Carnegie-Rochester Conference Series on Public Policy 32 (1990) 63–82
North-Holland

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

JAMES H. STOCK*
Kennedy School of Government
Harvard University

I. INTRODUCTION

Christiano and Eichenbaum are to be congratulated for presenting an impressive array of results concerning unit roots in economic time series in general and in real GNP in particular. Some of their results elucidate the considerable literature on the existence and interpretation of a unit root in real GNP. Other results -- for example, those in the final section of their paper, in which stochastic productivity in a real business cycle model is modified to have both permanent and stationary components -- mark the beginning of an interesting and potentially rich research program.

It would be beyond the scope of these brief remarks to examine every aspect of Christiano and Eichenbaum's paper. Instead, I make two main points. The first, methodological, focuses on Christiano and Eichenbaum's finding that "we do not know" if GNP has a unit root. An obvious question is whether this result is special to the time series they have chosen to analyze (postwar U.S. GNP) or whether it is intrinsic to their procedure and indeed to the question being asked. I will argue that it is largely the latter, that is, with high probability Christiano and Eichenbaum's econometric procedure for answering "do we know?" will result in the answer "we do not know." The reasoning, spelled out in Section II, is simple: hypothesis tests cannot reject alternative models that are "close to" -- formally, within a local neighborhood of -- the null model. These local alternatives cannot be consistently distinguished from the null model: the

*I thank Larry Christiano, Robin Lumsdaine and Mark Watson for helpful discussions during the preparation of these comments. I am particularly grateful to John Cochrane and Danny Quah for their extensive comments on an earlier draft of these remarks. This research was supported in part by NSF grants SES-86-18984 and SES-89-10601.

consistency of any test, including tests for unit roots, hinges on placing a suitable distance between the null and the alternative. These local alternatives continue to exist even as the sample size grows, and Christiano and Eichenbaum's procedure will in fact select precisely these local alternatives that the data cannot reject. The observation that the power of tests against local alternatives is less than one applies to all tests of nested hypotheses: it is an implication of the fact that, for any finite sample size, a confidence interval centered around the null contains points under the alternative. What is special about the case of trend-stationary (TS) vs. difference-stationary (DS) models is not just that for each DS model there is a local neighborhood of TS models, but that the converse is also true.[1] This idea lies behind the work of Clark (1988), Cochrane (1988b) and Blough (1988), who examine the relationship between TS and DS models in finite samples of a fixed length. The contribution of this comment is to provide a formal extension of this result to increasingly large samples. The practical implication of these unrejectable local alternatives is that one must first be precise about why the TS/DS distinction matters before deciding which model better describes the data.

The second main point is that, from the perspective of certain applications of this decision, the answer to "do we know?" is far less ambiguous than Christiano and Eichenbaum suggest. In Section III, I suggest three reasons why one might be interested in the TS/DS distinction for GNP: for forecasting, for drawing econometric inferences, and for sharpening macroeconomic theories. The answers to "do we know" and "do we care" are linked, and they differ in each of these cases. For forecasting, I argue that whether U.S. GNP is modeled as TS or DS is in one sense unimportant, in another sense potentially important but as yet unresolved. For econometric inference, the presence of integrated regressors can sharply alter how one interprets empirical evidence; in this case I will argue that the evidence for the adequacy of the DS approximation for GNP is more compelling than Christiano and Eichenbaum suggest. For macroeconomic theory, to the extent that the permanent effect of a shock plays a critical role, Christiano and Eichenbaum are correct in suggesting that the evidence simply does not resolve this issue. Indeed, the theoretical result in

---

[1]These local neighborhoods are readily constructed: in the DS case by replacing the unit autoregressive root with a root of ρ_T, in the TS case by introducing a unit autoregressive root and a moving average root of θ_T, where $\rho_T \to 1$ and $\theta_T \to 1$ at appropriate rates.

Section III suggests that this is an intrinsic problem with the question being asked. There are, however, alternative related questions that from a practical perspective might be more relevant, such as what is the effect of a shock over three years? Christiano and Eichenbaum make headway in answering this question, and their results suggest that this three-year response is large, larger than had generally been accepted ten or even five years ago. This is an important conclusion, even if it does not quickly distinguish between broad classes of macroeconomic theories.

