How Big Is the Random Walk in GNP?

John H. Cochrane

The Journal of Political Economy, Vol. 96, No. 5 (Oct., 1988), 893-920.

Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198810%2996%3A5%3C893%3AHBITRW%3E2.0.CO%3B2-P

The Journal of Political Economy is currently published by The University of Chicago Press.

How Big Is the Random Walk in GNP?

John H. Cochrane

University of Chicago

This paper presents a measure of the persistence of fluctuations in GNP based on the variance of its long differences. That measure finds little long-term persistence in GNP. Previous research on this question found a great deal of persistence in GNP, suggesting models such as a random walk. A reconciliation of this paper's results with previous research shows that conventional criteria for time-series model building can produce misleading estimates of persistence.

I. Introduction

Macroeconomists once viewed fluctuations in gross national product as temporary deviations from a trend. The economic theory of business cycles described temporary deviations from "potential GNP," which was assumed to evolve smoothly over time, and data were routinely detrended prior to analysis. A body of recent empirical work (described below) has questioned this time-honored view. By using a variety of time-series models, it finds that fluctuations in GNP are permanent—that a decline in GNP today lowers forecasts of GNP into the infinite future.

This paper reexamines the long-run properties of GNP and argues that GNP does, in fact, revert toward a "trend" following a shock. However, that reversion occurs over a time horizon characteristic of business cycles—several years at least. Therefore, the short-run properties of GNP are consistent with a model with very persistent shocks,

I thank Eugene Fama, Lars Hansen, John Huizinga, Robert Lucas, James Stock, Robert Shiller, an anonymous referee, and the editors of this *Journal* for many helpful comments and suggestions.

[*Journal of Political Economy*, 1988, vol. 96, no. 5]
 0022-3808/88/9605-0003$01.50

and one can incorrectly infer a great deal of long-horizon persistence by fitting a time-series model to this short-run behavior.

The class of time-series model most commonly used to describe temporary deviations about trend is

$$y_t = bt + \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \tag{1}$$

where y_t stands for log GNP, bt describes the trend, and ϵ_t is a random disturbance.[1] Fluctuations in y_t are temporary if $\Sigma a_j \epsilon_{t-j}$ is a stationary stochastic process (y_t is then called "trend stationary"). For $\Sigma a_j \epsilon_{t-j}$ to be stationary, the a_j must approach zero for large j. As a result, a decline in GNP below trend today has no effect on forecasts of the level of GNP, $E_t(y_{t+j})$, in the far future, and it implies that growth rates of GNP must rise above their historical average for a few periods until the trend line is reestablished.

The simplest time-series model that captures permanent fluctuations in GNP is a random walk with drift:

$$y_t = \mu + y_{t-1} + \epsilon_t. \tag{2}$$

Fluctuations in a random walk are permanent in the following sense: suppose that $\epsilon_t = -1$, so that y_t falls one unit below last period's expected value. Then, since $y_{t+j} = y_t + j\mu + \epsilon_{t+1} + \ldots + \epsilon_{t+j}$, forecasts $E_t(y_{t+j})$ fall by one unit for the indefinite future. Also, a low or negative growth rate today implies nothing about growth rates in the future, and there is no tendency for future levels of GNP to revert to a trend line. The random walk is also nonstationary.

The distinction between a random walk (2) and a trend-stationary series (1) is extreme. Long-range forecasts of a random walk move one for one with shocks at each date, while long-range forecasts of a trend-stationary series do not change at all. There are two related ways to think about a series that lies between these two extremes.

First, one can ask how much long-term forecasts respond to shocks. In one interpretation, the measure of this paper asks the question, How much does a one-unit shock to GNP affect forecasts in the far future? If by one unit, it finds a random walk; if by zero, it finds a trend-stationary process like (1). It can also find numbers between zero and one, characterizing a series that returns toward a "trend" in the far future, but does not get all the way there, or it can find a number greater than one, characterizing a series that will continue to

[1] Simple univariate time-series models like (1) should be thought of as a way of capturing the dynamic behavior of y_t that results from a rich multivariate world. They are not "structural" in any way.

diverge from its previously forecast value following a shock. Campbell and Mankiw (1987) originated and emphasize this interpretation.

Second, one can model a series whose fluctuations are partly temporary and partly permanent as a combination of a stationary series and a random walk. The random walk carries the permanent part of a change and the stationary series carries the temporary part of a change. Then, one can ask how important the permanent or random walk component is to the behavior of the series. In a second interpretation, the measure of this paper asks the question, How large is the variance of shocks to the random walk or permanent component of GNP compared with the variance of yearly GNP growth rates? Or, equivalently, How big is the random walk in GNP?

If the variance of the shocks to the random walk component is zero, the series is trend-stationary, and long-term forecasts do not change in response to shocks. If the variance of the shocks to the random walk component is equal to the variance of first differences, the series is a pure random walk. As before, there is a continuous range of possibilities between zero and one and beyond one.

A model consisting of a random walk plus a stationary component may seem quite special. However, I show below that we can think of *any* series whose growth rates or first differences are stationary (any series with a unit root) as a combination of a stationary series plus a random walk. The decomposition into stationary and random walk components is a convenient way of thinking about the properties of a time series, but it adds no structure. I also show that the response to innovations is proportional to the square root of the variance of shocks to a random walk component, so we can freely transform between these two interpretations.

The idea that GNP may contain a random walk goes back to Irving Fisher's "Monte Carlo hypothesis," examined by McCulloch (1975). There is now a large literature following the first half of Nelson and Plosser (1982) that applies the Dickey and Fuller (1979, 1981) and subsequent tests for unit roots to aggregate time series. Since a series with a unit root is equivalent to a series that is composed of a random walk and a stationary component, tests for a unit root are attempts to distinguish between series that have no random walk component (or for which the variance of shocks to the random walk component is zero) and series that have a random walk component (or for which the variance of shocks to the random walk component is between zero and infinity). Stated this way, it is clear why tests for a unit root have low power: it is hard to tell a stationary series from a stationary series plus a very small random walk. This paper and the related literature cited in it go beyond testing for the presence or absence of a unit root

or random walk component and measure how important the unit root or random walk component is to the behavior of a series.

Implications of the Random Walk in GNP

The size of a random walk in GNP is important from a purely statistical viewpoint. Many statistical procedures rely critically on the distinction between series that do not contain a random walk component (1), which we can and should detrend, and first-difference stationary series—(3) below, or series that do contain a random walk component—which we should first-difference prior to analysis. Hypothesis tests that rely on asymptotic distribution theory are an important example because that distribution theory is often quite sensitive to the presence of a random walk component. A measurement of the size of the random walk component can be a better guide to the proper procedure than a unit root test because if the random walk component is small but still nonzero, then an asymptotic distribution theory based on trend stationarity may provide a better approximation in a given small sample than the theory based on a unit root.

The size of a random walk in GNP has been cast as a direct test between competing models of the economy. For example, Nelson and Plosser (1982) interpreted their result that GNP has a large random walk component as evidence for stochastic equilibrium models over traditional monetary or Keynesian business cycle models. They argued that traditional models produce only temporary deviations from trend, while models that find the ultimate source of GNP variability in technology shocks can produce permanent fluctuations.

With the advantages of hindsight, it now seems that the size or existence of a random walk component in GNP cannot directly distinguish broad classes of economic theories of the business cycle at their present stage of development. The Kydland and Prescott (1982) and Long and Plosser (1983) stochastic equilibrium models were constructed precisely to generate temporary fluctuations about trend. On the other hand, King et al. (1987) show that one can modify these models to produce a random walk component by introducing a random walk in the technology shocks or a linear technology for human or physical capital accumulation. Presumably, the same modifications would introduce a random walk component into monetary or "Keynesian" models as well.

Furthermore, the results of this paper are compatible with a variety of random walk components. I show below that an AR(2) about a deterministic trend, which has no random walk component, and a model with a random walk whose variance is 0.18 times the variance of first differences of log GNP account equally well for the results of

this paper. Also, the standard errors in this paper are large, and I argue that this is unavoidable. I conclude that the existence or size of a random walk component in GNP is not a precisely measured "stylized fact" that we should require any reasonable model to reproduce.

The most promising direct use for the point estimates of the size of a random walk component in this paper may be the calibration of a given model rather than a test that can distinguish competing classes of models. If a model (like the ones cited above) produces a random walk in GNP, the results of this paper suggest that the parameters of that model should be picked to also generate interesting short-run dynamics of GNP, so that the variance of yearly changes in GNP is much larger than the variance of shocks to its random walk component.

