THE NONSTATIONARITY OF AGGREGATE OUTPUT:
A MULTI-COUNTRY PERSPECTIVE

Working Paper #558
Roger C. Kormendi*
Philip Maguire**
The University of Michigan

ABSTRACT

We compute the scaled variogram (the variances of successive long differences scaled by the variance of first differences) for the log of annual per capita real aggregate output as measured by a) the long series on US GNP and UK GDP; b) Maddison's [1982] long series of GDP index numbers for 12 countries; and c) the postwar IFS data on GDP (GNP) for 32 countries. We use simulations to show that, contrary to previous results for the US alone, the scaled variograms are consistent with the presence of a substantial unit root component in the univariate representation of these series. We also show that the power of the variogram to discriminate between parsimonious trend (TS) and difference stationary (DS) processes in the IFS data is somewhat greater than that of a Dickey-Fuller-like F test.

KEY WORDS: Variogram, unit root, stationarity, time series analysis, GNP, simulation

I.   INTRODUCTION

Whether the univariate time series representation of real GNP contains a unit root instead of a linear trend has been much debated of late. The straightforward application of Box-Jenkins identification techniques strongly suggests the presence of a unit root, in that the autocorrelations of the first differences of the log of real GNP are all smaller than 2 (asymptotic) standard errors after the second lag or so. The tests derived by Dickey and Fuller concur in this conclusion.[1] The classic reference is of course Nelson and Plosser [1982], who find that a unit root dominates a linear trend in nearly every long US macroeconomic series they examine. Similar conclusions, for US GNP alone, have also been reported by Harvey [1986], Rose [1986], Stock and Watson [1986], Deaton [1986], Campbell and Mankiw [1987], Campbell and Deaton [1987] and Schwert [1987]. Finally, Kormendi and Meguire [1984] found that postwar M1 and real GDP from 47 countries contain unit roots.

In a stimulating paper that uses a radically different technique from these authors, Cochrane [1987] reaches a contrary conclusion. He finds that less than 20% of the variance of the growth rate of the log of annual real per capita US GNP (henceforth simply GNP) over the period 1869-1984 can be attributed to the presence of a unit root.[2] Cochrane contends that previous evidence in favor of the presence of a unit root in US GNP is misleading. He argues that the evidence from the autocorrelation function of first differences has been misinterpreted because the high-order autocorrelations are ignored. These are all small but mostly negative, which may be is consistent with overdifferencing. As for tests of the Dickey-Fuller type, Cochrane restates the well-known objection that they

---

[1] See Dickey, Bell and Miller [1986] for a discussion and review of these tests, with references and applications.

[2] Watson [1986], applying an unobserved components decomposition to US data, reaches a similar conclusion.

lack power against stationary but highly autocorrelated alternatives. More generally, Cochrane argues that it is invalid to infer, using what he terms "ad hoc identifying restrictions," the long run dynamics of a series, such as the presence of a unit root, from the short run dynamic information revealed by standard parsimonious Box-Jenkins identification procedures. His preferred in-sample model for US GNP among those he considered is a stationary AR(2) about a deterministic linear trend.

Cochrane's procedure for detecting a unit root in a time series is quite simple, and derives from the following property of random walks. For any time series $\{X_t\}$, let $\sigma^2(k) \equiv \sigma^2(X_t - X_{t-k})$. Then if $\{X_t\}$ follows a random walk, $\sigma^2(k) = k\sigma^2(1)$ holds in population for all positive integers k. We will call the successive overlapping quantities $X_t - X_{t-k}$ the kth differences of the series $\{X_t\}$. We will call the sequence $V(k) \equiv \sigma^2(k)/k$, $k=1,\ldots,n$ the variogram of $\{X_t\}$.[3] Borrowing from the frequency domain, we call the single term $\sigma^2(k)/k$ the kth ordinate of the variogram. Likewise, by analogy with the time domain, we call the degree of differencing k, the lag of the ordinate. A plot of the variogram of a random walk against its lags is then a horizontal line. More generally, the variogram of a series with a unit root tends to some fixed positive value as k grows large.

Cochrane's estimated variogram for US GNP over the period 1869-1984, k =

---

[3] The variogram as defined here is a one-dimensional (time) special case of a more general technique, known for some years in the geostatistics and meteorological literatures, used to analyze spatial nonstationarity and autocorrelation in 2 and 3 dimensions. We have appropriated the term variogram from this literature. See Cressie [1986] and references therein, especially Matheron [1971]. We thank Noel Cressie for bringing these historical matters to our attention. Cochrane credits Robert Lucas with the insight that the variance of kth differences of a time series with a unit root should be k times the variance of first differences. While use of the variogram is quite new in economics and finance, it is spreading rapidly. Cf. Lo and MacKinley [1987a,b], Fama and French [1986], and Huizinga [1986].

1,...,30, is shown in Fig. 1. For lags 20 through 30, the estimated variogram is roughly flat at a value of about $.4\sigma^2(1)$. He also reported simulations showing that the variogram of US GNP was consistent with GNP following either a nonstationary ARIMA(15,1,0) or a stationary AR(2) about a deterministic trend. In the former case, the unit root would cause the variogram to asymptote to $.18\sigma^2(1)$. In the latter case, the variogram would asymptote to 0 because, as k increases, the variances of the long differences would tend to 0. Cochrane concludes "...the AR(2) about a deterministic trend best replicates the behavior of the variance of kth differences of GNP in a large class of ARMA models." Cochrane does not articulate the precise metric implied by the term "best replicates." However if imposing a unit root requires 15 AR terms in order to fit the actual variogram as well as 2 AR terms with a trend, then mere parsimony militates against the presence of a unit root.

Our interest in whether or not GNP possesses a unit root is more than merely statistical, but derives from a need to understand the time series properties of output (income) and other aggregate time series in order to address a number of macroeconomic questions. One such basic question is the relationship between measured income (GNP) and measured consumption. If GNP has a unit root, then innovations in GNP have a permanent effect on future GNP, and hence on rational forecasts thereof. The effect of innovations in GNP on consumption then depends crucially on the presence or absence of a unit root, so that a proper understanding of the time series properties of income is essential to testing capital-theoretic models of the aggregate consumption function.[4] A properly specified consumption function is also needed to analyze aspects of fiscal policy

---

[4] This point is central to the approach in Mankiw and Shapiro [1985], Deaton [1986] and Kormendi and LaHaye [1986]. See also Kormendi, LaHaye and Meguire [1986], Campbell and Mankiw [1987] and Campbell and Deaton [1987].

such as the possible substitutability between public and private consumption, the

differential effects of tax and deficit financing of government outlays, and the

aggregate effects of transfer payments.[5] Whether or not income and money have

unit roots is also important in estimating the demand for money (Fama [1982],

Mankiw and Summers [1986]), and the effects of monetary policy under rational

expectations (Kormendi and Meguire [1984], Rush [1986]).[6]

In section II, we define our estimator of the scaled variogram and study its

sampling distribution under the null of a Gaussian random walk. In section III,

we estimate the scaled variogram for a number of series on per capita real output:

US GNP (NIPA extended using Romer [1986]) and UK GDP (Feinstein [1972]) back to

1870; the GDP indeces for 12 countries from Maddison [1982], again extending back

---

[5] Cf. Kormendi [1983], Kormendi and Meguire [1986], Kormendi, LaHaye and Meguire
[1986]. Brunner [1986] proposes that any regression test of fiscal policy using
aggregate time series data should be run in differences as well as levels,
because of the adverse statistical consequences, discussed in the next paragraph
of the text, of estimating regressions over levels data containing possible unit
roots. We would apply Brunner's specific recommendation to any statistical
analysis of data where unit roots could be present. This is consistent with
Plosser, Schwert and White's [1982] advocacy of differencing as a Hausman-type
test of specification.

[6] Independently of any particular economic hypothesis, the failure to account for
possible unit roots by differencing may be a fundamental specification error
with serious statistical consequences. Regressions estimated from levels data
with unit roots can have nonstationary residuals, and inferences drawn therefrom
can be fraught with Type II errors (the "spurious regression" phenomenon; cf.
Granger and Newbold [1974], and Plosser and Schwert [1978]). A deeper problem
with such regressions is that the usual estimators may not have asymptotic
sampling distributions, especially when the regressors are not exogeneous, as
the required matrix probability limits may fail to exist. This insufficiently
appreciated point is well articulated on p. 48 in Brunner [1986], who references
Grenander [1954], Meese and Singleton [1982] and Stulz and Wasserfallen [1985].
For a recent analytical discussion of this point, see Phillips [1986], who shows
that, for spurious regressions, the usual t and F statistics do not possess any
limiting distributions. Stock [1984] and Engle and Granger [1987] argue,
however, that economic time series with unit roots may be cointegrated. Under
certain conditions, regressions relating cointegrated variables and estimated in
levels may in fact be well specified. Finally, Nelson and Kang [1981] show that
detrended data with unit roots contain spurious periodicities.

to 1870; and the postwar <u>IFS</u> data on GDP and GNP from 32 countries. In section

IV, we fit both a trend stationary (TS) AR(3) and a difference stationary (DS)

ARIMA (3,1,0) to these series. Conditional on the estimated parameters, we simu-

late the sampling distribution of selected ordinates of the variogram, and use

these distributions to compute the relative probabilities that the actual vario-

grams could have been generated by such parsimonious TS and DS models. In section

V, we compare, again using simulations, the relative power of the variogram and an

F-test in the manner of Dickey and Fuller to discriminate between TS and DS. Sec-

tion VI offers a brief summary with concluding remarks.

II. THE SCALED VARIOGRAM AND ITS SAMPLING PROPERTIES.

1. The Scaled Variogram

Let $x_0, \ldots, x_n$ be a realization from the stochastic process $\{X_t\}$. We then

compute V(k) as follows. For any positive integer k, first let

$$\sigma^2(k) \equiv \sum_{t=k}^{n} [x_t - x_{t-k} - \frac{k}{n}(x_n - x_0)]^2 \div (n-k) \qquad (1)$$

be an unbiased estimator of the variance of the kth difference of $\{X_t\}$. Cochrane

multiplies $\sigma^2(k)$ by $n/(n-k)$ as a tacit adjustment for the finite sample bias

caused by the overlap in the differences. He then divides $\sigma^2(k)$ by k so that a

plot of the variogram as a function of k has, in the case of a random walk, slope

0 instead of 1. V(k) is then[7]

$$V(k) \equiv \frac{n}{k(n-k)} \sigma^2(k) \quad . \qquad (2)$$

We go one step further and scale each variogram ordinate by the variance of first

---

[7] The variogram can also be defined in terms of the variance of (linear) filtered
first differences, where the filters are of successive length k and have equal
weights.

differences, V(1), to obtain the <u>scaled variogram</u>, with typical element R(k) ≡

V(k) ÷ V(1).[a]

The scaled variogram should be a fundamental addition to the time series

analyst's tool box, and one of the purposes of this paper is to explore its

properties and to advocate its (proper) use. The scaled variogram is a

dimensionless quantity independent of the units in which the underlying data x_t

are measured. It is a function only of the length, n, of the series, and of the

autocorrelations of its first differences. Its sampling properties do not depend

on any other nuisance parameters. The scaled variograms of all series of length n

can be readily compared, even plotted together along the same set of coordinates.