II. STATISTICAL PROPERTIES OF CHRISTIANO AND EICHENBAUM'S PROCEDURE

The procedure used by Christiano and Eichenbaum to determine whether GNP is better described by a TS or a DS process has two steps: (1) Use the data to fit several ARIMA models; among these, select preferred DS and TS models; (2) Simulate data from the preferred DS and TS models, and use these simulated data to construct several statistics that might in principle distinguish between the DS and TS model, such as the likelihood ratio statistic. If data generated by the DS model are consistent with the statistics computed from the data under the TS model, but not vice versa, then the DS model is selected; if the reverse is true, the TS model is selected; if neither or both models explain the other, then the decision is "we do not know."

My claim is that this procedure will, with high probability, result in the conclusion "we do not know," even as the sample size tends to infinity. This argument is developed in three stages, and an interpretation follows.

A. The Case of Ordinary Least Squares (OLS) Regression

Suppose for the moment that the decision at hand is not whether real GNP is TS or DS, but whether or not the true regression coefficient β equals 0 in a standard linear regression equation with a single independent variable. Then Christiano and Eichenbaum's procedure reduces to: (1) fit a constrained and an unconstrained model; and (2) check whether the two models are capable of explaining each other. It is instructive to consider what would happen in the two cases that the null is true and that the null is false, where under the alternative the true value of β is a nonzero constant. For this problem, the natural statistic to consider is the t-statistic on the OLS estimator $\hat{\beta}$.

First suppose that the null is true. In the first step of the procedure, $\hat{\beta}$ will, of course, be nonzero, so the estimated unconstrained model will be an alternative model. With high probability the null model will explain the estimate obtained under the alternative: 95% of the time, the absolute t-statistic will be less than 1.96 in large samples. Whether the estimated alternative is capable of "explaining" the null depends on the distribution of the t-statistics generated under the estimated alternative in the second stage of the procedure. If most of the simulated t-statistics exceed, say, 1.96 in absolute value, then the alternative model cannot explain the null, and the null would be accepted; but if the t-statistics are small, then the estimated alternative explains the null and the conclusion is "we do not know."

The asymptotics here turn out to be straightforward. Given an estimate of β, the fraction of t-statistics simulated in the second stage that reject the null is given by the power function, which is a function of β. It is well-known that this power function has a well-defined limit in the case that the alternative is nested as $\beta_T=\delta/\sqrt{T}$ (where T is the sample size); let this limiting function be $g(\delta)$.[2] But under the null $\sqrt{T}\,\hat{\beta}$ has a limiting normal distribution. Thus the asymptotic probability that this procedure results in t-statistics that reject is $\int g(\beta^*)dF(\beta^*)$, where $F(\beta^*)$ is the limiting normal distribution of $\sqrt{T}\,\hat{\beta}$. For a two-sided test, the median power (over repeated samples of $\sqrt{T}\,\hat{\beta}$) is given by $g(\delta_{med})$, where δ_{med} is the median of the distribution of $|\sqrt{T}\,\hat{\beta}|$. Thus the alternative model will produce t-statistics that asymptotically have second-stage (simulated) power less than one, often much less than one. Moreover, both models will produce distributions of t-statistics that are consistent with the t-statistic computed from the data. Thus with high probability this procedure will result in the answer "we do not know."

Next suppose that the alternative is true, so β is a nonzero constant. In this case, the t-statistics generated under the alternative model will explain the t-statistics in the data. However, because the alternative is assumed to be fixed, in large samples the distribution of t-statistics simulated under the null will explain neither the t-statistic computed from the data nor those generated by the alternative model. Thus the procedure

[2]The limiting power function $g(\delta)$ is the probability that a noncentral χ_1^2 with noncentrality parameter $\delta V_0(\hat{\beta})^{\frac{1}{2}}$ exceeds the relevant critical value, where $V_0(\hat{\beta})$ is the asymptotic variance of $\sqrt{T}\,\hat{\beta}$ under the null.

would conclude that the alternative model is preferred.

In OLS regressions, then, if the alternative is true this will be detected by Christiano and Eichenbaum's procedure, but if the null is true, then with high probability the conclusion will be "we do not know." The reason is simple: if the null is true, alternative values within a finite confidence region cannot be consistently distinguished from the null. However, if the alternative is true and fixed, in the case of OLS regression it can be distinguished consistently.

B. A Stylized Version of the Christiano-Eichenbaum Procedure: the TS/DS Decision

The TS/DS decision differs from the β=0 decision of the previous subsection because for every DS model there are TS models local to it and vice versa. This leads to the proposition that, whether the true model is TS or DS, with high probability Christiano and Eichenbaum's procedure will result in "we do not know."

