NBER WORKING PAPER SERIES

UNIT ROOTS IN REAL GNP: DO WE KNOW, AND DO WE CARE?

Lawrence J. Christiano

Martin Eichenbaum

Working Paper No. 3130

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1989

We are grateful to Christopher Sims for criticisms and suggestions which led to substantial improvements on an earlier draft of this paper. We also thank R. Anton Braun, John Campbell, John Cochrane, Pierre Perron, Danny Quah, James Stock, Mark Watson, and Ken West for their comments. We are indebted to Kathy Rolfe for exceptional editorial assistance. This research was supported in part by a grant from the National Science Foundation. This paper is part of NBER's research program in Economic Fluctuations. The views expressed herein are those of the authors and not those of the Federal Reserve Bank of Minneapolis, the Federal Reserve Bank of Chicago, the Federal Reserve Bank System, or the National Bureau of Economic Research.

NBER Working Paper #3130
October 1989

UNIT ROOTS IN REAL GNP: DO WE KNOW, AND DO WE CARE?

ABSTRACT

No, and maybe not.

Lawrence J. Christiano
Research Department
Federal Reserve Bank of
Minneapolis
Minneapolis, MN 55480

Martin Eichenbaum
Department of Economics
Northwestern University
Evanston, IL 60208

Macroeconomists have traditionally viewed movements in aggregate output as representing temporary fluctuations about a deterministic trend. According to this view, innovations to real gross national product (GNP) should have no impact on long-run forecasts of aggregate output. Increasingly, however, this view of aggregate fluctuations has been challenged. Following the important work of Nelson and Plosser (1982), numerous economists have argued that real GNP is best characterized as a stochastic process that does not revert to a deterministic trend path. Under these circumstances, innovations to real GNP should affect output forecasts into the indefinite future. In pursuing this interpretation of the data, various researchers have tried to measure the long-run response of real GNP to a shock. Estimates of this response are often referred to as the persistence of shocks to real GNP.

Not surprisingly, the literature on persistence has become intertwined with recent controversies over the empirical plausibility of two important classes of statistical univariate time series models: trend stationary and difference stationary models. Ultimately, proponents of the view that shocks to real GNP are persistent must build their case on the empirical plausibility of the hypothesis that real GNP is difference rather than trend stationary or, in other words, that real GNP has a unit root.

To us, the possibility of providing a compelling case that real GNP is either trend or difference stationary seems extremely small, certainly on the basis of postwar data. This is because there is only one difference between these two types of processes and that difference is completely summarized by the answer to the question, How much should an innovation to real GNP affect the optimal forecast of real GNP into the infinite future? If the answer is zero, then real GNP is trend stationary. If the answer is not zero, then real GNP is difference stationary. The competing hypotheses have no other testable differences. Once we pose the question in this way, it seems clear that economists ought to be extremely skeptical of any argument that purports to support one view or the other. Simply put, it's hard to believe that a mere 40 years of data contain any evidence on the only experiment that is relevant.

Notwithstanding these obvious difficulties, over the past decade there has been an explosion of empirical research on whether macroeconomic time series are best viewed as trend or difference stationary. Even more surprising are the strong conclusions that seem to mark this literature. For example, Schwert (1987, p. 99) writes that "following Nelson and Plosser, many authors have found that many aggregate output series . . ., aggregate price level series . . ., and other aggregate nominal series . . . contain a unit root." Blanchard and Quah (1988, p. 1) simply begin their analysis by stating that "in response to an innovation in GNP of 1%, one should revise one's forecast by more than 1% over long horizons. This fact is documented by Campbell and Mankiw (1987a), building on earlier work by Nelson and Plosser (1982)." And Campbell and Mankiw (1987b, p. 111) write that "much disagreement remains over exactly how persistent are shocks to output. Nonetheless, among investigators using postwar quarterly data, there is almost unanimity that there is a substantial permanent effect."

The first part of this paper argues that the new consensus about the presence and size of the unit root in real GNP is not supported by an analysis of postwar U.S. GNP data. The data simply do not discriminate between the trend stationary and the difference stationary views of U.S. real GNP. Given the nature of the difference between these two types of stochastic processes, any argument in favor of one or the other necessarily relies on strong identifying restrictions. Unfortunately, inference turns out to be extremely sensitive to exactly which identifying restrictions are made. Moreover, the relevant sets of identifying restrictions are not the sort which economic theory has anything to say about.

The paper argues this point in two ways. Initially, we investigate the problem using the parametric approach proposed by Campbell and Mankiw (1987a). The basic strategy is to estimate the long-run response of real GNP to an innovation using a particular parametric, autoregressive moving average (ARMA) representation of the growth rate of real GNP. For the postwar U.S. data, the relative plausibility of the trend and the difference stationary hypotheses depends critically on the precise order of the ARMA representation chosen. Campbell and Mankiw (1987a) emphasize an ARMA(2,2) representation of postwar U.S. real GNP. Indeed, if one conditions on that precise representation of the data, then real GNP is plausibly argued to be difference stationary. Unfortunately, very small perturbations in the order of the ARMA representation turn out to have a very large impact on this inference. For example, if one conditions on an ARMA(3,3) representation of the data, then real GNP appears to be trend stationary. Neither the data nor economic theory can convincingly discriminate between these competing representations of real GNP.

Since inference is very sensitive to particular parametric assumptions, we also examine the problem using the nonparametric methods developed by Cochrane (1988a). We show that, when applied to postwar data, these methods are completely uninformative about the relative plausibility of the trend and the difference stationary hypotheses. How much should one's forecast of real GNP over long horizons be revised in response to a 1 percent innovation? Taken together, our results lead us to answer, We don't know. Indeed, we argue that this is the right answer even if by long horizons we mean periods as short as five years. Taking any other position simply seems unwarranted by the available evidence.

The second part of this paper investigates the consequences of not knowing. Here we ask the question, Do we care if real GNP has a unit root? Our answer is, Maybe not. Some authors have argued that we should care because the degree of persistence in real GNP can be used to infer what the principal impulses driving business cycles are (for example, Long and Plosser, 1983; De Long and Summers, 1988). This line of reasoning presumes that if real GNP is highly persistent, then the shocks must be principally to technology, whereas if there is little persistence, then the shocks must be principally to aggregate demand, such as innovations to monetary and fiscal policy. Campbell and Mankiw (1987a), Cochrane (1988a), and West (1988b) argue persuasively against this view by pointing out that in a variety of plausible models these presumptions are wrong.

A more plausible reason for caring revolves around the possibility that the implications of dynamic economic models depend sensitively on the presence or absence of a unit root in the forcing variables to agents' environments. Our results suggest that the degree of sensitivity is minimal. What economic agents care about is the relative importance of temporary versus permanent shocks to their environments--and this is, at best, only loosely related to the unit root issue.

We reach this conclusion by considering two examples which have been used to argue for the importance of unit roots. These are Deaton's (1986) and Hansen's (1989) analyses of the permanent income hypothesis (PIH) and real business cycle (RBC) models. Both authors argue that the dynamic properties of their respective models are extremely sensitive to the presence or absence of unit roots in income. Indeed, Deaton and Hansen argue that the degree of sensitivity is large enough to affect inference about the overall plausibility of the models.

Taken at face value, these examples do create a presumption that we should care about the unit root issue. However, this presumption turns out to depend on a key maintained assumption of both analyses, namely, that the forcing variables of concern to agents are driven by a single shock. This assumption implies a sharp dichotomy between trend and difference stationary specifications which is not tied in any logical way to the unit root issue. With only one shock to agents' environments, a trend stationary specification implies that all shocks have purely temporary effects, whereas difference stationary specifications necessarily incorporate the opposite extreme; all shocks have purely permanent effects. But without this assumption, difference stationarity does not imply the absence of temporary shocks to agents' environments.

To see why, suppose we actually knew that a particular random variable was difference stationary. Indeed, assume that we actually knew the univariate, difference stationary, Wold representation of the random variable. For each such representation, the variable can be decomposed into permanent and temporary components in an infinite number of ways (Quah, 1988). While each decomposition implies precisely the same univariate time series representation, each decomposition embodies different assumptions about the relative importance of permanent and temporary shocks to the variable. Simply knowing the univariate time series representation of a random variable provides no information about which of the infinite decompositions agents may be observing and responding to. At the same time, which decomposition is chosen is critical since agents' actions differ depending on whether they are responding to a permanent or a temporary shock. Consequently, the properties of dynamic models will in general depend sensitively on the relative importance of permanent and temporary shocks. Agreeing on the presence of a unit root in the law of motion for some variable, or even the variable's univariate time series representation, imposes almost no restrictions on one's view of this issue since there always exists a decomposition which makes the permanent component arbitrarily small.

Allowing for the presence of a temporary component in the difference stationary representation of the forcing variables to agents' environments breaks the sharp dichotomy implicit in the analyses of Deaton and Hansen. Quah (1989) shows this quite dramatically in his discussion of Deaton's results. When agents see only the univariate, difference stationary representation for labor income, consumption is predicted to be about 1.8 times as volatile as income; when the trend stationary specification is adopted, consumption is predicted to be 0.2 times as volatile as income. Despite these sharp differences, Quah is able to display a temporary/permanent decomposition of Deaton's difference stationary model of labor income which has the following property: If agents observe both components separately, then the predicted relative volatility of consumption coincides with the predictions of the trend stationary model of labor income. This is true despite the fact that the univariate time series representation implied by Quah's components model is precisely the same as that implied by Deaton's difference stationary model.

In the second part of this paper, we illustrate Quah's results in the context of RBC models. In particular, we show that the implications of an unobserved components, difference stationary version of Hansen's RBC model closely resemble those of a trend stationary specification. Our tentative conclusion is that once we admit the possibility that agents are responding to both temporary and permanent shocks, the unit root question loses much of its importance.

Our paper is organized as follows. First we discuss the concepts of trend and difference stationarity and two statistical procedures which have been used to distinguish between them empirically. In the next two sections, we argue that, using postwar data, one cannot determine the long-run effect of an innovation to real GNP based on two leading statistical procedures, the ARMA method of Campbell and Mankiw (1987a,b) and the nonparametric method of Cochrane (1988a). Then we address the issue of whether unit roots matter from an economic perspective. In the last section, we make some concluding remarks.

A SELECTIVE OVERVIEW OF THE LITERATURE

Recent research aimed at analyzing the persistence of shocks to real GNP has been conducted almost exclusively within the confines of atheoretical time series models. Much of the debate has centered on efforts to support or refute the traditional view that fluctuations in real GNP reflect temporary deviations from a deterministic trend path. At issue is the relative plausibility of two important classes of statistical univariate time series models: trend stationary and difference stationary models. We begin by reviewing these models.

Consider the time series variable y_t, which we assume is measured in logarithms. According to the trend stationary model, y_t is covariance stationary about a deterministic trend. If the growth rate of y_t is a stationary stochastic process, the deterministic trend component must be linear. The following univariate time series representation for y_t reflects these assumptions:

$$y_t = \gamma t + a(L)\varepsilon_t. \tag{1}$$

Here $a(L) = 1 + a_1L + a_2L^2 + \ldots$ is a polynomial in the lag operator L, γ is a scalar constant, t denotes time, and ε_t is the zero mean, serially uncorrelated time t innovation to y_t. We denote the variance of ε_t by σ_ε^2. Since y_t is covariance stationary, $\sum a_j^2$ and σ_ε^2 are both finite. For convenience, we assume that $\sum|a_j|$ is finite.

According to the difference stationary model, the first difference of y_t is a covariance stationary process which we write as

$$\Delta y_t = \mu + b(L)u_t. \tag{2}$$

Here Δ denotes the first-difference operator, $b(L) = 1 + b_1L + b_2L^2 + \ldots$ is a polynomial in the lag operator L, μ is a scalar constant, $\sum|b_j| < \infty$, and u_t is the zero mean, serially uncorrelated innovation to y_t. We denote the variance of u_t by σ_u^2. In addition, we impose the requirement that $b(1) = \sum b_j \neq 0$. Without this requirement, there is no difference between trend stationary and difference stationary processes. This follows from the fact that any trend stationary process, (1), can be represented in the form of (2). To see this, simply first-difference both sides of (1) to obtain

$$\Delta y_t = \gamma + A(L)\varepsilon_t \tag{3}$$

where $A(L) = (1-L)a(L)$. This process satisfies all of the conditions imposed by the difference stationary model except for one. The sum of coefficients on current and lagged ε_t's in (3) is given by $A(1)$. Under our assumptions, $A(1) \equiv 0$. But this violates the condition that the sum of the moving average coefficients in (2) is not equal to zero. Evidently, the condition that $b(1) \neq 0$ is all that distinguishes trend and difference stationary processes.

Two widespread interpretations of $b(1)$ revolve around its role in determining the degree of persistence in y_t. One measure of persistence centers on the response to u_t of the optimal forecast of y_t into the infinite future. Let E_t denote the time t expectations operator conditioned on the information set $\{u_t, u_{t-1}, \ldots\}$. Beveridge and Nelson (1981) show that

$$\lim_{k\to\infty} [E_t y_{t+k} - E_{t-1} y_{t+k}] = b(1)u_t. \tag{4}$$

It follows that b(1) completely characterizes the revision in the long-run outlook for y_t induced by a time t unit innovation to y_t. If y_t is difference stationary, then $b(1) \neq 0$, so that an innovation at time t ought to affect our forecast of y_t into the infinite future. However, if y_t is trend stationary--say, as given by (3)--then $A(1) = 0$. Consequently, an innovation to y_t should have no impact on our forecast of y_t into the infinite future. This simply reflects the fact that eventually a trend stationary process always returns to its deterministic trend path.

