Further Investigation of the Uncertain Unit Root in GNP

Yin-Wong Cheung; Menzie D. Chinn

Journal of Business & Economic Statistics, Vol. 15, No. 1 (Jan., 1997), 68-73.

Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199701%2915%3A1%3C68%3AFIOTUU%3E2.0.CO%3B2-5

Journal of Business & Economic Statistics is currently published by American Statistical Association.

Further Investigation of the Uncertain Unit Root in GNP

Yin-Wong Cheung and Menzie D. Chinn
Department of Economics, University of California, Santa Cruz, CA 95064

A more powerful version of the augmented Dickey–Fuller test and a test that has trend stationarity as the null are applied to U.S. gross national product. Simulated critical values generated from plausible trend- and difference-stationary models are used to minimize possible finite-sample biases. The discriminatory power of the two tests is evaluated using alternative-specific rejection frequencies. For postwar quarterly data, these two tests do not provide a definite conclusion. When analyzing annual data over the 1869–1986 period, however, the unit-root null is rejected, but the trend-stationary null is not.

KEY WORDS: Output persistence; Sample-specific critical value; Stationarity test; Unit-root test.

Output persistence is one of the most debated issues in macroeconomics. In the wake of the seminal work by Nelson and Plosser (1982), a large literature testing for unit roots was spawned, including the work of Stock and Watson (1986), Perron and Phillips (1987), Campbell and Mankiw (1987), and Evans (1989), to name but a few studies that have failed to reject the presence of a unit root in gross national product (GNP). Recently, concern has arisen regarding the low power of conventional unit-root tests, such as the augmented Dickey–Fuller (ADF) test, and consequently the apparent finding of a unit root in GNP data using these tests. For instance, Christiano and Eichenbaum (1990), Stock (1991), Rudebusch (1992, 1993), and DeJong, Nankervis, Savin, and Whiteman (1992) showed that the ADF test has low power to differentiate between the trend- and difference-stationary properties of GNP.

This article adopts a different approach to the study of the persistence of U.S. GNP. First, instead of the standard ADF test, we use the ADF-GLS$^{\tau}$ test of Elliott, Rothenberg, and Stock (1992). They showed that the modified ADF test is more powerful than the original ADF test and is approximately uniformly most power invariant.

Second, the commonly used ADF test has the unit root, or $I(1)$, process as the null hypothesis. In addition to the aforementioned power consideration, the use of ADF tests also gives the unit-root specification the benefit of a doubt. In particular, we reject the unit-root specification *only* if there is strong evidence against it. To account for this asymmetric treatment, we also examine the results from a unit-root test that has trend stationarity, or $I(0)$, as the null. The test employed is Leybourne and McCabe's (1994) version of the Kwiatkowski, Phillips, Schmidt, and Shin (1992) test [hereafter KPSS(LM)].

Third, simulated critical values generated from plausible trend and difference stationary models for GNP data are used to minimize the possible biases induced by nuisance parameters in finite samples. The ability of these two tests to discriminate against a plausible alternative is evaluated using alternative-specific rejection frequencies.

Fourth, to evaluate the implication of extending the span of the data on the ability to make clear inferences regarding the presence of unit roots, we examine both postwar quarterly data and a longer annual series spanning the period 1869 to 1986.

Stock (1994) outlined at least four motivations for research in the area, which we shall repeat very briefly. First, one is always interested in characterizing the data—what are the autoregressive properties of GNP, the degree of persistence, and so forth. Second, the particular manner in which univariate forecasting is conducted will depend on the answers to the first question. Third, knowledge of the time series properties of the data is crucial to conducting proper inference in a multivariate context. Fourth, the presence or absence of a unit root may have implications for assessing certain economic theories.

