that "if the resulting estimate of ρ is close to one, as it should be in this case, differencing would be indicated leading to the correct model...." They do not, however, consider the effects on the estimation of β of concluding one's investigation with the estimate provided by the levels regression that takes account of AR disturbances--which is the situation corresponding to the presumed behavior of the investigator in the overdifferencing case. This asymmetry in discussion prevents them from giving a comparison of the relative costs of underdifferencing vs. overdifferencing when the investigator is intelligently taking account of the serial correlation properties of the disturbances.

Some Monte Carlo results relevant to this type of procedure have, however, been obtained by Harvey (1980) and Nelson and Kang (1984). The latter of these papers is devoted primarily to emphasizing various ways in which investigators could be led to misleading results if they estimate underdifferenced relationships and do not correct for serially correlated residuals, but it briefly reports (on pp. 79-80) results of testing a true hypothesis analogous to $\beta = 0$ in (22) with β and ρ estimated jointly. With T = 100 and a significance level of 0.05, the frequency of rejection in 1000 replications is 0.067, which compares favorably with the Plosser-Schwert results for the case with overdifferencing. The study by Harvey (1980) compares mean-squared-error (MSE) values[34] for estimates of β in (22) with $\rho = 1.0$ when estimated with first differences and when estimated jointly with ρ using levels data (i.e., with underdifferencing and an autocorrelation correction). Two specifications regarding the behavior of the exogenous variable x_t are considered by Harvey. In one of these the x_t process is stationary; in that case the MSE value for the estimator of β is 0.310 with the (correct) first-difference specification and 0.309 with underdifferencing (and autocorrelation correction).[35] In the other case, which features strongly trending x_t data, the analogous MSE figures are 0.061 and 0.078.[36]

Also of relevance, though not conforming to the symmetric contrast provided by our specifications (18) and (20), is evidence provided by Harvey relating to the estimation of a relation like (22) but with $\rho = 0.9$. The alternative estimators are based on application of maximum likelihood to the (correct) levels specification and OLS to the first-differenced specification, the latter amounting to estimation with $\rho = 1.0$ by constraint.[37] The T = 100 MSE values are 0.263 and 0.262 for the two estimators with stationary x_t's, and 0.009 vs. 0.018 with trending x_t's.

On the basis of the described Monte Carlo experiments, the appropriate

conclusion would seem to be that neither overdifferencing nor underdifferencing leads to serious estimation or testing mistakes in regression models with exogenous regressors, provided that the investigator takes intelligent account of serial correlation present in the regression residuals. It is perhaps worth noting, given the tenor of their discussion, that this conclusion is not contradicted in the least by the four studies involving actual data (and unknown specifications) that are explored by Plosser and Schwert (1978).

Specifically, in each of these four cases the authors conclude that first differencing is probably appropriate, but the point estimates and standard errors (for the parameter analogous to β) that are provided by regressions with undifferenced data are virtually the same when the Cochrane-Orcutt procedure is used to account for $\rho \neq 0$. In their Table 1 regression of (log) income on the (log) money stock, for example, the slope coefficient (and standard error) values are 1.127 (0.122) for the Cochrane-Orcutt levels regression and 1.141 (0.126) in the differenced case. The OLS regressions with data that have been differenced twice give estimates that do not agree as well, but in each of these cases there is evidence of uncorrected serial correlation in the residuals. In Table 1, for example, the first residual autocorrelation is -0.36.

It is additionally worth noting that Plosser and Schwert (1978, p. 638) also take the view that "the real issue is not differencing, but an appropriate appreciation of the role of the error term in regression models." Thus our disagreement with Plosser and Schwert seems to be whether the "representative investigator" will, or will not, recognize and take steps in response to the presence of autocorrelated residuals.

The foregoing evidence relates, however, principally to relations with exogenous regressors. In practice, it is much more common for equations of

interest to include one or more lagged endogenous variables. But if Δy_{t-1} were to appear as an additional regressor in (18), then the situation regarding estimation of β would be quite different. In order to obtain a bit of evidence as to the validity of the suggestion--that the presence or absence of differencing is not crucial when serial correlation corrections are applied--in situations in which lagged endogenous variables are present, let us consider results pertaining two example relationships (that may be of some substantive interest).

Specifically, let us first consider estimation of the rudimentary single-equation model of aggregate demand utilized in McCallum (1987), that is, an equation relating growth of nominal GNP to growth rates of the monetary base. Notationally, let x_t and b_t denote logarithms of nominal GNP and the base, respectively, for period t and consider quarterly observations, seasonally adjusted, for the sample period 1954.1 - 1991.3.[38]