The other common measure of persistence revolves around the fact that the long-run forecast of a difference stationary process is always changing. According to (4), the time t revision to the long-run forecast of y_t is the random variable $b(1)u_t$. A natural measure of the amount of variation in this variable is its variance, $[b(1)]^2\sigma_u^2$. If y_t is trend stationary, then fluctuations in u_t induce only transitory movements in y_t; that is, the long-run outlook is deterministic. Consequently, the variance of the revision to the long-run forecast of a trend stationary random variable, $[A(1)]^2\sigma_\varepsilon^2$, is zero.

Under the first interpretation of b(1), the issue of trend versus difference stationarity reduces to the question, How much should an innovation to the stochastic process y_t at time t affect our forecast of y_t into the infinite future? Under the second interpretation of b(1), the issue of trend versus difference stationarity reduces to the question, How variable is the optimal forecast of y_t in the infinite future? Posed in these terms, neither question is answerable on the basis of any finite data set. There simply are no observations on the experiment. The relevant issues then become, What identifying restrictions have been made in the literature to answer questions about persistence? And how sensitive are the answers to different identifying assumptions?

To review the literature from this perspective, we adopt the following generic representation for Δy_t:

$$\Delta y_t = \alpha + C(L)\eta_t \tag{5}$$

where η_t is the white noise innovation to y_t, with variance σ_η^2. When $C(1) = 0$, (5) is a trend stationary representation. Otherwise, it is difference stationary. The literature on the persistence of U.S. real GNP can be roughly divided into two categories. One category focuses simply on the question of whether or not $C(1) = 0$, that is, whether real GNP is trend or difference stationary. Two strategies are taken here, with one adopting $C(1) = 0$ as the null hypothesis and the other adopting $C(1) \neq 0$. For example, Campbell and Mankiw (1987a) use the first strategy; Nelson and Plosser (1982), the second. This type of analysis only addresses the narrowly defined question, Are output fluctuations temporary or permanent?

The other category of research aims to quantify the persistence of y_t by focusing on the two measures discussed above. Watson (1986), Campbell and Mankiw (1987a,b), Clark (1987), and Campbell and Deaton (1988) estimate C(1) using postwar U.S. real GNP data. Campbell and Mankiw (1987a,b) and Campbell and Deaton (1988) also measure $[C(1)]^2\sigma_\eta^2$ using an estimator proposed by Cochrane (1988a), who implements his procedure on per capita real GNP data covering the pre- and postwar period.

How can these authors make inferences about either C(1) or $[C(1)]^2\sigma_\eta^2$ using a finite amount of data? As Cochrane (1988a) emphasizes, inferences about the persistence of y_t are made possible only by imposing identifying restrictions on C(L) and η_t. The point of these restrictions is to allow the econometrician to make inferences about the long-run dynamics of y_t--say, as measured by C(1)--from its short-run dynamics. The different types of identifying assumptions in the literature fall into two categories which are closely linked to different strategies for actually estimating objects like C(1). One strategy amounts to fitting parsimoniously parameterized ARMA models for Δy_t and then drawing inferences about persistence from the resulting parameter estimates. The other strategy is less parametric in nature.

Consider first the more parametric strategy. Here, two different approaches for achieving parsimony have been pursued. Authors like Campbell and Mankiw (1987a) achieve parsimony by limiting the orders of the autoregressive and moving average components of the ARMA representation of the data. In particular, C(L) is assumed to be of the form

$$C(L) = \frac{\theta(L)}{\phi(L)} \tag{6}$$

where $\theta(L)$ and $\phi(L)$ are polynomials in the lag operator of order q and p, respectively. This implies an ARMA(p,q) representation for Δy_t. No additional restrictions are imposed on the model. To test the null hypothesis that y_t is trend stationary [that is, $C(1) = 0$], Campbell and Mankiw obtain both an unconstrained estimate of C(L) and an estimate of C(L) subject to the constraint that $C(1) = 0$. Given some metric for judging the empirical plausibility of the constraint, they calculate $C(1) = \theta(1)/\phi(1)$ using the preferred model.

A second strategy for achieving parsimony is to work within the confines of an unobserved components model. Here the idea is to model y_t as the sum of permanent and temporary components: $y_t = z_t + c_t$, where z_t and c_t are difference and trend stationary stochastic processes, respectively. An important advantage of this approach is that parsimonious representations for z_t and c_t will imply ARMA representations for Δy_t with high autoregressive and moving average components. For example, Watson (1986) assumes that z_t is a pure random walk and c_t has a second-order autoregressive representation. In addition, he assumes that z_t and c_t are orthogonal processes. Under these assumptions, his model generates an ARMA(2,2) representation for Δy_t that is completely described by four parameters: the variances of the innovations to z_t and c_t and the autoregressive parameters of c_t.

We think of unobserved components models of Δy_t as simply devices for achieving parameter parsimony in ARMA models. From this perspective, the particular decomposition adopted need not be structural in any interesting economic sense. Indeed, this approach for achieving parsimony may be quite useful for forecasting purposes even if the restrictions are false. However, as a device for obtaining the true cyclical component of the data, this approach is more problematic. This is because there exists an uncountable number of decompositions for y_t, even if we impose the assumption that z_t and c_t are orthogonal processes (Quah, 1988). Each decomposition implies precisely the same univariate time series representation for y_t. Obviously, strong restrictions are required to identify z_t and c_t. Unfortunately, the economic motivations behind the decompositions used in the unobserved components literature are often left unspecified and are at best problematic when viewed from the perspective of economic theory.

For example, the Beveridge and Nelson (1981) decomposition assumes that the innovations to z_t and c_t are perfectly correlated. This is clearly incompatible with real business cycle models in which there is more than one shock (such as in Christiano and Eichenbaum, 1988b; Braun, 1989; or McGratten, 1989). At the other extreme, the assumption that z_t and c_t are orthogonal processes is also incompatible with these models. RBC models often do lead to unobserved components representations for y_t (as in King, Plosser, Stock, and Watson, 1987). However, nothing inherent in the models underlying these representations implies that y_t ought to be difference stationary (as opposed to trend stationary) or, if so, that z_t ought to be a random walk. Finally, and most importantly, when agents' environments have more than one shock, z_t and c_t will be imperfectly correlated. But these are precisely the circumstances under which unobserved components models are not identified (Watson, 1986).

The set of restrictions imposed on C(L) by a particular unobserved components model clearly influences inference about C(1). One way to see this is to compare Campbell and Mankiw's (1987a) results with those of Watson (1986). Working with an unconstrained ARMA(2,2) model for Δy_t, Campbell and Mankiw estimate C(1) to be 1.52. But Watson's unobserved components, constrained ARMA(2,2) model generates a quite different estimate: 0.57. Aside from a minor discrepancy in sample period, the only difference between the two procedures is the constraints imposed by the unobserved components model. The point of this comparison is not to evaluate the relative plausibility of these two parametric procedures for estimating C(1). Instead, we only want to emphasize--as Watson (1986) does--that inference about C(1) can be very sensitive to different identifying assumptions about C(L).

In contrast to the parametric approaches discussed above, Cochrane (1988a) proposes the following statistic to measure persistence:

$$V^k \equiv \frac{\mathrm{var}(y_t - y_{t-k})}{k\sigma^2_{\Delta y}} = 1 + 2\sum_{j=1}^{k-1} \frac{k-1}{k}\rho_j. \tag{7}$$

Here $\sigma^2_{\Delta y} = \mathrm{var}(y_t - y_{t-1})$, $\rho_j = \mathrm{cov}(\Delta y_t, \Delta y_{t-j})/\sigma^2_{\Delta y}$, and $k = 2, 3, \ldots$. [See Cochrane, 1988a, for a proof of the equality in (7).] For a given value of k, Campbell and Mankiw (1987a) estimate V^k by replacing the population moments in (7) with their sample analogs. (For an alternative estimator, see Cochrane, 1988a.) We denote the sample estimator of V^k by $\hat{V}^k$.

To motivate the usefulness of V^k as a measure of persistence, we use the fact that

$$V = \lim_{k\to\infty} V^k = S_{\Delta y}(1)/\sigma^2_{\Delta y}. \tag{8}$$

Here $S_{\Delta y}(1)$ is the spectral density of Δy_t evaluated at frequency zero. Let $z = e^{i\omega}$ for $\omega \in [0,2\pi]$. If y_t has the law of motion given by (5), then

$$S_{\Delta y}(z) = C(z)C(z^{-1})\sigma^2_\eta. \tag{9}$$

Consequently,

$$V = [C(1)]^2\sigma^2_\eta/\sigma^2_{\Delta y}. \tag{10}$$

Combining (8) and (10), we see that the issue of whether a time series is trend stationary [that is, C(1) = 0] or difference stationary [C(1) ≠ 0] is equivalent to asking whether the value of its spectral density at frequency zero (z = 1) is zero or nonzero, respectively.

To implement this estimator, one must choose a value of k. Consequently, the crucial identifying assumption underlying this measure of persistence is the assumption that whatever value of k is chosen, the higher autocorrelations are of negligible importance. Cochrane (1988a) argues that for real GNP a good value for k is in the region of 20-30 years. The key point is that, unlike parametric ARMA approaches, Cochrane's procedure does not exploit information about the short-run dynamics of y_t to measure long-run dynamics.

UNIT ROOTS IN REAL GNP: DO WE KNOW?

Parametric Measures of Persistence

In this section, we analyze the persistence of U.S. real GNP using the parametric methods discussed above. We begin by estimating a variety of parsimoniously parameterized ARMA models for the first difference of the log of quarterly real GNP using data from 1948:1 to 1985:4. To estimate the models, we used Ansley's (1979) exact maximum likelihood procedure. As Campbell and Mankiw (1987a) do, we restrict ourselves to ARMA(p,q) models with p = 0, 1, 2, 3 and q = 0, 1, 2, 3, but do not consider the case in which both p and q equal zero. In addition to estimating the unconstrained ARMA models, we estimated the models (those for which q ≥ 1) subject to the constraint of trend stationarity [$\theta(1) = C(1) = 0$].

Our estimation results are in Table 1. There we report twice the difference of the log likelihood values associated with the unconstrained and constrained versions of each ARMA model. The corresponding number in parentheses is the probability value of the associated likelihood ratio statistic implied by the chi-square distribution with one degree of freedom. These probability values are included as a convenient benchmark only. Standard justifications for interpreting the likelihood ratio statistic as a realization from an asymptotic chi-square distribution rule out unit moving average roots under the null hypothesis. (See Kohn, 1979, and Plosser and Schwert, 1977.) Each estimated ARMA model generates an estimate of C(1), which we denote by $\hat{C}(1)$. The sample estimator of C(1) is simply $\hat{\theta}(1)/\hat{\phi}(1)$. Since $\hat{C}(1) = 0$ by construction for the constrained models, Table 1 reports the value of $\hat{C}(1)$ corresponding to only the unconstrained models.

We want to emphasize four key features of our results:

1. Generally, imposing the constraint $C(1) = 0$ causes a greater deterioration in the likelihood function of the more parsimoniously parameterized models. For example, the drop is very dramatic in the $p = 0$ models, where the likelihood ratio statistic exceeds 150. The smallest drop occurs in the models with $p \geq 1$ and $q = 3$, where the deterioration in the likelihood function is trivial. Indeed, using conventional sampling theory, we cannot reject the null hypothesis that $C(1) = 0$ at the 30 percent significance level.

2. An important exception to the general pattern that more parsimony implies a smaller likelihood ratio statistic is the ARMA(2,2) model. In the class of models with $p \geq 2$, this is the only model in which, using conventional sampling theory, we can reject the null hypothesis that $C(1) = 0$ at the 5 percent significance level.

3. With the exception of the ARMA(3,3) model, all of the estimated values of $C(1)$ are substantially greater than 1. At the same time, for most of the models, the likelihood ratio statistic suggests a great deal of uncertainty about the true value of $C(1)$.

4. In the ARMA(3,3) model, the only specification for which the global maximum of the likelihood function occurs at $\hat{C}(1) = 0$, one of the autoregressive roots equals 0.949, so that there is near parameter redundancy.

An alternative way to represent our results is to display the graph of the maximized value of the likelihood surface of the different ARMA models as a function of $C(1)$. This representation makes even clearer the first three features discussed above. For a fixed value of $C(1)$ equal to k, the parameters of the ARMA model must satisfy the restriction $\theta(1) - k\phi(1) = 0$. To generate the desired likelihood surface, we computed the maximized value of the likelihood function for all the ARMA(p,q) models with $p \geq 1$, $q \geq 0$, subject to this restriction. In so doing, we chose values of k belonging to the grid defined by the boundaries zero and 2 and fixed grid size 0.01.

The resulting likelihood surfaces for the ARMA(0,q), ARMA(1,q), ARMA(2,q), and ARMA(3,q) models are displayed in Figures 1a-1d, respectively. The lowest, middle, and highest curves in each figure correspond to q = 1, 2, and 3, respectively. According to these figures, all of the likelihood surfaces have a local maximum at a large value of C(1). Moreover, all of the surfaces corresponding to models with a nontrivial moving average component flatten out as C(1) goes to zero. This is a manifestation of the well-known fact that the slope of the exact likelihood function is zero on the unit circle.[1]

To see our first result, compare the global maximum of the likelihood function with its value at C(1) = 0. Generally, the distance between these two values is smaller for the more profligately parameterized models. For example, whenever $p \geq 1$ and q = 3, the global maximum of the likelihood function is very close to its alternative values. Consistent with our second result, the ARMA(2,2) model stands out as an exception to this pattern. More typical are the ARMA(1,1) and ARMA(2,3) models, in which the global maximum is extremely close to the value of the likelihood function at C(1) = 0. Finally, consistent with findings in Plosser and Schwert (1977), the graphs in Figures 1a-1d indicate that conventional estimates of the standard error of $\hat{C}(1)$ are likely to overstate the precision with which C(1) is estimated. This is because conventional methods for computing standard errors are based on the local curvature of the likelihood function at $\hat{C}(1)$. For most of our models, there is substantial curvature around $\hat{C}(1)$ but little difference between the value of the likelihood function at $\hat{C}(1)$ and C(1) = 0. Interestingly, conventional methods for computing standard errors do not give misleading results for the ARMA(1,3) and ARMA(3,3) models. In both cases, the likelihood function is basically constant over the whole range of values for C(1) which we considered.