To anticipate our results, we find that for quarterly data, these two unit-root tests do not provide a definite conclusion regarding the existence of a unit root in GNP data. Neither the trend- nor difference-stationary null hypotheses can be rejected when using null-specific critical values. We also observe, however, that the alternative-specific power of both tests is low so that no unambiguous conclusions can be made. Hence, we confirm the Rudebusch (1993) results for this dataset. In contrast, when analyzing annual data over the 1869–1986 period, we obtain very sharp results: The unit-root null is rejected, but the trend-stationary null is not. Moreover, the alternative-specific power for the test with a trend-stationary null is fairly high. We conclude that with a longer span of data one can obtain strong evidence of trend stationarity in per capita GNP.

It is of interest to compare our approach and results with those obtained by Burke (1994). He contrasted the results from the standard ADF test and a test with a trend-stationary null (Kwiatkowski et al. 1992, hereafter KPSS) when applied to various macroeconomic time series. The agreement of the tests is taken to be "confirmatory." He argued on the basis of simulation evidence that such confirmatory data analysis can be a useful guide to making inference regarding the time series properties of macroeco-

Journal of Business & Economic Statistics
January 1997, Vol. 15, No. 1

nomic variables. A related approach was pursued in Cheung and Chinn's (1996) analysis of GNP series in 126 countries.

The outline of the article is as follows. In Section 1, the methodology is described. Empirical results are presented in Section 2. Section 3 interprets the pattern of results. Section 4 concludes.

1. METHODOLOGY

1.1 Overview

First, we identify the most plausible trend-stationary autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) representations for the GNP data. Second, both the estimated trend-stationary ARMA process and the estimated difference-stationary ARIMA process are used to generate the empirical distributions of the ADF-GLS$^{\tau}$ and the KPSS(LM) tests. The empirical distribution of the ADF-GLS$^{\tau}$ [KPSS(LM)] statistic computed from the estimated difference-stationary ARIMA (trend-stationary ARMA) process provides the null-specific critical values to test the unit-root (trend-stationary) null against the trend-stationary (unit-root) alternative. Information on the ability of these two tests to reject the plausible alternative dynamic specification is given by the other two empirical distributions. The size-adjusted power of the test is then obtained using the null-specific critical value.

If the ADF-GLS$^{\tau}$ rejects the unit-root null and the KPSS(LM) test fails to reject the stationary null, this result is considered strong evidence in favor of a trend-stationary specification for the GNP data. If, in contrast, the ADF-GLS$^{\tau}$ fails to reject but KPSS(LM) rejects, we obtain strong evidence in support of a difference-stationary GNP process. If both tests fail to reject their respective null hypotheses, we then conclude that the data do not contain sufficient information to discriminate between the difference- and trend-stationary hypotheses. A more complicated situation occurs when both tests reject their respective null hypotheses. This outcome may indicate that the underlying data-generating mechanism is more complex than that captured by standard linear time series models.

1.2 Identification

The first step is to identify and estimate the ARMA and ARIMA processes that best describe the respective trend- and difference-stationary hypotheses. For the first case, various ARMA processes are fitted to the data,

$$y_t = \mu + \beta t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{1}$$

where $\{y_t\}$ is log real per capita GNP. The final specification [here, and in Eq. (3)] is chosen from models with the lag parameters p and q ranging from 0 to 5 using the Schwarz (1978) information criterion (SIC). As long as the true lag parameters are less than 5, the SIC will select the true model with probability 1 in large samples (Hannan 1980). Hall (1994) showed that the use of lag-selection criteria such as the SIC can improve both the size and the power of conventional unit-root tests. The Box–Ljung statistic is used to ensure that there is no significant serial correlation in the residuals of the selected model specification.

For the second case, the relevant series is first-differenced, and then an ARMA process is fit to the differenced series,

$$(1-L)y_t = \mu + \sum_{i=1}^{p} \phi_i (1-L) y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{2}$$

where L is the lag operator. The same selection criterion is applied. We label the selected specifications as the TS and DS models.