The population value of R(k) computed from a random walk is 1, for all k. As

stated in section I, if a series contains a unitroot, then the scaled variogram

flattens out to an asymptote. If this asymptote lies in the [0,1] interval, it

can be interpreted as the fraction of the variance of a series due to shocks whose

effects are permanent.[9]

---

[a] Equations (1) and (2) are equivalent to (A-3) in Cochrane [1987], except that
Cochrane multiplies the sum of squared kth differences by the incorrect factor
$n/(k(n-k)(n-k-1))$, forgetting that n is the number of first differences.
Instead of multiplying the sum of squared kth differences by $n/(n-k)^2$, Fama and
French [1986] derive the factor, $1/(n - 2k + ((k^2 -1)/3(n-k)))$. According to
their simulations, the resulting estimator for the variogram is also mean
unbiased. We will not evaluate here the relative merits of these two methods of
adjusting the degrees of freedom for bias.

[9] If $\{X_t\}$ has a more general ARIMA representation of the form ARIMA(p,1,q), then
the variogram will still possess an asymptote, at least for values of k large
enough for the effects of any AR and MA terms to have died out. This also holds
for a series following a general ARIMA(p,d,q) if the variogram is computed over
the d-1 difference of the original series. A series that requires <u>d</u> differences
in order to attain stationarity should be differenced <u>d-1</u> times before having
its (scaled) variogram computed; the resulting variogram can then be interpreted
as usual. Hence the variogram can, with suitable qualifications, serve not only
as a diagnostic procedure for determining the presence of a unit root, but also
as an a an aid to identifying the degree of differencing needed to transform to
stationarity any time series with a canonical ARIMA representation.

The variogram summarizes the covariance properties of a time series. Hence

it is related to the correlogram and to the sample spectrum. Cochrane emphasizes

that $\lim_{k \to \infty} V(k)$ is a consistent estimate of the power of $(1-L)x_t$ at frequency 0.

Also

$$R(k) = 1 + 2\sum_{j=1}^{k-1} \frac{(k-j)}{k} \, rho(k) \tag{3}$$

holds in population, so that $R(k)$ is a linear combination of the first k-1

autocorrelations (the $rho(k)$) of the differenced series.[10]

2. The Sampling Distribution of R(k) Computed from a Pure Random Walk.

Since $R(k)$ is a ratio of variances divided by their degrees of freedom, it

would appear that $R(k)$ is distributed as $F(n-k,n)$. This happy result does not

obtain because $V(k)$ and $V(1)$ are not statistically independent, as the kth

differences are linear combinations of the first differences.[11] Nonetheless, the

sampling distribution can be easily explored by simulation, as it has only two

parameters, n and k, both of which take on discrete values only. Table 1A shows

the sampling distribution of $R(k)$, k = 3, 5, 10, 20, 30, 50 and 75, computed from

---

[10] The variogram is not as directly informative as the correlogram and sample
spectrum about the detailed correlation properties of a time series. When unit
roots are present however, the interpretation of the correlogram and sample
spectrum are not straightforward; in fact the spectrum of a series with a unit
root is not even defined in population. Only the variogram remains easily
interpretable in the presence of a unit root. The variogram is also defined
for stochastic process for which the autocovariance function does not exist,
e.g., the Wiener process (see Cressie [1986]).

[11] As is well known, the overlapping kth differences are not statistically
independent, even when the first differences are. In fact, the kth differences
follow an MA(k-1) when the first differences are white noise. At a later time,
we intend to examine how well $R(k)$ conforms to $F(n-k,n)$ in spite of the lack of
independence of the numerator from the denominator.

1000 simulated random walks of length 117, corresponding to the period 1869-1985.[12]

The skewness of R(k) is blatant—for instance, note how the third quartile is considerably wider than the second one—and increases with k. Whereas the sample means of the R(k) are essentially 1, so that the formula in (2) is unbiased, the sample medians are biased downwards. Although the median of any skewed distribution is necessarily less than the mean, this bias increases with k. The dispersion of the estimates, as evidenced by the interquartile and interdecile ranges (not shown), is also increasing in k. Taking FRAC5 and FRAC95 as the lower and upper bounds of a 90% "confidence interval" under the random walk null, this interval is quite wide, especially for the larger k, e.g., about [.4, 1.9] for k = 15, and about [.2, 2.5] for k = 50. For k ≥ 30, ordinates as small as .1 and as large as 5 can be observed in 1000 replications. This large sampling variability limits the power of the scaled variogram in testing for a unit root null.[13]

Lo and MacKinley [1987a] derive analytically as well as simulate the properties of a statistic very similar to the scaled variogram.[14] However they

---

[12] The innovations are iid Gaussian, with variance, drift parameter and starting values all set to their actual values for US GNP. Again, this is immaterial because R(k) is invariant with respect to such parameters. With the exception of the calculations underlying Figs. 2 and 3, and Table 2, which were done using version 5.1 of Minitab, the simulations and estimates reported in this paper were performed using version 2.05 of RATS on a microcomputer. The Minitab and RATS programs, as well as the data, are available upon request in machine readable form from the authors. We have been unable to determine the exact algorithm embodied in the RATS function RAN that creates random normal deviates.

[13] Note that R(k) and R(k+l) (where l is an integer such that k+l ≥ 1) are definitely not statistically independent, so that it is not obvious how to distill the variogram into a "portmanteau" statistic like the Box-Pierce Q. In section V, we will use R(20) as a test statistic for making power comparisons.

[14] R(k) differs by 1 from their $M_r(q)$, as can be seen from their eqs. (9) and (14c) when k is substituted for q and n for nq, q=1,...,nq-1. They also number the first observation 1 instead of 0, so that they have n-k-1 instead of n-k kth differences. Their simulations confirm the presence of a unit root in several broad aggregates of US stock prices [1987b], for differencing intervals

fail to recognize the strong right skewness of (scaled) variogram ordinates

computed from finite samples.[15] This skewness means that it is misleading to

evaluate (as Cochrane and Lo and MacKinley do) theoretical, simulated or actual

(scaled) variograms by means of symmetric standard error bands. The resulting

"confidence" intervals do not have maximum posterior density, and hence do not

possess the natural Bayesian interpretation as intervals in which, under the null,

estimates of R(k) will be found with probability α, as would be the case with a

symmetric sampling distribution. For these reasons, we will summarize our

simulations by reporting means and selected fractiles of the empirical pdfs. All

inferences will also be based on fractiles rather than on sample means and

standard errors.

III. THE VARIOGRAM OF REAL OUTPUT

1. The Long Series on US GNP and UK GDP.

All data sources and computations are described in the Data Appendix. For

the US, we use the new and less volatile data for the period 1869-1928 recently

computed by Romer ([1986], [1987]). Fig. 2 shows that for lags 1 through 30, the

shape of the scaled variogram computed from US GNP is quite similar to that of the

(unscaled) variogram in Fig. 1 taken from Cochrane [1987]. Substantially above 1

at low lags, the variogram slopes down more or less continuously after a peak at

lag4. Variogram ordinates for those lags appearing in Table 1 are shown in the

---

up to 16 months long. Their results do not address the findings of Fama and
French [1986, Table 3], who found that the R(k) computed from similar data were
substantially below 1 for differencing intervals in the range of 5 to 10 years.

[15] The extensive simulations (20,000 replications) of Lo and Mackinley (see their
Tables 2a,b) do reveal skewness and excess kurtosis that are several standard
errors in size, especially when k/n exceeds .1 or so, even for n as large as
1024. This skewness is also suggested by the F distribution with finite
degrees of freedom.

first row of Table 2. This variogram is certainly not the one expected from a

pure iid Gaussian random walk, but instead appears consistent with Cochrane's

preference for modelling US GNP as a stationary process about a deterministic

trend. The evidence is not conclusive, however. The $R(k)$ for $k \leq 30$ in Table 2

are generally greater than those implied by Fig. 1, and fall outside of the first

decile shown in Table 1A. Moreover, $R(50)$ lies at about the 25% fractile.

However the levelling off of the variogram at about .4, which Cochrane assumed to

begin around $k = 20$, is not borne out; by lag 75, $R(k)$ declines to .15, the 5%

fractile in Table 1A.

The variogram of the long series for (log real per capita annual) GDP in the

United Kingdom, mainly due to Feinstein [1972], is shown in Fig. 3.[16] While it is

questionable whether this series can be described as a pure random walk, the

scaled variogram strongly suggests the presence of a unit root. In fact, $R(k)$

never falls below 1, although only the values for lags 3, 5 and 50 exceed the 90%

fractile shown in Table 1A. We have also plotted in Fig. 3 the 95% fractile from

Table 1A, from which it can be seen that this fractile almost forms an upper

envelope for the estimated variogram.

2. Subperiod Sensitivity and Deflation by Population.

Table 2 also explores the sensitivity of the variogram to the choice of

sample period and to the use of per capita data. We believe that per capita data

are not unambiguously preferable, because annual population estimates are largely

interpolated from vital statistics between census benchmarks. For each subperiod,

the second row in Table 2 gives variogram ordinates calculated from raw data. The

---

[16] In view of a) the existence of personal and corporate income taxation in the UK
throughout the later Victorian era; and b) the non-trivial assumptions needed
to estimate US GNP for years prior to 1909 (Romer [1986]), it is possible that
the long data for the UK are of better quality overall than those for the US,
especially prior to World War I.

per capita series generally have smaller ordinates, so that raw ouput appears more nonstationary than per capita ouput.

Table 2 also shows that the variogram is highly variable across sampling periods. For both the US and the UK, the pre-1914 period yields lower values than the other periods. The sample period used by Nelson and Plosser [1982], 1909-70, yields the largest ordinates of any subperiod shown except at lag 20. This may explain their unambiguous conclusion in favor of a unit root. Also note that only those periods which include the Depression and World War II yield high values of R(k) at low lags.[17]

Table 2 also sheds new light on recent work by Campbell and Mankiw [1987]. In a paper that is cognizant of Cochrane's work but reaches an opposite conclusion, these authors propose to measure the persistence of an innovation to a time series in two ways: as the sum of the coefficients in the (possibly infinite) moving average representation of the differenced series, and as V(k) estimated by

---

[17] The following summary statistics show that significant positive serial correlation in the growth rates of real ouput characterizes only those subperiods that include the Great Depression and one or both World Wars.

Summary Statistics by Subperiod for Log Growth Rates of:

	US GNP				UK GDP		
Period	SD	r_1		Period	SD	r_1
1869-1985	.053	.37		1870-1985	.033	.23
1869-1929	.039	-.03		1870-1913	.026	-.27
1930-1954	.096	.57		1914-1947	.048	.40
1955-1985	.024	.16		1948-1985	.019	.02

NOTE: SD = standard deviation. r_1 = first order sample autocorrelation.

As Cochrane [1987] and Lo and Mackinley [1987a] have recognized, the sampling distribution of the variogram for a random walk may be sensitive to departures from homoskedasticity in the random walk innovations. The above standard deviations by subperiod reveal such departures. Note that the high variance periods are also those with large positive serial correlation.