Overall, the results in Table 1 and Figures 1a-1d show that inference about C(1) is extremely sensitive to the choice of ARMA model, that is, the choice of p and q. Some of the ARMA models support the difference stationary perspective since they indicate that C(1) is large and precisely estimated. Models in this category include the ARMA(1,1), ARMA(2,2), and ARMA(0,q) specifications, q = 1, 2, 3. But other ARMA models are consistent with the trend stationary perspective because they indicate that either C(1) = 0 or little can be said about its value. Models in this category include the ARMA(2,1), ARMA(3,1), and ARMA(p,3) models, p = 1, 2, 3. The key question is, Which perspective is best able to account for these apparently conflicting results?

To answer this question, we adopt the following approach. First, we ask whether an empirically plausible trend stationary data-generating mechanism exists that can explain those results in Table 1 which appear to support the difference stationary perspective. Then we ask the analog question for difference stationary data-generating mechanisms.

This general strategy for selecting between competing explanations of apparently contradictory statistical results was described and implemented in Christiano and Ljungqvist (1988).[2] To apply this strategy, each explanation must be formalized as one or more fully specified data-generating mechanisms. To represent the trend stationary perspective, we chose two models. The first is our estimated ARMA(3,3) model. In addition, we considered an ARMA(1,3) model in order to ensure that any results based on the ARMA(3,3) model are not sensitive to the fact that it has a relatively large number of parameters. Although our estimated ARMA(1,3) model implies a large value for C(1), the likelihood function hardly deteriorates when we impose the constraint that C(1) = 0 (Figure 1d). Indeed, when Campbell and Mankiw (1987a) estimate the ARMA(1,3) model using a slightly different data set and a slightly different estimation method, the global maximum actually occurs at $\hat{C}(1) = 0$.[3] Consequently, we chose as our second trend stationary mechanism the Campbell-Mankiw ARMA(1,3) model.[4]

To represent the difference stationary perspective, we chose our estimated ARMA(2,2) model. The choice of only one model reflects our need to economize on computational costs. This particular model was chosen because we think it has the best chance of accounting for the evidence in Table 1 that C(1) = 0 or is imprecisely estimated. In addition, focusing on the ARMA(2,2) model has the important advantage of making our results directly comparable to those of Campbell and Mankiw (1987a).

We analyze the relative plausibility of the trend and difference stationary perspectives using two questions: Can the ARMA(3,3) and the trend stationary ARMA(1,3) specifications account for the high likelihood ratio statistic and the high value of $\hat{C}(1)$ obtained using the ARMA(2,2) model? And can the ARMA(2,2) model account for the low likelihood ratio statistics and the value of $\hat{C}(1)$ obtained using the ARMA(1,3) and the ARMA(3,3) models?

Evaluating the Trend Stationary Perspective

To assess the plausibility of the trend stationary perspective, we conducted the following Monte Carlo experiments. We generated 2,000 data sets, each 151 observations long, using our estimated ARMA(3,3) model and the Campbell-Mankiw ARMA(1,3) model.[5] For each realization of 151 observations, we estimated both constrained and unconstrained ARMA(2,2) models and then computed a likelihood ratio statistic to test the null hypothesis that C(1) = 0. The frequency distributions of these likelihood ratio statistics, as well as the distribution of the chi-square statistic with one degree of freedom, are displayed in Figure 2a.

In Table 1, we reported that the ARMA(2,2) model produced a likelihood ratio statistic of 4.356 when we tested the null hypothesis that C(1) = 0. Our Monte Carlo evidence reveals that if the true data-generating mechanism was our ARMA(3,3) model, then a likelihood ratio statistic greater than or equal to 4.356 would actually occur 74 percent of the time. If the true data-generating mechanism was Campbell and

Mankiw's (1987a) ARMA(1,3) model, then this would occur 38 percent of the time. Obviously, these numbers are much larger than the frequency of 4 percent predicted by the conventional chi-square distribution. Indeed, Figure 2a reveals that the likelihood ratio statistic of 4.356 is close to the central tendency of that statistic for ARMA(1,3) and ARMA(3,3) models. Evidently, both of these trend stationary models can easily account for the high likelihood ratio statistic associated with testing the null hypothesis of trend stationarity obtained with the ARMA(2,2) model.

By-products of the preceding Monte Carlo studies are simulated frequency distributions for the values of $\hat{C}(1)$ obtained using the ARMA(2,2) model in data generated by the ARMA(3,3) and ARMA(1,3) models. These frequency distributions are plotted in Figure 2b. The main characteristics of the two distributions are very similar. Both are bimodal. The larger mode is centered about a value of C(1) substantially greater than 1, while the smaller mode is centered about zero. The unconditional means of $\hat{C}(1)$ are 1.43 and 1.23 when the data are generated by the ARMA(3,3) and ARMA(1,3) models, respectively. The corresponding standard errors are 0.39 and 0.57. Comparing these values with the value of $\hat{C}(1)$ produced by the ARMA(2,2) model (1.53), we conclude that both the ARMA(3,3) and the ARMA(1,3) models can easily account for the estimate of C(1) obtained with the ARMA(2,2) model.

Campbell and Mankiw report that, with the ARMA(2,2) model, $\hat{C}(1) = 1.52$, with a standard error of 0.16. While our reported standard errors are larger, this may partly reflect the tendency of standard errors based on the local curvature of the likelihood function to overstate the precision with which C(1) is estimated. We conjecture that standard errors based on local curvature of the likelihood function will on average correspond to the standard deviation of $\hat{C}(1)$ conditional on being in the larger mode of the bimodal frequency distribution. Some support for this conjecture is provided by the fact that the standard errors of $\hat{C}(1)$ conditional on being in the larger mode of Figure 2b are 0.20 and 0.23 when data are generated by the ARMA(3,3) and ARMA(1,3) models, respectively. [The corresponding mean values of $\hat{C}(1)$ are 1.51 and 1.46.]

Specification Error and Near Parameter Redundancy. So far, we have established that the ARMA(2,2) results can easily be accounted for by both of our trend stationary models. At first glance, this may seem surprising, for at least two reasons. First, in both data-generating mechanisms, C(1) = 0 by construction; yet the estimated value of C(1) is typically quite large. Second, it is well known that, when a moving average root is near the unit circle, maximum likelihood parameter estimates have positive mass on the unit circle (Sargan and Bhargava, 1983). This is referred to as the pileup phenomenon. Our data-generating mechanisms have an exact unit moving average root. As Campbell and Mankiw (1987a) emphasize, the pileup phenomenon suggests that, other things equal, conventional sampling theory substantially understates the evidence against the null hypothesis that C(1) = 0. In sharp contrast, our Monte Carlo experiments reveal that the conventional probability value of 4 percent associated with the likelihood ratio test of C(1) = 0 substantially overstates the evidence against the null hypothesis of trend stationarity.

Now we show that two key factors account for the results in Figures 2a and 2b: (1) the specification error arising from the fact that the ARMA(2,2) model is misspecified from the perspective of either the ARMA(1,3) or the ARMA(3,3) model and (2) the near parameter redundancy problem which arises from the fact that both trend stationary models have an autoregressive root near the unit circle. We begin by providing the underlying intuition using large-sample arguments. Then we present the results of a suitably chosen Monte Carlo experiment.

Our large-sample argument is based, in part, on the probability limit (plim) of the ARMA(2,2) model when we assume that the true data-generating process corresponds to the estimated ARMA(3,3) model. This plim was calculated by first simulating a realization of 20,000 observations from the estimated ARMA(3,3) model and then estimating an ARMA(2,2) model on the synthetic data set. The resulting ARMA(2,2) model is given by

$$(1 - 0.6249L + 0.4614L^2)\Delta y_t = (1 - 0.3143L + 0.5934L^2)\eta_t. \qquad (11)$$

The model summarized by (11) implies that C(1) = 1.53, a value very different from zero, the true value of C(1) in the data-generating process. Evidently, the impact of specification error is to substantially bias the estimate of C(1) away from zero.

A striking feature of the previous result is that the plim of $\hat{C}(1)$ actually corresponds (up to three decimal places) to the value of $\hat{C}(1)$ obtained from the ARMA(2,2) model estimated with the actual U.S. data (Table 1). To understand this result, recall the near redundancy of the parameters describing the ARMA(3,3) model. As reported in Table 1, the autoregressive and moving average roots of that model are (0.299 ± 0.565i, 0.949) and (0.133 ± 0.747i, 1), respectively. Since 0.949 is close to 1, the ARMA(2,2) model defined by stripping away both these roots has roughly the same covariance structure as the estimated ARMA(3,3) model. Equation (11) reveals that this is indeed what the probability limit of the misspecified ARMA(2,2) model amounts to. This is because the autoregressive and moving average roots in (11) are (0.312 ± 0.603i) and (0.157 ± 0.754i), respectively. These are very close to the corresponding roots of the ARMA(2,2) model estimated using the postwar U.S. data: (0.293 ± 0.614i) and (0.139 ± 0.776i). Since C(1) is determined entirely by these two roots, the two ARMA(2,2) models generate essentially identical values.

The previous reasoning takes as given that a maximum likelihood estimator of the misspecified ARMA(2,2) model wants to ignore the maximal autoregressive and moving average roots of the ARMA(3,3) model, even though this has the effect of converting the model into one with a large value of C(1). Why should this be so? With a large amount of data, maximum likelihood selects a theoretical spectral density that matches as closely as possible the true spectral density of the data. When the model to be estimated is correctly specified, the estimated spectral density will, in population, coincide with the true spectral density matrix. However, since the ARMA(2,2) model is misspecified, the estimated spectral density cannot match the true spectral density matrix at all frequencies. From this perspective, the relevant question is, Which frequencies will bear the brunt of the specification error?

An implication of results in Christiano and Eichenbaum (1987) and Cochrane (1988a) is that maximum likelihood seeks to minimize the average percentage error of the discrepancy between the theoretical spectral density matrix of the misspecified model and the true spectral density matrix. Let $S_{\Delta y}(e^{-i\omega})$ denote the spectral density of the true model at frequency $\omega \in [0,2\pi]$. The spectral density of the estimated model is $C(e^{-i\omega};\phi,\theta)C(e^{i\omega};\phi,\theta)\sigma_{\eta}^{2}$. Here we have modified our notation slightly in order to explicitly reflect the dependence of C on the autoregressive parameters, ϕ; the moving average parameters, θ; and the innovation variance, σ_{η}^{2}. In population, the maximum likelihood estimator of ϕ and θ minimizes[6]

$$\int_0^{2\pi} S_{\Delta y}(e^{-i\omega})/\left[C(e^{-i\omega};\phi,\theta)C(e^{i\omega};\phi,\theta)\right] d\omega. \tag{12}$$

Notice, first, that when the true data-generating process is trend stationary, $C(1) = 0$. This implies that $S_{\Delta y} = 0$ at frequency zero [equation (9)]. By continuity, $S_{\Delta y}$ will be small in a neighborhood around frequency zero. Consequently, other things equal, the method of maximum likelihood will sacrifice accuracy in a neighborhood of frequency zero in order to achieve a better fit over intervals of higher frequencies. This suggests the possibility that the impact of specification error will fall heavily on the object of interest, C(1). Second, notice that errors over any small band of frequencies do not contribute in an important way to the criterion function which maximum likelihood is minimizing.

To understand the combined impact of these considerations, examine Figure 3a, which displays the spectra of Δy_t implied by the estimated ARMA(3,3) model and the ARMA(2,2) model of equation (11). Two features of the first spectral density are worth noting: (1) There are two regions in which the level of the spectral density undergoes substantial change--the area around frequency zero and the area around the seasonal frequency, 1.5. (2) The first region is smaller than the second because of the very steep slope of the spectrum near zero. This reflects the near parameter redundancy of the ARMA(3,3) model.

Figure 3a shows that the misspecified ARMA(2,2) model described by (11) succeeds in closely mimicking the high-frequency properties of the ARMA(3,3) spectrum. Indeed, the two spectra are virtually identical for values of ω exceeding 1, that is, time periods less than 6.3 quarters. The brunt of specification error is heavily borne by the low frequencies.

To better understand the nature of the trade-offs involved in fitting the misspecified ARMA(2,2) model, consider the following experiment. Suppose that we forced the ARMA(2,2) to match the low-frequency behavior of the ARMA(3,3) model. What would the cost be? To answer this question, we computed the plim of the ARMA(2,2) model subject to the constraint that $C(1) = 0$.[7] The resulting spectrum is shown in Figure 3b, where that of the ARMA(3,3) is repeated, from Figure 3a. As can be seen, the constrained ARMA(2,2) model succeeds in capturing the behavior of the ARMA(3,3) model in the neighborhood of $\omega = 0$. However, to accomplish this, it must set one of the autoregressive roots close to 1. This root and the unit moving average root have a negligible effect on the spectrum at higher frequencies because they cancel each other out at those frequencies. Consequently, the ARMA(2,2) model has only two parameters left to match the relatively complicated seasonal and high-frequency dynamics of the ARMA(3,3) model. The best that the constrained ARMA(2,2) model can do is to ignore the seasonal dip and draw a smoothed version of the ARMA(3,3) spectrum at the higher frequencies. The misspecified ARMA(2,2) model simply does not have enough flexibility to capture the dynamics of the ARMA(3,3) spectrum at both the zero and the seasonal frequencies. Our analysis indicates that when forced to choose which dynamics to mimic, the unconstrained ARMA(2,2) simply gives up on the long-run dynamics.