1.3 The ADF-GLS$^{\tau}$ Test

The ADF-GLS$^{\tau}$ test is carried out using the following regression:

$$(1-L)y_t^{\tau} = a_0 y_{t-1}^{\tau} + \sum_{j=1}^{p} a_j (1-L) y_{t-j}^{\tau} + \omega_t, \tag{3}$$

where y_t^{τ}, the locally detrended data process under the local alternative of $\bar{\alpha}$, is given by

$$y_t^{\tau} = y_t - \tilde{\beta}' z_t, \tag{4}$$

with $z_t = (1, t)'$ and $\tilde{\beta}$ being the regression coefficient of $\tilde{y}_t$ on $\tilde{z}_t$, for which

$$\begin{aligned}(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_T) &= (y_1, (1-\bar{\alpha}L)y_2, \ldots, (1-\bar{\alpha}L)y_T)\\ (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_T) &= (z_1, (1-\bar{\alpha}L)z_2, \ldots, (1-\bar{\alpha}L)z_T).\end{aligned}$$

The ADF-GLS$^{\tau}$ test statistic is given by the usual t statistic testing $a_0 = 0$ against the alternative of $a_0 < 0$ in Regression (4). Elliott et al. (1992) recommended that the parameter $\bar{c}$, which defines the local alternative through $\bar{\alpha} = 1 + \bar{c}/T$, be set equal to -13.5. Critical values for the ADF-GLS$^{\tau}$ test statistic were provided by Elliott et al. (1992, table 1) using the Monte Carlo method. Cheung and Lai (1995) provided finite-sample critical values for this test. It can be shown that the ADF-GLS$^{\tau}$ test can achieve a substantial gain in power over conventional unit-root tests.

For this exercise, we generate null specific critical values using the selected DS specification. If the ADF-GLS$^{\tau}$ statistic exceeds the null specific critical value, then we reject the difference-stationary null. If the test fails to reject the null, then it is important to assess the "size-adjusted" power of the test. This can be done, given an empirical size of a test, by inspecting the empirical distribution of the TS model and calculating the proportion of times the ADF-GLS$^{\tau}$ test statistic exceeds the null specific critical value. Both the null specific critical value and the alternative-specific power are generated based on 10,000 replications of the relevant process.

1.4 The Leybourne and McCabe Test

To examine the dynamic properties of GNP in a symmetric manner, we apply the KPSS(LM) procedure to test the trend-stationary null hypothesis against the unit-root alternative. The Leybourne and McCabe (1994) test is implemented because it is comparable to the ADF-GLS$^{\tau}$ test, which uses the parametric autoregressive model to account for serial correlation in constructing the test statistic. Leybourne and McCabe also asserted that their test provides more robust inference (in particular they argued that their test statistic converges at a rate faster than the KPSS test statistic). The procedure assumes that the time series is the sum of a deterministic trend, a random walk, and a stationary error and hence is a Lagrange multiplier test for the null hypothesis that the error variance in the random-walk component of the series is 0.

Consider the model

$$\begin{aligned} \Phi(L)y_t &= \alpha_t + \beta t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + \eta_t, \end{aligned} \tag{5}$$

where ε_t iid$(0, \sigma_\varepsilon^2)$ and η_t iid$(0, \sigma_\eta^2)$. This expression can be rewritten as

$$\begin{aligned} \Phi(L)(1-L)y_t &= \eta_t + \beta + \varepsilon_t - \varepsilon_{t-1} \\ &= \beta + (1 - \theta L)\zeta_t, \qquad 0 < \theta < 1, \end{aligned} \tag{6}$$

with ζ_t distributed as iid$(0, \sigma_\zeta^2)$; further assume $(1 - \theta L)$ is not a factor of $\Phi(L)$. Notice that, when $\sigma_\eta^2 = 0$, then Equation (6) reduces to

$$\Phi(L)y_t = \alpha + \beta t + \varepsilon_t. \tag{6$'$}$$

Hence, the hypothesis to be tested is

$$\begin{aligned} H_0\colon \sigma_\eta^2 &= 0; \quad \text{that is, ARMA}(p, 0) \\ H_A\colon \sigma_\eta^2 &> 0; \quad \text{that is, ARIMA}(p, 1, 1). \end{aligned}$$

The first step in the test procedure is to obtain the maximum likelihood estimates ($\hat{\phi}_i$'s) of the ϕ_i's in (6). Under both the null and the alternative, the estimates thus derived are asymptotically unbiased.

