Journal of Economic Dynamics and Control 12 (1988) 255–296. North-Holland

MULTIVARIATE ESTIMATES OF THE PERMANENT COMPONENTS OF GNP AND STOCK PRICES*

John H. COCHRANE and Argia M. SBORDONE

University of Chicago, Chicago, IL 60637, USA

Received November 1987, final version received January 1988

The economic assumption that a present value relation holds between consumption and income and between stock prices and dividends, or the statistical assumption that the consumption/income ratio and the dividend/price ratio are stationary imply that the permanent or random walk component in each series of a pair must be the same as the permanent component in the other series of a pair. Either assumption then allows us to estimate the variance of the permanent component of one series (GNP, stock prices) from the variance of the permanent component of the other (consumption, dividends), or from the covariance of the two series' permanent components. This paper presents such estimates, and finds that the permanent components are about half those estimated by similar univariate methods.

1. Introduction

A series of recent papers has examined the importance of a potential unit root to the behavior of real GNP [Campbell and Mankiw (1986, 1987), Cochrane (1986), Clark (1987)] and stock prices [Fama and French (1986), Lo and MacKinlay (1986), Poterba and Summers (1986, 1987)]. The simplest interpretation of these papers is that they all ask the question: if we observe a unit innovation in a series (GNP or stock prices) today, how much does that event raise our long-term forecasts of the series? If by one unit, then the series is a pure random walk; if by 0, the series is stationary. The response can be any real number, which is the innovation in these estimates: instead of forcing us to ask the dichotomous question 'is there a unit root or not', they allow us to continuously measure the importance of a potential unit root.

The actual estimating technique used in these papers varied somewhat. Cochrane, Poterba and Summers, and Lo and MacKinlay examined the variance of long differences. The idea here is that if a series is a random walk, the variance of its k differences will be k times the variance of its first differences; if a series is stationary, the variance of k differences will tend to

*A precursor to this paper appeared under the title 'Spectral Density Estimates of Unit Roots'. We thank Masanao Aoki, Lars Hansen, Masao Ogaki, and Chris Sims for helpful suggestions. The computer programs and data used in this paper are available on request from the authors.

twice the unconditional variance of the series. If a series is a combination of a stationary and a random walk component, $1/k$ times the variance of k differences tends to the variance of the random walk or permanent component.

It turns out that we can think interchangeably of series which contain a unit root and series which are composed of a stationary and a random walk component, and that the variance of the random walk component can be directly related to the effect of a univariate innovation on long horizon forecasts.

These papers all provided interesting point estimates and unfortunately large standard errors. The problem is that there are few nonoverlapping 'long runs' of data available, so unless strong restrictions are imposed on the form of the estimated time series process, the response of 'long-run' forecasts to a unit innovation will necessarily be imprecisely measured.

A natural response is to try to examine more series, in the hope that observations of several series over a short time horizon can in some sense proxy for the observation of one series over a long horizon. In this vein, Campbell and Mankiw (1987) examined GNP from many countries, and Poterba and Summers (1987) examined stock prices from many countries. The defect of this approach in its present implementation is that it is not clear how much, if any, independent information is contained in estimates from several countries at the same time.

This paper brings multivariate evidence to bear in the following way: consider a pair of time series, with the property that the ratio of the two series is stationary. For example, it is possible that the consumption/GNP ratio is stationary, even if log consumption and log GNP each have random walk components, or that the dividend/price ratio is stationary even if log dividends and log stock prices each have random walk components.

If the ratio of two series is stationary, the random walk component of the two series must be *exactly the same* – we can express each series as a sum of a common random walk component and separate stationary components. If we couldn't do this, the ratio of the two series would contain a random walk.

More precisely, if $\log(X_t)$ and $\log(W_t)$ must be differenced to obtain stationary series (if they contain random walk components), yet $\log(X_t/W_t)$ is stationary, then there must be a representation

$$\log(X_t) = z_t + c_{xt}, \qquad \log(W_t) = z_t + c_{wt}, \tag{1}$$

where z_t is a random walk and c_{xt} and c_{wt} are stationary. If the z_t entered with different coefficients or if there was a random walk component in one series not present in the other, then $\log(X_t/W_t)$ would contain a random walk.

Now, if the two series can be expressed as the sum of a common random walk component and distinct stationary components, then $1/k$ *times the* *variance of k differences of each series, as well as 1/k times the covariance of k differences of the two series, must tend to exactly the same number, the innovation variance of the common random walk component.* In the representation (1), all these quantities are $1/k \ \mathrm{var}(z_t - z_{t-k}) = \mathrm{var}(z_t - z_{t-1})$ for large enough k.

Then, we can estimate the variance of the permanent or random walk component of one series (GNP, stock prices) from $1/k$ times the variance of k differences of the other series (consumption, dividends), or $1/k$ times the covariance of k differences of the two series.

How do we decide what k is 'large enough', or that the limit has been adequately reached? The choice of k is exactly the choice of a window width of a spectral density estimator: larger k (smaller window) gives less bias but more uncertain estimates, while smaller k (larger window) gives more precise but more biased estimates. Operationally, we stop at a k large enough that 'business cycle' fluctuations are ironed out and only the 'long run' remains, on the order of 20 or 30 years. We also stop at $k =$ half the sample size, which amounts to taking a variance based on two data points. Since we have no reason to prefer $k = 20$ or $k = 30$, etc., we present results for a variety of k, and hope that the results are robust in a range of k.

Figs. 1 and 2 present $1/k$ times the variance and covariance of k differences for stock prices and dividends, and GNP and consumption, respectively. Once we impose the assumption that the consumption/GNP ratio or the dividend/price ratio is stationary, the three lines in each figure are each estimates of the variance of one underlying random walk component.