Other Estimates

Several authors have estimated the persistence of fluctuations in GNP, and their estimates vary greatly. Nelson and Plosser (1982) matched a model consisting of permanent and temporary components to a stylized autocorrelation function for growth rates of GNP and concluded that the permanent component was more important than the temporary component. Watson (1986) and Clark (1987) estimated different unobserved components models and found a small permanent component. Campbell and Mankiw (1987) estimated the effect of a shock on long-term forecasts of GNP from the parameters of low-order autoregressive, moving average (ARMA) representations of postwar GNP and found a large random walk component.

Several authors have examined the persistence of fluctuations in other time series using a variety of methods. Rose (1986) presents a survey of papers that find large random walk components in various macroeconomic time series. In finance, conventional wisdom favored the random walk model while macroeconomists favored the trend-stationary model. Poterba and Summers (1987), Fama and French (1988), and Lo and MacKinlay (1988) use variance ratio estimators similar to the one used in this paper and related estimators to document a temporary component in stock prices. Huizinga (1987) uses a closely related estimator to document a temporary component in real exchange rates. Cochrane and Sbordone (1988) present a multivariate extension.

This Paper's Technique

In this paper, I measure the size of a random walk component in GNP from the variance of its long differences. The intuition behind this

measure comes from the following argument: Imagine that log GNP, denoted y_t, is a pure random walk (model [2]). Then the variance of its k-differences grows linearly with the difference k: $\text{var}(y_t - y_{t-k}) = k\sigma_\epsilon^2$. On the other hand, if log GNP is stationary about a trend (model [1]), the variance of its k-differences approaches a constant, twice the unconditional variance of the series: $\text{var}(y_t - y_{t-k}) \to 2\sigma_y^2$. Now plot $(1/k)\text{var}(y_t - y_{t-k})$ as a function of k. If y_t is a random walk, the plot should be constant at σ_ϵ^2. If y_t is trend-stationary, the plot should decline toward zero.

Next, suppose that fluctuations in GNP are partly permanent and partly temporary, which we can model as a combination of a stationary series and a random walk. Now the plot of $(1/k)\text{var}(y_t - y_{t-k})$ versus k should settle down to the variance of the shock to the random walk component.

If fluctuations in GNP are partly temporary—if the random walk component is small and a shock today will be partially reversed in the long run—that reversal is likely to be slow, loosely structured, and not easily captured in a simple parametric model. The variance of k-differences can find such loosely structured reversion, whereas many other approaches cannot. I show in Section IV that this difference can reconcile the results of this paper with other measures of the permanence of fluctuations in GNP.

Results

Figure 1 and table 1 present $(1/k)\text{var}(y_t - y_{t-k})$ for log real per capita GNP, 1869–1986. Pre-1939 data are taken from Friedman and Schwartz (1982). I use real per capita GNP to eliminate possible nonstationarity induced by inflation or population growth. (Henceforth, I will refer to log real per capita GNP as just "GNP.") Figure 1 and table 1 also include asymptotic standard errors, discussed below. Table 1 also presents $1/k$ times the variance of k-differences divided by the variance of first differences (the variance ratio). The units in table 1 and figure 1 are annual percentage growth.

Since $1/k$ times the variance of k-differences settles down to about one-third of the variance of first differences, figure 1 and table 1 suggest that the innovation variance of the random walk component is about one-third of the variance of year-to-year changes: annual growth rates of GNP contain a large temporary component. In fact, I show below that the pattern of figure 1 is consistent with a deterministic trend, which has *no* permanent or random walk component, and whose fluctuations are entirely temporary.

Figure 2 presents the log of real per capita GNP. Notice that this data set looks as if it has a trend in it. Fluctuations occur, but the level

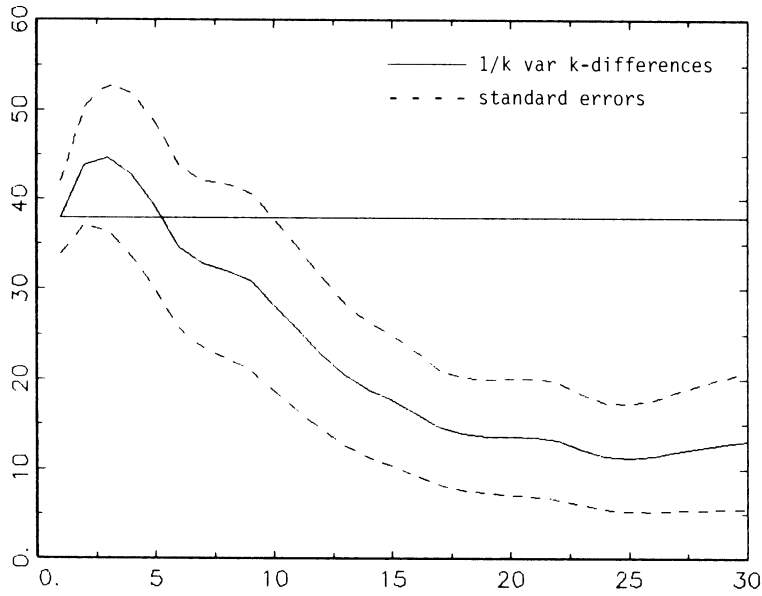

FIG. 1.—$1/k$ times the variance of k-differences of log real per capita GNP, 1869–1986, with asymptotic standard errors.

of the series always returns to the "trend line." Furthermore, that trend line is linear: there are no "waves" of low-frequency movement. These characteristics drive the finding of a small random walk component. (Note that low-frequency movement generated by a nonlinear trend, a shift, etc. would show up as a large random walk component in this and most other estimation techniques based on linear time-series models.)

Prewar GNP data are more variable than postwar data, and one might suspect that this characteristic drives the result. However, figure 3 and table 1 present $1/k$ times the variance of k-differences for postwar GNP, and the same pattern is evident. Both the variance of first differences and the variance of the random walk component are lower, but their proportions do not change much.[2]

[2] The pattern of fig. 2 is sensitive to the precise specification of the variables. First, the variance of quarterly differences of seasonally adjusted GNP is less than one-fourth the variance of yearly differences, so the variance ratio is higher if one uses quarterly rather than annual differences in the denominator. This observation explains most of the difference between fig. 2 and the results reported by Campbell and Mankiw (1988), who use a similar technique on quarterly data. Second, taking the variance of overlapping k-year differences of quarterly data vs. the variance of k-year differences of annual averages, including or excluding population growth, taking logs or not, and even changing the sample by a few years can all change the variance ratio by about one standard error.

TABLE 1

1/k TIMES THE VARIANCE OF k-DIFFERENCES OF GNP

	k (Years)									
	1	2	3	4	5	10	15	20	25	30
	1869–1986									
$\hat{\sigma}_k^2$	40.0	43.8	44.6	42.8	39.2	28.2	17.7	13.6	11.3	13.1
	(4.1)	(6.6)	(8.2)	(9.1)	(9.4)	(9.5)	(7.3)	(6.5)	(6.0)	(7.7)
$\hat{\sigma}_k^2/\hat{\sigma}_1^2$	1.00	1.15	1.17	1.13	1.03	.74	.47	.36	.30	.35
	(.11)	(.17)	(.22)	(.24)	(.25)	(.25)	(.19)	(.17)	(.16)	(.20)
$\hat{\sigma}_k$	6.1	6.6	6.7	6.5	6.3	5.3	4.2	3.7	3.4	3.6
	1947–86									
$\hat{\sigma}_k^2$	7.0	8.2	8.0	7.3	6.5	4.5	2.9	2.6		
	(1.3)	(2.2)	(2.6)	(2.8)	(2.8)	(2.7)	(2.1)	(2.2)		
$\hat{\sigma}_k^2/\hat{\sigma}_1^2$	1.00	1.17	1.14	1.05	.93	.63	.42	.38		
	(.19)	(.31)	(.37)	(.39)	(.39)	(.38)	(.30)	(.32)		
$\hat{\sigma}_k$	2.7	2.9	2.8	2.7	2.6	2.1	1.7	1.6		

NOTE.—$\hat{\sigma}_k^2$ is 1/k times the sample variance of k-differences. Standard errors (in parentheses) are the Bartlett standard error, with $\hat{\sigma}_1^2$ used for the random walk component; i.e., standard error is $(4k/3T)^{.5}\hat{\sigma}_k^2$. $\hat{\sigma}_k^2/\hat{\sigma}_1^2$ and its standard error are the same quantities divided by $\hat{\sigma}_1^2$. $\hat{\sigma}_k$ is the square root of $\hat{\sigma}_k^2$; its units are percentage growth rates.