Specification error will not always result in a biased estimate of C(1). The fact that the ARMA(2,2) model generates a large value of $\hat{C}(1)$ depends critically on the near parameter redundancy in the ARMA(3,3) model. A simple way to see this is to repeat our previous experiment but assume a trend stationary ARMA(3,3) model in which the problem of near parameter redundancy is less severe. Figure 3c displays the spectrum of the ARMA(3,3) model obtained by replacing the largest autoregressive root, 0.949, in our estimated ARMA(3,3) model with 0.500. Comparing Figures 3a and 3c, we see that the primary effect of reducing the maximal autoregressive root is to expand the band about frequency zero in which the level of the spectrum undergoes substantial change. This suggests that maximum likelihood will give greater weight to matching the low-frequency dynamics of the modified ARMA(3,3) model. Indeed, the plim of the ARMA(2,2) model fit to data generated by the modified ARMA(3,3) model turns out to imply a value of $\hat{C}(1)$ precisely equal to zero. Figure 3c also displays the spectrum of this ARMA(2,2) model. Notice that the spectrum exactly coincides with that of the modified ARMA(3,3) model at $\omega = 0$ and matches its low-frequency behavior quite closely. At the same time, it does quite poorly with respect to the high-frequency behavior. As before, the misspecified ARMA(2,2) model does not have enough flexibility to capture the dynamics of the ARMA(3,3) spectrum at both the zero and the seasonal frequencies. But now, when forced to choose which dynamics to mimic, the ARMA(2,2) model chooses to mimic the long-run dynamics. By reducing the severity of the near parameter redundancy problem, we have increased the cost of ignoring the long-run dynamics.

The previous large-sample considerations suggest that, without near parameter redundancy, the trend stationary perspective could not have accounted for the large value of $\hat{C}(1)$ associated with the ARMA(2,2) model. To show that this is indeed true, we repeated the Monte Carlo study of Figure 2a using the modified ARMA(3,3) model which has a maximal autoregressive root of 0.5. When we did this, we found that only 3 percent of the simulated likelihood ratio statistics exceed the value reported in Table 1, 4.356. This stands in sharp contrast to Figure 2a, where 74 percent of the simulated likelihood ratio statistics exceed that value. Taken together, these results establish that both specification error and the near parameter redundancy problem allow the trend stationary models to account for the high and apparently precise estimate of C(1) obtained using the ARMA(2,2) model.[8]

To summarize, from the trend stationary perspective, the high value of $\hat{C}(1)$ associated with the ARMA(2,2) model reflects the choice of maximum likelihood to model the area around the seasonal dip in the spectrum of the data. This dip may reflect the effects of the procedure used by the U.S. Department of Commerce to seasonally adjust the data (Granger and Newbold, 1977, p. 66). This, in turn, raises the possibility that measures of persistence generated from low-order ARMA models could be very sensitive to different seasonal adjustment procedures. Jaeger and Kunst (1989) obtain precisely this result. Our analysis provides a possible explanation for their result.

Reconciling Our Results With Campbell and Mankiw's. Our results contrast sharply with those of Campbell and Mankiw (1987a). Using Monte Carlo methods, they reach the conclusion that trend stationary models cannot account for the results obtained with the ARMA(2,2) model. The reason for the difference is that Campbell and Mankiw's (1987a) Monte Carlo study assumes that the ARMA(2,2) model is correctly specified. In particular, their data-generating mechanism is the model estimated by Blanchard (1981), according to which y_t is a second-order autoregression about a linear trend, with autoregressive roots equal to 0.5 and 0.84. This implies an ARMA(2,1) representation for Δy_t with a moving average root of 1.

Taking as given the estimated Blanchard model, Campbell and Mankiw generated 20 data sets, each 151 observations long. Then, using the ARMA(2,2) model, they computed a likelihood ratio test of the null hypothesis that $C(1) = 0$ for each data set. Their calculations lead to the dramatic result that the likelihood ratio statistic does not exceed 4.356 in any of the 20 artificial data sets.[9] Obviously, under their maintained assumptions, the likelihood ratio test is not biased toward rejecting the null hypothesis that $C(1) = 0$. Indeed, the test rejects the null hypothesis considerably less often than it should. Based on this evidence, Campbell and Mankiw infer that Blanchard's model cannot account for the ARMA(2,2) results.

Campbell and Mankiw (1987a, p. 871) summarize their findings this way: "We conclude from our literature review and our small Monte Carlo study that while there are some statistical difficulties with our estimator, there is no reason to think that these bias us toward rejecting stationarity." This conclusion seems warranted if the ARMA(2,2) model is correctly specified. However, both our results and theirs point to other empirically plausible trend stationary models relative to which the ARMA(2,2) model is misspecified. According to these models, there is reason to think that tests based on the ARMA(2,2) model are severely biased toward rejecting stationarity.

Evaluating the Unit Root Perspective

Now we consider whether the ARMA(2,2) model can account for the key features of Campbell and Mankiw's estimated ARMA(1,3) model as well as our estimated ARMA(3,3) model. The salient characteristic of both of these models is that, according to the likelihood ratio statistic, there is very little evidence against the hypothesis that C(1) = 0. Suppose, in fact, that the true data-generating mechanism is given by our estimated ARMA(2,2) model. What should we expect if we estimate an ARMA(1,3) or an ARMA(3,3) model? Campbell and Mankiw (1987a) conjecture that the high likelihood value associated with the test that C(1) = 0 reflects the pileup phenomenon.

To investigate that conjecture, we performed the following Monte Carlo experiment. Using our estimated ARMA(2,2) model, we generated 2,000 data sets, each 151 observations long. For each data set, we estimated constrained and unconstrained versions of the ARMA(1,3) and ARMA(3,3) models, thus generating 2,000 likelihood ratio statistics for testing the null hypothesis that C(1) = 0. Consistent with Campbell and Mankiw's conjecture, we found that, for the ARMA(1,3) and ARMA(3,3) models, 37 and 47 percent, respectively, of the likelihood ratio statistics are identically zero. Figure 4 displays the frequency distributions of the estimated values of C(1) corresponding to the ARMA(1,3) and ARMA(3,3) models. In both cases, the estimated C(1)'s pile up at zero. This, in turn, corresponds to a pileup of likelihood ratio statistics at zero.

Among the estimated C(1)'s that exceed zero, the vast majority are greater than 1. This suggests that the typical likelihood surface in the Monte Carlo study resembles the surfaces depicted in Figures 1b-1d in two respects. First, there is a local maximum in the region of large values of C(1). Second, the likelihood surface is increasing as C(1) declines toward zero. These results are consistent with the notion that, across synthetic data sets, the likelihood surface tilts back and forth, with the global maximum shifting between extreme values at C(1) = 0 and C(1) > 1.

Overall, the empirical values of $\hat{C}(1)$ implied by the ARMA(1,3) and ARMA(3,3) models are clearly consistent with the bimodal distribution for the simulated $\hat{C}(1)$'s when the data-generating mechanism is the ARMA(2,2) model. We conclude that the ARMA(2,2) model can account for the salient characteristics of the estimated ARMA(1,3) and ARMA(3,3) models.

Summary

We have argued that the parametric methods of Campbell and Mankiw (1987a) do not provide a basis for taking a strong position on whether shocks to real GNP are best characterized as having temporary or permanent effects. Perhaps the best way to conclude this section is to consider Figure 5, which displays the impulse response functions of real GNP implied by a subset of the ARMA models that we estimated. Included are the impulse response functions implied by the Blanchard ARMA(2,1) model estimated using his data up to 1980 (old Blanchard), the updated version of that model estimated using our larger data set (new Blanchard), as well as our estimated ARMA(2,2) and ARMA(3,3) models.[10] Notice that all of the impulse response functions have very similar shapes for the first 5-10 quarters. Only after this is the impulse response function of the difference stationary ARMA(2,2) model radically different in shape from that of the trend stationary models. To sharply differentiate among these models would require reasonably precise information about the higher-order autocorrelations. Needless to say, these are not estimated very precisely with postwar U.S. real GNP data.

Campbell and Mankiw concluded that a 1 percent innovation in real GNP ought to induce a revision in the long-run forecast of real GNP of more than 1 percent.

Suppose by the long run we mean anything more than four years. Figure 5 indicates that this conclusion is supported only by the ARMA(2,2) model. None of the trend stationary models support it, and one of these models, the ARMA(3,3), is at least as plausible as the ARMA(2,2).

Figure 5 also reveals that all of the ARMA models we investigated have impulse response functions above the old Blanchard model. Suppose we accept Blanchard's (1981, p. 150) assertion that this model summarizes macroeconomists' views in 1980 about the nature of the dynamics in real GNP. On this premise, it seems fair to conclude that macroeconomists must now revise upward their point estimate of the half-life of an innovation in real GNP. This is true regardless of whether they take a trend or a difference stationary perspective. Less obvious is the idea that the increased point estimate has any economic significance. We know of no interesting case in which the differences in persistence among the three trend stationary models in Figure 5 are important.[11] Later we will discuss whether there are interesting economic issues at stake in adopting a unit root perspective.

Nonparametric Measures of Persistence

Above we argued that one cannot distinguish between the competing null hypotheses that postwar U.S. real GNP is trend stationary or difference stationary using the parametric methods of Campbell and Mankiw (1987a). A natural response to this problem is to examine the persistence of real GNP using the nonparametric methods of Cochrane (1988a).

Table 2 reports Cochrane's variance ratio statistic $\hat{V}^k$ for the values of k used by Campbell and Mankiw (1987a). The numbers in parentheses are asymptotic standard errors computed using the formula of Priestly (1982, p. 463): s.e. $(\hat{V}^k) = V^k\{(3/4)[T/(k+1)]\}^{1/2}$. To make our results comparable with those above, we also report nonparametric estimates of C(1), obtained from $\hat{V}^k$, using a transformation discussed in Campbell and Mankiw (1987a). Let $R^2 = 1 - \sigma_\eta^2/\sigma_{\Delta y}^2$. Then (8) can be written as

$$C(1) = \left[V/(1-R^2)\right]^{1/2}. \tag{13}$$

Here R^2 is the fraction of the variance in Δy_t that is predictable using all lagged values of Δy_t. Let ρ denote the first-order autocorrelation of Δy_t. Campbell and Mankiw's (1987a) nonparametric estimator of C(1), $\hat{C}(1)^k$, is defined by

$$\hat{C}(1)^k = \left[\hat{V}^k/(1-\hat{\rho}^2)\right]^{1/2} \tag{14}$$

where $\hat{\rho}$ is the sample estimate of the first-order autocorrelation of Δy_t.

In practice, to calculate Cochrane's variance ratio statistic, we must choose a value of k. As we stressed earlier, the key identifying assumption underlying this measure of persistence is the assumption that, whatever value of k is chosen, the higher autocorrelations are of negligible importance. This suggests that k should not be chosen too small. To see this, notice that $\hat{V}^1 = 1 + \rho_1$. As long as Δy_t is positively autocorrelated, $\hat{V}^1$ will exceed 1 even if the process is trend stationary. But when $k = T - 1$, $\hat{V}^k = 0$ by construction. Clearly, k should not be chosen too large relative to the sample size, T. Table 2 reports the values of $\hat{V}^k$ and $\hat{C}(1)^k$ for a variety of values of k. Comparing Tables 1 and 2, we see that, roughly, the nonparametric estimates of C(1) are lower than the parametric estimates of C(1). But the reported standard errors, calculated using the Priestly formula, are all quite large relative to the point estimates.

According to Table 2, distinguishing between the classes of trend stationary and difference stationary models using Cochrane's nonparametric measure of persistence is difficult. Unfortunately, we cannot formally test the null hypothesis of trend stationarity since, under that null hypothesis, Priestly's formula for the standard error of $\hat{V}^k$ equals zero for all k. We can, however, ask whether the representative models discussed in the last section are consistent with the computed $\hat{V}^k$'s.

To investigate this question, we performed the following Monte Carlo experiment. We considered three data-generating mechanisms: the Campbell and Mankiw ARMA(1,3) model and our estimated ARMA(2,2) and ARMA(3,3) models. For each of these, we generated 2,000 data sets, each with 151 observations. We then calculated $\hat{V}^k$ for $k = 1, 2, \ldots, 75$ using each of the data sets. The results of our experiment are reported in Table 2. To understand these numbers, consider a particular column, say, the one labeled ARMA(1,3). Any given row in that column corresponds to a particular value of k. The corresponding entry in the row is the fraction of times (out of 2,000) that the k-lag variance ratio statistic calculated from the simulated ARMA(1,3) data exceeds the corresponding empirical value of $\hat{V}^k$ reported in column (3). The numbers in the columns labeled ARMA(2,2) and ARMA(3,3) are constructed in an analogous way. Notice that the probability values in columns (4)-(6) lie between 0.18 and 0.73, so that the empirical $\hat{V}^k$'s in column (3) can be accounted for by each of our three ARMA models. We conclude that, even from this limited perspective, the $\hat{V}^k$'s do not let us discriminate between trend and difference stationary representations of the data.