Which of the three estimates, or which combination of the three is the best estimate of the underlying random walk component? As long as the difference k is finite, each variance or covariance is a biased estimate of the variance of the underlying random walk component, because some stationary components are still included in the k-differenced series. On the other hand, though each variance or covariance contains no independent information about the variance of the one common random walk component, the stationary components are not perfectly correlated, so there is some independent movement in each variance or the covariance in finite samples. The optimal combination of the variance and covariance is thus a tradeoff between the extra bias of series with stronger stationary components, against the reduction in variance that occurs when you combine estimates with some independent information.

Below, we will argue that for pairs of series like GNP and consumption or stock prices and dividends, in which one series (consumption, dividends) is nearly a random walk (has a flat graph of $1/k$ times the variance of k differences), the reduction in standard error is not worth the increase in bias, so the best combined estimate is provided by just looking at $1/k$ times the variance of k differences of the series which is closest to a pure random walk, consumption and dividends.

Hence, the best multivariate estimate of the variance of the random walk component of GNP is provided by the variance of k differences of consumption; the best estimate of the variance of the random walk component of stock prices is provided by the variance of k differences of dividends. These estimates are about half previous univariate estimates. (See figs. 1 and 2 or tables 4 and 5: the consumption and dividends lines are about half of the price and GNP lines).

1.1. Some comments on the estimation strategy

The major advantage of using several time series in this way is that it reduces the bias associated with finite k differencing of the time series of interest (GNP, stock prices). The standard errors are mostly associated with the standard error of measuring the common random walk component and not the individual stationary components; hence these bivariate estimates do not significantly reduce the standard errors associated with univariate estimates.

Hence, it does not seem useful to generalize the procedures of this paper to multiple time series that are all cointegrated, for example to include many other components of GNP. The other components of GNP have more stationary components than consumption (their graph of $1/k$ times the variance of k differences starts higher and slopes down more than the consumption graph, winding up slightly higher than $1/k$ times the variance of k differences of GNP), so including them will only bias the estimate of the common random walk component while adding little independent information.

In order to use multivariate information to reduce the standard errors, we would have to find other series that are *not* cointegrated with GNP (stock prices), but have the same variance of a random walk component or variance ratio. Then, observations of several series will add some independent observations on the variance of a random walk component. For example, if the GNP or stock prices of several countries are not cointegrated (if the ratios of their prices are not stationary, or if the squared covariance of their k differences is not equal to the product of the variances of their k differences), and yet each follows the same process (has the same variance of k differences at all k), then a pooled estimate can reduce the standard errors. We leave this as a suggestion for future research.

1.2. Plan of the paper

The rest of this paper formalizes these arguments, and presents our results for stock prices and dividends, and for consumption and GNP. In section 2, we discuss the decomposition of first difference stationary series into stationary and random walk components, and relate that decomposition to the property that ratios of the series may be stationary. In particular, we derive the representation (1) above.

In section 3 we discuss estimation, and present some of the properties of the $1/k$ times the variance of k differences technique. In particular, this section argues that just looking at the variance of the series closest to a random walk is the best combined estimate in cases like ours.

Section 4 presents our results. First we prove that the present value relation implies that the dividend/price and consumption/GNP ratios should be stationary. We also apply univariate variance ratios to these series to check this stationarity assumption. Then, we impose the assumption that the dividend/price and consumption/GNP ratios are stationary to measure the random walk components in GNP and stock prices from $1/k$ times the variance of k differences of consumption and dividends.

2. Representation of time series with unit roots

2.1. Decomposition into stationary and random walk components

Let y_t be an N-dimensional vector time series. Throughout we will assume that y_t is stationary in first differences; in particular we will assume that $(1-L)y_t$ has a moving average representation

$$(1-L)y_t = \mu + A(L)\varepsilon_t. \tag{2.1}$$

L is the lag operator ($Ly_t = y_{t-1}$); μ is an N-dimensional vector of means; $A(L)$ is an $N \times N$ matrix of lag polynomials with $A(0) = I_N$, the $N \times N$ identity matrix [i.e., $A(L)\varepsilon_t = I_N\varepsilon_t + \sum_{j=1}^{\infty} A_j \varepsilon_{t-j}$]; and ε_t is an N-dimensional vector of innovations. $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t\varepsilon_t') = \Sigma$ (Σ is a positive definite matrix) and $\mathrm{E}(\varepsilon_t\varepsilon_{t-j}') = 0$ for $j \neq 0$.

We can construct a decomposition of y_t into a stationary component and a random walk component with a multivariate generalization of the Beveridge and Nelson (1981) decomposition:

$$y_t = z_t + c_t,$$

$$(1-L)z_t = \mu + A(1)\varepsilon_t, \tag{2.2}$$

$$-c_t = A^*(L)\varepsilon_t \quad \text{where} \quad A_j^* = \sum_{k=j+1}^{\infty} A_k \quad \text{and} \quad A(1) = \sum_{k=0}^{\infty} A_k.$$

In this decomposition, $z_t = \lim_{k \to \infty} \mathrm{E}_t y_{t+k} - k\mu$, so z_t can be interpreted as a 'stochastic trend' or 'permanent component' of y_t. Also, $A(1)\varepsilon$ gives the limiting response of $\mathrm{E}_t y_{t+k}$ to an ε innovation at time t. [Stock and Watson (1986) derive this and some other representations.]