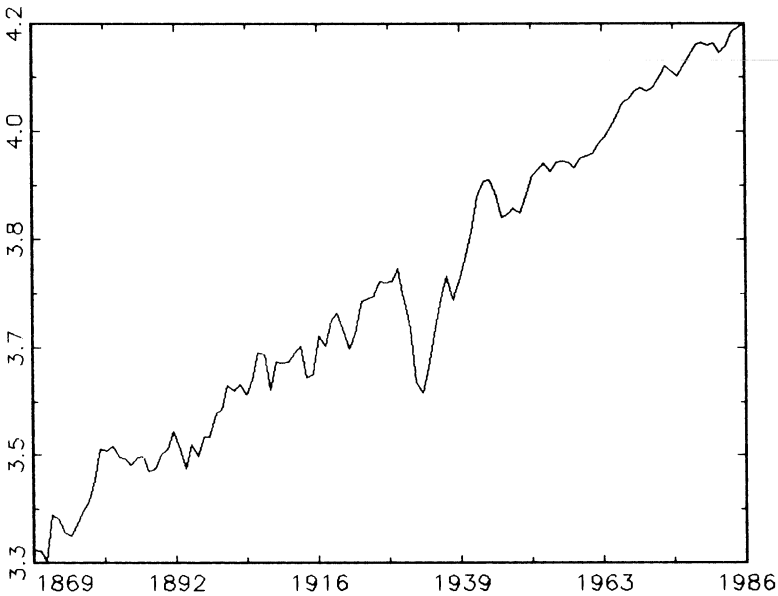

FIG. 2.—Log real per capita GNP, 1869–1986

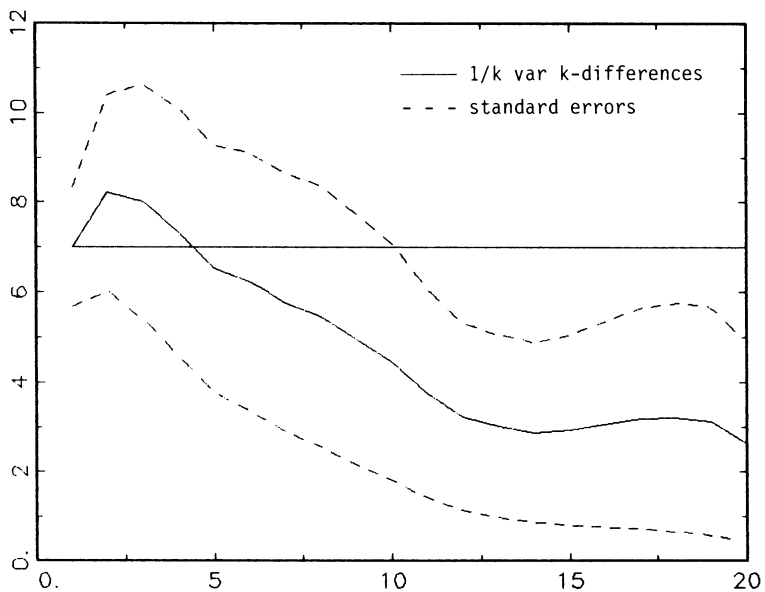

FIG. 3.—$1/k$ times the variance of k-differences of log real per capita GNP, 1947–86, with asymptotic standard errors.

Romer (1986) argued that prewar GNP data overstate the actual cyclical variability of GNP. This possibility will not bias the estimate of the variance of the random walk component. Taking k-differences acts as a filter that ignores cyclical fluctuations and concentrates on the variability of longer "runs," so a different GNP data set will have a different variance of k-differences if the early GNP has a significantly different and more variable trend line, not if its cyclical fluctuations are different. A graph similar to figure 1, using Romer's adjusted early GNP series, produces a variance of a random walk component very similar to that of figure 1. It should because Romer kept the decade trends the same in her corrections for cyclical volatility. Her criticism, or the seasonal adjustment of quarterly data, will affect the variance of first differences, so the variance *ratio* can be biased by excessive volatility or smoothness of the first differences.

The presence of a splice in 1947 also does not drive the result. Every long series of GNP data contains at least one splice. The wide surveys used to construct later data are simply not available for earlier periods, so some projection using a restricted set of industries is unavoidable. However, forcing the *levels* of the "old" and "new" GNP series to match at a certain date does not bias the variance of k-differences. It is biased only if the old series has different growth rates over long horizons.

The body of this paper consists of an investigation of $1/k$ times the variance of k-differences as an estimate of the random walk component in GNP. Section II provides several interpretations of a random walk component. Section III discusses estimation. Section IV reconciles these results with previous research that found a large random walk component by showing how conventional time-series estimation techniques can provide misleading estimates of a random walk component. Section V contains a summary and concluding remarks.

II. Unit Roots and Random Walk Components

This section discusses and documents several claims in the Introduction about the representation of time series. It shows that first-difference stationary time series or time series with a unit root are equivalent to time series that are composed of a stationary and a random walk component. It argues that the variance of shocks to a random walk component is just a convenient interpretation of the parameters of an arbitrary first-difference stationary series, but it requires no additional structure. It shows how to transform between the variance of a random walk component and the response of long-term forecasts to a shock.

Assume that log GNP follows a first-difference stationary linear process; that is, growth rates of GNP are stationary. In this case, log GNP has a moving average representation of the form

$$\Delta y_t = (1 - L)y_t = \mu + A(L)\epsilon_t = \mu + \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \tag{3}$$

which I take as the starting point; L is the lag operator, $Ly_t = y_{t-1}$. The first equality defines the notation Δy_t and $(1 - L)y_t$ for first differences of y_t. The last equality defines the lag polynomial notation $A(L)$. The ϵ_t are independent identically distributed (i.i.d.) error terms with common variance σ_ϵ^2.

The random walk process (2) obviously has a representation of the form (3). The trend-stationary process (1) is a limiting case of (3): if $\mu = b$ and if the lag polynomial $A(L)$ in (3) has a unit root—that is, we can express $A(L) = (1 - L)B(L)$—we recover (1) by canceling the terms $(1 - L)$. Many unobserved components models are first-difference stationary and hence have a representation (3). Nelson and Plosser (1982) and Watson (1986) are examples. On the other hand, (3) does not include nonlinear processes such as Quah (1986), a process with a nonlinear trend, or second-difference stationary processes (the growth rates of GNP follow a random walk) as in Clark (1987).

Given the representation (3), we have the following fact.

FACT 1. Any first-difference stationary processes can be represented as the sum of stationary and random walk components.

To show that a representation as stationary plus random walk components exists, we simply construct it from the representation (3). This decomposition comes from Beveridge and Nelson (1981). Let

$$y_t = z_t + c_t, \tag{4}$$

where

$$z_t = \mu + z_{t-1} + \left(\sum_{j=0}^{\infty} a_j\right)\epsilon_t,$$

$$-c_t = \left(\sum_{j=1}^{\infty} a_j\right)\epsilon_t + \left(\sum_{j=2}^{\infty} a_j\right)\epsilon_{t-1} + \left(\sum_{j=3}^{\infty} a_j\right)\epsilon_{t-2} + \ldots.$$

This decomposition is constructed so that $\lim_{k\to\infty} E_t y_{t+k} = z_t + k\mu$; that is, long-term forecasts of y_t converge to z_t plus $k\mu$. In this sense, z_t is the permanent component of y_t. Beveridge and Nelson call it a stochastic trend. Long-term forecasts of y_t are unaffected by c_t, the temporary component.

The innovation variance of the random walk component $\sigma_{\Delta z}^2$ is a natural measure of the importance of the random walk component.

From the definition (4) we can write the variance of the random walk component $\sigma^2_{\Delta z}$ in terms of the moving average representation (3):

$$\sigma^2_{\Delta z} = (\Sigma a_j)^2\sigma^2_\epsilon = |A(1)|^2\sigma^2_\epsilon \tag{5}$$

(sums without indices run from zero to infinity).