The effects of the pattern of heteroskedasticity shown above for the US on

substituting sample autocorrelations into (3). Both measures, when applied to quarterly US GNP (not deflated by population) over the period 1952:1-1984:3, yield estimates that are at least as great as the means of their respective simulated sampling simulations under a pure random walk null. Table 2 shows why these results favoring a unit root could be expected. First, the raw series is always more nonstationary than the per capita one, at least for the US. Second, their sample period is close to the period 1953-85, when the R(k) are fully consistent with a random walk. Note, however, that including the 5 year period 1948-52 substantially lowers the variogram estimated over the postwar era. Finally, their maximum value for k of 60 corresponds to a k of 15 with annual data, and for (annual) k in the range of 3 to 15, the R(k) for most US subperiods are consistent with a unit root.

3. Maddison's Long Series.

These radically different results for the data from two selected economies suggest two competing explanations. One is that the time series properties of aggregate output in the US and the UK are fundamentally different. The other explanation is that the US and UK series are different realizations from the same class of stochastic processes, e.g., TS or DS. In particular, the sampling dispersion of the variogram is so great that both US and UK real output could even be realizations of random walks. These possibilities motivate an exploration of data from more countries.

We turn, therefore, to the index numbers for real GDP in 12 countries

---

the sampling distribution of the scaled variogram, again for realizations of length 117, are shown in Table 1B. Relative to the homoskedastic case, the lower fractiles decrease, and the dispersion increases, but both only slightly. Overall, allowing for the observed heteroskedasticity does not materially change the probability that US GNP or UK GDP has a unit root.

computed by Maddison [1982].[18] We have used every country with continuous data

for the period 1900-79, with 10 out of 12 series extending back to 1870. These 12

countries include the US and UK, but Maddison's data for the period 1870-1947 are

not identical to those already analyzed.[19] Maddison also made a serious effort to

assure that these data were as comparable as possible across countries and through

time. The real output data are deflated by Maddison's data on annual population

at midyear.

Values of R(k) at selected lags are shown in Table 3. Two facts immediately

stand out. First, about 75% of the ordinates reported are greater than 1, with

only 4 being below .8, three of these being for the US. Second, the US is not

typical in that its variogram ordinates at longer lags are clearly the smallest of

any country in the sample.[20] Viewing these data as a whole, however, it is likely

that these 12 series contain significant unit root components.

4. IFS Postwar Data

The final data we examine are the postwar annual series on real output (GDP

or GNP) for 32 countries, taken from the International Monetary Fund's data

---

[18] These countries do not make up a representative sample from all over the globe; instead they consist of the four current principal Europeean economies, the Scandinavian nations and the Netherlands, and the three major English-speaking former colonies of the UK. Further details on data and sources are given in the Appendix.

[19] A puzzle in Romer's [1986] data is the 16.6% increase in GNP for 1872. This is the largest one year increase in the long US series except during World War II. Maddison's series only increases by 7.3% in 1872, a value we find more plausible.

[20] We discount R(50) = .35 for the Netherlands because k/n exceeds .5. The US ordinates in Table 3 are somewhat smaller than those given in Table 2. For the UK, Maddison's data yields a variogram with a less marked peak between lags 20 and 50.

[21] A country was included if it had continuous annual data on real output and population for the period 1950-83. The countries in our sample include most of the industrial/OECD countries (the exceptions being Japan, Belgium-Luxembourg, Spain and New Zealand) and a number of developing countries, especially in Latin America. Further details on data and sources can be found in the Data

base.[21]These data are likewise deflated by the IMF's annual data on midyear population.

Given that this data set includes at most 38 (annual) data points per country, one could question how such short series can address a long run dynamic property such as the presence of a unit root. In particular, Cochrane [1987] states that

> "The number of nonoverlapping 'long runs' are a rough guide to the number of degrees of freedom... With a 10 to 20 year 'long run' there are no more than ... 2 to 4 observations in postwar data. Obviously, using more frequently observed data doesn't help."

While we agree with this quotation when the focus is on a single country, we nevertheless take the view that by treating the postwar data on real output from a panel of countries as multiple realizations from either the TS or DS class, these data can shed light on whether real output contains a unit root. This is because the scaled variogram is free of all country-specific nuisance parameters. This approach was implicitly taken with Maddison's data, where short data length was not an issue. The added power from such cross-sectional pooling depends on the degree of interdependence among the country-specific growth rates. In the case of pure independence, an ergodic property should hold whereby the pooling of many short series may be as informative as one long series with the same number of data points. In the case where growth rates are perfectly correlated across countries, there is no benefit from pooling. More generally, the greater the dependence among countries' growth rates, the smaller the effective degrees of freedom and hence the lower is the power of cross-sectional analysis. We plan to investigate the extent of this dependence in future work.

Table 4, which should be compared to Table 1A, shows the sampling distribution of R(k) for a series of length 36 under a pure random walk null. The

---

Appendix.

length was chosen to correspond to the period 1950-85, the typical period for which real GDP is reported by IFS. The dispersion, left skewness, and downward bias in the median are all more pronounced for n=36 than for n=117. This is to be expected because the sampling distribution of R(k) is further from asymptotic normality when n=36 than when n=117.

Selected variogram ordinates computed from the annual postwar IFS data on 32 countries are given in Table 5. The findings here are even stronger than those that emerged from Maddison's data. The vast majority of ordinates shown are greater than 1; in fact, for several countries, the R(k) all lie above 2, the 90% fractile or better in Table 4, and are increasing in k. Once again, the US has the lowest variogram among all countries considered, which now include nearly all of its industrial peers. Recall from Table 2, however, that the R(k) for the US over the period 1953-85 are close to 1, and hence are much more consistent with the R(k) for the other countries in Table 5.

IV. DISCRIMINATING BETWEEN THE TS AND DS CLASSES.

1. The Simulation Methodology

Dramatic as the results in Tables 2 through 5 are, we wish to draw more formal inferences about the presence of unit roots in our data, while also entertaining hypotheses richer than the simple random walk. To this end, we define two classes of univariate models: a trend stationary class (TS) consisting of parsimonious ARMA deviations from linear trends, and a difference stationary (DS) class consisting of parsimonious ARIMA (p,1,q) models. By parsimonious we mean at most a half-dozen or so parameters to be estimated. We simulate each estimated model from the TS and DS classes with iid Gaussian disturbances and compute selected ordinates of the scaled variogram for each simulation. Our estimate of the sampling distribution of R(k) is the empirical distribution (pdf)

of the simulated R(k).

To compare any two estimated models, we first use the empirical pdfs to determine the corresponding fractiles of the R(k) computed from the actual data. The distance of these fractiles in probability units from the medians of their empirical pdfs forms our basis for comparing alternative models; the greater the distance, the less probable the model. Let the fractile of the empirical pdf for R(k) that contains the actual R(k) (conditional on some model Ω), be $F(R(k)|\Omega)$. Then our measure of the probability that R(k) is a realization from a series whose univariate model is Ω is

$$P(R(k)|\Omega) = 1 - 2\|F(R(k)|\Omega) - Med(R(k)|\Omega)\| \qquad (4)$$

where "$\|$" denotes absolute value, and $Med(R(k)|\Omega)$, the median of the empirical pdf. If $F(R(k)|\Omega)$ equals this median, then $P(\cdot|\Omega) = 1$. If R(k) lies outside the range of observed simulated values, $P(\cdot|\Omega) = 0$. Thus the area under that part of the empirical pdf for R(k) that lies at a greater distance from its median than the actual R(k), is our measure of the relative probability that a particular model is true. Note that this probability is akin to a two-tailed test of the null that R(k) was computed from a series generated by the model Ω, with parameters set to their (approximate and Gaussian) maximum likelihood estimates.

For simplicity, we restrict ourselves to just one model from each class: an AR(3) about a linear trend, and an ARIMA (3,1,0). For the data used in this study, three AR terms generally sufficed to yield white noise residuals. For the TS class, an AR(3) about trend subsumes an AR(2), Cochrane's preferred model for US GNP. With respect to the DS class, preliminary estimates of low-order ARIMA models with MA components, revealed results similar to the ARIMA(3,1,0) case, at least for those countries where the estimation converged successfully in 100 iterations or less. In addition, models with two unit roots typically showed

evidence of overdifferencing. Using standard diagnostic methods, no <u>low order</u>

ARIMA model seemed superior overall to the (3,1,0).

2. Empirical Results.

Estimates of the TS and DS models for the US and the UK are given in Table 6.

Table 7 gives the actual $R(k)$ (from Table 2), and the corresponding simulated

$Med(R(k)|\Omega)$ and $P(R(k)|\Omega)$. For the US, a model with a unit root fits the

variogram better through lag 10, but worse thereafter. Remembering that in

population, the $R(k)$ are linear combinations of the first k-1 autocorrelations,

this finding is consistent with Cochrane's observation that the identifying

restrictions implied by low order ARIMA models fit the low order autocorrelations

at the expense of higher order ones. Even for $k \leq 10$, TS does not perform all

that badly. Overall, if the objective is to do equal justice to high as well as

low order autocorrelations, these results do not by themselves overturn Cochrane's

preference for modelling US GNP as stationary AR deviations about a deterministic

trend.

For the UK, the pattern of the probabilities is the opposite of that for the

US; the TS model fits the variogram better at lags ≤ 10, while DS generally

genrally does so at lags greater than 10. Except at lag 10, $P(\cdot|DS)$ roughly

equals or exceeds $P(\cdot|TS)$, so that the presence of a unit root seems likely. Even

TS suggests a unit root, as the sum of the estimated AR coefficients from the TS

model in Table 6 is .97, which is very close to the value of 1.0 characteristic of

a unit root. Correspondingly, $Med(R(k)|TS)$ is near 1 for all k.

For Maddison's data, Table 8 repeats the estimated R(20) and R(50) shown from

Table 3, and gives the corresponding simulated medians and probabilities.[22] At

---

[22] For the Maddison and IFS data, the estimated TS and DS models from which the
$P(\cdot|\Omega)$ were computed are available in a separate Appendix available from the
authors.

lag 20, the DS model dominates in all but the United Kingdom (where both models are tied at 90%) and the United States. At lag 50, DS is superior in all countries except the US and the Netherlands (about whose low value of R(50) we have reservations). In fact, although the probabilities for the other lags shown in Table 4 are not exhibited, the bottom of Table 8A shows that DS yields a higher probability than TS in at least 9 of the 12 countries at all lags.

Table 8A also contains a frequency count of these probabilities over a richer set of lags. At lag 3, both TS and DS seem to fit the data equally well, but at all higher lags the superior fit of the DS model emerges. This result is different from Cochrane's finding for the US that the superiority of a parsimonious TS model emerges when fitting the higher-order lags. Furthermore, TS has low probability at all lags in Norway and Germany. On the other hand, DS never does very badly in any country, although it does worst in the US. For these data, DS is a clearly superior model for aggregate output.

Table 9 shows that the results for the postwar IFS data are even stronger than those from Maddison's data. Table 9 focuses on lag 20 because longer lags may yield too few observations for efficient variance estimation given sample sizes around 35. DS dominates in all but 3 countries, one of these being the United States.[23] Table 9A contains a frequency count of the $P(\cdot | \Omega)$ for the richer set of lags shown in Table 5. The superior fit of DS at all lags is clear, especially at longer lags.[24] This is again dramatically different from Cochrane's finding that the superiority of TS as a model for US GNP emerges at the higher lags. The fit of TS also steadily deteriorates with increasing lag, e.g., the

---

[23] Table 2 shows that the high-order ordinates for the US rise substantially when data for the years 1948-52 are omitted from postwar sample.