From (2.1), the spectral density of $(1-L)y_t$ at frequency 0 is

$$S_{(1-L)y_t}(\mathrm{e}^{\mathrm{i}0}) = S_{(1-L)y_t}(1) = A(1)\Sigma A(1)',$$

so the variance–covariance matrix of changes in the permanent components $(1-L)z_t$ is the same as the spectral density of $(1-L)y_t$ at frequency 0, and is the same as the spectral density matrix of $(1-L)z_t$ as well. To avoid repeating these three interpretations, we will denote this matrix by Ψ:

$$\Psi = A(1)\Sigma A(1)' = S_{(1-L)y_t}(1) = \mathrm{var}((1-L)z_t) = S_{(1-L)z_t}(\mathrm{e}^{-\mathrm{i}0}).$$

We'll denote the elements of Ψ by ψ_{ij}, so that ψ_{ii} is the variance of the permanent component in element i of y_t and ψ_{ij} is the covariance of the ith and jth permanent components.

Although it is derived in the context of one of many possible decompositions, the matrix Ψ gives a complete characterization of the unit root or cointegration properties of a series in a finite sample. Given the spectral density of $(1-L)y_t$ at frequencies other than 0, we can always construct a trend stationary series by changing the value of the spectral density to be 0 at frequency 0. [Cochrane (1987) discusses this point in detail.]

2.2. Cointegration

In this system, the time series y_t are said to be *cointegrated* if there is an $N \times M$ matrix α (rank M) such that $\alpha' y_t$ is stationary. [These concepts and terminology are due to Engle and Granger (1987).] The columns of α are called the *cointegrating vectors*. Now, $\alpha' y_t = \alpha' z_t + \alpha' c_t$, $\alpha' c_t = -\alpha' A^*(L)\varepsilon_t$, and since linear combinations of stationary series are stationary, this term imposes no restrictions on $A(L)$. However, $\alpha' z_t = \alpha' z_{t-1} + \alpha'\mu + \alpha' A(1)\varepsilon_t$. For this term to be stationary, we require

$$\alpha'\mu = 0 \quad \text{and} \quad \alpha' A(1) = 0. \tag{2.3}$$

In turn, $\alpha' A(1) = 0$ implies $\alpha' A(1)\Sigma A(1)' = \alpha'\Psi = 0$, so both $A(1)$ and Ψ must be of rank $N-M$ for the system to be cointegrated. Since we will use logs of series, cointegration amounts to the statement that ratios of the series are stationary.

Granger and Engle (1987) show that, if y_t is cointegrated, there is an equivalent error correction representation:

$$(1-L)y_t = B\mu - \gamma\alpha' y_{t-1} + H(L)\varepsilon_t,$$

where B is a matrix of constants and γ is a $N \times M$ matrix. [See Granger and Engle for the construction of this representation from (2.1).] This representation shows how changes in y_t depend on how far away $\alpha' y_{t-1}$ is from its equilibrium value.

Fama and French's (1987) regressions of future returns on dividend price ratios are estimates of this error correction form; our variance based estimates properties of the original representation. Since either representation can be derived from the other, we are after the same phenomenon, just as Fama and French's (1986) regressions of future returns on past returns are measures of the same phenomenon as Poterba and Summer's univariate variance ratios.

2.3. *A representation of cointegrated series that measures the importance of cointegration by the size of random walk components*

Cochrane (1986) emphasized that the univariate Ψ [spectral density at 0, variance of random walk component, and $(\Sigma a_j)^2\sigma_\varepsilon^2$] is a useful as well as a complete characterization of the unit root properties of a series, because it allows us to measure the importance of a unit root on a continuous scale from 0 to ∞ rather than just ask 'is there or isn't there a unit root?' For a vector of time series y_t, we can ask the further question: how many unit roots ('common trends' in the Stock–Watson language) are there? This section derives a rewriting of the Beveridge–Nelson representation that allows us to quantify the importance of the N potential unit roots, rather than ask simply what is their number.[1]

Express the spectral density matrix at frequency 0 as

$$\Psi = S_{(1-L)y_t}(1) = Q \Lambda Q^{-1}, \tag{2.4}$$

where Λ is a diagonal matrix of eigenvalues of the spectral density matrix, organized from highest to lowest,

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}, \tag{2.5}$$

and Q is a corresponding orthogonal matrix of eigenvectors. Since Ψ is symmetric, it has a full set of linearly independent eigenvectors, so this

[1] With this representation, multivariate variance–covariance ratio estimates (or other multivariate spectral density estimates) can play the same role with regard to Stock and Watson's tests for the number of random walk components that the univariate variance ratio tests do to the Dickey–Fuller, etc. tests for the presence or absence of a single unit root – they allow one to estimate the same quantities while imposing less additional structure.

representation exists. When the spectral density matrix Ψ is of rank $N-M$ less than N, M of the λ_j are 0.

Define a new N-dimensional series of innovations by $\nu_t = Q'A(1)\varepsilon_t$. Then, we can rewrite the random walks z_t in terms of these innovations as

$$(1-L)z_t = \mu + Q\nu_t \quad \text{where} \quad \mathrm{E}(\nu_t)=0, \quad \mathrm{E}(\nu_t\nu_t') = \Lambda. \tag{2.6}$$

Since the variances of the last $N-M$ elements of ν are 0, this representation expresses y_t in terms of $N-M$ random walks (or common trends), whose innovation variances are the eigenvalues of Ψ.