This proposition is demonstrated in a stylized version of their procedure, in which I consider one TS and one DS model as the potential true models, a set of four estimated models (which to simplify things contains both hypothesized true models), and a second-stage decision criterion based on a Dickey-Fuller (1979) type test for a unit root. The advantage of considering this stylized version of Christiano and Eichenbaum's procedure is that analytic asymptotic results either exist or can be derived to demonstrate the proposition.

Suppose that four different ARIMA models are estimated in the first stage, all of which are special cases of an ARMA(1,1) model in levels:

ARIMA(0,0,0): $$y_t = \varepsilon_t \tag{1}$$

ARIMA(1,0,0): $$(1-\alpha L)y_t = \varepsilon_t \tag{2}$$

ARIMA(0,1,0): $$\Delta y_t = \varepsilon_t \tag{3}$$

ARIMA(0,1,1): $$\Delta y_t = (1-\theta L)\varepsilon_t \tag{4}$$

where ε_t is white noise.

I consider two cases for the true model: the DS model (3) and the TS model (1). These might reasonably be thought of as extremes, one being a random walk, the other serially uncorrelated. As in the previous subsection, I consider the two cases -- the DS and TS cases -- and under each see what conclusion would be drawn from this procedure. To make things concrete, assume that tests are conducted using 5% significance levels.

First suppose that the DS model $\Delta y_t = \varepsilon_t$ is true. Asymptotically, of the four models fit in the first stage of the procedure, model (1) will be rejected, model (3) will be admitted as a possibility, model (4) will usually (95% of the time) be recognized as redundant, and model (2) will be admitted as a possibility. With high probability, $\hat{\alpha}$ in model (2) will be less than one (if a constant and a time trend are included in the regression, this probability exceeds 99%). Thus model (2) will be chosen as the candidate TS model and model (3) as the candidate DS model. Because model (3) is true by assumption and is nested in model (2), model (3) will explain model (2) in the sense of Christiano and Eichenbaum.

The only question, then, is whether Dickey-Fuller statistics computed under model (2) will explain the t-statistic computed from the data, that is, whether the Dickey-Fuller statistics simulated in the second stage using estimated model (2) will fail to reject with high probability. An argument analogous to that used in the OLS problem indicates that these simulated Dickey-Fuller statistics typically will not reject the unit root null. Specifically, Dickey and Fuller (1979) showed that if model (3) is true, then $T(\hat{\alpha}-1)$ has a nondegenerate asymptotic distribution. It is also well-known, however, that if the true model is $(1-(1-\delta/T)L)y_t=\varepsilon_t$, then conditional on δ the asymptotic power function of the Dickey-Fuller test has a well-defined limit; call this $h(\delta)$. (This function does not have a convenient closed-form expression, but can be expressed variously as an integral with respect to a functional of a diffusion process or as a function of normal variates; see Cavanagh, 1986; Chan and Wei, 1987; Chan, 1988; and Phillips, 1987.) For this local nesting, the power will be less than one, and often the power will be small indeed. The <u>average</u> power of this two-stage procedure is $\int h(\alpha^*)dG(\alpha^*)$, where $G(\alpha^*)$ is the asymptotic distribution of $T(\hat{\alpha}-1)$ under model (3). This average power, calculated by Monte Carlo simulation for various values of T, is summarized in Table 1. The values in the table indicate that, on average over draws of the data from the true model (3), the power of the Dickey-Fuller test using data generated by the estimated alternative model (2) is slight, having an asymptotic limit of approximately 15% for tests at the 5% significance level. In addition, the convergence to the asymptotic limit is rapid. Thus if model (3) is true, with high probability this procedure will yield the answer "we do not know."

Table 1

Integrated Power of the Dickey-Fuller t-Statistic Against an Estimated ARIMA(1,0,0) Model Local to an ARIMA(0,1,0) Model

	------ Significance Level ------		
T	10%	5%	1%
50	26.3%	17.7%	6.8%
100	26.7	17.7	6.2
200	26.3	16.4	5.1
400	25.5	14.7	4.4

Notes: The entries are the mean probability of rejection of the unit root null against the TS alternative using a Dickey-Fuller t-statistic computed with data generated from the model $y_t=\hat{\alpha}y_{t-1}+\varepsilon_t$, ε_t i.i.d. N(0,1), where $\hat{\alpha}$ is the first autocorrelation of x_t estimated from data generated according to $\Delta x_t=\eta_t$, η_t i.i.d. N(0,1). The series $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independently distributed. The results are based on Monte Carlo simulations with 2000 replications and the sample sizes indicated in the first column. The critical values used are those of the Dickey-Fuller $\hat{\tau}_\mu$ statistic: -2.57 (10%), -2.86 (5%), and -3.43 (1%). The Dickey-Fuller t-statistic is computed from a regression of Δy_t on 1 and y_{t-1}.