As a starting point, consider the following updated version of the specification emphasized in McCallum (1987):

$$\Delta x_t = \underset{(.002)}{0.0078} + \underset{(.073)}{0.3248}\Delta x_{t-1} + \underset{(.104)}{0.3190}\Delta b_t$$

$$R^2 = 0.196 \quad SE = 0.0097 \quad DW = 2.12 \quad Q(10) = 8.3 \tag{23}$$

Here parameter standard errors are shown in parentheses while the reported statistics are the unadjusted R^2, the estimated standard deviation of the disturbance term, the Durbin-Watson statistic, and the Box-Pierce Q statistic based on the first ten autocorrelation terms.[39] These statistics give no evidence of residual autocorrelation and it is the case that Δx_{t-2} would not provide additional explanatory power. As it happens, however, inclusion of Δb_{t-1} would provide additional explanatory power and would make Δb_t

insignificant. Accordingly, let us switch our attention to the variant of (23) in which Δb_t is replaced by Δb_{t-1}, a variant also used in McCallum (1987). The 1954.1 - 1991.3 estimates are as follows:

$$(24) \quad \Delta x_t = \underset{(.002)}{0.0076} + \underset{(.075)}{0.2845}\Delta x_{t-1} + \underset{(.105)}{0.3831}\Delta b_{t-1}$$

$$R^2 = 0.215 \quad SE = 0.0096 \quad DW = 2.07 \quad Q(10) = 8.0$$

Here there is no evidence of residual autocorrelation and additional lagged values of Δx_t and Δb_t would not enter significantly. The important properties of the estimated relation are that Δx_t is mildly autoregressive and is positively related to Δb_t, with a moderately large elasticity value that is not significantly different from 0.5.

The first question to be answered, then, is "what would we have found if we had estimated this same relation in (log) levels, using series with deterministic trends removed?" To develop an answer, first consider equation (25), where the detrending is effected by inclusion of time as an additional regressor:

$$(25) \quad x_t = \underset{(.072)}{0.0273} + \underset{(.0002)}{0.00021}t + \underset{(.020)}{1.0160}x_{t-1} - \underset{(.014)}{0.0321}b_{t-1}$$

$$R^2 = 0.9999 \quad SE = 0.0104 \quad DW = 1.40 \quad Q(10) = 23.1$$

Here the results are entirely different from those in (24), but there is distinct evidence of residual autocorrelation. Re-estimation with the disturbance term assumed to follow an AR(1) process yields

$$(26)\quad x_t = \underset{(34.1)}{5.857} - \underset{(.060)}{0.0067}t + \underset{(.081)}{0.2763}x_{t-1} + \underset{(.150)}{0.592}b_{t-1} + \underset{(.023)}{0.996}e_{t-1}$$

$$R^2 = 0.9999 \quad SE = 0.0095 \quad DW = 2.14 \quad Q(10) = 9.0,$$

where e_{t-1} is our estimate of ε_{t-1}. Now, with the AR(1) disturbance specification, we estimate the autocorrelation parameter to be very close to 1.0 and the magnitude of the coefficients attached to x_{t-1} and b_{t-1} revert to the neighborhood of the corresponding values in the differenced relation (24).[40] The trend term is insignificant, as was the constant in (24), and qualitatively the relation in (26) is quite similar to the version estimated in differences.

Next, we move in the opposite direction by differencing the variables one more time than in the reference case (24). Let $\Delta\Delta x_t \equiv \Delta(\Delta x_t)$ for brevity. Then with the disturbance treated as white noise, the result is

$$(27)\quad \Delta\Delta x_t = \underset{(.0009)}{0.0002} - \underset{(.074)}{0.3993}\Delta\Delta x_{t-1} + \underset{(.158)}{0.363}\Delta\Delta b_{t-1}$$

$$R^2 = 0.182 \quad SE = 0.0110 \quad DW = 2.12 \quad Q(10) = 18.3$$

Here the estimated parameter on the lagged GNP variable is entirely unlike that in (24), but the Q statistic gives borderline evidence of serial correlation. Estimated with a MA(1) specification for the disturbance, the results change to:

$$(28)\quad \Delta\Delta x_t = \underset{(.0008)}{0.00001} + \underset{(.065)}{0.1666}\Delta\Delta x_{t-1} + \underset{(.139)}{0.3571}\Delta\Delta b_{t-1} - \underset{(.032)}{0.946}e_{t-1}$$

$$R^2 = 0.370 \quad SE = 0.0097 \quad DW = 1.89 \quad Q(10) = 9.5$$

Now the figures are again quite close to those in the once-differenced specification (24). Not only the estimated parameter values, but also the standard errors are approximately the same -- and there is no evidence of serial correlation. Thus the results are similar for regressions using detrended levels, differences, and second differences of the x_t and b_t variables, provided that autocorrelation corrections are used.

A second example concerns spot and forward exchange rates. In a recent paper (McCallum, 1992), I have summarized some empirical regularities for the post-1973 floating rate period, focussing on \$/£, \$/DM, and \$/Yen rates over the time span 1978.01 - 1990.07. Letting s_t and f_t denote logs of the spot and 30-day forward rates at the end of month t, one of the observed regularities is that OLS regression of s_t on f_{t-1} provides a tight fit with a slope coefficient very close to 1.0 -- see the estimates reported in the first panel of Table 4. When Δs_t is regressed on Δf_{t-1}, however, the relationship disappears and the estimated slope coefficient becomes insignificantly different from zero--see the second panel of Table 4. That contrast would seem to contradict the argument of the preceding paragraphs since there is little indication of serial correlation in the residuals in either case.