In the Beveridge and Nelson decomposition (4), the innovations in the random walk and stationary components are identical. In a more general combination of random walk and stationary components, the innovations may be correlated:

$$\begin{aligned} y_t &= z_t + c_t, \\ z_t &= \mu + z_{t-1} + \eta_t, \\ c_t &= B(L)\delta_t, \quad E(\eta_t\,\delta_t) \text{ arbitrary.} \end{aligned} \tag{6}$$

If we start with a process (6), Δy_t is stationary, and so the process has a representation of the form (3). Most processes of the form (3) can be decomposed into a variety of processes (6), with varying correlation between the innovations; but only the decomposition (4) is *guaranteed* to exist.[3]

Since a variety of decompositions into stationary and random walk components of the form (6) exist for any given stationary process (3), a measure based on the variance of the random walk component would be in serious trouble if it depended crucially on which arbitrary decomposition we choose. Fortunately, it does not, as seen in the following fact.

FACT 2. In every decomposition of a process (1) into stationary and random walk components (6), the innovation variance of the random walk component is the same: $\sigma^2_{\Delta z} = (\Sigma a_j)^2\sigma^2_\epsilon$.

To show fact 2, start with an arbitrary decomposition (6). The corresponding moving average representation of the form (3) is

$$(1 - L)y_t = \mu + \nu_t + (1 - L)B(L)\delta_t \equiv \mu + A(L)\epsilon_t. \tag{7}$$

The last equality defines the parameters $A(L)$ of a moving average representation from the parameters $B(L)$ of (6). Now form the Beveridge and Nelson decomposition of both sides of the last equality in (7). Since the processes on both sides of the last equality are the same, they must have the same variance of a random walk component, so we must have[4] $|A(1)|^2\sigma^2_\epsilon = \sigma^2_\nu$. The correlation between ν_t and

[3] Watson (1986) derives this fact. For example, if we seek a representation with uncorrelated innovations, the spectral density of the combination can be no less than the spectral density of each component; thus such a representation exists only if the spectral density of the first differences has a global minimum at zero.

[4] This statement can be more compactly derived by noting that for the processes on each side of the last equality in (7) to be the same, their spectral densities must be the same at all frequencies, and zero in particular.

δ_t is irrelevant for this argument, so the innovation variance of every decomposition (6) of the same moving average representation (3) must have the same variance of shocks to the random walk component. This argument demonstrates fact 2.

There is one more interpretation, which will be useful in the next section. The spectral density[5] of Δy_t is, by (1), $S_{\Delta y}(e^{-i\omega}) = |A(e^{-i\omega})|^2\sigma_\epsilon^2$. Therefore, we have the following fact.

FACT 3. The innovation variance of the random walk component is equal to the spectral density of Δy_t at frequency zero, that is,

$$\sigma_{\Delta z}^2 = (\Sigma a_j)^2\sigma_\epsilon^2 = S_{\Delta y}(e^{-i0})\sigma_\epsilon^2 \tag{8}$$

or, dividing by the variance of first differences,

$$\frac{\sigma_{\Delta z}^2}{\sigma_{\Delta y}^2} = \frac{(\Sigma a_j)^2}{\Sigma a_j^2} = \frac{S_{\Delta y}(e^{-i0})}{\sigma_{\Delta y}^2}. \tag{8'}$$

Equations (8) and (8′) summarize three equivalent ways of looking at the long-run properties of a series: we can break it into permanent (random walk) and temporary (stationary) components, we can examine the response of long-term forecasts to an innovation, or we can examine the spectral density at frequency zero of its first differences. All three interpretations allow us to think of the permanence of the fluctuations in a series as a continuous phenomenon rather than a discrete choice. Furthermore, equations (8) and (8′) show that the quantity $\sigma_{\Delta z}^2$ or $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2$ defined from the Beveridge and Nelson decomposition (3) is no more than a useful interpretation of the sum of the moving average coefficients Σa_j. The decomposition into stationary and random walk components adds no structure.

The variance of shocks to the random walk component or spectral density at frequency zero of first differences also captures *all* the effects of a unit root on the behavior of a series in a finite sample. As a sample of T observations of a series is completely characterized by its $T - 1$ autocovariances, it is also completely characterized by $T - 1$ periodogram ordinates. By changing the periodogram ordinate at frequency zero of first differences without changing the others, we can make a stationary series into a series with a unit root or random walk component and vice versa.[6]

Since the size of a random walk component is a continuous choice, any test for trend stationarity ($\sigma_{\Delta z}^2 = 0$ or $S_{\Delta y}(e^{-i0}) = 0$) must have arbitrarily low power against the alternative of a small enough ran-

[5] I use the notation $S(e^{-i\omega})$ for the spectral density at frequency ω and, hence, $S(e^{i0})$ for the spectral density at $\omega = 0$.

[6] With an infinite sample, or in population, this proposition does not hold. The spectral density is defined only almost everywhere; and in some cases we can bound the variation of the population spectral density function with very weak assumptions.

dom walk component $\sigma^2_{\Delta z}$. As a result, efforts to categorize series as trend-stationary or difference-stationary and read great things into the difference between the two will not be very fruitful.

III. Estimation

I claimed in the Introduction that the variance of k-differences could be used to estimate the innovation variance of a random walk component. To document that claim and to provide standard errors, this section discusses the statistical properties of the variance of k-differences.

Asymptotic Properties

Let σ^2_k denote $1/k$ times the population variance of k-differences of y_t, $\sigma^2_k = k^{-1} \operatorname{var}(y_t - y_{t-k})$; σ^2_k is related to the autocorrelation coefficients of Δy_t by

$$\sigma^2_k = \left(1 + 2\sum_{j=1}^{k-1} \frac{k-j}{k} \rho_j\right) \sigma^2_{\Delta y}, \tag{9}$$

where $\sigma^2_{\Delta y} = \operatorname{var}(y_t - y_{t-1})$ and $\rho_j = \operatorname{cov}(\Delta y_t \Delta y_{t-j})/\sigma^2_{\Delta y}$. The derivation is straightforward but tedious, so it is presented in the Appendix. Equation (9) shows that the limit of σ^2_k is indeed the innovation variance of the random walk component:

$$\lim_{k\to\infty} \sigma^2_k = \left(1 + 2\sum_{j=1}^{\infty} \rho_j\right) \sigma^2_{\Delta y} = S_{\Delta y}(e^{-i0}) = \sigma^2_{\Delta z}. \tag{10}$$

The second equality is the definition of spectral density, while the third is reproduced from equation (8).

Equation (9) suggests that we could also estimate $1/k$ times the variance of k-differences by using sample autocorrelations $\hat{\rho}_j$ in the place of their population values ρ_j. (Huizinga [1987] and Campbell and Mankiw [1988] perform the calculation this way.) The right-hand side of (9) with $\hat{\rho}_j$ in place of ρ_j is the definition of the Bartlett estimator of the spectral density at frequency zero (Anderson 1971, p. 511). Hence, $1/k$ times the variance of k-differences is asymptotically equivalent to the Bartlett estimator.[7]

[7] $1/k$ times the variance of k-differences and the conventional Bartlett estimate are not identical in small samples. The estimates of sample autocorrelations implied by the sample variance of k-differences underweight observations k dates away from the endpoints, compared with the usual estimates of autocorrelation. The difference disappears asymptotically but may be important in small samples. Also, the conventional Bartlett estimate is not unbiased in small samples, as the corrected $1/k$ times the variance of k-differences $\hat{\sigma}^2_k$ is for a random walk. I thank John Huizinga for pointing this out.

The properties of the Bartlett estimator are well known, so we can establish the asymptotic properties of $1/k$ times the variance of k-differences by reference to those of the Bartlett estimator. In particular, (1) if $k/T \to 0$ as $T \to \infty$, where T is the sample size, $1/k$ times the sample variance of k-differences is a consistent estimate of the spectral density at frequency zero; (2) the asymptotic variance of σ_k^2 is $4kS^2(e^{-i0})/3T$ (Anderson 1971, p. 531).

The equivalence between $1/k$ times the variance of k-differences and the Bartlett estimator provides a useful interpretation of the variance of k-differences for readers familiar with spectral density estimation; in turn, the variance of k-differences is a useful and intuitive time domain counterpart to the Bartlett spectral density estimator. To use the Bartlett estimator, we have to decide what k to use: how many autocovariances or autocorrelations to include in (9) or how many periodogram ordinates to smooth. The choice of k requires a trade-off between bias and efficiency, and it is usually made arbitrarily. In this context, a plot of $1/k$ times the variance of k-differences versus k is an experimental determination of the proper k or window width.

Small-Sample Properties

In small samples, $1/k$ times the variance of k-differences and the Bartlett estimator can be biased, and the asymptotic standard errors may be a poor approximation to the actual standard errors. In this subsection, I discuss corrections for small-sample bias, and I present some Monte Carlo experiments to evaluate standard errors.