Consider the problem of finding $M+L$ nearly cointegrating vectors when there are really M. In a large sample, we want to find $M+L$ vectors α_j, $j=1,2,\ldots,M+L$, that minimize

$$\min \operatorname{var}\big(\alpha_j'(1-L)z_t\big) = \alpha_j'\Psi\alpha_j = \alpha_j'Q\,\Lambda\,Q'\alpha_j \tag{2.7}$$

(subject to an arbitrary normalization for the vector α_j). The answer is,[2] pick for α_j the eigenvectors corresponding to the $M+L$ lowest eigenvalues of the spectral density matrix. The (minimized) variance of the permanent component of the corresponding linear combinations of y_t are then

$$\operatorname{var}(1-L)\alpha_j'z_t = \lambda_j. \tag{2.8}$$

The representation (2.6) and its interpretation (2.7)–(2.8) provide a multivariate extension of the quantity $a(1)^2\sigma_\varepsilon^2$ used for univariate time series. If there are M cointegrating vectors, M of the λ_j elements of Λ are 0, there are M linear combinations of y_t that are stationary, and so the N-dimensional series y_t can be expressed as a sum of only $N-M$ random walk components plus stationary components. As the series y_t becomes closer and closer to being cointegrated with $M+1$ cointegrating vectors, the $(M+1)$th λ_j will approach 0, and the $(N-M+1)$th random walk $\nu_{N-M+1,t}$ will contribute less and less to the variance of the long-term forecasts z_t.[3] In applications where it is more interesting to measure the size of a univariate random walk component rather than test for its presence or absence of a unit root, this representation suggests we estimate the variance of the N potential ν_{it} rather than test for their number.

[2]Take the normalization as $|\alpha_j|=1$. Then, we want to minimize $(\alpha_j'Q\Lambda Q'\alpha_j/\alpha_j'\alpha_j)$. Any linear algebra textbook [e.g., Strang (1976, p. 253) under 'Rayleigh's Quotient'] shows that this quotient is minimized by taking α_j as one of the eigenvectors or columns of Q, and the value of the minimized quotient is the corresponding eigenvalue or element of Λ.

[3]Phillips and Ouliaris (1986) derive the asymptotic distribution of eigenvalues of the spectral density matrix at frequency 0, under the null that the matrix is in fact of full rank.

2.4. Effect of cointegration on Ψ

Later on, we will impose the restriction that the consumption/income ratio and the dividend/price ratios are stationary. This assumption implies that the logs of these series are cointegrated with cointegrating vector $[1 \; -1]'$. In this section, we ask: what effect does this assumption have on the matrix Ψ?

Consider a pair of time series, $y_t = [x_t \; w_t]'$, cointegrated with cointegrating vector $\alpha = [1 \; -\alpha_1]'$ ($x_t - \alpha_1 w_t$ is stationary). (Think of x as GNP and w as consumption, or x as stock prices and w as dividends.) Cointegration implies $\alpha' A(1) = 0$ and hence $\alpha' A(1) \Sigma A(1)' = \alpha' \Psi = 0$. This requirement means, for the elements of Ψ,

$$(1/\alpha_1) = \psi_{12}/\psi_{11}, \qquad \alpha_1 = \psi_{12}/\psi_{22}, \tag{2.9}$$

$$(\psi_{11}/\psi_{22}) = \alpha_1^2, \tag{2.10}$$

$$\psi_{12}^2 = \psi_{11}\psi_{22}. \tag{2.11}$$

(2.9) shows that α_1 can be found from an OLS regression of the permanent components; (2.10) shows that the relative variance of the random walk components is determined by the cointegrating vector α; and (2.11) reminds us that Ψ is singular, so that there is effectively only one random walk component. We can rewrite (2.9)–(2.11) as

$$\Psi = \begin{bmatrix} \alpha_1^2 & \alpha_1 \\ \alpha_1 & 1 \end{bmatrix} \text{ (constant)}, \tag{2.12}$$

or, in a common trends representation,

$$\begin{aligned} x_t &= \alpha_1 z_t + c_{xt}, \\ w_t &= z_t + c_{wt}, \\ z_t &= z_{t-1} + \eta_t, \quad \eta_t \text{ i.i.d.} \end{aligned} \tag{2.13}$$

If $\alpha = [1 \; -1]'$, then $(x_t - w_t)$ is stationary; or, with $x_t = \log(X_t)$ and $w_t = \log(W_t)$, $\log(X_t/W_t)$ is stationary. Otherwise, $\log(X_t/W_t^{\alpha-1})$ is stationary.

When $\alpha = [1 \; -1]'$ and $\log(X_t/W_t)$ is stationary, then (2.13) reduces to

$$\begin{aligned} x_t &= z_t + c_{xt}, \\ w_t &= z_t + c_{wt}, \\ z_t &= z_{t-1} + \eta_t, \quad \eta_t \text{ i.i.d.} \end{aligned} \tag{2.14}$$

This is eq. (1) in the introduction. In words, x_t and w_t can be expressed as the sum of a common random walk component and a stationary component. In this case

$$\Psi = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{(constant)}. \tag{2.15}$$

We will use $1/k$ times the variance of k differences of x and w to estimate the diagonal elements of Ψ, and $1/k$ times their covariance to estimate the off-diagonal elements. From (2.15), *if the* $\log(X_t/W_t)$ *is stationary*, $1/k$ *times the variance of k differences of* x_t, $1/k$ *times the variance of k differences of* w_t, *and* $1/k$ *times the covariance of their k differences all approach the same number as k grows*. We can read the variance of the random walk component of income, for example, from the variance of the k differences of consumption, and from the covariance of the k differences of consumption and income as well as from the variance of k differences of income.

3. Estimation

3.1. $1/k$ *times variance of k differences as a spectral density estimator*

Most tests for cointegration, like tests for unit roots, are based on parsimonious VAR representations. For examples, see Stock and Watson (1986) or Granger and Engle (1987). In retrospect, this is surprising, because the cointegration and unit root properties of a series are entirely a function of Ψ, the spectral density at 0, and spectral density is usually not estimated from parsimonious VAR or VARMA representations.