The results in panel three, however, support our previous argument. There the levels equation relating s_t and f_{t-1} is reestimated with an AR(1) specification for the disturbance process, even though the DW and Q statistics in the top panel do not clearly call for such a step. And for all three exchange rates the result is the same -- the AR parameter ϕ is estimated to be close to 1.0 with the slope coefficient on f_{t-1} becoming indistinguishable from zero. The results in panel three, in other words, are essentially equivalent to those in panel two, even though differenced data are used in the latter and not in the former.

Table 5

Spot on Forward Exchange Rate Regressions

Sample Period 1978.01 - 1990.07

Rate	Variables	Estimates (std. errors) Const.	Slope	AR(1)	MA(1)	R^2	Statistics SE	DW	Q(10)
$/DM	s_t on f_{t-1}	-0.009	0.990			0.963	0.0362	2.05	9.6
		(.012)	(.016)						
$/£	" "	0.014	0.977			0.960	0.0359	1.82	5.9
		(.009)	(.016)						
$/Yen	" "	-0.046	0.991			0.975	0.0380	1.84	7.7
		(.067)	(.013)						
$/DM	Δs_t on Δf_{t-1}	0.002	-0.063			0.004	0.0358	1.96	7.5
		(.003)	(.081)						
$/£	" "	0.000	0.024			0.000	0.0352	1.99	3.4
		(.003)	(.082)						
$/Yen	" "	0.003	0.038			0.001	0.0374	1.99	4.8
		(.003)	(.082)						
$/DM	s_t on f_{t-1}	-0.575	-0.057	0.991		0.964	0.0359	1.96	8.1
		(.510)	(.083)	(.015)					
$/£	" "	0.509	0.035	0.979		0.962	0.0351	1.99	3.8
		(.143)	(.084)	(.017)					
$/Yen	" "	-4.743	0.046	0.989		0.976	0.0374	2.00	5.0
		(.681)	(.083)	(.013)					
$/DM	$\Delta\Delta s_t$ on $\Delta\Delta f_{t-1}$	0.000	-0.206		-0.887	0.526	0.0361	1.86	4.1
		(.003)	(.057)		(.059)				
$/£	" "	0.000	-0.032		-0.933	0.472	0.0357	1.96	3.1
		(.003)	(.060)		(.035)				
$/Yen	" "	0.000	-0.010		-0.956	0.468	0.0379	1.94	4.0
		(.003)	(.060)		(.027)				

Data source: Bank for International Settlements

In addition, the specification using second differences together with a MA(1) disturbance is implemented in the fourth panel of Table 5. There the DM case differs slightly from the previous results, the slope coefficient on f_{t-1} being estimated as about 0.21 and significant, but for both the £ and Yen rates the previous results are obtained again -- the slope coefficient is close to zero with the MA parameter being estimated in the vicinity of -1.0. For five out of the six comparisons with the reference case of panel two, then, the results are in strong agreement despite contrasting treatment in terms of differencing. And even in the sixth case, the extent of disagreement is relatively minor. Thus the evidence is again supportive of the general argument that the extent of differencing is not crucial, in the context of detrending of variables prior to econometric analysis, provided that residual autocorrelation corrections are utilized.[41]

VI. Cointegration

Now suppose that x_t and y_t are two time series variables generated by DS processes--i.e., their univariate series have AR unit roots--that are dynamically related by a distributed-lag relation with a stationary disturbance. In (29), for example, we assume u_t to be stationary:[42]

$$(29) \quad y_t = \alpha + \beta(L)x_t + u_t.$$

Under these conditions y_t and x_t are said to be cointegrated, the term arising because DS variables are referred to by many time series analysts as "integrated."[43] Now it is a striking fact that when y_t and x_t are cointegrated, then an OLS regression of y_t on x_t alone--with no lags--will yield a slope coefficient b that is a consistent estimator of the "long run" effect $\beta(1) = \beta_0 + \beta_1 + \ldots$. This result would appear to be of practical importance, as it promises to provide a simple way of discovering features of long-run relationships between variables. To demonstrate the result, let us express the residual $e_t = y_t - bx_t$ as

$$(30) \quad e_t = \alpha + \beta(L)x_t + u_t - bx_t = \alpha + [\beta(L) - b]x_t + u_t.$$

But with x_t an integrated (DS) variable, e_t will then be integrated unless $\beta(1) - b = 0$. And if e_t were integrated, then the sum of squared e_t values would increase without limit as the sample size goes to infinity, so the OLS criterion of picking b to minimize this sum forces b toward $\beta(1)$.

There are numerous additional theoretical results concerning cointegrated variables including extension to multivariate settings[44] and close connections between cointegration and the existence of "error correction" forms of dynamic models.[45] For present purposes, however, the

main item of interest concerns the frequently expressed contention that if two (or more) DS variables are not cointegrated, then there exists no long-run relationship between (or among) them. On the basis of this notion, various researchers have concluded that purchasing-power-parity fails even as a long-run tendency [see, e.g., Taylor (1988) and McNown and Wallace (1988)] whereas others have drawn analogous conclusions regarding traditional money demand relations--see, e.g, Engle and Granger (1987).[46] Cuthbertson and Taylor (1990, p. 295) have stated the matter thusly: "If the concept of a stable, long-run money demand function is to have any empirical content whatsoever, then m_t..." (log money) "must be cointegrated" with log prices, log income, and interest rates.