Examining the same empirical variance ratio statistics, Campbell and Mankiw (1987a, p. 875) conclude that "the nonparametric estimates thus confirm our finding that postwar quarterly real GNP appears to be more persistent than a random walk." The reason Campbell and Mankiw give for reaching this conclusion is that "$\hat{V}^k$ for the real GNP data are consistently . . . larger than one would expect to find for a random walk in a sample of this size" (pp. 874-75). This can be seen in Figure 6, which plots the mean values of the $\hat{V}^k$'s implied by the random walk model, together with the $\hat{V}^k$'s calculated using the data. What Campbell and Mankiw's reasoning ignores is that a large class of trend stationary models, including those which they consider, imply mean values for the $\hat{V}^k$'s which closely mimic their empirical counterparts. This also can be seen in Figure 6, which displays the mean values of $\hat{V}^k$ implied by our estimated ARMA(3,3) model.

A different way to state our objection to Campbell and Mankiw's argument is that it implicitly assumes that the direction and magnitude of the bias in $\hat{V}^k$ is relatively insensitive to the underlying data-generating mechanism. Unfortunately, this assumption is not true. Suppose that the underlying data-generating process is a random walk. Campbell and Mankiw (1987a), among others, stress that, in this case, the variance ratio statistic is severely downward biased. In contrast, suppose the true generating mechanism is our estimated ARMA(3,3) process, so that V = 0. According to Figure 6, the $\hat{V}^k$'s are consistently much larger than zero, even for k = 75. We infer that, at least for this data-generating mechanism, the variance ratio statistic is very substantially upward biased.

We conclude that, given the sample size of our data set on real GNP, the variance ratio statistic gives us almost no reliable information. Our nonparametric estimates of V and C(1) are no doubt consistent with the view that postwar U.S. real GNP is more persistent than a random walk. But they are at least as consistent with the view that postwar U.S. real GNP is less persistent than a random walk. Which view is true? We can't tell.

UNIT ROOTS IN REAL GNP: DO WE CARE?

Suppose we knew the answer to the question, How much should one revise a long-run forecast of aggregate output in response to an innovation in U.S. real GNP? Would we care? At one level, the answer is obvious: unit roots per se cannot be very important. The existence of a unit root means only that $C(1) \neq 0$; that does not preclude a value of C(1) arbitrarily close to zero. We do not know of any model in which agents' decision rules are discontinuous in C(1). Therefore, it seems likely that for any trend stationary specification of the forcing variables in agents' environments, some difference stationary specification will imply arbitrarily similar dynamic behavior.

In practice, however, this is not the perspective of concern to economists. Typically, the analyst's problem is not one of selecting between different specifications with arbitrarily similar values for C(1). Usually, the decision to model a time series as difference or trend stationary leads the analyst to adopt specifications with very different implications for C(1). For example, in Deaton's (1986) analysis of the PIH, the difference and trend stationary specifications for measured labor income imply values of C(1) near 1.5 and zero, respectively. Similarly, in Hansen's (1989) analysis of RBC models, the difference and trend stationary specifications for technology shocks imply values of C(1) equal to 1 and zero, respectively. In both cases, the dynamic properties of the endogenous variables behave very differently depending on which specification is chosen.

Here we argue that the dramatic results obtained by Deaton and Hansen do not reflect model sensitivity to unit roots per se or even the value of univariate measures of persistence like C(1). Rather, they reflect the assumption that the forcing variables of concern to agents are driven by a single shock. Under these circumstances, the assumption of difference stationarity implies that all of the shocks to agents' environments have purely permanent effects. Once the unit root issue is decoupled from the temporary/permanent issue, the unit root issue loses much of its quantitative significance.

Unit Roots and the Permanent Income Hypothesis

Because of its simplicity, the PIH is a convenient vehicle for illustrating both why unit roots seem to matter and why they may not matter after all. We use the PIH for illustrative purposes only. Unobserved component models of labor income will not remedy the empirical shortcomings of the PIH (West, 1988a). More generally, this type of modification to the basic model cannot account for the fact that the orthogonality conditions implied by the PIH are violated by the data (Flavin, 1981; Campbell and Deaton, 1988; Campbell and Mankiw, 1989; and Christiano, Eichenbaum, and Marshall, 1989).

According to the PIH, the level of consumption depends on both asset and labor income. However, given a constant real interest rate, the only thing that induces households to set date t consumption, c_t, to a value different from c_{t-1} is news about current or expected future labor income, y^{ℓ}_t. When such news arrives, households adjust consumption by the annuity value of the resulting revision to expectations about y^{ℓ}_{t+s} for s = 0, 1, ... This annuity value, computed using the constant interest rate r, is the change in consumption that can feasibly be maintained indefinitely in an expected value sense. Formally, the PIH posits that

$$\Delta c_t = \frac{1}{1+r} \sum_{i=0}^{\infty} \left(\frac{1}{1+r}\right)^i \{E_t y_{t+i}^{\ell} - E_{t-1} y_{t+i}^{\ell}\}. \tag{15}$$

Hall (1978) describes a partial equilibrium consumer optimization problem with a solution which implies (15), while Christiano (1987), Hansen (1987), Sargent (1987), and Christiano, Eichenbaum, and Marshall (1989) discuss general equilibrium environments which rationalize relation (15).

We now consider the case that has been made for the view that unit roots matter. Suppose that agents only see y_t^{ℓ} and that its univariate time series representation is $\Delta y_t^{\ell} = C(L)\varepsilon_t$. Here ε_t is the innovation in y_t^{ℓ}; that is, $\varepsilon_t = y_t^{\ell} - E_{t-1} y_t^{\ell}$. To see how sensitive Δc_t is to C(L), suppose that y_t^{ℓ} is a first-order autoregression about a trend with autoregressive parameter ϕ, so that $C(L) = (1-L)/(1-\phi L)$. When $\phi = 1$, y_t^{ℓ} is simply a random walk [C(L) = 1]. Notice that C(1) drops discontinuously to zero for values of ϕ less than 1. With this general specification of C(L), equation (15) implies that

$$\Delta c_t = C\left(\frac{1}{1+r}\right)\varepsilon_t = \frac{r}{1+r-\phi}\,\varepsilon_t. \tag{16}$$

One way of measuring the sensitivity of the model's implications to different specifications of ϕ is to examine the relative volatility of consumption, $V_{\Delta c}$. Deaton (1986) and others define $V_{\Delta c}$ as the ratio of the standard deviation of changes in consumption, $[E(\Delta c_t)^2]^{1/2}$, to the standard deviation of the univariate innovation in y_t^{ℓ}, σ_ε. For simplicity, suppose r = 0.01. Then, according to (16), $V_{\Delta c} = 1$ when $\phi = 1$. However, $V_{\Delta c} = 0.5$ when $\phi = 0.99$. Obviously, the impact on consumption volatility of a change in ϕ is very large for values of ϕ in a neighborhood of 1. This sensitivity reflects the fact that households place substantial weight on expected income in the distant future. And this is precisely where small differences in ϕ in a neighborhood of 1 have large effects.[12]

At first glance, then, this example seems to provide powerful motivation for the view that economists should care about unit roots. However, we think the example is misleading. To demonstrate why, we use an argument in Quah (1989) which draws out the implications of the well-known fact that there exists an infinite number of orthogonal decompositions of difference stationary processes into persistent and transient components. Let y_{1t} and y_{0t} denote the time t values of two orthogonal difference and trend stationary stochastic processes, respectively, which constitute such a decomposition, so that

$$y_t^{\ell} = y_{1t} + y_{0t}. \tag{17}$$

Assume that the time series representations of y_{1t} and y_{0t} are given by

$$\Delta y_{1t} = C_1(L)\varepsilon_{1t} \tag{18}$$

$$\Delta y_{0t} = (1-L)C_0(L)\varepsilon_{0t} \tag{19}$$

where ε_{it} is the white noise innovation to y_{it} for $i = 0, 1$. Also, ε_{0t} and ε_{1t} are orthogonal at all leads and lags. Consistent with its definition as the trend stationary component of y_t^{ℓ}, the sum of the coefficients in the moving average representation for Δy_{0t} is zero.

For illustrative purposes, consider the random walk case, $\Delta y_t^{\ell} = \varepsilon_t$. One class of orthogonal decompositions of this process is given by

$$C_1(L) = 1 + \psi L \tag{20}$$

$$C_0(L) = 1 \tag{21}$$

$$\sigma_{\varepsilon_1}^2 \equiv E\varepsilon_{1t}^2 = (1+\psi)^{-2}\sigma_{\varepsilon}^2 \tag{22}$$

$$\sigma_{\varepsilon_0}^2 \equiv E\varepsilon_{0t}^2 = \psi E\varepsilon_{1t}^2 \tag{23}$$

for all $0 \leq \psi \leq 1$. To prove that (20)-(23) is a valid decomposition of y_t^ℓ, we need only verify that $E(\Delta y_t^\ell) = \sigma_\varepsilon^2$ and $E(\Delta y_t^\ell \Delta y_{t-s}^\ell) = 0$ for $s \geq 1$.

Under the assumption that households observe y_{1t} and y_{0t} separately, relation (15) implies that Δc_t evolves according to

$$\Delta c_t = C_1\left(\frac{1}{1+r}\right)\varepsilon_{1t} + \frac{r}{1+r} C_0\left(\frac{1}{1+r}\right)\varepsilon_{0t}. \tag{24}$$

Thus, the time t change in consumption equals the annuity value of the innovation to the permanent component of labor income, ε_{1t}, plus the annuity value of the innovation to the temporary component, ε_{0t}. Substituting (20)-(23) into (24) we obtain

$$\Delta c_t = \left(1 + \psi/(1+r)\right)\varepsilon_{1t} + \frac{r}{1+r}\varepsilon_{0t}. \tag{25}$$

Therefore, the relative volatility of consumption is given by

$$V_{\Delta c} = \left[\left(\frac{1 + \psi/(1+r)}{1 + \psi}\right)^2 + \left(\frac{r/(1+r)}{1 + \psi}\right)^2 \psi\right]^{1/2}. \tag{26}$$

This expression is minimized for $\psi = 1$, in which case $V_{\Delta c} = 0.995$. The assumption that agents react to this particular orthogonal decomposition of y_t^ℓ results in only a trivial reduction in consumption volatility relative to the case in which agents only observe y_t^ℓ. Specifications of $C_0(L)$ and $C_1(L)$ that reduce $V_{\Delta c}$ to the empirically plausible value of 0.5 are described below and by Quah (1989). However, all the intuition for understanding how Quah's examples work is contained in our example.

There are at least two ways to understand why $V_{\Delta c}$ is less than 1 when ψ exceeds zero. First, note from (20)-(23) that, as ψ increases, the variance of innovations to the permanent component falls relative to the variance of the innovations to the temporary component, so that, loosely speaking, an increasing proportion of news is about the stationary component of labor income. This is important for determining the relative volatility of consumption because the response of consumption to an innovation in the temporary component of y_t^ℓ is much smaller than the corresponding response to an innovation in the permanent component.

Second, consider the response of y_t^ℓ to a one standard deviation increase in ϵ_{1t}, that is, $\sigma_\varepsilon/(1+\psi)$. The dynamic response of Δy_{t+s}^ℓ to such an impulse is given by $\sigma_\varepsilon/(1+\psi)$, $\psi\sigma_\varepsilon/(1+\psi)$ for $s = 0, 1$, respectively, and zero for $s > 1$. The corresponding response of y_{t+s}^ℓ is $\sigma_\varepsilon/(1+\psi)$ for $s = 0$ and σ_ε for $s > 0$. Notice that, irrespective of the value of ψ, the long-run response of y_t^ℓ to a one standard deviation impulse in the permanent component is $C(1)\sigma_\varepsilon$, which equals σ_ε in our example.[13] While the eventual impact of a typical permanent shock is invariant to ψ, the path by which one gets there is not. With $\psi = 0$ (the no-component case), the response of y_{t+s}^ℓ is equal to σ_ε for all $s \geq 0$, so that the long-run impact on labor income of a typical innovation is realized immediately. In contrast, with $\psi > 0$, the long-run impact on y_t^ℓ is not realized until one period later. Since $r > 0$, the present value of a permanent standardized innovation to y_{1t} is decreasing in ψ. Therefore, the response of consumption to such an innovation is also decreasing in ψ. Consistent with this intuition, $V_{\Delta c}$ is invariant to ψ when $r = 0$.

To pursue this line of reasoning, we consider the following class of decompositions for y_t^ℓ:

$$C_1(L) = (1-\rho L)^{-d} \tag{27}$$

for $d = 1, 2, 3$. In addition, we set $\rho = 0.98$ and $\sigma_\varepsilon = 1$. The remaining elements of the decomposition--σ_{ϵ_1}, σ_{ϵ_0}, and $C_0(L)$--are determined by the requirement that $y_{1t} + y_{0t}$ is a random walk with innovation variance σ_ε^2. The scalar σ_{ϵ_1} is determined by the requirement that the long-run impact of an innovation in y_{1t} of magnitude σ_{ϵ_1} must equal σ_ε. This condition requires that σ_{ϵ_1} equal $(1-\rho)^d\sigma_\varepsilon$. Since $|\rho| < 1$, σ_{ϵ_1} is a decreasing function of d. This in turn suggests that $V_{\Delta c}$ ought to be decreasing in d. The remaining elements of the unobserved components model are obtained using the methods described by Quah (1989). These are given by

$$\underline{d = 1}: \quad C_0(L) = 1/(1-\rho L) \tag{28}$$

$$\sigma_{\epsilon_0}^2 = \rho\sigma_\varepsilon^2 \tag{29}$$

$$\underline{d = 2}: \quad C_0(L) = (1-aL)/(1-\rho L)^2 \tag{30}$$

$$\sigma^2_{\varepsilon_0} = \rho^2\sigma^2_{\varepsilon}/a \tag{31}$$

where a satisfies $a^2 - \gamma a + 1$, $|a| \leq 1$, and $\gamma = 2[1+(1-\rho)^2/\rho]$, and

$$\underline{d = 3}: \quad C_0(L) = (1-a_1L)(1-a_2L)/(1-\rho L)^3 \tag{32}$$

$$\sigma^2_{\varepsilon_0} = \rho^3\sigma^2_{\varepsilon}/(a_1a_2) \tag{33}$$

where $|a_i| \leq 1$ for i = 1, 2.[14]

Figure 7 plots the first 500 coefficients in the polynomial in L, $\sigma_{\varepsilon_1}C_1(L)/(1-L)$ for d = 1, 2, 3. Each curve represents the impulse response of y^{ℓ}_t to a one standard deviation innovation in ε_{1t}. Consistent with our previous example, the long-run response of a typical shock is invariant to d. This can be seen in Figure 7 by noting that all of the impulse response functions converge to 1, the long-run impact of a standardized innovation in the no-component version of the model (d = 0). At the same time, the value to households of a one standard deviation shock to ε_{1t} is not invariant to d. This is because forward-looking agents care about the intermediate-term impact of permanent shocks. Since those are a decreasing function of d, the annuity value of a standardized innovation in ε_{1t} is decreasing in d. This annuity value equals 1, 0.82, 0.45, and 0.31 in the d = 0, 1, 2, and 3 decompositions, respectively.