Table 1. Time Series Representations for Quarterly GNP Per Capita

	DS model	*TS model*
Constant	.0028	−.2378
	(.0008)	(.0090)
Time (×1000)		.2230
		(.0885)
ϕ_1	.3745	1.3456
	(.0693)	(.0692)
ϕ_2		−.3970
		(.0694)
SER	.0095	.0094
$Q(10)$	7.65	7.29
$Q(20)$	15.78	13.90
Roots	1.0000	.9087
	.3745	.4368

NOTE: The sample is 1948.1—1993.2. The dependent variable is log real per capita GNP. The DS model selected is an ARIMA(1, 1, 0) model with a constant. The TS model selected is an ARIMA(2, 0, 0) model with constant and trend. ϕ_i is the estimate of the ith-order autoregressive coefficient. Time (×1000) is the coefficient on time, multiplied by 1,000. SER is the standard error of regression. $Q(j)$ is the Ljung—Box Q statistic for serial correlation of the 1st to jth residuals. "Roots" are the roots of the AR polynomial.

In the second step, the variable y_t^* is generated, where y_t^* is defined as

$$y_t^* = y_t - \sum_{i=1}^{p} \hat{\phi}_i y_{t-i}. \tag{7}$$

For the tests of mean stationarity and trend stationarity, respectively, one calculates the following residuals:

$$\begin{aligned} \hat{\varepsilon}_t \quad \text{from} \quad y_t^* &= c + \varepsilon_t \\ \hat{\varepsilon}_t \quad \text{from} \quad y_t^* &= c + bt + \varepsilon_t. \end{aligned} \tag{8}$$

The test statistics are generated in the third step. Let $\hat{\sigma}_\varepsilon^2 = (\hat{\varepsilon}'\hat{\varepsilon})/T$. Then Leybourne and McCabe (1994) showed that the two test statistics $\hat{s}_\alpha$ and $\hat{s}_\beta$, respectively, are given by

$$\begin{aligned} \hat{S}_\alpha &= \hat{\sigma}_\varepsilon^2 T^{-2} \hat{\varepsilon}' V \hat{\varepsilon} \\ \hat{S}_\beta &= \hat{\sigma}_\varepsilon^2 T^{-2} \hat{\varepsilon}' V \hat{\varepsilon}, \end{aligned} \tag{9}$$

where $V = \{v_{ij}\}_{i,j=1,\ldots,T}$ and $v_{ij} = \min(i, j)$. These test statistics have the same asymptotic critical values as the KPSS $\hat{\eta}_t$ test statistics. The major difference between the two procedures lies in the manner in which serial correlation is addressed. The Leybourne and McCabe procedure deals with it parametrically, but the KPSS procedure uses a nonparametric robust estimator to account for serial correlation. The results based on the latter approach, reported by Cheung and Chinn (1995), are qualitatively the same as those obtained using the Leybourne and McCabe procedure.

From the simulated empirical distribution of the TS model, we obtain the null specific critical values. For each critical value, we can then obtain the DS alternative-specific power.

2. EMPIRICAL RESULTS

Data on quarterly U.S. GNP in 1987$ and total population from CITIBASE are used to construct the real per capita output series. The data span 1948:1–1993:2. For the TS model [see Eq. (1)], an ARMA(2, 0) with constant and trend was selected. For the DS model [(Eq. (2)], an ARIMA(1, 1, 0) process with constant is selected by SIC. This model is also selected by the Akaike information criterion (AIC). In general, the models reported are chosen by both AIC and SIC. These model estimates are reported in Table 1. In both cases, the Ljung–Box Q statistics indicate insignificant serial correlation in the residuals. It is interesting to note that the largest characteristic root of the ARMA(2, 0) process is approximately .91, which is substantially less than unity. Both of the estimated models closely resemble those obtained by Campbell and Mankiw (1987) and Rudebusch (1993), so our results are not specific to the dataset we used.