Now clearly there is a technical sense in which these suggestions are correct: if y_t and x_t are both DS but not cointegrated, then the disturbance entering any linear relationship between them must (by definition) be nonstationary. So they can drift apart as time passes. The present author would argue, however, that it is highly misleading to conclude that in any practical sense long-run relationships are therefore nonexistent. My argument is entirely interpretive; it includes no suggestion of technical error in the literature criticized. But its importance is not thereby diminished.

To develop the argument at hand, let us take the example of a traditional money demand function of the form

$$(31) \quad m_t - p_t = \beta_0 + \beta_1 y_t + \beta_2 R_t + \eta_t,$$

where $m_t - p_t$ is the log of real money balances, y_t the log of a real transactions variable (such as GDP), and R_t is an opportunity-cost variable relevant to the measure of money being used. Let us suppose for the purpose

of the argument that $m_t - p_t$, y_t, and R_t are all DS variables. And let us suppose that $m_t - p_t$, y_t, and R_t have all been processed by removal of a deterministic trend.[47] Then the cointegration status of the relationship depends upon the properties of the disturbance η_t--if its process is of the DS type, the variables in (31) will not be cointegrated.

It is my contention that the traditional view of money demand theory, represented for example by the New Palgrave entry by McCallum and Goodfriend 1987), would actually suggest that the variables in (31) are unlikely to be cointegrated. The reason is that the rationale for (31) depends upon the transactions-facilitating function of money, but the technology for effecting transactions is constantly changing. And since technical progress cannot be well represented by measurable variables, the effects of technical change not captured by a deterministic trend show up in the disturbance term, η_t. But the nature of technological progress is such that changes (shocks) are typically not reversed. Thus one would expect there to be an important permanent component to the η_t process, making it one of the DS type.

In such a situation, however, the "long run" messages of traditional money demand analysis may continue to apply. Provided that the magnitude of the variance to the innovation in η_t is not large in relation to potential magnitudes of Δm_t values, it will still be true that inflation rates will be principally determined by money growth rates, that long-run monetary neutrality will prevail, that superneutrality will be approximately but not precisely valid, etc. That the disturbance term in the money demand relationship is of the DS class is simply not a source of embarrassment or special concern for supporters of the traditional theory of money demand.[48]

Much the same can be said, furthermore, in the context of PPP doctrine. Nominal exchange rates are probably not cointegrated with relative price levels because technological and taste shocks affecting real exchange

rates have permanent components.[49] But major differences among nations in money growth and inflation rates may nevertheless dominate other effects on nominal exchange rates over long spans of time, leaving the practical messages of the PPP doctrine entirely valid as a long run matter.[50] That such is the case in actuality is indicated by the evidence collected by Galliot (1970) and Officer (1976).

In both of the preceding examples, it was argued that one should expect the disturbance term in a relation among levels of economic variables to include both permanent and transitory components, and therefore to possess an autoregressive unit root. This argument--which is an application to disturbance terms of the unobserved-components perspective put forth in Section IV--would seem to be applicable quite broadly; indeed, to the disturbances of most behavioral relations. That point of view implies, unfortunately, that cointegrating relationships will be rare[51] and so the potentially beneficial estimation result mentioned in the first paragraph of this section will not be forthcoming.[52]

The argument of the present section has a natural counterpart, it might be added, in the context of debates concerning non-trend-stationarity of the price level. Some commentators, including Barro (1986) and Haraf (1986), have emphasized uncertainty concerning future values of the price level and have accordingly suggested that it is highly undesirable for p_t (log of the price level) to be generated by a unit root process. The point of view expressed here emphasizes, by contrast, the relative unimportance of p_t nonstationarity per se, given the existing magnitude of the disturbance variance for the p_t process, in comparison with recent values of the trend growth rate.[53] One way to express the point is to hypothesize that citizens and policy-makers in the United States would view price level performance as highly satisfactory if it were generated (in quarterly terms) as

(32) $p_t = \delta + p_{t-1} + c_t$

if $\delta = 0$ and c_t were white noise with $\sigma_c^2 = 0.00002$. (The latter figure approximately equals the one-quarter forecast variance over 1954-1991.) Looking twenty years ahead, the forecast variance of p_t would be 80 (0.00002) = 0.0016 so a 95 percent confidence interval would be the current value plus or minus 0.08 (or ± 8 percent in terms of the price level). That figure pales into insignificance in comparison with the expected change in p_t over twenty years if δ were nonzero and equal to (e.g.) 0.011, a figure that corresponds to a 4.5 percent annual rate of inflation.

VII. Conclusions

In this final section we shall conclude the arguments. The discussion will not be a summary of what has gone before--which is itself largely a condensation of other work -- but instead will attempt to reach conclusions in the sense of "logical consequences" of what has gone before. In developing our arguments it will be useful to distinguish the two different purposes of trend analysis that were mentioned above: (i) isolating trend from cyclical components and (ii) trend removal for the purpose of obtaining series suitable for econometric analysis. We begin with subject (ii).