I corrected for two sources of small-sample bias in the sample variance of k-differences. These corrections produce an estimator of σ_k^2 that is unbiased when applied to a pure random walk with drift. First, I used the sample mean of the first differences to estimate the drift term μ at all k rather than estimate a different drift term at each k from the mean of the k-differences. Second, I included a degrees of freedom correction $T/(T - k + 1)$. Without this correction, $1/k$ times the variance of k-differences declines toward zero as $k \to T$ for any process because you cannot take a variance with one data point.

I will use the notation $\hat{\sigma}_k^2$ to denote $1/k$ times the bias-corrected sample variance of k-differences. The formula for $\hat{\sigma}_k^2$ is presented in the Appendix as equation (A3). The Appendix also contains a proof that $\hat{\sigma}_k^2$ is unbiased when y_t is a random walk with drift.

Table 2 presents standard errors from a Monte Carlo experiment using 100 observations of a random walk with drift. I picked the innovation variance of this random walk $\sigma_\epsilon^2 = \sigma_{\Delta z}^2 = 1$. The mean of $\hat{\sigma}_k^2$ was very close to one at all k in this experiment, confirming the bias corrections for a pure random walk. The table presents the standard

TABLE 2

MONTE CARLO STANDARD ERRORS FOR $1/k$ TIMES THE VARIANCE OF k DIFFERENCES
Model: $y_t = 1 + y_{t-1} + \epsilon_t$; $\sigma_\epsilon^2 = 1$ ($T = 100$, 500 trials)

	$100k/T$									
	1	2	3	4	5	10	20	30	40	50
Monte Carlo	.137	.160	.200	.231	.263	.409	.607	.772	.888	.896
Bartlett*	.115	.163	.200	.231	.258	.365	.516	.632	.730	.816

* This row gives $(4k/3T)^{.5}$.

errors from the Monte Carlo experiment and the corresponding Bartlett standard errors for comparison. The Bartlett errors slightly understate the Monte Carlo errors at large k/T, but the difference is small compared to the size of the standard errors. Monte Carlo experiments with different sample sizes and random walk variance confirm that the standard errors of table 2 scale with k/T and the innovation variance of the random walk.

What about processes that are more complicated than a pure random walk? The Appendix presents a derivation of $E(\hat{\sigma}_k^2)$ for a first-order moving average: $(1 - L)y_t = \mu + (1 + \theta L)\epsilon_t$. It shows that $E(\hat{\sigma}_k^2)$ approaches $\sigma_{\Delta z}^2$ for large k, so $\hat{\sigma}_k^2$ can recover the variance of the random walk component for this process as well.

I ran several further Monte Carlo simulations to examine whether the variance of k-differences is robust when applied to more complicated processes for GNP. I fit a variety of ARMA processes to first differences of log real per capita GNP, simulated 118 observations of each process, and computed $\hat{\sigma}_k^2$ in 100 trials. In each case, the mean of $\hat{\sigma}_k^2$ at $k = 30$ was equal to the variance of the random walk component implied by the estimated ARMA processes—$k = 30$ was large enough to identify the random walk from the stationary components—and the standard errors at large k were close to those implied by table 2, scaled to the variance of the random walk component.

All the low-order ARMA processes produced $\hat{\sigma}_k^2$ lines that *rise* for k from 1 to 5 and then are flat at the variance of the random walk component from $k = 10$ on, unlike figure 1. They implied $\sigma_{\Delta z}^2 > \sigma_{\Delta y}^2$. Two processes that do capture the behavior of figure 1 are an AR(15), figure 4, and AR(2) about a deterministic trend, figure 5. In the next section, I will discuss why the low-order ARMA models failed to capture the behavior of figure 1. For now, note that since they replicate the behavior of $\hat{\sigma}_k^2$ for GNP, figures 4 and 5 can provide small-sample standard errors. These standard errors are similar to the asymptotic standard errors used in figure 1.

Figures 4 and 5 also include $\hat{\sigma}_k^2$ for GNP from figure 1, marked

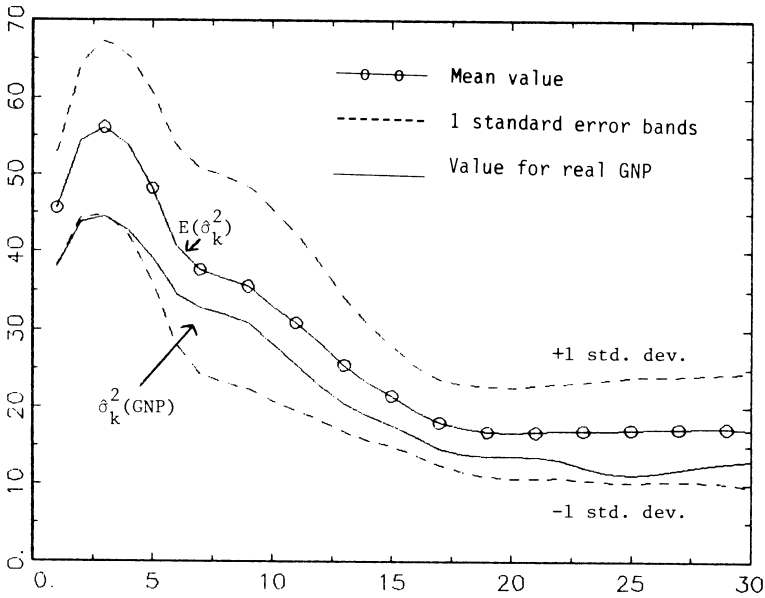

FIG. 4.—Monte Carlo simulation of an AR(15)

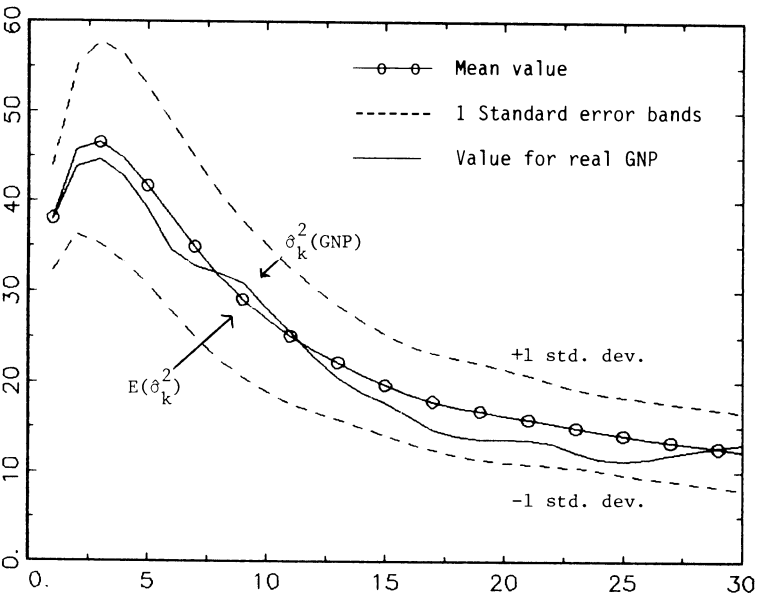

FIG. 5.—Monte Carlo simulation of an AR(2) with a linear trend

$\hat{\sigma}_k^2$(GNP). Since the $\hat{\sigma}_k^2$(GNP) line falls inside the one-standard-error bands, neither model can be rejected for real GNP. However, the standard errors from the random walk (table 2) or any of the other low-order ARMA processes are large enough that we cannot reject them at 5 percent either. (Note that the standard errors scale with the size of the random walk component. Under the hypothesis of a random walk, the standard errors are bigger than indicated in fig. 1.) A confidence interval includes both $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2 = 0$ and 1.

While this is unfortunate, I will argue below that estimates of a random walk component are limited by the number of nonoverlapping "long runs" in the data set, so that large efficiency gains are not possible without imposing additional structure on the time-series process for GNP. As a result, this and related exercises can provide a point estimate of the size of a random walk component with associated standard errors but will not provide useful tests to discriminate between models that imply various sizes of the random walk component.

The parameters of the AR(15) model imply that the variance ratio $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2 \simeq .18$, while the AR(2) about a trend implies $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2 = 0$. Hence, the simulations behind figures 4 and 5 also reveal an *upward* bias in $\hat{\sigma}_k^2$ as an estimate of the random walk component when the series has a small random walk component or is trend-stationary.