Next consider the TS case, and suppose that model (1) is true. In the first stage of the procedure, model (1) will be admitted, model (2) will typically (95% of the time) be recognized as being redundant (the true value of α is 0), and model (3) will be rejected. There are two possibilities for model (4). With asymptotic probability .6575, $\hat{\theta}=1$ (the pile-up phenomenon: Sargan and Bhargava [1983], Corollary 2) and model (4) reduces to model (1). With probability .3425, $\hat{\theta}<1$, but in this event $\hat{\theta}$ will fall in a 1/T neighborhood of 1 (Sargan and Bhargava [1983], Proposition 2). If $\hat{\theta}=1$, then there is no acceptable DS model. If, however, $\hat{\theta}<1$, then model (4) will be chosen as the candidate DS model.

The issue thus is whether the candidate DS model (4) with $\hat{\theta}<1$ will generate Dickey-Fuller statistics that are consistent with the data and the TS model (1). Because $T(\hat{\theta}-1)$ is $O_p(1)$ in this case, the question reduces to the behavior of the statistic used to discriminate between the DS and TS models, when the TS model is the limit of a sequence of the form $\Delta y_t = (1-(1-\delta/T)L)\varepsilon_t$, given $\delta>0$. Although this nesting has not been studied as extensively as the case of AR roots local to one, some results have been obtained for special cases. Pantula (1988) examined several tests for a unit root under various local nestings and reached the interesting conclusion that the local neighborhoods of different statistics vanish at different rates. For the 1/T neighborhood here, the Dickey-Fuller statistic with no additional lags of Δy_t in the regression will reject the DS null with probability one (Pantula (1988), Theorem 3.3(a)). One might object to the use of this statistic here, since it does not allow for possible MA terms. One modification is to consider a test based on $\bar{\alpha}-1=(\Sigma\Delta y_t y_{t-2})/(\Sigma y_{t-2}^2)$, which is asymptotically equivalent (under the fixed DS null) to the instrumental variables estimator proposed by Pantula and Hall (1988), minus one. In this case, Pantula's (1988) techniques can be used to show that, under the 1/T local DS model, $\sqrt{\tfrac{1}{2}T}\,(\bar{\alpha}-1)$ has a standard limiting normal distribution. Thus the probability of rejection using a test based on $T|\bar{\alpha}-1|$ tends to one. The same result, rejection with probability one, obtains for a test based on $T|(\Sigma y_t y_{t-2})/(\Sigma y_{t-2}^2) -1|$. The implication is that for such statistics the local DS model typically will result in simulated second-stage statistics that reject the DS hypothesis. Thus this local DS model will be able to "explain" the results in the data (the rejection of the DS null) and the simulated distributions obtained under the TS model (1). With high probability the conclusion is again that "we do not know."

In both cases, then, this procedure results in the asymptotic answer, "we do not know," even though the two models considered could not be clearer examples of DS and TS processes. The main caveat to this statement concerns the treatment of the case that the estimated MA root is 1 when the true model is TS. If in this event additional ARMA models were estimated until a DS model appeared, then the conclusion of this discussion arguably would still apply. If instead the decision is to stop and to decide that no DS model described the data, then the probability that Christiano and Eichenbaum's procedure reaches the conclusion "we do not know" is reduced to approximately 34%.

C. Christiano and Eichenbaum's Actual Procedure

The procedure used by Christiano and Eichenbaum differs from that examined in the previous subsection in three important respects. First, they follow Campbell and Mankiw (1987a) and consider not just four but many ARMA models, in principle perhaps an unbounded number. Second, they use a variety of statistics, none being Dickey-Fuller statistics, as criteria for distinguishing between the models. Third, and most importantly, there is no guarantee that the true model is nested in the sequence they consider. These features conspire to place an analytic examination of the properties of their procedure beyond the scope of this comment.