These arguments do not imply that $V_{\Delta c}$ necessarily falls as d increases. This is because consumption also adjusts in response to ε_{0t}. Not surprisingly, as d increases and permanent shocks become less important, temporary shocks become more important. The annuity value of a standardized innovation in ε_{0t} is 0, 0.330, 0.422, and 0.478 in the d = 0, 1, 2, and 3 decompositions, respectively. However, the increasing contribution of temporary shocks to $V_{\Delta c}$ is smaller than the reduced impact of permanent shocks. We derive the relative volatility of consumption, $V_{\Delta c}$, from these annuity values by squaring them and taking the square root of the resulting sum. Doing so, we find that $V_{\Delta c}$ equals 1, 0.88, 0.62, and 0.57 in the d = 0, 1, 2, and 3 decompositions. Interestingly, when d = 3, $V_{\Delta c}$ is very close to the empirically plausible value of 0.5.

To summarize our findings, consider three of the models presented above: the ϕ = 1, no-components model; the ϕ = 1, components model with d = 3; and the ϕ = 0.99, no-components model. The implications of the first model for the behavior of consumption are very different from those of the other two models. At the same time, the implications of those other models for consumption dynamics are very similar. The key feature that distinguishes the first model from the others is the absence of temporary shocks.

This suggests that what is important for the dynamics of consumption is not the value of ϕ per se. Rather, it is the relative importance of temporary and permanent shocks. A tight link between ϕ and the dynamics of consumption exists only under the strong assumption that agents do not see and respond to different components of labor income. Without this assumption, the assertion that the exogenous driving variables faced by agents contain a unit root--or that $C(1)\sigma_\epsilon$ has a particular value--does not have important implications for the dynamics of consumption.

A key feature of the two-component model of labor income is that agents' information sets are larger than the econometrician's. We can build on this fact to reinforce the intuition about the driving force underlying our results.[15] Consider the extreme example where economic agents actually know the entire future path of their labor income. The econometrician does not. From the perspective of agents, the innovation variance of labor income equals zero, so that the change in consumption always equals zero. This would be true even if there were a large innovation variance and a unit root in the univariate labor income process. The examples we have discussed above can be viewed as less extreme illustrations of this point.

Unit Roots and the Real Business Cycle Model

Now we use the results above to analyze a second example which, according to Hansen (1989), suggests that model dynamics appear to be sensitive to unit roots. The model he considers is one in which a representative agent chooses consumption c_t, capital k_{t+1}, and hours worked n_t to maximize $\sum_{t=0}^{\infty}\{\ln(c_t) + \gamma \ln(T-n_t)\}$ subject to the resource constraint $c_t + k_{t+1} - (1-\delta)k_t = y_t$. Here y_t, time t gross output, is produced according to the Cobb-Douglas production function, $y_t = (z_t n_t)^{(1-\theta)} k_t^{\theta}$. The random variable z_t is a technology shock that satisfies

$$\Delta\log(z_t) = \lambda + C(L)\varepsilon_t \tag{34}$$

where $C(L) = (1-L)/(1-\phi L)$.

Hansen's results indicate that the volatility of hours worked, n_t, relative to productivity, y_t/n_t, is very sensitive to values of ϕ near 1. The basic intuition behind this result can be described as follows. Given our production function, the marginal productivity of labor is proportional to average productivity, y_t/n_t. Other things equal, both are an increasing function of z_t. When ϕ is positive and less than 1, a positive innovation in z_t is associated with a smaller upward revision in the outlook for z_{t+1}. Under these circumstances, the returns from working at time t are unusually high, thus triggering a strong intertemporal substitution effect on n_t. When $\phi = 1$, the outlook for future z_t moves one-for-one with innovations in z_t, since $E_t z_{t+i} = z_t$ for all i, t > 0. Not surprisingly, here agents have less incentive to intertemporally substitute labor over time.

In analyzing this example, we consider two measures of the volatility in hours worked, σ_n. One measure is the standard deviation of $\Delta\log(n_t)$, while the other is the standard deviation of $\log(n_t)$, after the Hodrick and Prescott filter has been applied (Prescott, 1986). Similarly, we have two measures of the volatility of productivity, $\sigma_{y/n}$. One is the standard deviation of $\Delta\log(y_t/n_t)$; the other, the standard deviation of $\log(y_t/n_t)$, after the Hodrick-Prescott filter has been applied.

All model parameter values--aside from those pertaining to $\Delta\log(z_t)$--coincide with those used in Christiano and Eichenbaum (1988b).[16] The method used to approximate the solution to the model is also described there. An important feature of the solution is that it implies a linear bivariate time series representation for $\log(n_t)$ and $\log(y_t/n_t)$. In addition, both $\log(n_t)$ and $\log(y_t/n_t)$ depend partly on the present discounted value of expected future values of $\log(z_t)$. This allows us to use the intuition developed earlier.

Table 3 reports results based on the first-difference filter.[17] Comparing columns (1) and (2), we see that $\sigma_n/\sigma_{y/n}$ rises more than 40 percent when ϕ drops from 1 to 0.99. Clearly, this is due primarily to an increase in the volatility of n_t. As ϕ drops below 1, fluctuations in z_t go from being 100 percent permanent to being 100 percent temporary. As indicated above, employment responds more to temporary than to permanent shocks because agents intertemporally substitute hours worked toward periods in which the returns to working are relatively high.

The sharp difference between these models reflects the maintained assumption that there is only one source of shocks to agents' environments. The analysis above suggests that if we abandon this assumption and adopt the d = 3 components representation of z_t given by (32)-(33), then the volatility of employment should rise toward the value implied by the trend stationary model (ϕ = 0.99). This is because a substantial component of the shocks to the agents' environments will then be transitory. And these are the types of shocks that induce large changes in labor supply.

The results of this experiment are reported in column (3) of Table 3. Notice that the relative volatility of hours is now roughly equal to the value which emerges from the ϕ = 0.99 model. Also, the other moments of the unobserved components model match the corresponding moments of the ϕ = 0.99 model reasonably well. Table 4 reports results for the same three model economies but for which the Hodrick-Prescott filter has been used to induce stationarity. The same general pattern observed in Table 3 emerges in Table 4.

These calculations roughly confirm the findings in our analysis of the PIH. However, there are at least two respects in which a more complete analysis is required before firm conclusions can be reached. First, we have only studied a small subset of the second-moment properties of the model. The $\phi = 0.99$ and the difference stationary components models may differ substantially in other dimensions. Second, we wonder whether a components representation of z_t can mimic a trend stationary RBC model with substantially lower values of ϕ.

CONCLUSION

In this paper, we have argued that macroeconomists should not take strong positions on whether postwar U.S. real GNP is trend or difference stationary. As we emphasized in the introduction, there are strong a priori reasons for being suspicious of claims in favor of difference or trend stationary representations of real GNP. A simple way to see this is to consider parametric ARMA representations of the first-differenced data. Blough (1988) and Cochrane (1988b) have pointed out that every trend stationary ARMA model has a difference stationary ARMA model local to it, and vice versa. Distinguishing between these on the basis of a finite data set is surely an impossible task. This a priori line of reasoning can be used to dismiss Campbell and Mankiw's (1987a) rejection of the hypothesis of trend stationarity in favor of the difference stationary ARMA(2,2) model. To do so, one need only consider a trend stationary ARMA(3,3) model with autoregressive and moving average roots identical to those of their ARMA(2,2) model plus an autoregressive root of $1 - \varepsilon$ and a moving average root of 1. For $\varepsilon > 0$ but sufficiently small, it must be true that there is no detectable difference between the competing models.

A similar line of a priori reasoning could be used to dismiss almost any argument in favor of a given difference stationary model of real GNP. One need only select a trend stationary model that is arbitrarily close to it. But suppose this were the only way to salvage the trend stationary perspective. At best, this would be a Pyrrhic victory for that perspective since, for all practical purposes, the selected trend stationary model would coincide with the given difference stationary model. This is because the two models' impulse-response functions would be virtually identical at all but infinite horizons. An analogous set of observations applies to arguments in favor of a given trend stationary model of real GNP. In this paper, we go beyond the a priori line of reasoning by showing that one cannot distinguish between difference and trend stationary models with impulse response functions which are substantially different at horizons even as short as three years. On this issue, the a priori line of reasoning summarized above is moot.

It is useful to contrast our results with those in the literature on the power properties of stationarity tests. A variety of authors have concluded that existing tests of whether time series are difference stationary or trend stationary have extremely poor power properties (for example, DeJong, Nankervis, Savin, and Whiteman, 1988). Power issues are of interest in the context of tests of the unit root null hypothesis because this hypothesis is typically not rejected for postwar U.S. real GNP data. However, power issues are obviously of less interest when the null hypothesis is rejected. This is precisely the relevant case in the context of testing the trend stationary null hypothesis. The major result in the literature is the strong rejection of trend stationarity for postwar U.S. real GNP (Campbell and Mankiw, 1987a).

The principal focus of the first part of our paper is on this rejection. From this perspective, the issue of interest is the size of Campbell and Mankiw's (1987a) test, that is, the probability of rejecting the null hypothesis if the data-generating mechanism is in fact trend stationary. The size characteristics of their test are excellent if the analyst specifies the correct ARMA representation of the data. However, we show that their tests give extremely misleading results if that ARMA representation is misspecified in seemingly innocuous ways. This result complements those of DeJong, Nankervis, Savin, and Whiteman (1988), who show that specification error in the form of unmodeled residual correlation can lead to excessive rejection of the trend stationary null hypothesis.

Viewed as a whole, the results here are consistent with the view that one cannot discriminate, on the basis of postwar data, between the null hypotheses of trend and difference stationarity for U.S. real GNP. At the same time, economic theory offers no guidance on this question. As Sims (1988) has emphasized, in linear models, the trend behavior of endogenous variables is almost always determined by the analyst's assumptions about the trend behavior of the unobservable forcing variables in agents' environments. It does not emerge from some principle of economic theory. True, one can construct endogenous growth models in which the prediction of difference stationarity emerges from the production of human capital. (See, for example, King, Plosser, and Rebelo, 1988; King and Rebelo, 1986; and Christiano and Eichenbaum, 1988a.) Unfortunately, though, these implications depend very sensitively on particular functional form assumptions about which economic theory has little to say.

Should we despair at not knowing? We think not. Our results suggest that the implications of a broad class of dynamic models are reasonably robust to whether the forcing variables in agents' environments are modeled as trend or difference stationary. Existing examples which purport to document extreme sensitivity actually demonstrate sensitivity to the extreme assumptions that all shocks are either temporary or permanent. We think macroeconomists should care very much about the relative importance of permanent and temporary shocks to agents' environments. But conventional atheoretical measures of persistence convey little information about this question. And structural inferences based on such measures ought to be viewed with extreme skepticism. Convincing inference requires the use of economic theory in conjunction with the data.

NOTES

[1]To see this, suppose that the model being estimated is an ARMA(p,2) with $p > 0$. Denote the moving average roots by λ_1 and λ_2. For simplicity, assume these are real. Let ϕ denote the parameters of the p^{th}-order autoregressive component. It is well known that the exact likelihood function obtained after concentrating out the innovation variance, $L(\phi,\lambda_1,\lambda_2)$, has the property that $L(\phi,\lambda_1,\lambda_2) = L(\phi,\lambda_1,1/\lambda_2)$. Therefore, $L_3(\phi,\lambda_1,1) = 0$, where L_3 denotes the partial derivative of L with respect to its third argument. We can express L as a function of $C(1) = k$ by substituting out for λ_2 in terms of k: $L[\phi,\lambda_1,1-k\phi(1)/(1-\lambda_1)]$. Then the partial derivative of L with respect to k is $-L_3[\phi,\lambda_1,1-k\phi(1)/(1-\lambda_1)]\phi(1)/(1-\lambda_1)$, which equals zero for $k = 0$. Let $\mathcal{L}(k) = L[\phi,\lambda_1,1-k\phi(1)/(1-\lambda_1)]$, after maximizing out ϕ and λ_1. A simple envelope argument establishes that the derivative of $\mathcal{L}$ with respect to k is also zero.