We first examine the characteristics of the unit-root test. The simulated critical values for the 10%, 5%, and 1% marginal significance levels (MSL's) are reported in the top part of Panel A of Table 2. These critical values are quite similar to those tabulated by Elliott et al. (1992). In the bottom part of Panel A, the size-adjusted power for each MSL is reported, assuming the given trend-stationary ARMA(2, 0) alternative hypothesis.

Table 2. Empirical Size and Corresponding Alternative-Specific Power for Quarterly GNP Per Capita Data

	Marginal significance level			
	10%	5%	1%	Actual
	Panel A: ADF-GLS$^\tau$ test			
c.v.	−2.6550	−2.9587	−3.5223	−2.3401
Power	47.08%	27.99%	7.47%	
	Panel B: KPSS(LM) test			
c.v.	1.5270	1.8810	2.4958	1.3734
Power	48.78%	37.24%	22.38%	

NOTE: In Panel A, c.v. indicates the finite-sample critical value corresponding to the indicated MSL for the simulated difference-stationary null hypothesis, and Power is the empirical power associated with each MSL, for the specific simulated trend-stationary alternative. In Panel B, c.v. indicates the finite-sample critical value corresponding to the indicated MSL for the simulated trend-stationary null hypothesis, and Power is the empirical power associated with each MSL, for the specific simulated difference-stationary alternative. The statistics reported under the heading "Actual" are the sample statistics calculated from the data. The DS and TS models are described in Table 1.

The ADF-GLS$^\tau$ statistic is -2.3401, which is larger than the 10% critical value; hence, we fail to reject the null hypothesis of a unit root in per capita GNP. The statistic is computed from a lag 2 specification selected by the SIC. This is the same lag structure identified in Table 1. Using the 10% critical value, the alternative-specific power is less than 50%.

We now turn our attention to viewing the trend-stationary null test (see Panel B of Table 2). In the top half of Panel B, the null-specific critical values are presented (for the $\hat{s}_\beta$ statistic because a mean-stationary process is clearly irrelevant). In the bottom half of Panel B are the associated alternative-specific levels of power. The null specific critical values are quite different from those asymptotic critical values reported by KPSS (1992). The null specific critical values are given as 1.5270, 1.8810, and 2.4958 for the 10%, 5%, and 1% MSL's. The actual $\hat{s}_\beta$ statistic is 1.3734, so we fail to reject the trend-stationary null. If the KPSS (1992) asymptotic critical values (which are .119, .146, and .176 for the 10%, 5%, and 1% MSL's, respectively) are used, however, then we *would* reject at the 1% level. This contrast in results is indicative of the extreme sensitivity of this procedure and the consequent importance of adjusting for finite-sample biases.

As mentioned previously, the KPSS(LM) test statistic fails to reject the trend-stationary null at the 10% significance level; the corresponding size-adjusted power is again lower than 50%.

In sum, the ADF-GLS$^\tau$ test cannot reject the unit-root null hypotheses, and the KPSS(LM) test does not reject the trend-stationary null. We consider this outcome as evidence of the low power of the tests. Hence, the quarterly per capita GNP data series, which has a span of about 40 years, appears uninformative with regard to the presence or absence of a unit root.

We now turn our attention to the annual dataset, which spans the period from 1869 to 1986. (In the following section a more detailed discussion of the construction of these annual data and their possible implications on the test results is presented.) An ARIMA(1, 1, 0) specification is selected as the DS model, and an ARMA(2, 0) specification is selected for the TS model. The model estimates are reported in Table 3. The Box–Ljung statistic, again, indicates a satisfactory fit.

The null-specific critical values and alternative specific power are presented in Table 4. Consistent with the previous case, the ADF-GLS$^\tau$ test appears to be more robust than the KPSS(LM) test. The test results are quite different, however, from those obtained from the quarterly data. The two tests combined together provide strong evidence of a trend-stationary GNP series. For this historical annual data, the ADF-GLS$^\tau$ rejects the unit-root null at 1% MSL, and the KPSS(LM) test fails to reject the trend-stationary alternative. The alternative-specific and size-adjusted power levels are 90% for the ADF-GLS$^\tau$ test and 75% for the KPSS(LM) test. The relatively high power for the tests further reinforces the test result.