Nonetheless, there is reason to believe that each of these features tends to increase -- or at least not to reduce -- the probability of reaching the conclusion "we do not know." First, increasing the number of estimated models will result in models that better fit the data, and increasing the number of MA terms and AR terms admits the possibility of common roots (parameter redundancy) and of MA roots that equal one. This suggests an increased probability of estimating a TS model when the true model is DS, relative to the example.[3] Second, the close link between

---

[3]Note that, unlike the stylized example in Section II.B, Christiano and Eichenbaum's ARMA models specify GNP in growth rates. Thus a candidate DS model will always exist, but their procedure will find a candidate TS model only by estimating a unit MA root. One generalization of the result of the previous section to the Christiano and Eichenbaum procedure would be as follows. Suppose that the true process is ARMA(p_0,q_0), perhaps with a single unit AR root (it is innocuous to assume that the true process has no unit MA root). Run a sequence of ARIMA(p,1,q) models, for $p=0,\dots,p_1$, $q=0,\dots,q_1$, where p_1 and q_1 are large (and $p_1 \geq p_0$, $q_1 \geq q_0+1$). Among these models, use some information criterion to pick preferred DS and TS models (assuming that a TS model is estimated). Compare these two models using, say, a likelihood ratio statistic. The question is whether these two preferred models will with high probability explain each other in the Christiano and Eichenbaum sense, so that the conclusion will be "we do not know" whether the true model is DS or TS.

likelihood ratio statistics and Dickey-Fuller statistics suggests that they will have similar difficulties distinguishing between null hypotheses and local neighborhoods. Third, because finite order ARMA models surely involve specification error, this procedure will produce candidate DS and TS models that are equally close to the unknown model in the sense of one-step ahead linear predictions; TS and DS models that produce similar forecasts are arguably likely to be local to each other in the sense discussed above. These observations -- particularly the first concerning the estimation of multiple models -- suggest that the qualitative conclusions from the previous simplified analysis apply to the procedure Christiano and Eichenbaum actually use.

D. Interpretation

It is well-known that for each DS model there is a local TS model; this is the case of "roots local to unity" (e.g., Cavanagh, 1986). Clark (1988), Cochrane (1988b) and Blough (1988) have made the reverse argument in the context of finite samples. The preceding discussion extends these observations to large samples and to Christiano and Eichenbaum's decision procedure. These remarks therefore provide formal support for these authors' conclusions that it is unknowable whether C(1) is nonzero, because local alternatives (with opposite implications) always exist. This also accords with and reinforces Christiano and Eichenbaum's conclusion based on their empirical work. This conclusion might at first seem surprising, for unit root tests typically are consistent; but consistency by definition examines power against fixed (nonlocal) alternatives, while it is precisely the local alternatives that the Christiano and Eichenbaum procedure identifies. Moreover, the key difference between the DS/TS decision and more familiar examples of hypothesis tests is not the unusual rates of convergence nor the nonstandard distribution theory in the DS/TS case, but rather the existence of nuisance parameters that provide all the points in one hypothesis with local neighbors in the other.

This discussion raises the question of why Campbell and Mankiw's (1987a) Monte Carlo experiment reached the conclusion (according to the conjecture in Section II.C, the unlikely conclusion) that the TS "old Blanchard" model was inconsistent with the data. As Christiano and Eichenbaum point out, the answer lies in the way that Campbell and Mankiw implemented their Monte Carlo experiment. The theoretical discussion above refers to the case when all the ARMA models are estimated with the same data. However, Campbell and Mankiw adopted for their candidate TS model an estimate based on a subsample. Thus the theoretical discussion is consistent with both Campbell and Mankiw's rejection of the "old Blanchard" model and with Christiano and Eichenbaum's failure to reject their preferred TS model estimated over their full sample.

III. IMPLICATIONS FOR ECONOMETRIC PRACTICE AND EMPIRICAL MACROECONOMICS

Whether this unresolvable ambiguity -- these indistinguishable local models -- matters depends on whether these models have different implications for econometric practice and for our empirical understanding of the behavior of the macroeconomy. This section examines three possible cases in which the TS/DS distinction might be of importance.

A. Implications for Forecasting

Does the knowledge of whether GNP has or has not a unit root have any practical value to producers of GNP forecasts? One simple way to answer this question is to compare the forecasting performance of Christiano and Eichenbaum's candidate TS and DS models. Because the short-run forecasts of these models are so similar, it is clear that this question is only of interest for forecasts at least several quarters into the future. To be precise, I consider forecasts made with DS and TS models at the three-year horizon (for example, this is the relevant horizon for initial revenue forecasts used in developing Federal and most state budgets).