In summary, $1/k$ times the variance of k-differences $\hat{\sigma}_k^2$ provides an upward-biased point estimate of the variance ratio $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2$ of about .34, and two models with $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2 = .18$ and 0 replicate the behavior of the variance of k-differences of GNP. However, standard errors are large enough that we cannot statistically reject variance ratios between zero and one at conventional levels of significance.

IV. Reconciliation with Previous Estimates

Given the definition of the random walk component in terms of the parameters of a moving average representation, (4) or (8) above, the obvious thing to do is either to estimate a parsimonious time-series model for Δy_t and calculate Σa_j or to identify and estimate a simple parametric unobserved components model like (4). Campbell and Mankiw (1987) and Nelson and Plosser (1982) did just that, respectively, and both found large random walk components. Why do Nelson and Plosser and Campbell and Mankiw find large random walk components, while Watson (1986), Clark (1987), and I find small ones? Though there are small differences in definition—which quantities we look at to measure the importance of unit roots or random walk components—the major difference is in estimation strategies.

Nelson and Plosser specified an unobserved components model of the form

$$\begin{aligned} y_t &= u_t + v_t, \\ (1 - L)u_t &= \mu + A(L)\epsilon_t, \quad \epsilon_t \text{ i.i.d.}, \\ v_t &= B(L)\delta_t, \quad \delta_t \text{ i.i.d.} \end{aligned} \tag{11}$$

(ϵ_t and δ_t may be correlated). They identified the two components from a stylized autocorrelation function of GNP growth rates. If the first autocorrelation of Δy_t is positive but the others are zero, then the only model of the form (11) that works is $A(L) = 1$ and $B(L) = (1 + \theta L)$. By examining plausible parameter values for this restricted model, Nelson and Plosser concluded that $\sigma_\epsilon^2 > \sigma_\delta^2$.[8]

Campbell and Mankiw (1987) estimated parsimonious ARMA representations of log GNP, using seasonally adjusted quarterly postwar data. They measured the importance of the random walk component by $\Sigma a_j = A(1)$, the change in z_t (the long-term forecast) in response to a unit univariate innovation in GNP. They found values for $A(1)$ equal to or larger than one, which imply an innovation variance of the random walk component greater than the variance of first differences of GNP.[9]

[8] This measure of the importance of a random walk component has the conceptual disadvantage that it depends on which arbitrary unobserved components decomposition we choose. For example, since every series of the form (11) has a unique moving average representation, we could rewrite (11) as $(1 - L)y_t = \mu + C(L)\nu_t$, ν_t i.i.d., and eliminate the stationary component. Alternatively, we could use the Beveridge and Nelson decomposition of Sec. II to make the component with a unit root into a pure random walk:

$$\begin{aligned} y_t &= z_t + c_t, \\ (1 - L)z_t &= \mu + \nu_t, \quad \nu_t \text{ i.i.d.}, \\ c_t &= C(L)\zeta_t, \quad \zeta_t \text{ i.i.d.} \end{aligned}$$

These representations are observationally equivalent to the first form (11), but the measure $\sigma_\epsilon^2/\sigma_\delta^2$ changes according to which one we choose. In contrast, the innovation variance of a random walk component is invariant to the choice of decomposition (fact 2 in Sec. II). Also, the ratio of the innovation variance of the two components is not a good measure of their relative importance because the proportion of the variance of Δy_t explained by u_t and v_t depends on the coefficients of $A(L)$ and $B(L)$ as well as the ratio $\sigma_\epsilon^2/\sigma_\delta^2$.

[9] There are some conceptual disadvantages to scaling a persistence measure by the univariate innovations of y_t. The univariate innovations are not observable and must be inferred from a model; the univariate innovations do not correspond to the "surprise" movement because we live in a multivariate environment; a series may have small innovations but a large variance. For example, $\Delta y_t = 1.5\Delta y_{t-1} - .95\Delta y_{t-2} + \epsilon_t$. For this process, $\Sigma a_j = 2.22$ but $\sigma_{\Delta z}^2/\sigma_{\Delta y}^2 = (\Sigma a_j)^2/(\Sigma a_j^2) = 0.20$. However, for the GNP data used in this paper, there is little qualitative difference between the two definitions, and the difference in results must be explained by differences in estimation strategy.

In performing the Monte Carlo simulations of Section II, I also found that low-order ARMA models of GNP imply that σ_k^2 should rise with k, and they imply a large random walk component, while in fact σ_k^2 declines and the estimated random walk component is small. To replicate the behavior of σ_k^2 for GNP, I had to estimate an AR(15) or impose a deterministic trend.

To investigate this fact further, I fit a variety of ARMA processes to GNP growth rates, ranging from white noise out to an AR(15) (see table 3).[10] All representations past white noise are adequate by usual standards: the Durbin-Watson statistics are close to 2, the significance levels of the Q-statistic are around .5, the parameters of overfit models are statistically insignificant, and so forth. But the variance ratio and Σa_j start at about 1.2 for second-order processes and decline steadily to a variance ratio of .18 and $\Sigma a_j = .5$ for an AR(15). Low-order ARMA models systematically overestimate the random walk component of GNP, even though they adequately represent the series by all the usual diagnostic tests. The question is, why?

The innovation variance of a random walk component is a property of the very long-run behavior of a series alone. It is the spectral density at the frequency $\omega = 0$ corresponding to a period or "run" of infinity, it is related to the infinite sum of the moving average coefficients $|\Sigma a_j|^2$ or the autocorrelation coefficients $(1 + 2\Sigma\rho_j)$, and it corresponds to the effect of a shock today on forecasts into the infinite future. In theory, then, we should have to wait an *infinite* amount of time to get just one observation on the size of the random walk component!

In practice, we typically believe that the dynamic response of GNP to a shock is flat after a suitable long run has arrived.[11] This belief is implicit above: the graphs stop after the thirtieth difference, reflecting a belief that after 30 years the temporary effects of business cycles are over. The number of nonoverlapping long runs is a rough guide to the number of degrees of freedom (precisely, the number of periodogram ordinates) in this exercise. With a 10–20-year long run there are no more than five to 10 independent observations in 100 years of data and two to four observations in postwar data. Obviously, using more frequently sampled data does not help.

Estimating an unobserved components model or a parsimonious

[10] I used the RATS program to perform the estimation. Autoregressive models are estimated by ordinary least squares and moving average models by conditional maximum likelihood. The unreported moving average models did not converge.

[11] Precisely, if the coefficients of the moving average representation (1) are zero past a long-run value $M < \infty$, then the derivative of the spectral density of Δy at zero is bounded. If y_t is in fact trend-stationary and the spectral density of Δy at frequency zero is in fact zero, then the slope of the spectral density of Δy at zero is also zero.

TABLE 3

ARMA REPRESENTATIONS OF GNP

	White Noise	AR(1)	MA(1)	ARMA (1, 1)	AR(2)	MA(2)	AR(3)
Durbin-Watson statistic	1.67	1.97	1.97	1.99	1.91	1.99	1.96
Q-statistic*	34.82 (30)	27.68 (29)	27.58 (29)	27.43 (28)	26.95 (29)	27.75 (29)	24.58 (27)
Significance level	.25	.54	.54	.49	.52	.48	.60
Variance ratio $(\Sigma a_j)^2/(\Sigma a_j^2)$	1.00	1.39	1.34	1.40	1.23	1.27	.97
A(1) $(= \Sigma a_j)$	1.00	1.20	1.11	1.18	1.11	1.16	1.02
	MA(3)	AR(4)	ARMA (3, 1)	MA(4)	AR(5)	AR(10)	AR(15)
Durbin-Watson statistic	1.97	2.06	2.00	1.91	1.95	1.99	2.01
Q-statistic*	24.95 (27)	28.63 (26)	24.89 (26)	29.21 (26)	26.16 (25)	20.04 (20)	12.17 (15)
Significance level	.58	.33	.53	.30	.40	.46	.67
Variance ratio	.75	.74	1.07	.34	.53	.41	.18
A(1)	.89	.90	1.07	.60	.77	.69	.45

* Degrees of freedom are in parentheses.

ARMA model is an attempt to circumvent this problem. These models make identifying restrictions across frequencies: they draw inferences about the long-run (high-order autocorrelation or low-frequency) dynamics from a model fit to the short-run (low-order autocorrelation or high-frequency) dynamics. For an example that demonstrates how "effective" these procedures are, Campbell and Mankiw (1987) report estimates such as $A(1) = 1.306 \pm .073$ for the 20-year forecast of GNP. Since there are only two nonoverlapping 20-year forecasts in their data set, it is clear how heavily their estimates of $A(1)$ depend on the identifying assumption that the series follow a given low-order ARMA model.