To check the robustness of the trend-stationarity result to different data series of the same length, we also applied

Table 3. Identification of Time Series Representations for Annual GNP Per Capita

	DS model	TS model
Constant	.0138	.1523
	(.0059)	(.0352)
Time (×1000)	3.5183	
		(.8704)
ϕ_1	.2128	1.1092
	(.0911)	(.0896)
ϕ_2		−.3161
		(.0896)
SER	.0614	.0580
$Q(10)$	13.68	8.18
$Q(20)$	19.47	14.0
Roots	1.0000	.5959
	.2128	$\pm.0921i$

NOTE: The sample is 1869—1986. The dependent variable is log real per capita GNP. The DS model selected is an ARIMA(1, 1, 0) model with a constant. The TS model selected is an ARIMA(2, 0, 0) model with constant and trend. ϕ_i is the estimate of the ith-order autoregressive coefficient. Time (×1000) is the coefficient on time, multiplied by 1,000. SER is the standard error of regression. $Q(j)$ is the Ljung—Box Q statistic for serial correlation of the 1st to jth residuals. "Roots" are the roots of the AR polynomial.

Table 4. Empirical Size and Corresponding Alternative-Specific Power for Annual Per Capita Data

	Marginal significance level			
	10%	5%	1%	Actual
	Panel A: ADF-GLS$^\tau$ test			
c.v.	−2.6992	−2.9719	−3.5668	−4.1139
Power	99.83%	99.14%	89.77%	
	Panel B: KPSS(LM) test			
c.v.	.2326	.3865	.7586	.0334
Power	90.75%	86.28%	75.29%	

NOTE: In Panel A, c.v. indicates the finite-sample critical value corresponding to the indicated MSL for the simulated difference-stationary null hypothesis, and Power is the empirical power associated with each MSL, for the specific simulated trend-stationary alternative. In Panel B, c.v. indicates the finite-sample critical value corresponding to the indicated MSL for the simulated trend-stationary null hypothesis, and Power is the empirical power associated with each MSL, for the specific simulated difference-stationary alternative. The statistics reported under the heading "Actual" are the sample statistics calculated from the data. The DS and TS models are described in Table 3.

the same procedure to the historical data series that were reported by, for example, Romer (1989) and Balke and Gordon (1989), both also extended to 1986. Similar evidence in support of trend stationarity is obtained.

3. INTERPRETATION OF THE RESULTS

The strong contrast in the results obtained from the postwar quarterly data and the long span of annual data is consistent with several possibilities. The first possibility is that the method by which the GNP data was constructed during the prewar period artificially induces a finding of trend stationarity. Consider for instance the possibility that the GNP was linearly interpolated between benchmark years. This would certainly "smooth" the series so that a trend-stationary specification would seem to fit. It turns out that this is not a viable explanation.

The series used in the study represents a combination of work by Kuznets (1961), Kendrick (1961), and Gallman (1966) for the period 1869–1908, and the Kuznets series, as reported by the U.S. Department of Commerce (1975), is used for the period 1909–1928. Finally, the post-1928 data are the conventional figures from the National Income and Product Accounts (U.S. Department of Commerce 1986). Because all the alternative measures of GNP [i.e., those of Romer (1989) and Balke and Gordon (1989)] are the same for these years, it is the construction of pre-Depression data that is in question. The 1919–1928 data are generally regarded as accurate because income data were available. Prior to 1919, such income data were not available, so GNP had to be estimated. The following discussion of this estimation procedure draws on the work of Romer (1989), who has written extensively on the issue of GNP data construction during this early period.