The root mean square errors of 3-year forecasts made by these two models and a "rolling" variant are reported in Table 2. (Christiano and Eichenbaum consider two candidates for the TS specification, but for simplicity I examine only one, the ARMA(3,3) model.) Although a formal statistical analysis would be necessary to ascertain which if either model performs better, the qualitative difference between the two is clearly slight. This result is perhaps surprising, because one of the original motivations for developing procedures for identifying unit roots was the apparently poor performance of long-term forecasts from ARMA models that failed to impose a unit root when one was arguably present.

A striking feature of the results in Table 2 is that both of these models make very poor predictions of the 3-year growth in GNP, providing no real improvement over a constant forecast that simply equals the postwar mean. Thus a professional forecaster would eschew either of these

Table 2

Root Mean Square Errors (RMSE's) for k-Quarter Ahead Forecasts for the Christiano-Eichenbaum TS and DS Models

k-Quarter Growth in U.S. Real GNP, Annual Rates
1949:III - 1988:IV

	Estimated Over Full Sample		Estimated Over Rolling Sample		
k	DS	TS	DS	TS	Mean
1	.0406	.0409	.0417	.0417	.0454
4	.0296	.0297	.0272	.0281	.0305
8	.0212	.0211	.0209	.0219	.0210
12	.0160	.0157	.0159	.0169	.0158

Notes: The DS model is Christiano and Eichenbaum's ARMA(2,2) model, and the TS model is Christiano and Eichenbaum's ARMA(3,3) model. The first column reports the forecast horizon k. The second and third columns report the k-quarter ahead RMSE's (annual rates) for the forecasts computed using the TS and DS models based on Christiano and Eichenbaum's parameters (which they estimated over the full sample). The fourth and fifth columns report the RMSE's for these models estimated in a "rolling" regression fashion, i.e., in a given quarter, say 1977:I, the model was estimated through 1977:I and the k-quarter ahead forecast was made using this estimated model; in 1977:II, the model was reestimated. These forecast errors were computed over 1967:I-1988:IV and so correspond to a smaller sample than do the results in the previous two columns. Columns four and five thus are not directly comparable to the previous columns but can themselves be compared. The TS ARMA specification was estimated by imposing the unit MA root, i.e., by estimating an ARMA(3,2) model in the levels of log real GNP. The final column is the full-sample RMSE computed using only the mean k-period growth rate over the full sample as the forecast, that is, using no conditioning variables.

univariate models in favor of a multivariate forecast. However, the analysis here and in Christiano and Eichenbaum's paper is silent on the importance of the TS/DS choice in multivariate forecasting. An interesting aspect of this problem is whether forecasts can be improved in multivariate time series models by imposing the restriction that series be integrated and in particular cointegrated, compared with considering systems in levels that do not impose the integration or cointegration constraints. Although there is some initial work in this regard (e.g., Engle, Granger, and Hallman (1987)), whether multivariate aspects of the TS/DS decision improve forecasts substantially is largely unresolved.

B. Implications for Econometric Inference

Another reason for wanting to know whether to model real GNP as TS or DS is for econometric inference. It is well-known that whether a series is TS or DS has important implications for the asymptotic distribution (and thus interpretation) of certain statistics from regressions that involve levels of the series. Thus a relevant practical question is whether the TS or DS model will provide a better framework for approximating the distribution of statistics that involve the level of GNP.

For this purpose, in my judgment the evidence is clearly in favor of the DS approximation. The intuition is simple: precisely because the Christiano-Eichenbaum TS model (with its autoregressive root of .95) looks in finite samples so much like a DS model, it behaves in terms of econometric inference as if it were a DS model. As a specific example, consider the case in which log real GNP growth is regressed against p of its lags, a constant, a time trend, and lagged log real GNP. In the DS case, the t-statistic on lagged log real GNP (which is the Dickey-Fuller test statistic for a unit root) has a nonstandard asymptotic distribution. In the TS case, it tends towards minus infinity. Thus one check of which approximation works better is to check the percent rejections obtained by this Dickey-Fuller test under the two candidate models.