If the short- and long-run dynamics of GNP can both be captured by the assumed time-series model, these procedures can help estimation because we have much more data on high-frequency fluctuations. However, if the long-run dynamics cannot be captured in the model used to study the short run, these identification procedures bias conclusions about long-run behavior.

I offer two ways to see this fact. First, recall that the variance of the shock to the random walk component is related to the sum of the autocorrelations by

$$\frac{\sigma^2_{\Delta z}}{\sigma^2_{\Delta y}} = 1 + 2\sum_{j=1}^{\infty} \rho_j. \tag{12}$$

When we model short-run dynamics, we safely ignore high-order statistically insignificant autocorrelations or we slightly misspecify them by fitting a simple model. But all autocorrelations enter into (12) equally, so a large number of small high-order autocorrelations can offset a few large low-order autocorrelations.

Second, GNP growth has a positive autocorrelation at short lags and a small random walk component at long lags. A simple time-series model may not be able to capture both kinds of behavior. For example, if $(1 - L)y_t = \mu + (1 + \theta L)\epsilon_t$, we need $\theta > 0$ to capture positive first-order autocorrelation but $\theta < 0$ to capture a small random walk component. Faced with a choice, maximum likelihood estimates match the short-run behavior (they fit $\theta > 0$ in the example) and misrepresent the long-run behavior.

The Appendix contains a demonstration of this property of maximum likelihood estimates. It shows that maximum likelihood estimates of a model such as a low-order ARMA or a simple parametric unobserved components model pick parameters that match the model's and the actual spectral density over the entire frequency range. Therefore, maximum likelihood will sacrifice accuracy in the small region around $\omega = 0$ to better match spectral densities at higher frequencies.

In summary, the low-order ARMA approach of Campbell and Mankiw and the unobserved components approach of Nelson and Plosser cannot match the short-run dynamics and the small random walk component in the long-run dynamics at the same time. Faced with the choice, they capture the short-run dynamics and incorrectly imply large random walk components.

On the other hand, Clark's (1987) and Watson's (1986) decompositions can accommodate the behavior of GNP in both frequency ranges. (See, e.g., Watson's fig. 1*b*, in which he shows how his model can represent a large number of small high-order autocorrelations that a low-order ARMA cannot match.) Both Watson and Clark find a small random walk component. However, their decompositions also imply identifying restrictions to estimate long-run behavior from short-run dynamics. Since these restrictions are no more or less plausible than Nelson and Plosser's or Campbell and Mankiw's, they might not be able to capture the pattern of high-order correlations in other data sets as they seem to do for GNP.

Since the size of the random walk component is a property of the periodogram ordinate at frequency zero alone, any estimation technique must make some identifying restriction across the frequency range. The variance of k-differences assumes that past a certain k the random walk component is adequately identified, empirically determined as the point in which the graph (fig. 1) flattens out. Therefore, the variance of k-differences (or any other spectral window estimator) uses 10–20-year period information to identify the infinite-run property, the random walk component. The variance of k-differences does not use information about dynamics at business cycle frequencies to identify long-run movements, and this is its important advantage.

V. Conclusion

The variance of k-differences (fig. 1 and table 1) produced a point estimate that the innovation variance of the random walk component of GNP is about one-third the variance of yearly GNP growth rates. That estimate is upward biased for small random walk components: the parameters of two models that replicated the behavior of the variance of k-differences of GNP implied variance ratios of .18 (AR(15)) and 0 (AR(2) about a deterministic trend). I conclude that if there is a random walk component in GNP at all, it is small.

Another way to characterize these results, without reference to random walk components, is that GNP growth is positively autocorrelated at short lags, but there are many small negative autocorrelations at long lags. These bring future GNP back toward, if not all the way back to, its previously forecast value following a shock.

These results do *not* mean that "GNP follows an AR(2) about a deterministic trend." Our forecasts of the future may quite rightly be much more variable than the "trend" in GNP we have seen in the recent 118-year past might suggest.[12] These results *do* mean that an AR(2) about a deterministic trend or a difference-stationary ARMA process with a very small random walk component is a good in-sample characterization of the behavior of GNP.

In reconciling these results with previous research, I argued that conventional criteria for time-series model identification and estimation can produce misleading estimates of the random walk component of a series like GNP. The random walk component is a property of all autocorrelations taken together, but conventional procedures concentrate on the first few autocorrelations in order to parsimoniously capture short-run dynamics. When used to estimate the size of a random walk component, they impose identifying restrictions across the frequency range to infer the long-run properties of a series from its short-run dynamics. I argued that, in the absence of credible identifying restrictions, it is best to leave the short run out altogether, as the variance of k-differences or some other spectral window estimator does.

However, this view—that we should use only long-run properties of GNP data to estimate the long-run behavior of GNP—implies that standard errors of univariate estimates of the random walk component will remain large in century-long macroeconomic data and larger still in postwar macroeconomic data because there are inherently few nonoverlapping long runs available. These observations argue against the research strategy that says that the presence of a unit root and the size of a random walk component are crucial and well-documented stylized facts that any theoretical model must replicate.

Appendix

A. Derivation of Equation (9)

Start with

$$(1 - L)y_t = \mu + A(L)\epsilon_t = \mu + \sum_{j=0}^{\infty} a_j \epsilon_{t-j}. \tag{A1}$$

[12] A plausible model for GNP should have some random walk component. If GNP is truly stationary about a linear trend, then the variance of the forecast error of the level of GNP is the same for all dates in the far future. As long as there is some random walk component, the variance of forecast errors will grow unboundedly over the forecast horizon. However, only a very small random walk component is required to achieve this desirable property.

Using

$$(1 - L^k)(1 - L)^{-1} = (1 + L + L^2 + \ldots + L^{k-1}),$$
$$y_t - y_{t-k} = k\mu + \sum_{j=0}^{k-1}\left(\sum_{l=0}^{j} a_l\right)\epsilon_{t-j} + \sum_{j=k}^{\infty}\left(\sum_{l=j-k+1}^{j} a_l\right)\epsilon_{t-j}. \tag{A2}$$

Taking its variance,

$$\sigma_k^2 \equiv k^{-1}\operatorname{var}(y_t - y_{t-k}) = k^{-1}\left[\sum_{j=0}^{k-1}\left(\sum_{l=0}^{j} a_l\right)^2 + \sum_{j=k}^{\infty}\left(\sum_{l=j-k+1}^{j} a_l\right)^2\right]\sigma_\epsilon^2.$$

To simplify the algebra, express σ_k^2 as a difference equation

$$k\sigma_k^2 - (k-1)\sigma_{k-1}^2 = \left[\sum_{j=0}^{\infty}\left(a_j^2 + 2a_j\sum_{l=1}^{k-1} a_{j+l}\right)\right]\sigma_\epsilon^2,$$

$$\sigma_1^2 = \left(\sum_{j=0}^{\infty} a_j^2\right)\sigma_\epsilon^2,$$

so

$$1 + 2\sum_{j=1}^{k-1}\rho_j = \frac{k\sigma_k^2 - (k-1)\sigma_{k-1}^2}{\sigma_1^2},$$

where ρ_j = the jth autocorrelation of $(1 - L)y_t$, $\rho_j = \Sigma_{l=0}^{\infty} a_l a_{l+j}/\Sigma_{l=0}^{\infty} a_l^2$. Therefore,

$$\frac{\sigma_k^2}{\sigma_1^2} = k^{-1}[1 + (1 + 2\rho_1) + (1 + 2\rho_1 + 2\rho_2) + \ldots] = 1 + 2\sum_{j=1}^{k-1}\frac{k-j}{k}\rho_j.$$

B. Derivation of $\mathrm{E}(\hat{\sigma}_k^2)$ *for an MA(1)*

Assume that (A1) takes the form

$$(1 - L)y_t = \mu + (1 + \theta L)\epsilon_t$$

and assume that ϵ_t are i.i.d. normal. The data set is $T + 1$ observations of the levels of y_t or T observations of its first differences. By definition,

$$\hat{\sigma}_k^2 = \frac{T}{k(T-k)(T-k+1)}\sum_{j=k}^{T}\left[y_j - y_{j-k} - \frac{k}{T}(y_T - y_0)\right]^2. \tag{A3}$$

Equation (A2) specializes to

$$y_j - y_{j-k} = k\mu + \epsilon_j + \theta\epsilon_{j-k} + (1 + \theta)\sum_{l=1}^{k-1}\epsilon_{j-l},$$

and similarly for $y_T - y_0$. Collecting terms in ϵ_j and noting that $E(\epsilon_j\epsilon_k) = 0$ if $j \neq k$, we get (after some algebra)

$$E(\hat{\sigma}_k^2) = (1 + \theta)^2\sigma_\epsilon^2 - \frac{2\theta}{k}\frac{1 + (k^2/T^2) - [2k/T(T-k-1)]}{1 - (k/T)}\sigma_\epsilon^2.$$

Note that (1) as $T \to \infty$, $E(\hat{\sigma}_k^2) \to [(1 + \theta)^2 - (2\theta/k)]\sigma_\epsilon^2$; (2) as $k \to \infty$, $k < T$, $E(\hat{\sigma}_k^2) \to (1 + \theta)^2\sigma_\epsilon^2 = \sigma_{\Delta z}^2$; (3) for $\theta = 0$, $E(\hat{\sigma}_k^2) = \sigma_\epsilon^2 = \sigma_{\Delta z}^2$ for all k, T such that $k < T$.