[24] The only country where the $P(\cdot | DS)$ are uniformly very low is the Philippines, where TS performs almost as poorly.

number of countries with P(·|TS,k) < .1 going from 5 at k=3 to 15 at k=20. Again, the DS overwhelmingly dominates the TS as a parsimonious model for aggregate output during the postwar era.

Under a given null, the right tail areas (Marginal significance level (MSL) or p-value) under each country's empirical pdf for R(20) corresponding to that null should be uniformly distributed between 0 and 1 across countries. The two middle columns of Table 10 give a frequency count of the MSL (= 1 - F(R(20)|Ω) in the notation of (4)) for R(20) from the 32 countries in our panel. While the MSL under TS are clearly not uniformly distributed over [0,1], DS also yields too many MSLs in the lower half of [0,1]. Thus, as Cochrane argues, restricting the DS class to a parsimonious subset such as ARI(3,1) does not do justice to the longer run properties of the data. However, instead of being "too nonstationary", as Cochrane found for the US, parsimonious ARIMA models appear insufficiently nonstationary to simulate the large R(k) actually observed in our panel. In other words, the data show some "excess" nonstationarity relative to an ARI(3,1).

V. THE RELATIVE POWER OF ALTERNATIVE TEST STATISTICS

1. A Dickey-Fuller Type F Test.

To this point, we have used only the variogram to shed light on the presence of a unit root. An important alternative class of test statistics designed to elucidate the same question has been proposed by Dickey and Fuller (see Dickey et al. [1985]). We now compare the power of an F test in the spirit of Dickey and Fuller, with that of the variogram. Given the regression equation

$$X_t = a + b_0 T + b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} \quad , \tag{5}$$

where T is a time trend, the null hypothesis that X_t contains a unit root is equivalent to the joint hypothesis $b_0 = 0$ and $b_1 + b_2 + b_3 = 1$. Since this

hypothesis is a pair of linear restrictions on the coefficients of (5), it can be tested by means of an F statistic. We call this F statistic "DFF" (short for Dickey-Fuller F). As with other Dickey-Fuller statistics, the sampling distribution of DFF under the null is not a standard F; hence its distribution under both TS and DS will be simulated.

In this section we examine the relative ability of DFF and the variogram to distinguish between DS and TS alternatives in simulations of the IFS data. We also examine an estimator of the variogram derived by substituting sample estimates for population quantities in (3), as suggested by Cochrane [1987] and Campbell and Mankiw [1987]. All three statistics are computed from data simulated under both TS and DS.[25] For each set of $N(0,1)$ deviates used to simulate ϵ_t, the following two series are simulated:

$$X_t = b_0 T + b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + \epsilon_t.$$

$$X'_t = \alpha + (1+\beta_1)X'_{t-1} + (\beta_2-\beta_1)X'_{t-2} + (\beta_3-\beta_2)X'_{t-3} - \beta_3 X'_{t-4} + \epsilon_t.$$

The b_i are derived by estimating TS, while the β_i and α are derived by estimating DS.

The results of comparing DFF with the variogram are shown in Tables 11 and 12. As before, these Tables exhibit results for lag 20 only. The restriction to one lag biases the comparison in favor of DFF in that the information in the remainder of the variogram is being ignored. V20 is an estimate of R(20) computed from the rhs of (3). For each pair X_t and X'_t, DFF, R20 and V20 were computed and the empirical pdf of each was tabulated. The $P(\cdot|\Omega)$ were computed from (4) as

---

[25] In the case of TS, the intercepts of the simulated series are all set to 0. On the other hand, the intercept estimated from the ARIMA (3,1,1) model, α, has been included in the simulated DS series, because under TS, this intercept is an estimate of the trend coefficient.

before and are tabulated in Table 11.[26] The bottom row of this Table reveals that all three statistics find DS to be more probable than TS. However, the variogram estimates R20 and V20 reveal this by the overwhelming margins of 29 to 3 and 27 to 5, respectively.[27] In contrast, DFF shows DS to dominate TS by a margin of only 22 to 10, which suggests that DFF is somewhat less powerful than V20 and R20. Finally under TS, the $P(\bullet|\Omega)$ for all three statistics tend to cluster near 0, which is highly unfavorable to TS, while under DS this measure is more or less uniformly distributed over [0,1].[28]

2. Assessing Relative Power.

A final question is the extent to which the empirical pdfs of R20, V20 and DFF differ under TS and DS. In classical statistics, questions of this sort comes under the heading of the "power function." There is fortunately no need to study whole power functions because TS and DS are both point hypotheses by construction. We shall instead start by measuring the extent to which the simulated empirical pdfs for each statistic differ under the two hypotheses. We then define the power of a statistic as the probability that the statistic will correctly identify whether a series was generated by TS or DS. If the empirical pdf of a statistic is the same under both TS and DS, then it cannot distinguish between TS and DS.

---

[26] The actual values for each country are available from the authors.

[27] The only countries that favor TS using R20 are the Philippines, Turkey, and the US. The $P(\bullet|\Omega)$ value under TS for the Philippines is only .12. The estimated AR coefficients under TS for France and Venezuela summed to more than 1, so that the corresponding simulated series were "explosively" nonstationary.

[28] The marginal significance levels (MSL) for the three statistics under DS and TS are tabulated in Table 10. All three statistics exhibit under TS a sharp mode at that end of [0,1] which is least favorable to TS. Note that for R20 and V20, the MSL under DS also tend to cluster towards 0. This is not evidence against DS; rather this is consistent with the finding of "excess" nonstationarity discussed at the end of section IV.2.

Since the classification of a series would then be a toss-up, the probability of classifying a series correctly, i.e., the power, is then .5. If the smallest simulated value under one hypothesis is greater than the largest value under the other, so that there is no overlap between the empirical pdfs under the two hypotheses, the power is 1.0, conditional on a given finite number of replications.

More generally, let $\phi(n|\theta,\Omega)$ be the nth order statistic from the empirical pdf of θ computed over data generated under Ω. Also let MED = $[(NREP+1)/2]$, where NREP is the number of Monte Carlo replications (in this case, 1000). Now solve the following implicit equation for x:

$$\phi(MED + x|\theta,\Omega_1) = \phi(MED - x|\theta,\Omega_2)$$

where Ω_1 = DS when θ = DFF, and TS when θ = R20 or V20. Ω_2 is the other member of the pair (DS,TS). This is because when TS is true, R20 and V20 should be <u>lower</u> than when DS is true, and vice versa when estimating DFF. Let the (parametric) solution to this equation be $x(\theta,\Omega)$. We now posit the following decision rule: Classify a series as DS (TS) if the value of θ calculated from the series lies closer to the median of the empirical pdf of θ under DS (TS) than it does to the median under TS (DS). The probability of misclassifying a series is then

$$\mu(\theta) = \frac{MED - x(\theta,\Omega)}{NREP} \tag{6}$$

The probability that θ will correctly classify a series, i.e., the power of θ, is then $\pi(\theta) = 1-\mu(\theta)$.[29]

---

[29] Our use of the term power is related to the conventional one, namely the probability that the null will be rejected when the alternative is true, if the probabilities of Type I and Type II errors are equated, so that DS and TS are on an equal footing.

The frequency count of the $\pi(\theta)$ across countries is given in Table 12.[30] In 20 or more of the 28 countries for which a comparison is possible, $\pi(V20)$ or $\pi(R20)$ is higher than $\pi(DFF)$, with two ties at the second decimal place (Norway and the US). These frequency counts modestly favor R20 and V20 over DFF, in that the powers of R20 and V20 have medians and modes between .6 and .7, while those for DFF lie between .5 and .6. Yet only in 8 cases out of a possible 90 is π more than .8. The median value of $\pi(R20)$ - $\pi(DFF)$ is .04, and that of $\pi(V20)$ - $\pi(DFF)$ is .06, with only 8 and 7 values, respectively, out of 28 being negative. The median value of $\pi(V20)$ - $\pi(R20)$ is .03, with the largest absolute difference being .1 (the Philippines). Out of 30 values, there are only three negative ones and 4 ties at the second decimal place.

Viewed as a whole, $\pi(R20)$ and $\pi(V20)$ are neither all that close to 1 or all that far from $\pi(DFF)$, while V20 appears marginally more powerful than R20. All three statistics evidently lack power to distinguish between TS and DS in the context of a single country. In the context of classifying as a set the data from a group of countries, however, the modest incremental power of R20 and V20 over DFF may prove beneficial. At any rate, Table 11 shows that DFF, R20 and V20 all have sufficient power to distinguish DS from TS in the actual IFS data.

---

[30] The actual values for each country are available from the authors.

VII. CONCLUSION.

We have defined the scaled variogram and computed it for the series on annual real output from a number of countries to determine whether they contain a unit root. In contrast to Cochrane's previous estimates for the US, which we confirm, we find that the variograms of nearly all the 14 long and the 32 post-War real ouput series examined were consistent with the presence of a unit root. Among both long and short series, the variograms of the US data were the least consistent with a unit root. Confirming Campbell and Mankiw [1986], however, the post Korean War US data do reveal a unit root.

We use Monte Carlo methods to show that an ARIMA (3,1,0) (DS) was almost always a more "probable" model for the data than stationary AR(3) deviations from a linear trend (TS). We also show that neither the variogram nor a parametric Dickey-Fuller type F test have much power to discriminate between TS and DS in a single country, although for the bulk of countries in both the long and short data sets, DS dominates TS using either procedure. Hence, looking over a number of countries and time periods, our answer to the question "How big is the unit root in GNP?" would be "Rather substantial."

Although our results for other countries do not concur with Cochrane's for the US, we agree with Cochrane on a number of theoretical points. In particular, we agree that parsimonious ARIMA models tend to sacrifice goodness of fit at the lowest frequencies in exchange for a better average fit over the entire spectrum. Therefore such models, while doing justice to the short run dynamics of time series, may well misrepresent the long run behavior. As it turns out, however, the data do not bear out Cochrane's concern here. DS outperforms TS not only at the short lags, but even more so at longer lags. In fact, for a number of series, the estimated variogram is too large to be captured by an ARIMA (3,1,0), i.e., there appears to be "excess" nonstationarity. We wish to point out that parsimonious

ARIMA models do not nessarily do violence to the economically relevant part of the dynamics of a series. This is because the part of the impulse response function of a series implied by its long run dynamics can have negligeable impact on economic behavior, given "reasonable" discount rates. We plan to elaborate on this point in our future work.

We conclude by highlighting the implications of our results for the consumption function. As Mankiw and Shapiro [1985] have shown, if income has a unit root, then the usual finding (e.g., Flavin [1981]) that consumption is "too sensitive" to innovations in income is spurious. In fact Deaton [1986] (for the US) and Kormendi and LaHaye [1986] (for a panel of 30 countries), using differenced specifications, found undersensitivity of consumption to income innovations. More generally, a significant unit root component in output increases the potential for regressions that include output and are estimated in levels, with or without detrending, to be misspecified or "spurious."