These techniques impose restrictions across frequencies to estimate the spectral density at 0 from high frequency information in any finite sample. Most commonly, one has to impose that the errors are described by a short AR or MA process. This assumption essentially bounds the slope of the spectral density near zero. Direct estimation of the spectral density at frequency 0 (Ψ) often imposes fewer such restrictions. As in the univariate case, it can capture classes of time series behavior, such as long horizon mean reversion, that are precluded by the restrictions imposed by estimating parsimonious VAR representations.

This section shows how $1/k$ times the variance of k differences is a conventional spectral density estimate. In particular, it is a member of a class of spectral density estimates that we call *variance of filtered data* estimates. Estimates in this class are equivalent implementations of the usual weighted covariance and smoothed periodogram estimates, in the sense that for any member of one class there is an equivalent member of the other two, available by Fourier transformation.

Start with a stationary N-dimensional time series, Δy_t. Define its finite Fourier transform $x(\omega)$ as[4]

$$x(\omega) = T^{-1/2} \sum_{j=1}^{T} \mathrm{e}^{-\mathrm{i}\omega j} \Delta y_j, \qquad -\pi \leq \omega \leq \pi. \tag{3.1}$$

Its periodogram is defined for T frequencies between $-\pi$ and π,

$$I(\omega_j) = x(\omega_j) x(\omega_j)^*, \quad \omega_j = 2\pi j/T, \quad j = -(T-1)/2, \dots, T/2, \tag{3.2}$$

where $*$ means complex conjugation and transposition.

The periodogram is an unbiased but inconsistent estimate of the spectral density matrix at ω, because its variance does not decline to 0 as $T \to \infty$. Therefore, the spectral density is conventionally estimated as a weighted sum of nearby periodogram ordinates:

$$\hat{S}(\mathrm{e}^{-\mathrm{i}\omega}) = \sum_{j=-(T-1)/2}^{T/2} W(\omega_j - \omega) I(\omega_j). \tag{3.3}$$

(3.3) is the *smoothed periodogram* estimate of the spectral density. If we promise that the weighting function will approach a delta function as $T \to \infty$, but at a slower rate than T, (3.3) is a consistent estimate of the spectral density.

If we Fourier transform the weighting function

$$w_\omega(k) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\nu} W(\nu - \omega)\,\mathrm{d}\nu, \tag{3.4}$$

then the Fourier transform of (3.3) is

$$\hat{S}(\mathrm{e}^{-\mathrm{i}\omega}) = \sum_{k=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}k\omega} w_\omega(k) \hat{\gamma}(k), \tag{3.5}$$

where $\hat{\gamma}(k)$ is a consistent estimate of the kth autocovariance $\mathrm{E}(\Delta y_t \Delta y'_{t-k})$. This is the *weighted covariance* estimator.

The variance of long differences is an instance of a third equivalent class of estimators. Define a (two-sided) lag polynomial $F_\omega(\mathrm{e}^{-\mathrm{i}\nu})$ such that

$$|F_\omega(\mathrm{e}^{-\mathrm{i}\nu})|^2 = W(\nu - \omega). \tag{3.6}$$

[4] We chose to put the $2\pi^{-1}$ in the inverse Fourier transform rather than include a $(2\pi)^{-1/2}$ in both the forward and inverse transform, so that the identity between the spectral density at 0 and the variance of $(1-L)z_t$ is preserved, without an intervening $2\pi^{-1/2}$.

Then form a filtered version of the original series,

$$y^{\mathrm{f}}(t) = F(L)\,\Delta y_t. \tag{3.7}$$

The variance of this filtered series is

$$\operatorname{var}(y^{\mathrm{f}}(t)) = (2\pi)^{-1}\int_{-\pi}^{\pi} |F_\omega(\mathrm{e}^{-\mathrm{i}\nu})|^2 S_{\Delta y}(\mathrm{e}^{-\mathrm{i}\nu})\mathrm{d}\nu. \tag{3.8}$$

Therefore, the sample variance of filtered data is asymptotically equivalent to the weighted covariance and smoothed periodogram estimators of the spectral density (3.3) and (3.5).

We can express the variance of long differences estimate as the variance of filtered data:

$$\hat{\Psi} = \hat{S}(\mathrm{e}^{-\mathrm{i}0}) = \operatorname{var}\left(k^{-1}\sum_{j=1}^{k} \Delta y_j\right). \tag{3.9}$$

The corresponding smoothed periodogram estimate is [from (3.3)]

$$\hat{\Psi} = \hat{S}(\mathrm{e}^{-\mathrm{i}0}) = \sum_{\omega_j} \frac{1}{k}\frac{\sin^2|(k+1/2)\,\omega_j|}{\sin^2(\omega_j/2)} I(\omega_j), \tag{3.10}$$

and the weighted covariance estimate is

$$\hat{\Psi} = \hat{S}(\mathrm{e}^{-\mathrm{i}0}) = \sum_{j=0}^{k-1} \frac{k-|j|}{k}\hat{\gamma}_j, \tag{3.11}$$

which is the Bartlett estimate of the spectral density at frequency 0.

The usual procedure is to pick a k or window width and then calculate one of the above estimates. If k is too small the estimate will be biased from including too many far away periodogram ordinates. If k is too large it will have a large variance because few periodogram ordinates are included. A plot like figs. 1 and 2 represents the result of experimentation with different k or window widths. We hope to find a region in which the results are insensitive to the choice of the window width.