In the context of removing trends from time series so that relationships among these series can be studied by conventional econometric methods, we have seen that there is a tendency for similar results to be obtained from the two methods, provided that serial correlation corrections are applied to the residuals of the relationship being studied. This suggests that it is not crucial whether the analyst differences the data or removes deterministic trends. The recommended course of action would then be, evidently, to estimate the econometric model both ways--with differenced and (deterministically) detrended data--and hope that similar results will in fact be obtained. But emphasis in presentation will usually be given to one set of results or the other, and in some cases the results will not be similar. A natural basis for choice would then be to feature the results that require the smaller amount of correction to remove autocorrelation of the residuals. In the case of the GNP-monetary base example of Section V, for example, the preferred results would be those in equation (24), rather than (26) or (28). And in the exchange rate example of Table 5, the results in the second panel would be preferred, according to this criterion.

Now consider purpose (i), the estimation of trends so as to isolate trend from cyclical components of a series. In Sections III-V above we have

reviewed various results all of which indicate that there is no reliable method for distinguishing among alternative trend/cycle decompositions even when these have entirely different long run response characteristics and different implications about the relative importance of the two components. This seems, at first glance, a discouraging conclusion.

Reflection on the issue suggests, however, that it actually makes very little sense even to attempt to distinguish between trend and cycle on the basis of a variable's univariate time series properties alone. The reason is that the separation of trend and cycle will in most cases be desired because the analyst believes that the two components have different economic properties or significance. With regard to real GNP, for example, Nelson (1989, page 74) emphasizes that analysts "tend to think of the processes generating the two components as quite different," one being "due to growth in labor force and capital stock and to technological change" and the other "arising largely from monetary [and fiscal] disturbances." But such components will be neither independent nor perfectly correlated, as presumed by the two main trend estimation procedures described above. And without knowledge of the extent of correlation, they are not identified even under the assumption that the trend component is a random walk. This latter assumption, moreover, is itself rather unsatisfactory.

More generally, the distinction between trend and cycle is by many economists viewed as pertaining to movements that are socially desirable and undesirable, respectively. But whether such is the case clearly depends upon the economist's theory of cyclical fluctuations, for some of these--the real business cycle hypothesis, for example--will not view cyclical movements as something that policy should attempt to mitigate. The nature of the cycle vs. trend distinction, in other words, depends upon the theory of macroeconomic fluctuations adopted. But if that is the case, then it makes

little sense to attempt to separate out the cyclical component by means of a procedure that takes no account of alternative theories but relies merely on a variable's time series properties.[53]

The reader may have noticed that the remarks in this concluding section have pertained exclusively to trend analysis, with the term "unit roots" failing to appear. More generally, it may have been noted that there is no inevitable connection between the two concepts--unit roots may be present in a series that is entirely trendless (and vice-versa). But the presence of trends is a constant source of practical difficulties in the analysis of time series data, and the recent interest in unit roots has stemmed largely from the notion of stochastic trends. It is then for reasons of practicality that emphasis has here been given to the topic of trends. Our principal messages regarding unit roots <u>per se</u> are implicit in our conclusions regarding trends. But since those messages are somewhat negative concerning the value of unit root testing, it needs to be mentioned explicitly that introduction of the unit root concept, together with recognition that series are likely to include DS components, has been a valuable corrective to the earlier habit of presuming trend stationarity and has led to several analytical insights.[54]

References

Barro, Robert J., "Recent Developments in the Theory of Rules versus Discretion," Economic Journal 96 (Supplement, 1986), 23-37.

Blanchard, Olivier J., and Danny Quah, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," American Economic Review 79 (September 1989), 655-673.

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

Campbell, John Y., and N. Gregory Mankiw, "Are Output Fluctuations Transitory?" Quarterly Journal of Economics 102 (November 1987), 857-880. (a)

_____ and _____, "Permanent and Transitory Components in Macroeconomic Fluctuations," American Economic Review Papers and Proceedings 77 (May 1987), 111-117. (b)

Campbell, John Y., and Pierre Perron, "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," NBER Macroeconomics Annual 1991. Cambridge, MA: MIT Press, 1991.

Clark, Peter K., "The Cyclical Component of U.S. Economic Activity," Quarterly Journal of Economics 102 (November 1987), 797-814.

_____, "Nearly Redundant Parameters and Measures of Persistence in Economic Time Series," Journal of Economic Dynamics and Control 12 (June/September 1988), 447-461.

Cochrane, John H., "How Big is the Random Walk in GNP?" Journal of Political Economy 96 (October 1988), 893-920.

_____, "A Critique of the Application of Unit Root Tests," Journal of Economic Dynamics and Control 15 (_____1991), 275-284.

_____, "Comment," NBER Macroeconomics Annual 1991. Cambridge, MA: MIT Press, 1991.

Christiano, Lawrence J., and Martin Eichenbaum, "Unit Roots in Real GNP: Do We Know, and Do We Care?" Carnegie-Rochester Conference Series on Public Policy 32 (Spring 1990), 7-62.

Cuthbertson, Keith, and mark P. Taylor, "Money Demand, Expectations, and the Forward-Looking Model," Journal of Policy Modelling 12 (Summer 1990), 289-315.