The Kuznets GNP series for the period extending up to 1918 is based on annual commodity data from state and industry sources, adjusted so as to match the detailed figures obtained from the Census of Manufactures, the Census of Agriculture, and the Census of Mines. These estimates of commodity output are widely regarded as quite accurate. The problem is converting estimates of commodity output to GNP. The commodity output is valued at producer, rather than final, prices. Furthermore, all value added associated with services must be estimated. Kuznets assumed that the service value added moves one for one with commodity output. Although this assumption is clearly problematic for some questions, it is crucial to understand that this procedure does *not* smooth output; rather it accentuates the fluctuations and in fact imparts whatever time series properties commodity output has to the estimate of GNP. Elliott et al. (1992) indicated that conditional volatility tends to *reduce* the power of the unit-root test. Thus, it seems unlikely that the trend-stationarity result for the historical annual data is driven by the data-construction procedure.

The second possibility is that the true data-generating process may have changed between the postwar and prewar periods. On the surface, this interpretation is consistent with the change in the roots of the AR(2) of the trend-stationary specification. The estimated root for the postwar quarterly data is .9087, and that for the 1869–1986 period annual data is .6030. The degree of persistence, in annual terms, implied by the estimated root, however, is .6851 ($= .9087^4$), which is very close to that obtained from annual data. The two estimated roots, when placed in consistent terms, are therefore very close and do *not* indicate a substantial change in the time series process for GNP.

Another possibility is that the existence of structural breaks spuriously induces our results. Although there is some evidence of structural breaks, it does not necessarily argue against our interpretation of the statistical results. We applied the structural break test of Banerjee, Lumsdaine, and Stock (1992) and found mixed evidence for two breaks in trend in the annual data, in 1924 and 1942. Neither of these dates corresponds to changes in the way the GNP data are constructed. Furthermore, the fact that we reject the unit-root null using the ADF-GLS$^\tau$ test in the presence of possible structural breaks can be construed as very strong evidence of trend stationarity because such breaks typically bias unit-root tests against rejecting the null.

The third, and final, possibility is that the result is because a longer span of data allows one to detect trend reversion much more readily. To investigate this possibility, we applied our procedure to the Nelson and Plosser (1982) GNP series, which spans the period from 1909 to 1970. Neither the unit-root nor the trend-stationarity tests rejected their respective null hypotheses. Consequently, we cannot conclude from the Nelson and Plosser sample whether GNP has a unit root. This further emphasizes the importance of the time span of data in the study of persistence in GNP (see also Perron 1989; Shiller and Perron 1985).

As pointed out by one of the referees, the first-differencing of an $I(0)$ process should induce a moving average unit-root process. When estimating an ARIMA(2, 1, 1) for quarterly data, the moving average coefficient is $-.442$. This estimate, which is similar to other estimates reported in the literature, appears to be a robust feature of this quarterly data series [e.g., Campbell and Mankiw (1987) obtained an estimate of $-.455$]. In contrast, the ARIMA(2, 1, 1) estimate for annual data yields a moving average coefficient of 1.0258, very close to unity. This pattern of results is to be expected given the more obviously trend-stationary behavior of the annual series.

4. CONCLUDING REMARKS

This article reports the results of a study in which the issue of a unit root in GNP is examined from two perspectives–from the unit-root null as well as from the trend-stationary null. We have taken advantage of the latest econometric technology, including the more powerful unit-root test developed by Elliott et al. (1992), as well as the Leybourne and McCabe (1994) trend-stationarity test. Data-specific empirical distributions are used to mitigate possible finite-sample biases. We also explore the sensitivity of the results to data length.

Our exercise has shown that stationarity tests and unit-root tests can be used in a complementary manner to yield useful insights on persistence. Data-specific empirical distributions can be crucial in the making of inferences when the root considered is very close to unity. Furthermore, the ability to distinguish between an $I(1)$ or an $I(0)$ specification from GNP data depends on the power of the tests, which in turn is related to the information content of the data.

ACKNOWLEDGMENTS

We thank Tim Cogley, Francis X. Diebold, Alastair Hall, Louis Johnston, Kon S. Lai, Peter C. B. Phillips, Mark W. Watson (the editor), the associate editor, and three anonymous referees for useful comments. Financial support was provided by the University of California Pacific Rim Program.

[Received January 1995. Revised January 1996.]

REFERENCES

Balke, N. S., and Gordon, R. J. (1989), "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy*, 97, 38–92.