The results are summarized in Table 3 for a sample of size 140. The Dickey-Fuller statistic is best thought of as nested in a sequence of p's that tends to infinity with T, so the asymptotic approximation would not be expected to work exactly for fixed T and p. It is not clear that this approximation is any worse for the TS than the DS model; the power against the TS model slightly exceeds the asymptotic level, the size under the DS model is slightly less. Thus, even when the data are generated by the TS model, the DS approximation to the distribution is evidently better than

Table 3

Monte Carlo Percent Rejections by Dickey-Fuller t-Test for a Unit Root

Regression: Δy_t on 1, t, y_{t-1}, $\Delta y_{t-1}, \ldots, \Delta y_{t-p}$

p	Asymptotic significance level	- - - - Percent Rejections - - - - DS: ARMA(2,2)	TS: ARMA(3,3)
4	10%	5.4%	11.9%
	5%	2.5%	5.6%
8	10%	9.0%	15.3%
	5%	4.4%	8.0%

Notes: Based on 5000 Monte Carlo simulations of series of length T=140. The parameter values used are those reported by Christiano and Eichenbaum for their respective DS and TS models. The series were generated according to $a(L)y_t=b(L)\varepsilon_t$. The asymptotic critical values are -3.12 (10%) and -3.41 (5%).

the TS approximation (which would imply 100% rejections).

Inspection of the roots of Christiano and Eichenbaum's candidate TS and DS models indicates that the TS model is arguably closer to the DS boundary than is the DS model to the TS boundary: the largest AR root of the TS ARMA(3,3) model is .95, while the largest MA root in the DS ARMA(2,2) model is only .78. Although the size of the largest roots is only one factor in determining the finite sample distributions, this suggests that both the TS and DS models will exhibit DS properties in finite samples, and indeed that is the conclusion of the Monte Carlo simulation of the Dickey-Fuller tests. These observations strongly suggest that the DS model provides a better basis for approximating the distribution of estimators and test statistics than does the TS formulation.

C. Implications for Empirical Macroeconomics and Macroeconomic Theory

Two implications of the foregoing discussion for the distinction between TS and DS models bear emphasis. First, inference about the cumulative impulse response function C(1) cannot be made satisfactorily even with arbitrarily large samples: local to models with C(1)=0 are models with nonzero C(1) and vice versa. Thus macroeconomic theories which hinge crucially on whether or not C(1) is zero are bound to have ambiguous empirical support, that is, empirical evidence cannot resolve with certainty whether C(1) is zero or large. While one could take this as a weakness of econometrics, my interpretation is rather that the question is ill-posed.

In contrast, the outlook need not be as pessimistic if one's sights are lowered somewhat and attention is instead focused on, say, the impulse response function cumulated to 3 or 4 years. On a practical level, this horizon is still rather distant for the purpose of tracing out the effects of some macroeconomic shocks. This suggests that an interesting challenge is the estimation of a confidence interval for the 3- or 4-year cumulative response that properly incorporates model uncertainty. The gap between these values for the ARMA(2,2) and ARMA(3,3) models in Christiano and Eichenbaum's Figure 5 indicates that this interval will be wide, at least for horizons beyond 4 years. However, for shorter horizons -- even for 3 years -- as Christiano and Eichenbaum emphasize, these results indicate a substantial degree of persistence, with point estimates of this response ranging between .8 and 1.5 for the preferred models.

Second, implicit in Christiano and Eichenbaum's permanent-transitory decomposition is the observation that it is ultimately of greater interest to consider models, both empirical and theoretical, in which the information set is expanded beyond that of the errors in a univariate Wold representation. Following Quah (1989), Christiano and Eichenbaum consider some theoretical models in which agents are given information about which movements in the series are permanent and which are transitory. Not surprisingly, this changes the observed behavior of the models; perhaps surprisingly, these changes can be large and of economic significance. Rather than focus on the details, I simply note that this use of permanent-transitory decompositions raises several interesting questions for future research. For example, is it possible to derive a stochastic process for the permanent component from a model of technology diffusion? Does the microeconomic evidence on technology diffusion suggest any useful restrictions? Can the implied permanent and transitory components of productivity be extracted using an empirical model estimated with observable series? If so, what do they look like? These questions hint that ideas suggested by Christiano and Eichenbaum's paper could be a fertile area for further research.