[2]Essentially, the Christiano-Ljungqvist method is a model selection strategy. Alternative strategies are the sequential likelihood ratio tests and Akaike (1974) or Schwartz (1978) criteria. These model selection procedures may be quite useful for choosing among forecasting models when there is a clear gain to parameter parsimony. However, they may not be appropriate for our purposes. The Christiano-Ljungqvist procedure is closely related to methods for testing nonnested models, the encompassing principle discussed by Mizon and Richard (1986), as well as the selection criterion used by Sargent (1976) and Christiano and Eichenbaum (1987).

[3]In their analysis, Campbell and Mankiw (1987a) include data from 1947 and estimate their models using a Kalman-filtering algorithm.

[4]The log likelihood value associated with these parameter values is only 0.015 below the global maximum.

[5]All shocks were drawn from a normal distribution with mean zero and standard deviation 0.0100847.

[6]This formula can be obtained as follows. First replace the Gaussian likelihood function by its frequency domain approximation [for example, equation (45) in Christiano and Eichenbaum, 1987]. Then replace the periodogram in that formula by the spectral density of the true model. The formula in the text is obtained by concentrating out the innovation variance from the latter and by driving the number of observations to infinity. Christiano and Eichenbaum (1987) use the unconcentrated, multivariate version of this formula to analyze the large-sample consequences of maximum likelihood estimation of a misspecified model. The formula in the text is also an implication of equation (A4) in Cochrane (1988a). Both our derivation and Cochrane's assume that $C(1) \neq 0$. The following computational experiment makes us somewhat con-

fident that the formula also holds when C(1) = 0. We computed the plim of the misspecified ARMA(2,2) model in two ways. One corresponds to the method described in the text. The other is a discrete analog of the formula in the text which employs 1,000 equispaced points around the unit circle. Both methods yield virtually identical results.

[7] This model is $(1 - 1.5075L + 0.5296L^2)\Delta y_t = (1 - 1.1930L + 0.1930L^2)\eta_t$.

[8] We repeated all the experiments discussed in this section with the Campbell-Mankiw ARMA(1,3) model as the data-generating process. With respect to the asymptotic arguments, our results were virtually identical. The results of redoing the last Monte Carlo discussed in the text were similar, though less dramatic. The effect of reducing the autoregressive parameter from 0.95 to 0.5 is to reduce the percentage of likelihood ratio statistics exceeding 4.356 from 38 to 22.

[9] In fact, 14 of the 20 likelihood ratio statistics are exactly equal to zero, while the largest only equals 1.77.

[10] We obtained the following point estimates for the ARMA(2,1) model with C(1) = 0:

$$\Delta y_t = \underset{(0.074)}{1.357}\Delta y_{t-1} - \underset{(0.075)}{0.393}\Delta y_{t-1} + \eta_t - \eta_{t-1}.$$

Numbers in parentheses are standard errors obtained by taking minus the inverse of the second derivative of the log likelihood function.

[11] Once parameter uncertainty is taken into account, these differences may not even be statistically significant. For example, consider the old and new Blanchard models. Point estimates and associated standard errors for the latter are reported in note 10. They show that the lag 1 and 2 coefficients in the old Blanchard model are 0.22 and 0.36 standard errors, respectively, away from the corresponding coefficients in the new Blanchard model.

[12] For example, in (15), with a quarterly interest rate of 1 percent, changes in anticipated income even 40 years in the future receive a nonnegligible weight of $1.01^{-160} = 0.2$. See Christiano (1987) for an extended discussion of the view that unit roots may matter in the context of the PIH.

[13] This is a general property of temporary/permanent decompositions, not just orthogonal decompositions. For a proof, see Cochrane (1988a, p. 904).

[14] For the case d = 3, a_i also satisfies $(a_i)^2 - x_i a_i + 1 = 0$ for i = 1, 2. Here x_1 and x_2 are the solutions to $x^2 - (c_1+4)x + (c_0+2c_1+4) = 0$, where $c_1 = 3(1-\rho)^2/\rho$ and $c_0 = c_1^2/3$. When $\rho = 0.98$, $a_1 = 0.97 + 0.0067i$ and $a_2 = 0.97 - 0.0067i$, where $i = (-1)^{1/2}$.

[15]We thank Ken West for this argument.

[16]In particular, the parametric values are those reported in the "Divisible Labor" column of their Table 1.

[17]The second-moment properties reported in Tables 3 and 4 were obtained by applying the appropriate inverse Fourier transform to the spectral density of the filtered bivariate system. (For example, see Sargent, 1987, chap. 11, sec. 6.)

REFERENCES

Akaike, H. (1974) A New Look at Statistical Model Identification. IEEE Transactions on Automatic Control, AC-19: 716-23.

Ansley, C. F. (1979) An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process. Biometrika, 66: 59-66.

Beveridge, S., and Nelson, C. R. (1981) A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the "Business Cycle." Journal of Monetary Economics, 7: 151-74.

Blanchard, O. J. (1981) What is Left of the Multiplier Accelerator? American Economic Review (Papers and Proceedings), 71: 150-54.

Blanchard, O. J., and Quah, D. (1988) The Dynamic Effects of Aggregate Demand and Supply Disturbances. Manuscript, Massachusetts Institute of Technology.

Blough, S. R. (1988) On the Impossibility of Testing for Unit Roots and Cointegration in Finite Samples. Working Paper in Economics 211, Johns Hopkins University.

Braun, R. A. (1989) Taxes and the Postwar U.S. Business Cycle. Manuscript, University of Virginia.

Campbell, J. Y., and Deaton, A. (1988) Why is Consumption So Smooth? Manuscript, Princeton University.

Campbell, J. Y., and Mankiw, N. G. (1987a) Are Output Fluctuations Transitory? Quarterly Journal of Economics, 102: 857-80.

__________ (1987b) Permanent and Transitory Components in Macroeconomic Fluctuations. American Economic Review (Papers and Proceedings) 77: 111-17.

__________ (1989) Permanent Income, Current Income, and Consumption. Manuscript, Princeton University.

Christiano, L. J. (1987) Is Consumption Insufficiently Sensitive to Innovations in Income? American Economic Review (Papers and Proceedings) 77: 337-41.

Christiano, L. J., and Eichenbaum, M. (1987) Temporal Aggregation and Structural Inference in Macroeconomics. Bubbles and Other Essays, eds. K. Brunner and A. H. Meltzer. Carnegie-Rochester Conference Series on Public Policy, 26: 63-130. Amsterdam: North-Holland.

__________ (1988a) Human Capital, Endogenous Growth and Aggregate Fluctuations. Manuscript, Federal Reserve Bank of Minneapolis.

__________ (1988b) Is Theory Really Ahead of Measurement? Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations. Working Paper 2700, National Bureau of Economic Research.

Christiano, L. J.; Eichenbaum, M.; and Marshall, D. (1989) The Permanent Income Hypothesis Revisited. Manuscript, Federal Reserve Bank of Minneapolis.

Christiano, L. J., and Ljungqvist, L. (1988) Money Does Granger-Cause Output in the Bivariate Money-Output Relation. Journal of Monetary Economics, 22: 217-35.

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

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

__________ (1988b) A Critique of the Application of Unit Roots Tests. Manuscript, University of Chicago.

Deaton, A. S. (1986) Life Cycle Models of Consumption: Is the Evidence Consistent with the Theory? Advances in Econometrics. Fifth World Congress of the Econometric Society, ed. T. F. Bewley, 2: 121-46. Cambridge, Mass.: Cambridge University Press.

DeJong, D. N.; Nankervis, J. C.; Savin, N. E.; and Whiteman, C. H. (1988) Integration Versus Trend-Stationarity in Macroeconomic Time Series. Manuscript, University of Iowa.

De Long, J. B., and Summers, L. H. (1988) On the Existence and Interpretation of a "Unit Root" in U.S. GNP. Working Paper 2716, National Bureau of Economic Research.

Flavin, M. A. (1981) The Adjustment of Consumption to Changing Expectations About Future Income. Journal of Political Economy, 89: 974-1009.

Granger, C. W. J., and Newbold, P. (1977) Forecasting Economic Time Series. New York: Academic Press.

Hall, R. E. (1978) Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. Journal of Political Economy, 86: 971-87.

Hansen, G. D. (1989) Technical Progress and Aggregate Fluctuations. Department of Economics Working Paper 546, University of California, Los Angeles.

Hansen, L. P. (1987) Calculating Asset Prices in Three Example Economies. Advances in Econometrics. Fifth World Congress of the Econometric Society, ed. T. F. Bewley, 2: 207-43. Cambridge, Mass.: Cambridge University Press.

Jaeger, A., and Kunst, R. M. (1989) Seasonal Adjustment and Measuring Persistence in Output. Manuscript, Institute for Advanced Studies, Vienna, Austria.

King, R. G.; Plosser, C. I.; and Rebelo, S. T. (1988) Production, Growth and Business Cycles: II. New Directions. Journal of Monetary Economics, 21: 309-41.

King, R. G.; Plosser, C.; Stock, J.; and Watson, M. (1987) Stochastic Trends and Economic Fluctuations. Working Paper 2229, National Bureau of Economic Research.

King, R. G., and Rebelo, S. T. (1986) Business Cycles with Endogenous Growth. Manuscript, University of Rochester.

Kohn, R. (1979) Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models. Econometrica, 47: 1005-1030.

Long, J. B., Jr., and Plosser, C. I. (1983) Real Business Cycles. Journal of Political Economy, 91: 39-69.

McGratten, E. (1989) The Macroeconomic Effects of Tax Policy in an Equilibrium Model. Manuscript, Duke University.

Mizon, G. E., and Richard, J. F. (1986) The Encompassing Principle and Its Application to Testing Non-nested Hypotheses. Econometrica, 54: 657-78.

Nelson, C. R., and Plosser, C. I. (1982) Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. Journal of Monetary Economics, 10: 139-62.

Plosser, C. I., and Schwert, G. W. (1977) Estimation of a Non-Invertible Moving Average Process: The Case of Overdifferencing. Journal of Econometrics, 6: 199-224.

Prescott, E. C. (1986) Theory Ahead of Business Cycle Measurement. Federal Reserve Bank of Minneapolis Quarterly Review, 10: 9-22. Also in Real Business Cycles, Real Exchange Rates and Actual Policies. Carnegie-Rochester Conference Series on Public Policy, eds. K. Brunner and A. H. Meltzer, 25: 11-44. Amsterdam: North-Holland.

Priestly, M. B. (1982) Spectral Analysis and Time Series. London: Academic Press.

Quah, D. (1988) The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds. Working Paper 498, Massachusetts Institute of Technology.

__________ (1989) Permanent and Transitory Movements in Labor Income: An Explanation for "Excess Smoothness" in Consumption. Manuscript, Massachusetts Institute of Technology.

Sargan, J. D., and Bhargava, A. (1983) Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors When the Root Lies on the Unit Circle. Econometrica, 51: 799-820.

Sargent, T. (1976) Econometric Exogeneity and Alternative Estimators of Portfolio Balance Schedules for Hyperinflations: A Note. Journal of Monetary Economics, 2: 511-21.

__________ (1987) Macroeconomic Theory. 2d ed. New York: Academic Press.

Schwarz, G. (1978) Estimating the Dimension of a Model. Annals of Statistics, 6: 461-64.

Schwert, G. W. (1987) Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data. Journal of Monetary Economics, 20: 73-103.

Sims, C. A. (1988) Bayesian Skepticism on Unit Root Econometrics. Journal of Economic Dynamics and Control, 12: 463-74.

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

West, K. D. (1988a) The Insensitivity of Consumption to News About Income. Journal of Monetary Economics, 21: 17-33.

__________ (1988b) On the Interpretation of Near Random-Walk Behavior in GNP. American Economic Review, 78: 202-209.

Table 1

Results Based on Parametric Models of Persistence

Model p,q*	Likelihood Ratio Statistic** (p-value)	$\hat{C}(1)$	Roots: Autoregressive	Roots: Moving Average
0,1	568.719 (.00)	--	1.275	-.275
0,2	275.825 (.00)	1.577	--	-.153±.498i
0,3	170.775 (.00)	1.838	--	-.392, .025±.608i
1,0	--	1.602	.376	--
1,1	23.684 (.00)	1.746	.524	.168
1,2	11.274 (.00079)	1.772	.260	-.034±.491i
1,3	.110 (.74)	1.816	-.089	.012±.622i, -.451
2,0	--	1.830	.554,-.542	--
2,1	.840 (.36)	1.798	.508,-.542	-.365
2,2	4.356 (.04)	1.530	.293±.614i	.139±.776i
2,3	.990 (.32)	1.605	.168±.668i	.108±.848i, -.206
3,0	--	1.604	.432±.297i, -.518	--
3,1	.569 (.45)	1.365	.631±.340i, -.417	.514
3,2	3.020 (.08)	1.657	.249, .149 ±.720i	.108±.867i
3,3	.000 (1.00)	.000	.299±.565i, .949	.133±.747i, 1.000

*p,q = autoregressive and moving average order, respectively, of ARMA fit to Δy_t.

**Twice the difference between log likelihood values obtained when C(1) = 0 is and is not imposed. (The p-value is obtained using the chi-square distribution with 1 degree of freedom.)

Table 2

Nonparametric Estimates of Persistence

			Model		
k (1)	$\hat{C}(1)^k$ (2)	$\hat{V}^k$ (3)	ARMA(3,3) (4)	ARMA(1,3) (5)	ARMA(2,2) (6)
10	1.41	1.71 (5.49)	.29	.34	.46
20	1.24	1.32 (3.07)	.37	.41	.63
30	1.14	1.12 (2.14)	.34	.37	.65
40	1.00	.86 (1.43)	.38	.40	.68
50	.86	.64 (.95)	.44	.44	.73
60	.88	.67 (.91)	.31	.32	.62
75	.90	.70 (.85)	.18	.20	.48

Notes:

Columns (2) and (3): $\hat{C}(1)^k$ is defined in equation (14); $\hat{V}^k$, in (7).