The smoothed periodogram, weighted covariance and variance of filtered data estimates are asymptotically equivalent. In a finite sample, there are differences between the three estimates. For example, the variance of k differences estimate corresponds to a calculation of the autocovariances that underweights observations k away from the beginning and end of the data set compared to the conventional estimate of the autocovariance. These differ-

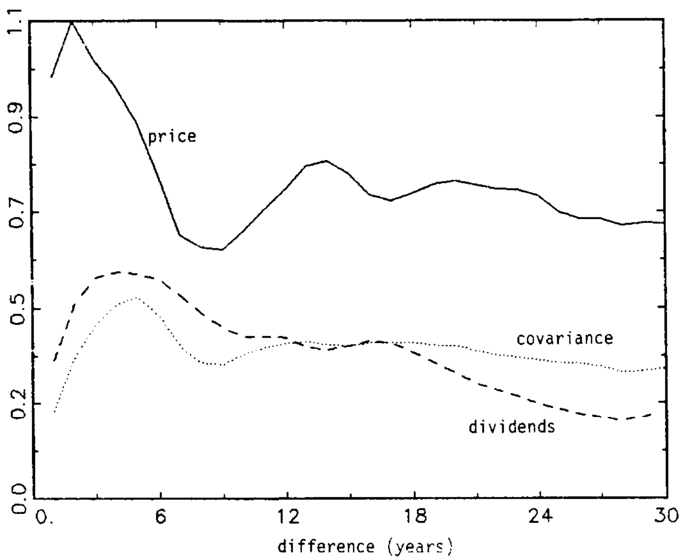


Fig. 1. $1/k \times$ variance and covariance of k differences for stock price and dividends. Units: variance of one-year change in price.

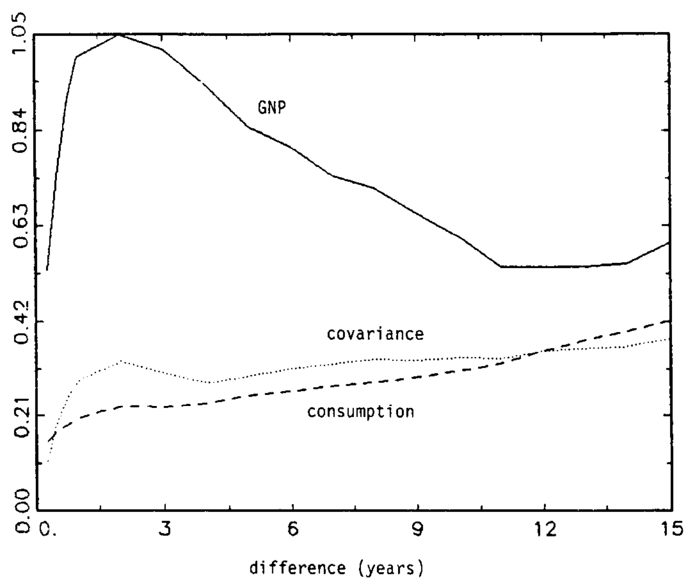


Fig. 2. $1/k \times$ variance and covariance of k differences of GNP and nondurable + services consumption. Units: variance of one-year change in GNP.

ences give rise to different small sample properties. In particular the small sample bias can differ.

The variance of k differences gives a clear picture of exactly what feature of the data drives the estimate. The essential characteristic that any estimate of the spectral density at 0 picks out is whether there is a great deal of variability left at long horizons. This is, in fact, the only distinguishing feature of a random walk component. By taking long differences of the levels of a series, which is the same as taking a long moving average of its differences, we are filtering out the high frequency variation in the series and leaving in only the long horizons.

In principle, one might get better (more efficient) results by using other filters than a simple unweighted moving average, because the unweighted moving average lets in a certain amount of high frequency information through 'side lobes' in its Fourier transform. (Other filters amount to other window shapes of smoothed periodogram estimates.) In experiments with these data sets, however, we have found very little difference in either results or standard errors from using other filters.

3.2. Standard errors

The asymptotic distribution of spectral density estimates is also standard. Koopmans (1974) gives the asymptotic variance of the Bartlett estimator at frequency 0 as

$$\text{S.e.} = \left[\frac{4k}{3T}\right]^{1/2} \hat{\psi}_k, \tag{3.12}$$

where $\hat{\psi}_k$ is $1/k$ times the sample variance of k differences, our estimate of the spectral density at frequency 0.

To assess the accuracy of the Bartlett formula in samples of our typical size, we ran a few Monte Carlo experiments, reported in table 1. (The mean value of the variance of k differences was close to 1.00 in all these experiments.) In the first row of table 1 we simulated 100 observations of a pure random walk with no drift, and we report standard errors of $1/k$ times the variance of k differences. [We calculated the variance of k differences using formula (3.19). That formula includes some corrections for small sample bias, discussed below.]

In the second row, we extend the sample to $T = 200$ observations. Note that the standard errors are very close to the same for equal values of k/T. This behavior is typical of all Monte Carlo experiments we ran.

The third row gives the results when the process includes, and we estimate, a drift. Here the process is a random walk with a drift term of 1, and the variance of k differences is calculated by (3.18) below.

Table 1

Monte Carlo standard errors for variance of k differences (500 trials).[a]

Model when no mean removed: $y_t = y_{t-1} + \varepsilon_t$; $\sigma_\varepsilon^2 = 1$

Model when mean removed: $y_t = 1 + y_{t-1} + \varepsilon_t$; $\sigma_\varepsilon^2 = 1$

$100\ k/T$	1	2	3	4	5	10	20	30	40	50
$T = 100$	0.137	0.160	0.200	0.231	0.263	0.409	0.607	0.772	0.888	0.896
$T = 200$, no mean	0.105	0.167	0.203	0.232	0.256	0.366	0.561	0.733	0.913	1.093
$T = 100$, no mean	0.139	0.178	0.210	0.246	0.271	0.379	0.563	0.710	0.853	0.992
Bartlett	0.115	0.163	0.200	0.231	0.258	0.365	0.516	0.632	0.730	0.816

[a] Sampling error in 500 trials is about 0.02.