DATA APPENDIX

The basic variable of this study is the log of annual real output per capita, computed from the output and population data described below.

Abbreviations:

IFS   International Financial Statistics monthly publication, Yearbook and computer tape.

NIPA   The National Income and Product Accounts of the United States, 1929-82: Statistical Tables.

SCB   Survey of Current Business.

Long US Data

Output: Annual real GNP in 1982 prices, rounded to the nearest $100 million. Data for the 1869-1908 segment in 1929 prices waere ratio spliced; all other data are in 1982 prices.

1983-85   Series 1.2.1, July 1986 SCB.

1929-82   Series 1.2.1, NIPA.

1909-28   Column labelled "Revised Estimates," Table 8, Romer [1987].

1869-1908   Table 3, Romer [1986].

Population: Data as of mid-year and rounded to the nearest 100,000:

1983-85   Last column, Table 8.2, July 1986 SCB.

1929-82   Last column, Table 8.2, NIPA.

1869-1928   Col. 5, Table 4.8, Friedman & Schwartz [1982].

Long British Data

Ouput: Annual real GDP in 1980 prices rounded to the nearest £100 million. Data segments ratio spliced to a 1948-85 base.

1948-85   IFS series 99bp.

1920-47   Col. 8, Table 5, Feinstein [1972]. In 1938 prices; excludes Ireland.

1913-19   Col. 8, Table 5, Feinstein [1972].   "   ; includes Ireland.

1870-1912   Col. 8, Table 5, Feinstein [1972]. In 1900 prices;   "

Population: Data as of midyear and rounded to the nearest 100,000. Data segments ratio spliced to a 1948-85 base.

1948-85	IFS series 99z.
1920-47	Col. 1, Table 55, Feinstein [1972]. Covers Great Britain and Ulster.
1870-1919	Col. 4, Table 55, Feinstein [1972]. Includes all of Ireland.

IFS Postwar Data

The main source is the IFS tape dated August 1986. All available data were used. Some additional significant digits were obtained from the IFS Yearbooks for years 1979 to 1985. Data for the 1980-5 period were checked against the April 1987 issue of IFS.

Output: Annual real GDP (series 99b.p or 99b.r) or real GNP (99a.p or 99a.r), whichever is available, in 1980 prices. The numerical difference between IFS nominal GNP and GDP is always very small; the difference between their real analogues is assumed small. US data cover the period 1948-85 and are as described under "Long US Data". Israeli data for 1950-3 are known to only one significant digit.

Population: Midyear estimate (series 99z). Data for Venezuela over the period 1975-85 were multiplied by .95 to correct for an apparent discontinuity starting in 1975. (The published data show an 8.6% increase over 1974-5.)

Maddison Data

Output: Index numbers for annual real GDP (1913=100), taken from Appendix Tables A6-8 in Maddison [1982] and ratio spliced to the available IFS real output series, using values for the first overlap year, usually 1950. The US series was spliced to the NIPA data for the period 1948-85 described under "Long US Data". Maddison constructed his series from a variety of government and historical sources, adjusting them to conform as much as possible to present boundaries.

Population: Mid-year estimate rounded to the nearest 100,000. Data taken from Appendix Tables B2-4 in Maddison [1982] and ratio spliced to IFS series 99z, using the first overlap year, typically 1948. To enhance the conformity of the data with postwar boundaries, further ratio splices were performed using overlapping data for those years in which boundaries changed. The US data are as described above under "Long US Data," and cover the period 1870-1985.

Fig. 1

Variance of k-Differences of Log Real Per Capita GNP
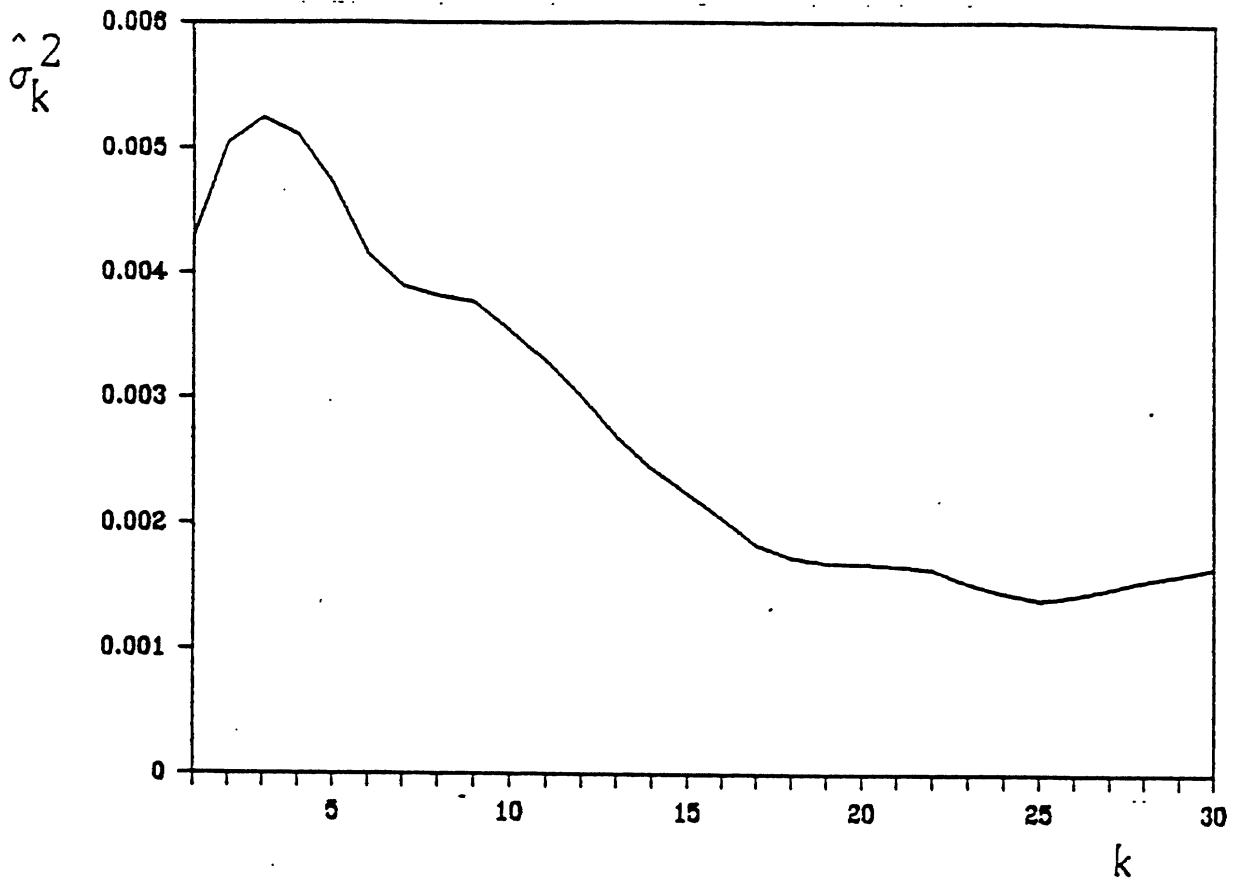
NOTE: This is an unscaled variogram with sample period 1869-1984, and is taken from Cochrane [1987]. Data are mainly from Friedman and Schwartz [1982].

FIG. 2

PLOT OF SCALED VARIOGRAM FOR LONG SERIES ON US GNP

```
 2.00+                   D
    -
    -
    -            D
    -      4
 1.50+    3 5
    -    2  D6
    -     D   7
    -          890
    -            1
 1.00+ 1 M M        2
    -          M          M          M
    -     D          3                              M
    -       D       45
    -           D    6789
 0.50+                 01234    8901234567 90                              M
    -                 D    567  D       8  123456789012345
    -                                  D      67890
    -                                            123456789   34D
    -                                             012
 0.00+
       +---------+---------+---------+---------+---------+---------+---------+-----
       0        10        20        30        40        50        60        70
                                         K
```
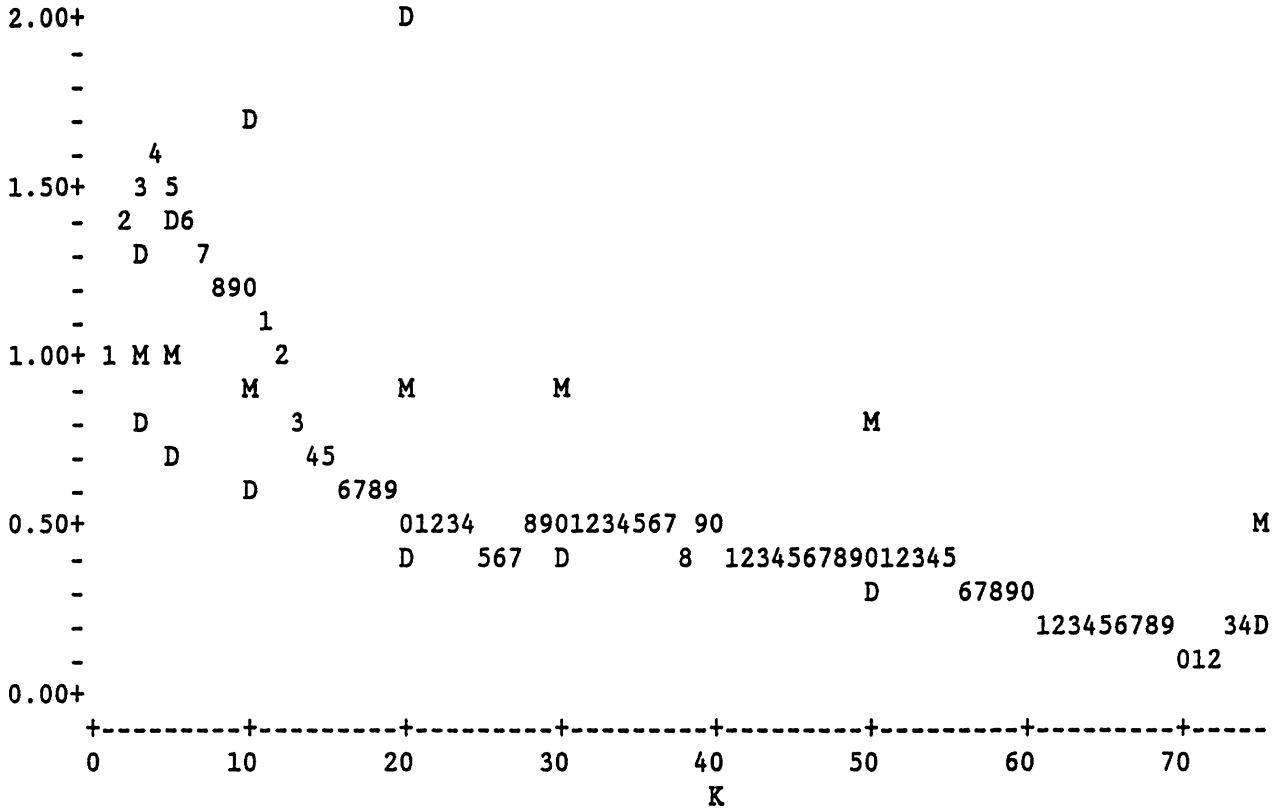
NOTE: Each value of R(k) plotted using the last digit of K.   50% fractile

from Table 1B plotted as "M" (Median); 10% and 90% fractile plotted (the

latter only for K ≤ 20) as "D" (Decile).