Columns (4)-(6): Frequency, in artificial data generated by the indicated ARMA model, that the simulated $\hat{V}^k$ exceeds the corresponding empirical value reported in column (3). The ARMA(3,3) and ARMA(2,2) models are those we estimated and reported in Table 1. The ARMA(1,3) model is the trend stationary model reported in Campbell and Mankiw (1987a).

Table 3

Results Based on First-Difference Filter

	No-Components Model		Components Model
	$\phi = 1$	$\phi = .99$	$\phi = 1$ $\rho = .98$
	(1)	(2)	(3)
$\sigma_n/\sigma_{y/n}$	.68	.97	.96
σ_n	.0065	.0086	.0085
$\sigma_{y/n}$	.0096	.0089	.0089
$cov(n,n_{-1})$	-.026	-.026	-.026
$cov(y/n,y/n_{-1})$	.060	.073	.084
$cov(n,y/n_{-1})$	-.072	-.074	-.072
$cov(y/n,n_{-1})$	.023	.031	.032

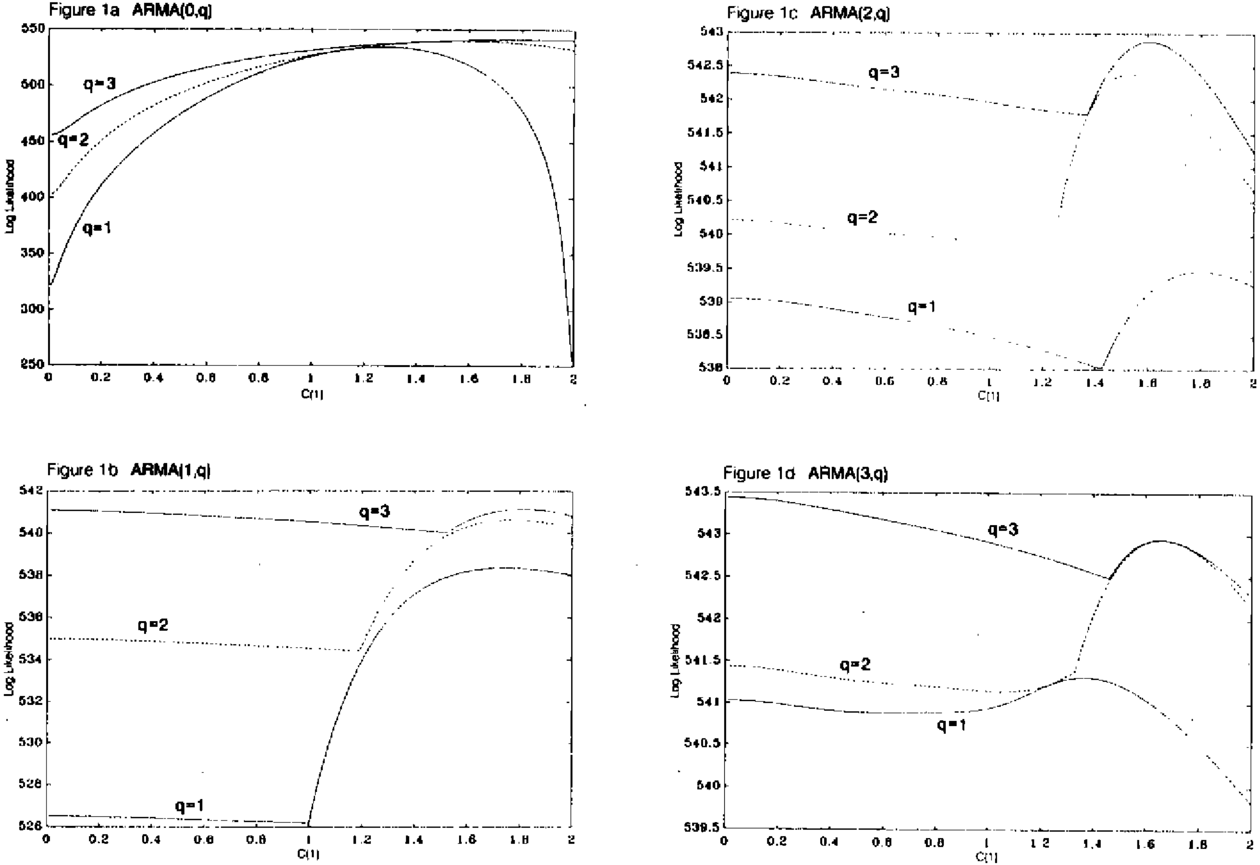
Table 4

Results Based on Hodrick-Prescott Filter

	No-Components Model		Components Model
	$\phi = 1$	$\phi = .99$	$\phi = 1$ $\rho = .98$
	(1)	(2)	(3)
$\sigma_n/\sigma_{y/n}$	.67	.94	.94
σ_n	.0084	.011	.011
$\sigma_{y/n}$	.013	.012	.012
$cov(n,n_{-1})$	.71	.71	.71
$cov(y/n,y/n_{-1})$	.74	.75	.75
$cov(n,y/n_{-1})$	.59	.57	.57
$cov(y/n,n_{-1})$	.75	.75	.75

Figures 1a–1d

Maximized Log Likelihood Function vs. C(1) for Various ARMA Models



Figures 2a–2b

Accounting for the ARMA(2,2) Results From the Trend Stationary Perspective

Figure 2a Frequency Distribution of the Likelihood Ratio Statistic for Testing C(1) = 0

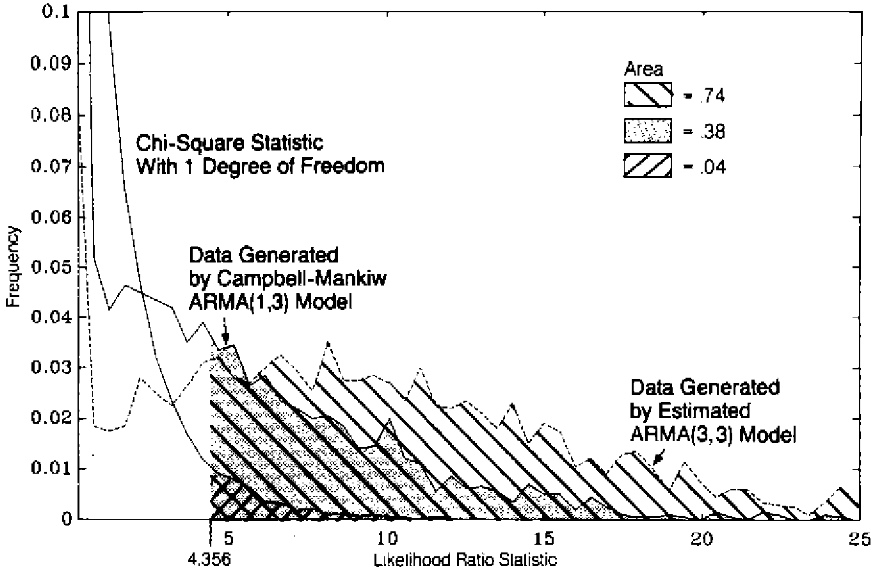
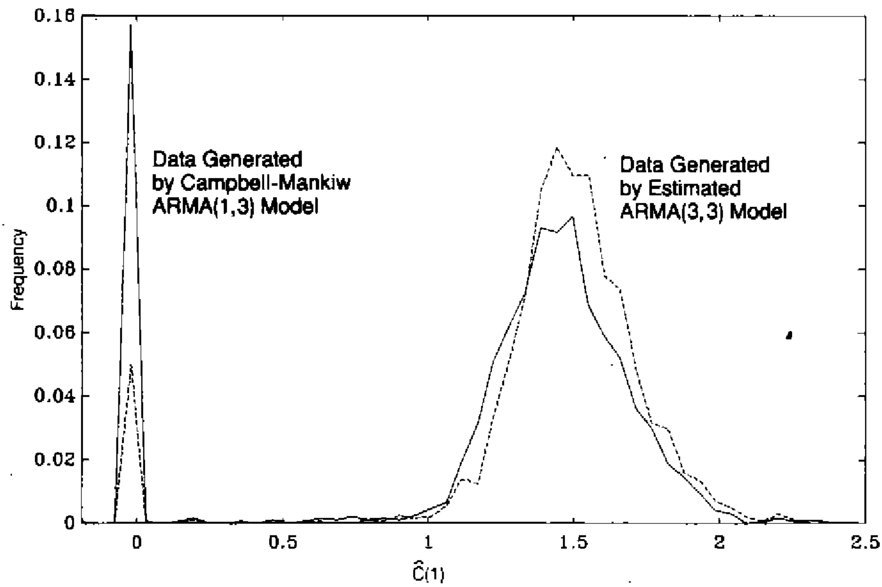


Figure 2b Frequency Distribution of $\hat{C}(1)$



Figures 3a–3c

Why the Trend Stationary Perspective Can Account for the ARMA(2,2) Results

A Large-Sample View

Figure 3a Unconstrained Results

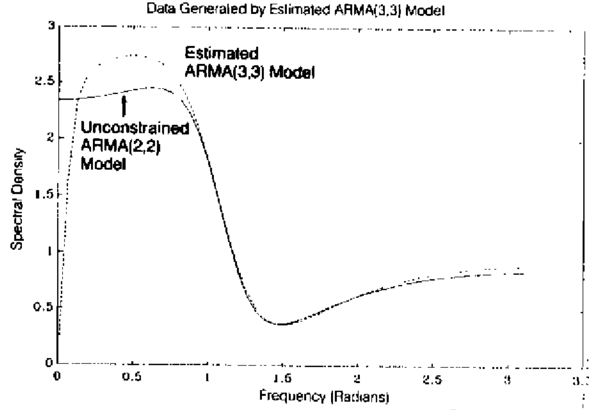


Figure 3b Constrained Results

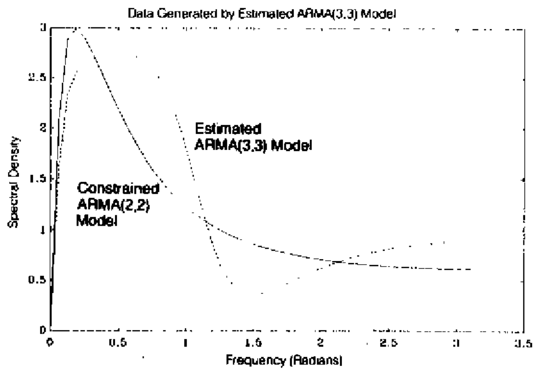
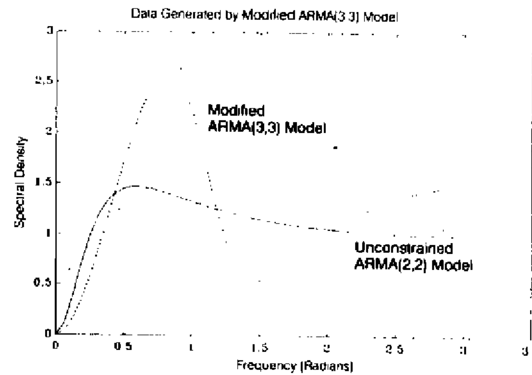


Figure 3c Unconstrained Results



Note. The curves in the figures above depict the spectral density of various models
Each spectral density has been scaled so that the innovation variance is unity

Figure 4

Accounting for the ARMA(1,3) and ARMA(3,3) Results From the Difference Stationary Perspective

Frequency Distribution of $\hat{C}(1)$
Using Data Generated by Estimated ARMA(2,2) Model

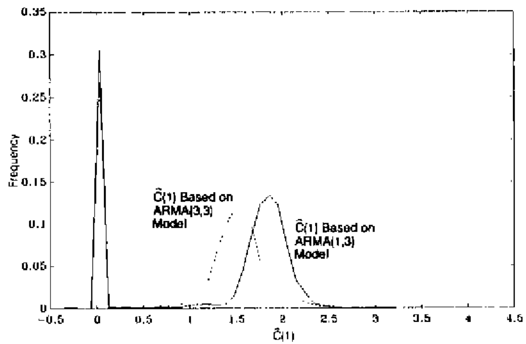


Figure 6

Nonparametric Measures of Persistence

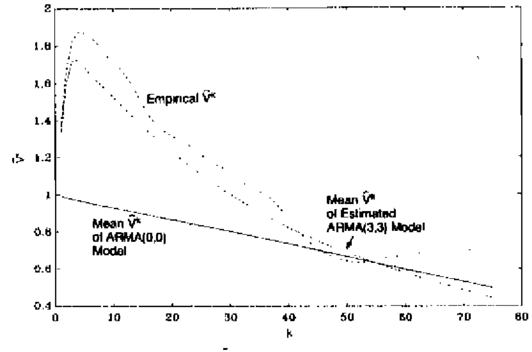


Figure 5

Impulse Response Functions Implied by Various ARMA Models

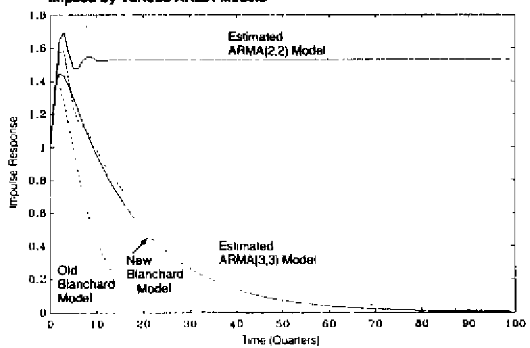


Figure 7

Dynamic Response of Permanent Component to a One Standard Deviation Shock

$a_u C_1(L)/(1-L)$ for d = 1, 2, 3