C. *How Maximum Likelihood Imposes Identifying Restrictions across Frequencies*

Let $x_t = (1 - L)y_t = A(L)\epsilon_t$. Assume that $A(0) = 1$, that $A(L)$ is one-sided and has zeros outside the unit circle, so that the spectral density of x is bounded away from zero, and that A has an inverse, so that x has an autoregressive representation $B(L)x_t = \epsilon_t$. Consider estimating $A(L)$ or $B(L)$ by maximum likelihood via a simple time-series or unobserved components model. For simplicity, assume infinite data, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$, and σ_ϵ^2 known. (The same point survives generalization to more complex estimation environments.) In this case, maximum likelihood is equivalent to

$$\min E[\hat{B}(L)x_t]^2 \quad \text{subject to } \hat{B}(L) \in \mathcal{B}, \tag{A4}$$

where $\hat{B}(L)$ is the autoregressive representation of the estimated model, and $\mathcal{B}$ is the restricted space of autoregressive representations allowed by the chosen time-series model. Since variance is the integral of spectral density, (A4) is the same as

$$\min(2\pi^{-1}) \int_{-\pi}^{\pi} |\hat{B}(e^{-i\omega})|^2 S_x(e^{-i\omega})\, d\omega \quad \text{subject to } \hat{B}(e^{-i\omega}) \in \mathcal{B}. \tag{A5}$$

The following expression is equivalent:

$$\min \int_{-\pi}^{\pi} |\hat{B}(e^{-i\omega}) - B(e^{-i\omega})|^2 S_x(e^{-i\omega})\, d\omega \quad \text{subject to } \hat{B}(e^{-i\omega}) \in \mathcal{B}. \tag{A6}$$

To see this, expand $|\hat{B} - B|^2$ and substitute $AA^*\sigma_\epsilon^2 = S_x$ (an asterisk denotes complex conjugation; I dropped the $e^{-i\omega}$'s). Then (A6) becomes

$$\min \int_{-\pi}^{\pi} (\hat{B}\hat{B}^* + BB^* - B\hat{B}^* - B^*\hat{B})AA^* d\omega. \tag{A7}$$

The first term is just (A5). Since $A^{-1} = B$, the second term is 2π, and the third and fourth are

$$\int_{-\pi}^{\pi} (\hat{B}A + \hat{B}^*A^*)\, d\omega.$$

Under the assumption that A and B are one-sided and that $A(0) = B(0) = 1$,

$$\int_{-\pi}^{\pi} \hat{B}A\, d\omega = \int_{-\pi}^{\pi} \left(1 + \sum_{j=1}^{\infty} \hat{b}_j e^{-i\omega j}\right)\left(1 + \sum_{j=1}^{\infty} a_j e^{-i\omega j}\right) d\omega;$$

since $\int_{-\pi}^{\pi} e^{-i\omega j} d\omega = 0$, $\int_{-\pi}^{\pi} \hat{B}A\, d\omega = \int_{-\pi}^{\pi} \hat{B}^*A^*\, d\omega = 2\pi$. Therefore, (A6) reduces to (A5) plus constants.

Equation (A6) is analogous to Sims's (1972) approximation formula, reproduced in Sargent (1979, p. 293). The message of (A6) is that maximum likelihood attempts to match the frequency response of the autoregressive representation across the entire frequency range, weighted by the true spectral density of x_t. The method of maximum likelihood will sacrifice accuracy

of the estimated $B(e^{-i\omega})$ at a point in the frequency range ($\omega = 0$) in order to achieve a better fit over an interval. Similarly, it will sacrifice accuracy in a small window (20 years to infinity is $\pi/10$ wide) to gain accuracy in a large window (2–4 years is $\pi/2$ wide). If $S_x(e^{-i\omega})$ is smaller near $\omega = 0$ than elsewhere, as the variance of k-differences suggests for GNP, then (A6) shows that maximum likelihood further deemphasizes accuracy in windows about zero.

References

Anderson, Theodore W. *The Statistical Analysis of Time Series.* New York: Wiley, 1971.

Beveridge, Stephen, and Nelson, Charles R. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle." *J. Monetary Econ.* 7 (March 1981): 151–74.

Campbell, John Y., and Mankiw, N. Gregory. "Are Output Fluctuations Transitory?" *Q.J.E.* 102 (November 1987): 857–80.

———. "International Evidence on the Persistence of Economic Fluctuations." Manuscript. Princeton, N.J.: Princeton Univ. Press, 1988.

Clark, Peter K. "The Cyclical Component of U.S. Economic Activity." *Q.J.E.* 102 (November 1987): 797–814.

Cochrane, John H., and Sbordone, Argia M. "Multivariate Estimates of the Permanent Components of GNP and Stock Prices." *J. Econ. Dynamics and Control* (1988), in press.

Dickey, David A., and Fuller, Wayne A. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *J. American Statis. Assoc.* 74, pt. 1 (June 1979): 427–31.

———. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49 (July 1981): 1057–72.

Fama, Eugene F., and French, Kenneth R. "Permanent and Temporary Components of Stock Prices." *J.P.E.* 96 (April 1988): 246–73.

Friedman, Milton, and Schwartz, Anna J. *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867–1975.* Chicago: Univ. Chicago Press (for NBER), 1982.

Huizinga, John H. "An Empirical Investigation of the Long Run Behavior of Real Exchange Rates." *Carnegie-Rochester Conf. Ser. Public Policy* 27 (Autumn 1987): 149–214.

King, Robert G.; Stock, James H.; Plosser, Charles I.; and Watson, Mark. "Stochastic Trends and Economic Fluctuations." Working Paper no. 2229. Cambridge, Mass.: NBER, April 1987.

Kydland, Finn E., and Prescott, Edward C. "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (November 1982): 1345–70.

Lo, Andrew W., and MacKinlay, A. Craig. "Stock Market Prices Do Not Follow a Random Walk: Evidence from a New Specification Test." *Rev. Financial Studies* (1988), in press.

Long, John B., Jr., and Plosser, Charles I. "Real Business Cycles." *J.P.E.* 91 (February 1983): 39–69.

McCulloch, J. Huston. "The Monte Carlo Cycle in Business Activity." *Econ. Inquiry* 13 (September 1975): 303–21.

Nelson, Charles R., and Plosser, Charles I. "Trends and Random Walks in

Macroeconomic Time Series: Some Evidence and Implications." *J. Monetary Econ.* 10 (September 1982): 139–62.

Poterba, James M., and Summers, Lawrence H. "Mean Reversion in Stock Prices: Evidence and Implications." Working Paper no. 2343. Cambridge, Mass.: NBER, 1987.

Quah, Danny. "What Do We Learn from Unit Roots in Macroeconomic Time Series?" Manuscript. Cambridge: Massachusetts Inst. Tech., 1986.

Romer, Christina D. "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869–1918." Working Paper no. 1969. Cambridge, Mass.: NBER, July 1986.

Rose, Andrew K. "Unit-Roots and Macroeconomic Models." Manuscript. Berkeley: Univ. California, 1986.

Sargent, Thomas J. *Macroeconomic Theory.* New York: Academic Press, 1979.

Sims, Christopher A. "The Role of Approximate Prior Restrictions in Distributed Lag Estimation." *J. American Statis. Assoc.* 67 (March 1972): 169–75.

Watson, Mark W. "Univariate Detrending Methods with Stochastic Trends." *J. Monetary Econ.* 18 (July 1986): 49–75.