FIG. 3

PLOT OF SCALED VARIOGRAM FOR LONG SERIES ON UK GDP

```
3.00+
     -
     -                                                                        *
     -                                          8901
     -                                        67    2
2.50+                                        5   *  3
     -                                       4        4
     -                                      3          5
     -                           *         12          67
     -                        67890              .        8
2.00+                      2345                              9
     -            *         01                              01
     -                     9
     -            0123   78                                   2
     -   4567 *       789    456                               3
1.50+      89     56                                          4
     -   3      01 3                                              5        5
     -    *       2                                              6      34
     -  *                                                        78  12
1.00+ 1 M M     M                                               90
     -            M
     -                  M                        M                          M
     -
     -
0.50+
     +---------+---------+---------+---------+---------+---------+---------+-----
     0        10        20        30        40        50        60        70
                                         K
```
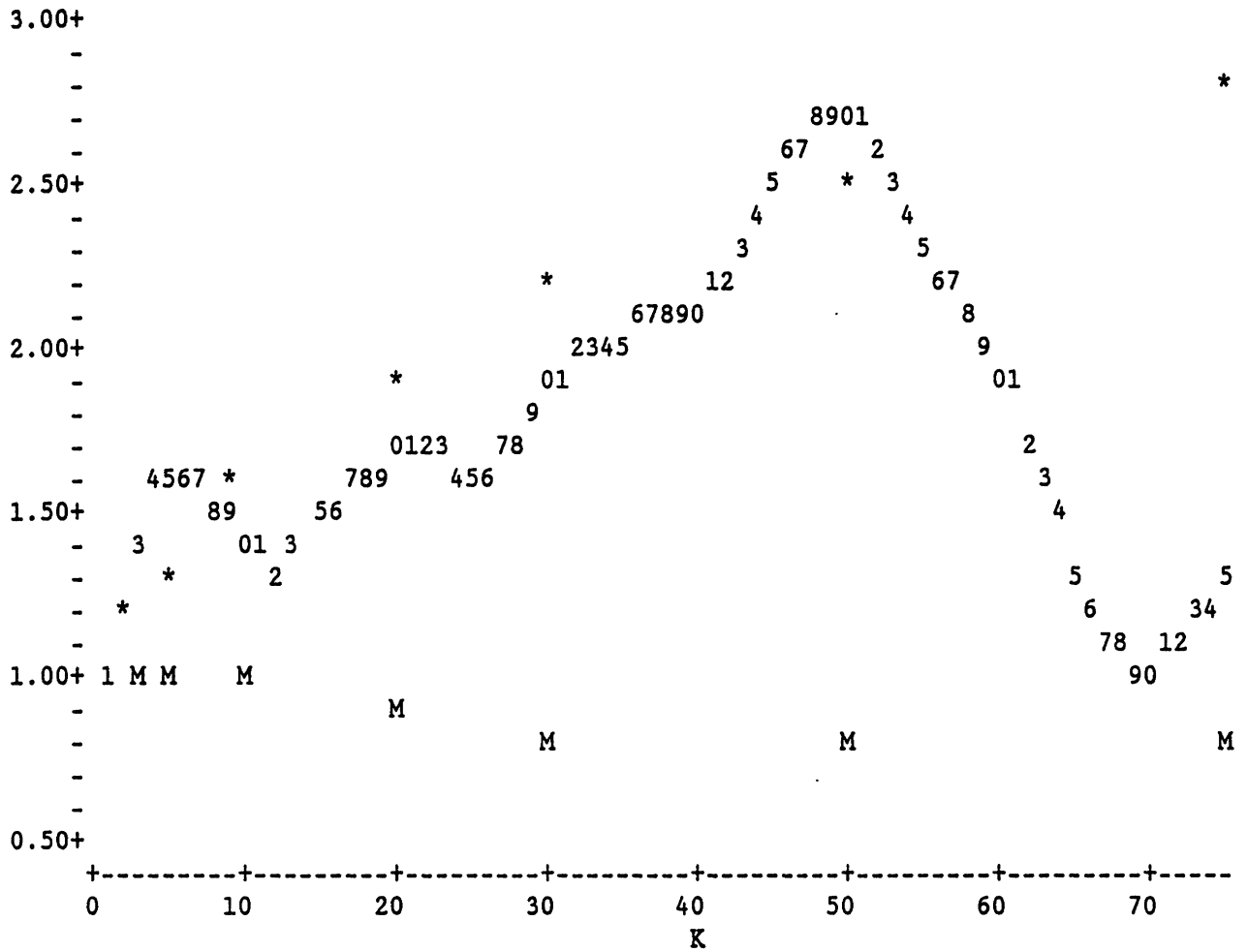
NOTE: Each value of R(k) plotted using the last digit of K.   50% fractile

from Table 1A plotted as "M" (Median); 95% fractile plotted as "*".

TABLE 1

SIMULATED SAMPLING DISTRIBUTION OF SCALED VARIOGRAM FOR RANDOM WALK
1000 REPLICATIONS OF LENGTH 117

1A.
HOMOSKEDASTIC DISTURBANCES

LAG K	MIN	5	10	25	FRACTILES 50	MEAN	75	90	95	MAX
3	.57	.79	.84	.90	.99	1.00	1.09	1.17	1.23	1.49
5	.49	.68	.73	.85	.98	.99	1.11	1.27	1.33	1.72
10	.35	.55	.62	.76	.95	.99	1.15	1.40	1.58	2.65
20	.16	.37	.46	.64	.87	.97	1.21	1.62	1.89	3.85
30	.12	.29	.36	.53	.81	.97	1.25	1.78	2.18	5.31
50	.10	.21	.28	.45	.76	.99	1.28	1.96	2.52	6.01
75	.07	.17	.25	.44	.75	1.06	1.45	2.22	2.77	5.27

1B.
HETEROSKEDASTIC DISTURBANCES

LAG K	MIN	5	10	25	FRACTILES 50	MEAN	75	90	95	MAX
3	.44	.71	.76	.86	1.00	1.01	1.15	1.26	1.36	1.75
5	.40	.61	.67	.80	1.00	1.02	1.19	1.43	1.57	2.23
10	.29	.48	.58	.72	.94	1.05	1.28	1.67	2.03	3.24
20	.19	.36	.43	.61	.90	1.09	1.37	1.97	2.54	4.77
30	.12	.29	.37	.52	.87	1.13	1.45	2.20	2.82	5.64
50	.07	.20	.26	.42	.81	1.11	1.43	2.36	3.19	6.16
75	.04	.11	.15	.26	.54	.83	1.06	1.96	2.56	6.11

NOTE: The model simulated is $X_t = \delta + X_{t-1} + k_t \epsilon_t$, where ϵ_t is a simulated draw-ing from a $N(0,1)$ distribution. In 1A, $k_t = .054$ for all t. In 1B, k_t is set to the standard deviation, given in footnote 17, of US GNP for the period that in-cludes t. For each simulated random walk, δ and X_0 were set equal to their estim-ated (actual) values using the log of US annual per capita real GNP. The same se-quence of disturbances was used to generate both 1A and 1B, so that none of the difference between them is due to sampling variation.

TABLE 2

LONG DATA FOR THE UK AND US

SELECTED SCALED VARIOGRAM ORDINATES OF AGGREGATE REAL INCOME FOR VARIOUS SUBPERIODS

FIRST ROW: LOG PER CAPITA   SECOND ROW: LOG RAW

PERIOD	3	5	10	K 20	30	50	75
US GNP							
1869-1985	1.54	1.51	1.15	.55	.50	.43	.15
	1.56	1.53	1.19	.62	.63	.47	.39
1869-1913	.81	.79	.80	.23	---	---	---
	.82	.80	.88	.35	---	---	---
1909-1970	1.82	1.87	1.53	.98	.62	---	---
	1.81	1.84	1.47	.95	.66	---	---
1948-1985	1.14	.89	.65	.35	---	---	---
	1.20	.95	.75	.61	---	---	---
1953-1985	1.10	1.02	.95	1.17	---	---	---
	1.11	1.04	1.19	1.07	---	---	---
UK GDP							
1870-1985	1.43	1.57	1.41	1.68	1.85	2.71	1.26
	1.41	1.53	1.28	1.36	1.29	1.57	.58
1870-1913	.56	.57	.51	.34	---	---	---
	.55	.58	.54	.27	---	---	---
1922-1985	1.49	1.59	1.22	.59	.41	---	---
	1.50	1.62	1.28	.64	.42	---	---
1948-1985	.92	.76	.71	.82	---	---	---
	1.03	1.03	1.30	1.57	---	---	---

TABLE 3

MADDISON'S DATA
SELECTED ORDINATES OF THE SCALED VARIOGRAM

COUNTRY	SAMPLE PERIOD	K					
		3	5	10	20	30	50
AUSTRALIA		1.05	1.09	1.17	1.45	1.62	2.35
CANADA		1.05	1.06	.88	.82	.82	.95
DENMARK		1.01	1.09	1.19	1.27	1.48	1.30
FINLAND	1900-85	1.12	1.07	.93	.96	.96	1.09
FRANCE		1.55	1.68	1.44	1.63	1.61	1.12
GERMANY		.98	1.04	1.19	1.50	1.77	1.72
ITALY		1.39	1.44	1.52	2.17	2.34	1.55
NETHERLANDS	1900-85	1.14	1.06	.95	1.09	1.32	.35
NORWAY		1.08	1.26	1.35	1.85	2.24	2.49
SWEDEN		1.15	1.21	1.06	.94	.89	1.11
UNITED KINGDOM		1.55	1.50	1.20	1.33	1.43	2.01
UNITED STATES		1.34	1.28	.95	.50	.48	.30

NOTE: Sample period is 1870-1985 unless otherwise noted.

TABLE 4

SIMULATED SAMPLING DISTRIBUTION OF SCALED VARIOGRAM FOR RANDOM WALK
1000 REPLICATIONS OF LENGTH 36

K	MIN	5	10	25	50	MEAN	75	90	95	MAX
3	.33	.62	.68	.82	.98	1.00	1.16	1.35	1.43	2.13
5	.21	.47	.56	.72	.94	1.00	1.21	1.53	1.71	2.84
10	.12	.31	.38	.55	.86	1.00	1.30	1.83	2.15	3.75
20	.08	.23	.31	.47	.80	1.01	1.34	2.00	2.59	5.06

NOTE: The model simulated is $X_t = X_{t-1} + \epsilon_t$, where ϵ_t is a simulated independent drawing from a N(0,1) distribution. For each simulated random walk, X_0 was set equal 0 and $\sigma^2(\epsilon) = 1$. The ϵ_t were identical to those in Table 1; hence none of the difference between these Tables is merely due to sampling variation.