The fourth row gives the asymptotic standard error from the Bartlett formula (3.12). These are close to the Monte Carlo values, and best for $k/T \ll 0.5$.

When applying the standard errors, we could choose to use the $\hat{\psi}_k$ at each k to scale the standard errors, or to choose one $\hat{\psi}_k$ – the $\hat{\psi}_k$ at the largest k – as the estimate of ψ for all k. We followed the latter choice in the applications. Since we will not use the covariance of k differences or estimates which are combinations of variances of k differences, we do not report the corresponding Monte Carlo results.

3.3. Optimal combinations: Bias

We have three estimates of the same quantity: $1/k$ times the variance of k differences of each series, and $1/k$ times the covariance of k differences. Which one, or which combination should we use?

The asymptotic distribution theory is of no help here because, asymptotically, all three estimates are identical. Eq. (2.14) shows that, when T and k are large, $1/k$ times the sample variance of k differences of each series and $1/k$ times their sample covariance all approach $1/k$ times the sample variance of k differences of z_t. Hence, the asymptotic variance–covariance matrix of the three estimates,

$$V\mathrm{cov}\, k^{-1}\begin{bmatrix} \mathrm{var}(x_t - x_{t-k}) \\ \mathrm{var}(w_t - w_{t-k}) \\ \mathrm{cov}(x_t - x_{t-k})(w_t - w_{t-k}) \end{bmatrix}', \tag{3.13}$$

is a constant – $\mathrm{var}(1-L)z_t$ – times a 3×3 matrix of ones.

For finite differences k and a finite sample T, the stationary components c in (2.14) still enter the variance of k differences, so the variance–covariance matrix of the sample variance and covariance of k differences is not singular.

Also, at finite k, the stationary components bias $1/k$ times the sample variance or covariance of k differences away from the innovation variance of the random walk component $\operatorname{var}(1-L)z_t$. In frequency domain terms, as long as the windows are of finite width, the estimates will be biased. Therefore, the optimal combination of the three estimates to use will depend on their finite k bias and finite sample variance–covariance matrix. We will examine the population bias (finite k, $T \to \infty$) for bivariate cointegrated MA(1)'s and apply the lessons learned to the actual data. We discuss small sample bias below.

Let the pair $[x_t \; w_t]$ have univariate MA(1) representations

$$\begin{aligned}
&(1-L)x_t = \mu_x + (1+\phi L)\varepsilon_t, \\
&(1-L)w_t = \mu_w + (1+\theta L)\nu_t, \\
&\mathrm{E}\begin{bmatrix}\varepsilon_t \\ \nu_t\end{bmatrix}[\varepsilon_{t-j}\nu_{t-j}] = \begin{bmatrix}\delta(j)\sigma_\varepsilon^2 & \sigma_\varepsilon\sigma_\nu\rho_j \\ \sigma_\varepsilon\sigma_\nu\rho_{-j} & \delta(j)\sigma_\nu^2\end{bmatrix},
\end{aligned} \tag{3.14}$$

where $\delta(j)=1$ for $j=0$ and 0 otherwise. For now, we will take $\mu_x=\mu_w=0$. The (population) variance–covariance of k differences for this model is

$$\begin{aligned}
&\frac{1}{k}\operatorname{var}(1-L^k)x_t = \left((1+\phi)^2 - 2\phi/k\right)\sigma_\varepsilon^2, \\
&\frac{1}{k}\operatorname{var}(1-L^k)w_t = \left((1+\theta)^2 - 2\theta/k\right)\sigma_\nu^2, \\
&\frac{1}{k}\operatorname{cov}\left[(1-L^k)x_t(1-L^k)w_t\right] \\
&= \sigma_\varepsilon\sigma_\nu\left[(1+\phi)(1+\theta)\sum_{j=-(k-1)}^{(k-1)}\frac{k-1-|j|}{k}\rho_j\right. \\
&\quad \left. + \frac{1}{k}\left((1+\theta\phi)\rho_0 + (1+2\theta+\theta\phi)\sum_{j=1}^{k-1}\rho_j + (1+2\phi+\phi\theta)\sum_{j=1}^{k-1}\rho_{-j}\right)\right].
\end{aligned} \tag{3.15}$$

To impose that x_t and w_t are cointegrated, we require that the limit of the matrix of the elements in (3.15) is singular as $k \to \infty$, which in turn implies

$$\sum_{j=-\infty}^{\infty}\rho_j = 1. \tag{3.16}$$

To impose that the cointegrating vector is $[1 \; -1]'$, all the elements of (3.15)

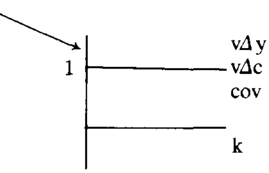
Table 2

Variance ratios for cointegrated MA(1)'s.[a]

Model:
$$\begin{cases} x_t = x_{t-1} + \varepsilon_t + \phi\varepsilon_{t-1} & E(\varepsilon_t \nu_{t-j}) = \sigma_\varepsilon \sigma_\nu \rho_j \\ w_t = w_{t-1} + \nu_t + \theta\nu_{t-1} \quad \sum_j \rho_j = 1 & \sigma_\varepsilon(1+\phi)^2 = \sigma_\nu(1+\theta)^2 = 1 \end{cases}$$