DeJong, David N., and Charles H. Whiteman, "Reconsidering 'Trends and Random Walks in Macroeconomic Time Series,'" Journal of Monetary Economics 28 (October 1991), 221-254.

_____ and _____, "Trends and Cycles as Unobserved Components in Real GNP: A Bayesian Perspective," Proceedings of the American Statistical Association Business and Economics Section 1989, 63-70.

Diebold, Francis X., Stephen Husted, and Mark Rush, "Real Exchange Rates Uncer the Gold Standard," Journal of Political Economy 99 (December 1991), 1252-1271.

Dhrymes, Phoebus J., Distributed Lags. San Francisco: Holden-Day, 1971.

Engle, Robert F., and C.W.J. Granger, "Co-integration and Error Correction: Representation, Estimation, and Testing," Econometrica 55 (March 1987), 251-276.

Fisher, Mark, and John J. Seater, "Long-Run Neutrality and Superneutrality in an ARIMA Framework," American Economic Review 83 (1993), forthcoming.

Gailliot, Henry J., "Purchasing Power Parity as An Explanation of Long-Term Changes in Exchange Rates," Journal of Money, Credit, and Banking 2 (November 1970), 348-357.

Goodfriend, Marvin, "Interest Rate Smoothing and Price Level Trend-Stationarity," Journal of Monetary Economics 19 (May 1987), 335-348.

Granger, C.W.J., "Co-Integrated Variables and Error-Correcting Models," Working Paper 83-13, University of California, San Diego, 1983.

Granger, C.W.J., and Paul Newbold, "Spurious Regressions in Econometrics," Journal of Econometrics (1974)), 111-120.

Haraf, William S., "Monetary Velocity and Monetary Rules," Cato Journal 6 (Full 1986), 641-662.

Hafer, R.W., and Dennis W. Jansen, "The Demand for Money in the United States: Evidence from Cointegration Tests," Journal of Money, Credit, and Banking 23 (May 1991), 155-168.

Harvey, Andrew C., "On Comparing Regression Models in Levels and First Differences," International Economic Review 21 (October 1980), 707-720.

_____, Time Series Models. Oxford: Phillip Allen, 1981.

_____, "Trends and Cycles in Macroeconomic Time Series," Journal Of Business and Economic Statistics 3 (1985), 216-227.

Hendry, David, "Econometric Modelling with Cointegrated Variables: An Overview," Oxford Bulletin of Economics and Statistics 48 (August 1986), 201-212.

Hoffman, Dennis L., and Robert H. Rasche, "Long-Run Income and Interest Elasticities of Money Demand in the United States," Review of Economics and Statistics 78 (1991), 841-858.

King, Robert, and Mark W. Watson, "Testing Long Run Neutrality," NBER Working Paper No. 4156, September 1992.

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin, "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," Journal of Econometrics 54 (1992), 159-178.

McCallum, Bennett T., "On 'Real' and 'Sticky-Price' Theories of the Business Cycle," Journal of Money, Credit, and Banking 18 (November 1986), 397-414.

_____, "The Case for Rules in the Conduct of Monetary Policy: A Concrete Example," Federal Reserve Bank of Richmond Economic Review 73 (September/October 1987), 10-18.

_____, "A Reconsideration of the Uncovered Interest Parity Relationship," NBER Working Paper No. 4113, July 1992.

McCallum, Bennett T., and Marvin S. Goodfriend, "Demand for Money: Theoretical Analysis," The New Palgrave: A Dictionary of Economics. London: Macmillan Press, 1987. Reprinted in Federal Reserve Bank of Richmond Economic Review 74 (January/February 1988), 16-24.

McNown, Robert, and Myles S. Wallace, "National Price Levels, Purchasing Power Parity, and Cointegration: A Test of Four High Inflation Economies," Journal of International Money and Finance (December 1989), 533-545.

Mehra, Yash P., "Some Further Results on the Sources of Shift in M1 Demand in the 1980s," Federal Reserve Bank of Richmond Economic Review 75 (September/October 1989), 3-13.

Miller, Stephen M., "Monetary Dynamics: An Application of Cointegration and Error-Correction Modelling," Journal of Money, Credit, and Banking 23 (May 1991), 139-154.

Nelson, Charles R., "Implicit Estimates of the National and Cyclical Components of Japan's Real GNP," Bank of Japan Monetary and Economic Studies 7 (August 1989), 73-91.

Nelson, Charles R., and Heejoon Kang, "Pitfalls in the Use of Time as an Explanatory Variable in Regression," Journal of Business & Economic Statistics 2 (January 1984), 73-82.

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

Officer, Lawrence H., "Effective Exchange Rates and Price Ratios Over the Long Run: A Test of the Purchasing-Power-Parity Theory," Canadian Journal of Economics 13 (May 1980), 206-230.

Phillips, Peter C.B., "To Criticise the Critics: An Objective Bayesian Analysis of Stochastic Trends," Journal of Applied Econometrics 6 (October-December 1991), 333-364.

Plosser, Charles I., and William G. Schwert, "Estimation of a Noninvertible Moving Average Process: The Case of Overdifferencing," Journal of Econometrics 6 (August 1977), 199-224.

_____ and _____, "Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing," Journal of Monetary Economics 4 (November 1978), 637-660.