TABLE 5

POSTWAR IFS DATA:

SELECTED ORDINATES OF SCALED VARIOGRAM (RK) FOR
LOG PER CAPITA AGGREGATE OUTPUT

COUNTRY	R3	R5	R10	R20	V20
AUSTRALIA	.80	.74	1.14	.82	.63
AUSTRIA	1.56	1.83	2.10	3.29	4.13
CANADA	1.03	1.14	1.26	.76	.72
COLOMBIA	1.88	1.87	2.56	2.05	1.88
DENMARK	1.04	1.49	2.38	2.76	2.21
DOMINICAN REPUBLIC	.97	.99	1.21	.80	.85
ECUADOR	1.65	1.83	2.09	1.56	1.30
FINLAND	1.20	.92	.67	.87	1.02
FRANCE	1.94	2.77	4.35	7.14	8.73
GERMANY	1.59	2.00	3.01	4.12	7.66
GREECE	1.51	2.08	2.94	4.21	5.20
GUATEMALA	1.33	1.24	1.59	4.28	2.72
HONDURAS	1.28	1.05	.55	.81	.89
ICELAND	1.16	.91	.67	1.15	.95
IRELAND	1.46	1.45	1.63	2.30	1.74
ISRAEL 1950-85	1.66	2.03	3.18	4.74	5.28
ISRAEL 1953-85	1.88	2.50	3.55	4.44	9.79
ITALY	1.69	1.95	3.43	4.08	7.52
MEXICO	1.43	.84	.69	1.40	1.70
NETHERLANDS	1.47	1.88	1.95	2.56	3.75
NORWAY	1.28	1.28	.56	1.13	.83
PANAMA	1.41	1.35	1.60	2.12	1.79

COUNTRY	R3	R5	R10	R20	V20
PARAGUAY	1.40	1.85	2.65	2.62	3.53
PHILIPPINES	1.67	1.67	1.17	.95	4.30
SOUTH AFRICA	1.07	1.18	1.63	1.71	1.76
SRI LANKA	1.12	1.60	2.29	2.06	3.88
SWEDEN	1.59	1.90	3.45	3.89	4.12
SWITZERLAND	1.04	1.15	1.62	1.96	1.95
THAILAND	1.10	1.44	1.41	2.27	2.24
TURKEY	1.06	.95	.70	.50	.63
UNITED KINGDOM	.92	.76	.71	.82	1.05
UNITED STATES	1.14	.89	.65	.35	.43
VENEZUELA	2.05	2.59	2.13	3.74	8.02

NOTE: V20 computed from (3) in text, with $k = 20$.

TABLE 6

LONG DATA FOR THE US AND THE UK

ESTIMATED COEFFICIENTS WITH T STATISTICS

COUNTRY	MODEL	AR AT LAG 1	AR AT LAG 2	AR AT LAG 3	TIME	SEE	Q (MSL)
United Kingdom 1870-1985:							
	AR(3)	1.17	-.10	-.10	.0005	.0320	28.1
		12.15	-.70	-1.05	2.04		.56
	ARI(3,1)	.21	.11	-.02	.0083	.0325	30.1
		2.13	1.08	-.24	2.44		.46
United States 1869-1985:							
	AR(3)	1.26	-.44	-.01	.0032	.0464	19.5
		13.25	-2.97	-.14	4.04		.93
	ARI(3,1)	.42	-.08	-.10	.0128	.0479	27.0
		4.58	-.85	-1.04	2.56		.62

NOTE: T statistics are given beneath their respective coefficients. AR(3) models were estimated with an intercept (not reported). Intercepts (with their t statistics) of the ARI(3,1) models are reported under TIME. SEE is the standard deviation of the residuals from the corresponding estimated model. Q is the Box-Ljung statistic computed over 32 lags of residual autocorrelations. Beneath Q is its marginal significance level under a white noise null.

TABLE 7

LONG DATA FOR THE UK AND US

ESTIMATED R(K) WITH SIMULATED MEDIANS AND PROBABILITIES, CONDITIONAL ON THE ESTIMATED TS AND DS MODEL IN TABLE 6

					K			
		3	5	10	20	30	50	75
United Kingdom: 1870-1985								
	Actual R(K)	1.44	1.58	1.41	1.69	1.86	2.72	1.26
TS	Median R(K)	1.36	1.50	1.50	1.33	1.16	1.01	.98
	Probability	.65	.78	.87	.54	.38	.15	.89
DS	Median R(K)	1.31	1.53	1.73	1.74	1.65	1.63	1.60
	Probability	.45	.88	.53	.96	.87	.49	.78
United States: 1869-1985								
	Actual R(K)	1.54	1.51	1.15	.55	.50	.43	.15
TS	Median R(K)	1.50	1.38	.82	.45	.35	.28	.34
	Probability	.74	.65	.18	.46	.20	.23	.06
DS	Median R(K)	1.53	1.53	1.40	1.25	1.22	1.03	1.10
	Probability	.95	.92	.58	.08	.16	.22	.01

NOTE: The actual R(k) are repeated from Table 2. The model simulated under TS is $X_t = aT + b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + \epsilon_t$, where T is a time trend. The model simulated under DS is $X_t = \alpha + (1+\beta)_1 X_{t-1} + (\beta_2 - \beta_1) X_{t-2} + (\beta_3 - \beta_2) X_{t-3} + \beta_3 X_{t-4} + \epsilon_t$. The b_i and a were taken from the rows of Table 6 labelled AR(3), while α and the β_i were taken from the rows labelled ARI(3,1). Both models were then simulated by generating 1000 sets of 165 N(0,1) deviates for ϵ_t. Start-up values of 0 were then assumed and the first 50 or so values of X_t were discarded. Hence for each country, none of the differences among the various probabilities reported reflects sampling variation. For each lag K are shown the medians of the simulated

TABLE 8

MADDISON DATA

ESTIMATED R(20) AND R(50), AND SIMULATED MEDIANS AND PROBABILITIES,
CONDITIONAL ON THE ESTIMATED TS AND DS MODELS.

COUNTRY	K	R(K)	TS Median R(K)	P(• I TS)	DS Median R(K)	P(• I DS)
		Actual	**Simulated**			
AUSTRALIA	20	1.45	.99	.42	.96	.49
	50	2.35	.77	.14	.87	.24
CANADA	20	.82	.65	.63	1.04	.72
	50	.95	.45	.26	.90	.97
DENMARK	20	1.27	.69	.14	.91	.46
	50	1.30	.49	.14	.82	.46
FINLAND	20	.96	.49	.21	.79	.73
	50	1.09	.38	.11	.72	.62
FRANCE	20	1.63	1.13	.39	1.55	.94
	50	1.12	.70	.50	1.28	.88
GERMANY	20	1.50	.65	.03	.83	.18
	50	1.72	.46	.02	.69	.24
ITALY	20	2.17	1.09	.11	1.27	.23
	50	1.55	.88	.34	1.15	.66
NETHERLANDS	20	1.09	.56	.16	.80	.56
	50	.35	.44	.75	.74	.27
NORWAY	20	1.35	.66	.01	1.17	.43
	50	2.49	.54	.01	1.11	.30
SWEDEN	20	.94	.59	.23	1.01	.91
	50	1.11	.37	.05	.85	.77
UNITED KINGDOM	20	1.33	1.24	.90	1.24	.90
	50	2.01	.84	.20	1.01	.46
UNITED STATES	20	.50	.39	.46	1.14	.10
	50	.30	.25	.63	.97	.10

The R(k) are repeated from Table 4. The TS and DS models are simulated as described in the Note to Table 7, except that the estimated parameters are available from the authors, and 250 sets of 165 N(0,1) were generated for ϵ_t. Start-up values of 0 were then assumed and the first 50 or so values of X_t were discarded. Hence for each country, none of the differences among the various probabilities reported reflects sampling variation. Probabilities were computed using (4) in the text, from the empirical fractiles of variogram ordinates derived from the simulated series.

TABLE 8A

MADDISON DATA
FREQUENCY COUNT OF SIMULATED PROBABILITIES AT ALL LAGS

		LAG (K)											
		3		5		10		20		30		50	
INTERVAL		TS	DS	TS	DS	TS	DS	TS	DS	TS	DS	TS	DS
.0	.1			1		1		2		5		3	
.1	.2	1				2		3	2	3	2	3	1
.2	.3			2		1		2	1		1	2	3
.3	.4			1		1	1	1		3	4	1	1
.4	.5			1		3	1	2	3	1			2
.5	.6			3	2	1	2		1			1	
.6	.7	6	1	1	2		1	1				1	2
.7	.8	1	6	1	5		4		2		2	1	1
.8	.9	2	4	2		1					2		1
.9	1.0	2	1		3	2	3	1	3		1		1

Number of
countries where
P(·|DS) > P(·|TS):   9   12   9   10   12   10

NOTE: All intervals are half open in that they include the lower but not the upper bound. All columns sum to 12.

TABLE 9

IFS POSTWAR DATA

ESTIMATED R(20) AND SIMULATED MEDIANS AND PROBABILITIES,
CONDITIONAL ON THE ESTIMATED TS AND DS MODELS.

	Estimated:		Simulated:			
			TS		DS	
COUNTRY	R(20)		Median R(20)	P(•|TS)	Median R(20)	P(•|DS)
AUSTRALIA	.82		.39	.23	.51	.44
AUSTRIA	3.29		.74	.04	1.56	.29
CANADA	.76		.52	.54	.91	.80
COLOMBIA	2.05		1.02	.46	1.27	.53
DENMARK	2.76		.64	.03	1.15	.21
DOMINICAN REPUBLIC	.80		.47	.54	.77	.97
ECUADOR	1.56		.89	.45	1.85	.86
FINLAND	.87		.56	.56	.88	.97
FRANCE	7.14		14.21*	.09	4.07	.30
GERMANY	4.12		.41	.00	1.39	.14
GREECE	4.21		1.55	.20	2.55	.40
GUATEMALA	4.28		.72	.01	2.15	.27
HONDURAS	.81		.46	.30	.90	.90
ICELAND	1.15		.41	.09	.77	.53
IRELAND	2.30		.72	.11	1.02	.29
ISRAEL	4.74		1.24	.02	2.13	.20
ITALY	4.44		.90	.02	2.05	.34
MEXICO	1.40		.47	.02	.87	.50
NETHERLANDS	2.56		.64	.05	1.91	.70
NORWAY	1.13		.47	.15	.98	.84
PANAMA	2.12		.93	.26	1.24	.42
PARAGUAY	2.62		1.12	.19	2.15	.77
PHILIPPINES	.95		3.91	.10	27.21*	.00
SOUTH AFRICA	1.71		.67	.15	.84	.27
SRI LANKA	2.06		.29	.00	1.47	.60
SWEDEN	3.89		.81	.00	1.45	.10
SWITZERLAND	1.96		.63	.08	1.19	.48
THAILAND	2.27		.41	.00	1.29	.44
TURKEY	.50		.50	.99	.87	.39
UNITED KINGDOM	.82		.42	.26	.55	.58
UNITED STATES	.35		.43	.75	.62	.36
VENEZUELA	3.74		15.28*	.07	4.81	.69

* The AR coefficients in the simulated model sum to more than 1.

NOTE TO TABLE 9

The R20 are repeated from Table 5. The TS and DS models are simulated as des-
cribed in the Note to Table 7, except that the estimated parameters are available
from the authors, and 250 sets of 85 $N(0,1)$ were generated for ϵ_t. Start-up val-
ues of 0 were assumed and the first 50 or so values of X_t discarded. Hence for
each country, none of the differences among the various probabilities reported re-
flects sampling variation. Probabilities were computed using (4) in the text,
from the empirical fractiles of variogram ordinates derived from the simulated
series.