Banerjee, A., Lumsdaine, R., and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10, 271–287.

Burke, S. P. (1994), "Confirmatory Data Analysis: The Joint Application of Stationarity and Unit Root Tests," Discussion Papers in Quantitative Economics and Computing No. 20, University of Reading, Dept. of Economics.

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

Cheung, Y.-W., and Chinn, M. D. (1995), "Further Investigation of the Uncertain Unit Root in GNP," working paper, University of California at Santa Cruz, Dept. of Economics.

——— (1996), "Deterministic, Stochastic and Segmented Trends in Aggregate Output: A Cross-Country Analysis," *Oxford Economics Papers*, 48, 134–162.

Cheung, Y.-W., and Lai, K. S. (1995), "Lag Order and Critical Values of a Modified Dickey–Fuller Test," *Oxford Bulletin of Economics and Statistics*, 57, 411–419.

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

DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992), "Integration Versus Trend Stationarity in Time Series," *Econometrica*, 60, 423–433.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1992), "Efficient Tests for an Autoregressive Unit Root," Technical Working Paper 130, National Bureau of Economic Research, Cambridge, MA.

Evans, G. W. (1989), "Output and Unemployment Dynamics in the U.S.," *Journal of Applied Econometrics*, 4, 213–237.

Gallman, R. E. (1966), "Gross National Product in the U.S., 1834–1909," in *Output, Employment and Productivity in the United States* (Studies in Income and Wealth, Vol. 30), New York: Columbia University Press for National Bureau of Economic Research, pp. 3–90.

Hall, A. (1994), "Testing for a Unit Root in Time Series With Pretest Data-Based Model Selection," *Journal of Business & Economic Statistics*, 12, 461–470.

Hannan, E. J. (1980), "The Estimation of the Order of an ARMA Process," *The Annals of Statistics*, 8, 1071–1081.

Kendrick, J. W. (1961), *Productivity Trends in the U.S.*, Princeton, NJ: Princeton University Press for National Bureau of Economic Research.

Kuznets, S. (1961), *Capital in the American Economy*, Princeton, NJ: Princeton University Press for National Bureau of Economic Research.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" *Journal of Econometrics*, 54, 159–178.

Leybourne, S. J., and McCabe, B. P. M. (1994), "A Consistent Test for a Unit Root," *Journal of Business & Economic Statistics*, 12, 157–166.

Nelson, C., and Plosser, C. (1982), "Trends and Random Walks in Macro-Economic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Perron, P. (1989), "Testing for a Random Walk: A Simulation Experiment When the Sampling Interval Is Varied," in *Advances in Econometrics and Modeling*, ed. B. Raj, Boston: Kluwer, pp. 47–68.

Perron, P., and Phillips, P. C. B. (1987), "Does GNP Have a Unit Root? A Re-evaluation," *Economic Letters*, 23, 139–145.

Romer, C. D. (1989), "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869–1908," *Journal of Political Economy*, 97, 1–37.

Rudebusch, G. D. (1992), "Trends and Random Walks in Macroeconomic Time Series: A Re-examination," *International Economic Review*, 33, 661–680.

——— (1993), "The Uncertain Unit Root in Real GNP," *American Economic Review*, 83, 264–272.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shiller, R. J., and Perron, P. (1985), "Testing the Random Walk Hypothesis: Power Versus Frequency of Observation," *Economic Letters*, 18, 381–386.

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

——— (1994), "Unit Roots, Structural Breaks and Trends," chap. 46 in *Handbook of Econometrics* (Vol. 4), eds. Z. Griliches and M. D. Intriligator, New York: North-Holland, pp. 2740–2841.

Stock, J. H., and Watson, M. W. (1986), "Does GNP Have a Unit Root?" *Economic Letters*, 22, 147–151.

U.S. Department of Commerce (1975), *Historical Statistics of the United States, Colonial Times to 1970*, Washington, DC: U.S. Government Printing Office.

——— (1986), *The National Income and Product Accounts of the United States, 1929–82*, Washington, DC: U.S. Government Printing Office.