Rudebusch, Glenn D., "The Uncertain Unit Root in Real GNP," American Economic Review 82 (1992), forthcoming.

Sims, Christopher A., "Bayesian Skepticism on Unit Root Econometrics," Journal of Economic Dynamics and Control 12 (1988), 463-474.

Sowell, Fallaw, "Modelling Long-Run Behavior with the Fractional ARIMA Model," Journal of Monetary Economics 29 (April 1992), 277-302.

Stock, James H., "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," Journal of Monetary Economics 28 (December 1991), 435-459.

Stock, James H., and Mark W. Watson, "Variable Trends in Economic Time Series, Journal of Economic Perspectives 2 (Summer 1988), 147-174.

Stockman, Alan C., "The Equilibrium Approach to Exchange Rates," Federal Reserve Bank of Richmond Economic Review 73 (March-April 1987), 12-30.

Taylor, Mark P. "An Empirical Examination of Long-Run Purchasing Power Parity Using Cointegration Techniques," Applied Economics 20 (1988), 1369-1381.

Watson, Mark W., "Univariate Detrending Methods with Stochastic Trends," Journal of Monetary Economics 18 (July 1986), 49-75.

West, Kenneth D., "On the Interpretation of the near Random-Walk Behavior in GNP," American Economic Review 78 (March 1988), 202-211.

Footnotes

[1]For other recent survey articles, see Stock and Watson (1988) and Campbell and Perron (1991).

[2]The proper term is "convolution." Any reader who desires a more rigorous description of lag operators may consult Dhrymes (1971).

[3]That is, non-stochastic.

[4]Throughout, our discussion will ignore seasonal components.

[5]A white-noise random variable is one generated by a process that specifies that each period's value, ε_t, is drawn from a population with mean zero and finite variance σ^2_ε, and is not dependent on previous values.

[6]This was the year in which the article by Nelson and Plosser (1982) was published. The popularity of differencing had been growing gradually, however, at least since the much earlier publication of Box and Jenkins (1970).

[7]That is, to include only a finite number of terms with non-zero coefficients. Any stationary stochastic process can be closely approximated by an expression of the form $\theta(L)\phi^{-1}(L)\varepsilon_t$.

[8]Perhaps a complex number.

[9]Consider, for example, the second-order case. Then $1 + \phi_1 z + \phi_2 z^2 = 0$ could equivalently be written as $(1-\alpha_1 z)(1-\alpha_2 z) = 0$, where $\phi_1 = -(\alpha_1+\alpha_2)$ and $\phi_2 = \alpha_1\alpha_2$. But the latter equation is satisfied by the two values $z^1 = 1/\alpha_1$ and $z^2 = 1/\alpha_2$. So the lag polynomial $1 + \phi_1 L + \phi_2 L_1^2$ could as well be expressed as $(1-\alpha_1 L)(1-\alpha_2 L)$. The roots of the polynomial $\phi(L)$ are said to be $1/\alpha_1$ and $1/\alpha_2$.

[10]Here $(1-\rho L)(\alpha_0 + \alpha_1 t) = \alpha_0 - \rho\alpha_0 L + \alpha_1 t - \rho\alpha_1 (t-1)$

$= \alpha_0(1-\rho) + \alpha_1 t - \rho\alpha_1 t + \rho\alpha_1 = \alpha_0(1-\rho) + \alpha_1(1-\rho)t + \rho\alpha_1$

[11]If $\rho>1.0$, then y_t will have explosive behavior of type that seems unlikely and that will become easily detectable after a few years.

[12]The difficulty was recognized, but not emphasized, by Nelson and Plosser (1982).

[13]See, e.g., McCallum (1986, pp. 405-406).

[14]It would, however, be possible to object that expressing trend stationarity as a zero-measure special case effectively biases the procedure in favor of a DS finding. Note, incidentally, that a unit root in the MA polynomial does not imply a process of the "unit root" type.

[15]I.e., in the limit as the sample size grows without bound.

[16]An ARMA model is one that admits both autoregressive and moving average polynomials. The notation (p,q) indicates how many terms (p and q) are included in the AR and MA polynomials. Sometimes the number of times d that the basic variable has been differenced is included in a (p,d,q) notation.

[17]Here the meaning of "model A is better than B" is that B is nested in A and can be rejected with a significance level of 0.05.

[18]Henceforth the words "per unit" will typically be omitted.

[19]Another useful measure has been featured in the work of Cochrane (1988). It is described below in text attached to fn. 26.

[20]Recall that they actually reported 16 cases, but that we are focussing on nine.

[21]It is sometimes said that they are highly "persistent," but that terminology is inappropriate for reasons clearly described by Cochrane (1991, pp. 206-7).

[22]With a test statistic designed to have a 0.05 significance level.

[23]I.e., a case represented by parameter values that would be graphically represented as a point (with zero area) in a region depicting all possible parameter values.

[24]Granger's Lemma says that the sum of two independent ARMA processes, one ARMA (p_1, q_1) and the other ARMA (p_2, q_2), is an ARMA (p^*, q^*) where $p^* \leq p_1 + p_2$ and $q^* \leq \max (p_1 + q_2, p_2 + q_1)$. For pure MA processes, then, $q^* \leq \max (q_1, q_2)$.