The observation that an expanded set of shocks might shed light on the importance of permanent components has stimulated some reduced-form empirical investigations that are worth mentioning. Recent empirical work suggests that a large fraction of the movements in output are permanent. For example, King, Plosser, Stock and Watson (1987) computed the fraction of the variance of the error in the 8-quarter ahead forecast of output attributable to movements in a "permanent" innovation, which was identified as being the innovation in the sole permanent component common to output, consumption and investment. In this three-variable model and in five- and six-variable models incorporating money, prices, and interest rates, this decomposition attributed roughly 20% to 40% of this variance to transitory innovations. In related work that supposes stationarity of the unemployment rate to identify an assumed permanent component of output, Blanchard and Quah (1989) attribute between 12% and 40% of this variance to temporary disturbances. Shapiro and Watson (1988), using different data and different identifying assumptions, estimate this fraction to be 20%. Finally, in a model identified through a mixture of long-run and contemporaneous covariance restrictions, Galli (1988) finds that transitory disturbances explain 27% of the variability of output at the 10-quarter horizon. Whether these models would stand up to Christiano and Eichenbaum's scrutiny is yet unknown. Still, despite the variety of specifications, variables, and sample periods, these authors reached strikingly similar conclusions about the importance of permanent components in explaining the movements in output. It certainly seems inappropriate to ignore this evidence and adopt the TS specification and its concomitant implications about the transitory nature of fluctuations.

IV. CONCLUSIONS

It is important to be clear why one wants to obtain a point estimate or to test a hypothesis before interpreting the results. Here, this means understanding what aspects of the DS or TS model are important for the objective at hand. If the goal is to estimate the infinite cumulative response C(1), I agree with Christiano and Eichenbaum that these results indicate a fundamental ambiguity: in retrospect, previous claims of accurate estimates of C(1) appear to have been excessively optimistic. For other applications, however, this ambiguity is not as severe. If the goal is to estimate the persistence of shocks over 3 years, for example, the point estimates reported by Christiano and Eichenbaum suggest substantial persistence in innovations to GNP whether or not the DS or TS model is adopted. This is reinforced by an examination of the multivariate evidence. Alternatively, if the goal is to approximate the distributions of regression estimators and test statistics when the regressors might be integrated, the evidence provided here and elsewhere suggests that the DS model of GNP is a substantially more reliable guide than the TS model. If instead the goal is to make long-run forecasts of output, it turns out these results are not germane, at least in the case of GNP: what matters is an investigation of these issues in the context of multivariate rather than univariate models.

REFERENCES

Blanchard, O.J. and Quah, D.

(1989) The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79: 655-673.

Blough, S.R.

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

Campbell, J.Y. and Mankiw, N.G.

(1987a) Are Output Fluctuations Transitory? *Quarterly Journal of Economics*, 102: 857-880.

Cavanagh, C.L.

(1986) Roots Local to Unity. Discussion Paper No. 1259, Harvard Institute of Economic Research.

Chan, N.H.

(1988) The Parameter Inference for Nearly Nonstationary Time Series. *Journal of the American Statistical Association*, 83: 857-862.

__________ and Wei, C.Z.

(1987) Asymptotic Inference for Nearly Nonstationary AR(1) Processes. *Annals of Statistics*, 15: 1050-1063.

Clark, P.K.

(1988) Nearly Redundant Parameters and Measures of Persistence in Economic Time Series. Research Paper No. 979, Graduate School of Business, Stanford University.

Cochrane, J.H.

(1988b) A Critique of the Application of Unit Root Tests. Manuscript, Department of Economics, University of Chicago.

Dickey, D.A. and Fuller, W.A.

(1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74: 427-431.

Engle, R.E., Granger, C.W.J., and Hallman, J.

(1987) Combining Short and Long Run Forecasts: An Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting. Manuscript, Department of Economics, University of California, San Diego.

Galli, J.

(1988) How Well Does the AD/AS Model Fit Postwar U.S. Data? Manuscript, Department of Economics, MIT.

King, R., Plosser, C., Stock, J.H., and Watson, M.W.

(1987) Stochastic Trends and Economic Fluctuations. NBER Discussion Paper No. 2229.

Pantula, S.G.

(1988) Asymptotic Distributions of the Unit-Root Tests when the Process is Nearly Stationary. Manuscript, Department of Statistics, North Carolina State University.

__________ and Hall, A.

(1988) Testing for Unit Roots in Autoregressive Moving Average Models: An Instrumental Variable Approach. Institute of Statistics Mimeograph Series No. 1924, North Carolina State University.

Phillips, P.C.B.

(1987) Time Series Regression With a Unit Root. *Econometrica*, 55: 277-302.

Quah, D.

(1989) Permanent and Transitory Movements in Labor Income: An Explanation for 'Excess Smoothness' in Consumption. Manuscript, Department of Economics, M.I.T.

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.

Shapiro, M.D. and Watson, M.W.
(1988) Sources of Business Cycle Fluctuations. *NBER Macroeconomics Annual*, 3: 111-156. Cambridge, Mass.: M.I.T. Press.