TABLE 9A

IFS POSTWAR DATA
FREQUENCY COUNT OF CONDITIONAL PROBABILITIES

INTERVAL		LAG (K) 3 TS	3 DS	5 TS	5 DS	10 TS	10 DS	20 TS	20 DS
.0	.1	5	1	9	1	11	2	15	1
.1	.2	6		4		9	1	5	2
.2	.3	1		1		4	4	4	6
.3	.4	1		1	2	1	2	1	4
.4	.5		1	5	1	2	5	2	5
.5	.6		4	1	7		7	3	4
.6	.7	5	12	4	5	2	4		2
.7	.8	2	5	3	7	1	3	1	2
.8	.9	3	6		3	2	3		3
.9	1.0	9	3	4	6		1	1	3

Number of
countries where
$P(\cdot|DS) > P(\cdot|TS)$:   18   26   26   29

NOTE: All intervals are half open in that they include the lower but not the
upper bound. All columns add to 32.

TABLE 10

IFS POSTWAR DATA
FREQUENCY COUNT OF RIGHT TAIL AREAS, MSL(θ|Ω) = 1 -F(θ|Ω)

| | | MSL(θ|Ω) | | | | | |
| INTERVAL | | DFF | | R20 | | V20 | |
θ:	Ω:	TS	DS	TS	DS	TS	DS
.0	.1	4	4	18	5	18	3
.1	.2	1	1	5	7	4	6
.2	.3	1	1	4	6	2	6
.3	.4		6		5	4	5
.4	.5	1			3	1	5
.5	.6	1	3	1	3	1	2
.6	.7	1	1	1			2
.7	.8	3	4		2		2
.8	.9	5	7				
.9	1.0	15	5	3	1	2	1

Number of
cases where
MSL(θ|DS) > MSL(θ|TS):   3   30   30

TABLE 11

IFS POSTWAR DATA
FREQUENCY COUNT OF P(θ|Ω)

| | | PROBABILITIES P(θ|Ω) | | | | | |
| INTERVAL | | DFF | | R20 | | V20 | |
θ:	Ω:	TS	DS	TS	DS	TS	DS
.0	.1	12	4	15	1	13	3
.1	.2	7	5	6	5	7	1
.2	.3	1	5	3	4	2	4
.3	.4	4	2	2	3	2	2
.4	.5	1	2	3	8	2	7
.5	.6	4	4	1		1	2
.6	.7	1	3		3		5
.7	.8		3	1	2	3	1
.8	.9	1	1		2	2	2
.9	1.0	1	3	1	4		5

Number of
cases where
P(θ|DS) > P(θ|TS):   22   29   27

TABLE 12

IFS POSTWAR DATA
FREQUENCY COUNT OF CLASSIFICATION POWER $\pi(\theta)$

θ

INTERVAL		DFF	R20	V20
.9	1.0	1		1
.8	.9	3	1	2
.7	.8	3	5	5
.6	.7	3	16	17
.5	.6	20	8	5
Undefined		2	2	2
Median Value:		.58	.63	.65
Number of cases where $\pi(\theta) > \pi(DFF)$: (28 possible)			20	22

NOTE TO TABLES 10 THROUGH 12

The estimates by country underlying these Tables are available from the authors upon request. All intervals except the last are half open in that they include the lower but not the upper bound. All columns add to 32. The simulated model under TS is $X_t = aT + b_1 X_{t-1} + b_2 X_{t-2} + b_3 X_{t-3} + \epsilon_t$. For DS the simulated model is $X'_t = \alpha + (1+\beta_1)X'_{t-1} + (\beta_2-\beta_1)X'_{t-2} + (\beta_3-\beta_2)X'_{t-3} - \beta_3 X'_{t-4} + \epsilon_t$. a, α, b and β were set to the values estimated from the actual data and given in Table A2. For each of 1000 replications, 85 N(0,1) deviates were generated for ϵ_t, start-up values were set to 0 and the first 50 or so values of X and X' were discarded. Since the same vector $\underline{\epsilon}$ was used to simulate X and X', none of the differences among the entries for a given country reflect sampling variation. DFF is the F statistic for jointly testing a=0 and $b_1 + b_2 + b_3 = 1$ in a regression of X (X') on a trend and three lags. R20 is the scaled variogram ordinate at lag 20. V20 is computed by substituting sample estimates for population parameters into (3) in text. $F(\theta|\Omega)$ is the empirical fractile of statistic θ derived from simulations of model Ω. The probabilities, $P(\cdot|\Omega)$, were computed from (4) in the text. $\pi(\theta) = 1-\mu(\theta)$, where $\mu(\theta)$ is computed from (6).

REFERENCES

Brunner, Karl, 1986, "Fiscal Policy in Macro Theory: A Survey and Evaluation," in

   The Monetary Versus Fiscal Policy Debate, R. W. Hafer ed. London: Rowman and

   Allanheld.

Bureau of Economic Analysis, 1986, The National Income and Product Accounts of the

   United States, 1929-82: Statistical Tables. Washington: US Government Print-

   ing Office.

Campbell, J. Y. and N. G. Mankiw, 1987, "Permanent and Transitory Components in

   Macroeconomic Fluctuations," American Economic Review 77 (May): 111-7.

---------------- and A. Deaton, 1987, "Is Consumption Too Smooth?" NBER Working

   Paper 2134.

Cochrane, J. H., 1987, "How Big Is the Random Walk in GNP?" Working Paper, Univ.

   of Chicago.

Cressie, N. A. C., 1986, "The Variogram," article forthcoming in Vol. 9 of the

   Encyclopedia of Statistical Science, N. L. Johnson and S. Kotz eds. New York:

   Wiley.

Deaton, A., 1986, "Life-cycle Models of Consumption: Is the Evidence Consistent

   with the Theory?" NBER Working Paper 1910.

Dickey, D. A., W. R. Bell and R. B. Miller, 1986, "Unit Roots in Time Series

   Models: Test and Implication," American Statistician 40: 12-26.

Fama, E. F., 1982, "Inflation, Output and Money," Journal of Business 55: 201-32.

----------- and K. French, 1986, "Permanent and Temporary Components of Stock

   Prices," CRSP Working Paper #178.

Feinstein, C. H., 1972, National Income, Expenditure and Output of the United

   Kingdom. Vol. 6 in Studies in the National Income and Expenditure of the Uni-

   ted Kingdom, series editor Richard Stone. London: Cambridge University Press.

Friedman, Milton and A. J. Schwartz, 1982, Monetary Trends in the United States and the United Kingdom. Chicago: Univ. of Chicago Press for NBER.

Granger, C. W. J. and P. Newbold, 1974, "Spurious Regressions in Econometrics," Journal of Econometrics 2: 111-20.

Grenander, V., 1954, "On the Estimation of Regression Coefficients in the Case of Autocorrelated Disturbances," Annals of Mathematical Statistics 25: 252-72.

Hall, Robert, 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," Journal of Political Economy 86: 971-987.

Harvey, A. C., 1985, "Trends and Random Walks in Economic Time Series," Journal of Business and Economic Statistics 3: 216-27.

Huizinga, J., 1986, "An Empirical Investigation of the Long Run Behavior of the Real Exchange Rate." Working Paper, GSB, Univ. of Chicago.

International Monetary Fund, International Financial Statistics (IFS). Washington: International Monetary Fund.

Kormendi, R. C., 1983, "Government Debt, Government Spending and Private Sector Behavior," American Economic Review 73: 994-1010.

-------------- and P. G. Meguire, 1984, "Cross-Regime Tests of Macroeconomic Rationality," Journal of Political Economy 92: 875-908.

-----------------------------, 1986, "Government Debt, Government Spending and Private Sector Behavior: Reply," American Economic Review 76: 1180-7.

------------- and L. LaHaye, 1987, "Cross-Regime Tests of the Permanent Income Hypothesis," Working Paper, Univ. of Michigan.

------------------------- and P. G. Meguire, 1986, "Cross-Country Evidence on the Effects of Government Spending," Working Paper, Univ. of Michigan.

Lo, A. and A. C. MacKinley, 1987a, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," Working Paper, Rodney L. White Center for Financial Research, Univ. of Pennsylvania.

-------------, 1987b, "A Simple Specification Test of the Random Walk

Hypothesis," Working Paper, Rodney L. White Center for Financial Research.

Maddison, Angus, 1982, Phases of Capitalist Development. London: Oxford University Press.

Mankiw, N. G. and M. Shapiro, 1985, "Trends, Random Walks and Tests of the

Permanent Income Hypothesis," Journal of Monetary Economics 16: 163-74.

------------ and L. Summers, 1986, "Money Demand and the Effects of Fiscal

Policies," Journal of Money, Credit and Banking 18: 415-429.

Matheron, G., 1971, The Theory of Regionalized Variables and Its Applications.

Cahiers du Centre de Morphologie Mathematique, No. 5, Fontainebleau.

Meese, R. A. and K. J. Singleton, 1982, "On Unit Roots and the Empirical Modelling

of Exchange Rates," Journal of Finance 57: 1029-35.

Nelson, C. R. and C. I. Plosser, 1982, "Trends and Random Walks in Macroeconomic

Time Series, Some Evidence and Implications," Journal of Monetary Economics

10: 139-62.

----------- and H. Kang, 1981, "Spurious Periodicity in Inappropriately Detrended

Time Series," Econometrica 49: 741-51.

Phillips, P. C. B., 1986, "Understanding Spurious Regressions in Econometrics,"

Journal of Econometrics 33: 311-40.

Plosser, C. I. and G. W. Schwert, 1978, "Money, Income and Sunspots: Measuring

Economic Relationships and the Effects of Differencing," Journal of Monetary

Economics 4: 637-660.

----------------------------- and H. White, 1982, "Differencing as a Test of

Specification," International Economic Review 23: 535-552.

Romer, C., 1986, "The Prewar Business Cycle Reconsidered: New Estimates of Gross

National Product, 1869-1918," NBER Working Paper #1969.

---------, 1987, "Gross National Product, 1909-1928: Existing Estimates, New Estimates, and New Interpretations of World War I and Its Aftermath," Working Paper, Princeton Univ.

Rose, A. K., 1986, "Four Paradoxes in GNP," Economics Letters 22: 137-41.

Rush, M., 1986, "Unexpected Money and Unemployment: 1920 to 1983," Journal of Money, Credit and Banking 18: 259-74.

Schwert, G. W., 1987a, "Tests for Unit Roots: A Monte Carlo Investigation," Working Paper WPB 87-01, W. E. Simon GSBA, Univ. of Rochester.

--------------, 1987b, "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," forthcoming in Journal of Monetary Economics 20.

Stock, J. H. and M. W. Watson, 1986, "Does GNP Have a Unit Root?" Economics Letters 22: 147-51.

Stulz, R. M. and W. Wasserfallen, 1985, "Macroeconomic Time Series, Business Cycles and Macroeconomic Policies," in Carnegie-Rochester Conference Series on Public Policy, Vol. 22.

US Department of Commerce, Survey of Current Business.

------------------------, 1986, The National Income and Product Accounts of the United States, 1929-82: Statistical Tables. Washington: US Government Printing Office.

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