[25]Both Clark (1988) and Cochrane (1988) have developed arguments suggesting that unconstrained ARMA models with differenced series tend to be poor at the job of estimating long-run properties such as A(1).

[26]As $k \to \infty$, the limit V of the V_k sequence is the long-run response measure proposed by Cochrane (1988) mentioned above in footnote 19. Its relation to A(1) is $V = (1-R^2)(A(1))^2$, where R^2 is $1-(\sigma_\epsilon^2/\mathrm{Var}\Delta y_t)$.

[27]But with more parameters in the DS component, the components model may become equivalent to an unconstrained ARMA.

[28]See Watson (1986, p. 53).

[29]Quite recently, Kwiatkowski, Phillips, Schmidt, and Shin (1992) have conducted tests of the hypothesis that the DS component is of negligible importance in a components formulation. For the real GNP series this hypothesis was found to be of borderline significance at the 0.05 level.

[30]In least squares regression analysis the inclusion of a time trend among the regressors is equivalent to the use of variables detrended by prior regression on the same time variable (i.e., using residuals, from these prior regressions on time, as the detrended variables).

[31]The standard formulae do not rely on asymptotic distribution theory in cases in which there are no lagged dependent variables in the system under study, but such cases are the exception in applied macroeconomics.

[32]A contrary argument is that differencing sacrifices information pertaining to levels or to long-run relationships. Estimation of a levels relationship after differencing will not, of course, provide any information about the constant term, but that is usually of little importance. The argument developed below suggests that little is lost with regard to long-run multipliers unless the variables are cointegrated, a topic that is taken up briefly in Section VII.

[33]This approach is used because the usual asymptotic distribution theory breaks down in cases with unit roots in either the MA or AR polynomial.

[34]Across 200 replications.

[35]Actually the estimator "with autocorrelation correction" involves full maximum likelihood estimation of (22).

[36]These values are for a sample size of T = 100; Harvey also gives results for T = 20 and T = 50.

[37]In the levels formulation, ρ is estimated jointly with β.

[38]Data for 1953-1990 are taken from the Citibase data set, while 1991 values come from the Survey of Current Business (GNP) and the Federal Reserve Bank of St. Louis (adjusted monetary base). Calculations are performed with version 7.0 of Micro TSP.

[39]Under the assumption that the disturbances are white noise, Q(10) has asymptotically a chi-squared distribution with 8 degrees of freedom; its critical value for a 0.05 significance level is therefore 18.3.

[40]The coefficient on the base variable is now somewhat larger than 0.5, rather than smaller, but the difference is less than two standard errors.

[41]It should be said explicitly that this argument is not being made with regard to Granger causality tests or variance decomposition statistics in vector-autoregression studies. It is my impression that these results are rather sensitive to the detrending procedure. But such results are, I believe, of less importance than impulse response patterns.

[42]It is not being assumed that x_t is necessarily a predetermined variable, i.e., that u_t is uncorrelated with x_t, x_{t-1}, ...

[43]If a variable must be differenced d times to render it stationary, it is said to integrated of order d, abbreviated I(d). The term "integrated" was popularized by Box and Jenkins (1970), its genesis being that a random-walk variable is at any time equal to the infinite sum ("integral") of all past disturbances. Cointegration analysis was developed by Granger (1983) and Granger and Engle (1987).

[44]See, for example, the expository piece by Dickey, Jansen, and Thornton (1991).

[45]See Hendry (1986).

[46]Other writers have apparently accepted this characterization prior to reaching the opposite empirical conclusion. A few examples are Mehra (1989), Hoffman and Rasche (1991), Miller (1991), Hafer and Jansen (1991), and Diebold, Husted, and Rush (1991).

[47]This step should not be at issue; the existence of technological change in the payments industry is widely accepted.

[48]Many of these supporters have been willing to estimate money-demand functions in first-differenced form, thereby implicitly assuming a DS disturbance process.

[49]As suggested, for example, by Stockman (1987).

[50]Here the interpretation of PPP is taken to agree with popular usage although a good case can be made for an alternative interpretation that expresses PPP as a form of a neutrality proposition.

[51]The s_t, f_t example in Section VI is, however, a case in which cointegration evidently does obtain.

[52]Campbell (1991, p. 218-219) argues against this suggestion by means of a reductio ad absurdum. The latter is not actually applicable, however, as my argument is directed only toward variables that enter agents' utility functions or budget constraints.

[53]With regard to the notable result of Goodfriend (1987), i.e., that interest rate smoothing behavior by the central bank can induce a unit root component into the price level process, the point is that it does not imply a positive drift term. Consequently, Goodfriend's result does not explain why prices are now several times as high as in 1950 (See McCallum, 1987) in all industrialized nations, rather than being substantially higher in some nations and lower in others.

[54]It should be noted that this argument does not imply that it is pointless to try to attempt to reach substantive macroeconomic conclusions on the basis of analyses such as that of Blanchard and Quah (1989), which utilizes multiple time series and relies upon explicit substantive assumptions for identification.

[55]A recent example is provided by the related analyses of Fisher and Seater (1993) and King and Watson (1992).