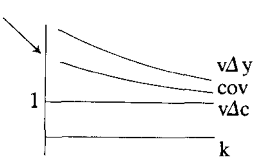
(A) Two random walks

ϕ	θ	ρ_{-1}	ρ_0	ρ_1	vΔx	vΔw	cov	vx10	vw10	cov10
0	0	0	1	0	1	1	1	1	1	1
0	0	1	0	0	1	1	0	1	1	0.9
0	0	0	0	1	1	1	0	1	1	0.9
0	0	0.5	0	0.5	1	1	0	1	1	0.9
0	0	−0.5	1	0.5	1	1	1	1	1	1
0	0	−1	1	1	1	1	1	1	1	1
0	0	1	−1	1	1	1	−1	1	1	0.8
0	0	0.75	−0.5	0.75	1	1	−0.5	1	1	0.85



(B) w_t a random walk, $(1-L)x_t$ an MA(1)

ϕ	θ	ρ_{-1}	ρ_0	ρ_1	vΔx	vΔw	cov	vx10	vw10	cov10
−0.1	0	0	1	0	1.25	1	1.1	1.03	1	1.01
−0.5	0	0	1	0	5	1	2	1.4	1	1.1
−0.9	0	0	1	0	181	1	10	19	1	1.9
−0.5	0	1	0	0	5	1	1	1.4	1	1
−0.9	0	1	0	0	181	1	0.9	19	1	1.8
−0.5	0	0	0	1	5	1	−1	1.4	1	0.8
−0.9	0	0	0	1	181	1	−9	10	1	0



must be equal as $k \to \infty$, so

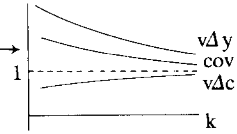
$$(1+\phi)^2\sigma_\varepsilon^2 = (1+\theta)^2\sigma_\nu^2. \tag{3.17}$$

Table 2 reports the variance of first and 1/10 times the variance of 10th differences for a variety of models of this form, and presents graphs of $1/k$ times the variance of k differences for a few of these models. We picked parameter values that give variance of k differences reminiscent of figs. 1 and 2; one series nearly a random walk and the other containing a strong stationary component. (Negative MA coefficients correspond to declining $1/k$ variance of k differences graphs, which accounts for their preponderance in the models of table 2.) We picked the innovation variances σ_ε^2 and σ_ν^2 so that the variance of the random walk component is 1 in each row of table 2.

Part A of the table reports results for the case in which both series are random walks. The $1/k$ variance of k differences is obviously flat for both series, while the covariance varies depending on the correlation of the innovations.

Table 2 (continued)

(C) Both $(1-L)w_t$ and $(1-L)x_t$ MA(1)										
ϕ	θ	ρ_{-1}	ρ_0	ρ_1	vΔx	vΔw	cov	vx10	vw10	cov10
−0.1	−0.1	0	1	0	1.25	1.25	1.25	1.03	1.03	1.03
−0.5	−0.5	0	1	0	5	5	5	1.5	1.5	1.5
−0.9	−0.5	0	1	0	181	5	29	19	1.4	3.8
−0.5	−0.1	0	1	0	5	1.25	2.3	1.4	1.03	1.13
−0.5	0.1	0	1	0	5	0.84	1.73	1.4	0.98	1.07 →
−0.5	0.5	0	1	0	5	0.56	1	1.4	0.96	1
−0.9	0.5	0	1	0	181	0.56	3.7	19	0.96	1.27
−0.1	−0.1	1	0	0	1.25	1.25	0	1.03	1.03	0.9
−0.5	−0.5	1	0	0	5	5	0	1.4	1.4	0.9
−0.9	−0.5	1	0	0	181	5	8	19	1.4	1.7
−0.5	−0.1	1	0	0	5	1.25	0.89	1.4	1.03	0.98
−0.5	0.1	1	0	0	5	0.84	1.09	1.4	0.98	1.01
−0.5	0.5	1	0	0	5	0.56	1.3	1.4	0.96	1.03
−0.9	0.5	1	0	0	181	0.56	9.3	19	0.96	1.83
−0.1	−0.1	0	0	1	1.25	1.25	0	1.03	1.03	0.89
−0.5	−0.5	0	0	1	5	5	0	1.4	1.4	0.89
−0.9	−0.5	0	0	1	181	5	−8	19	1.4	0.1
−0.5	−0.1	0	0	1	5	1.25	−0.89	1.4	1.03	0.81
−0.5	0.1	0	0	1	5	0.83	−1.1	1.4	0.98	0.79
−0.5	0.5	0	0	1	5	0.56	−1.3	1.4	0.96	0.77
−0.9	0.5	0	0	1	181	0.56	−9.3	19	0.96	−0.03



[a] VΔx, vΔw, and cov give the variance or covariance of first differences. Vx10, vw10, and cov10 give 1/10 the variance or covariance of 10th difference. The parameters are chosen so that the variance of the common random walk component is always 1.

Part B presents results for the case in which one series follows a random walk and the other follows an MA(1) in first differences. The variance of the random walk series is unbiased at all k, of course. The bias of the covariance is less than the bias of the variance of the MA(1) series, with absolute value that depends on the correlation of the innovations.

The least biased estimate was a foregone conclusion in parts A and B because one series was a pure random walk, and hence unbiased at all k. Part C explores the bias when both series are MA(1)'s, but one (w) is closer to a random walk than the other (x). The typical case is graphed in the table. With a few exceptions, the series closer to a random walk is the least biased. The only exceptions occur when the variance of x line declines, the variance of w line rises, and the correlation of the innovations is such that the covariance line lies between the two variance lines.

3.4. *Optimal combinations: Variance and covariance*

The optimal combination of the elements of the sample variance–covariance matrix of k differences depends on the small sample distribution as well as the finite k bias. One can reduce the mean squared error of a combination of estimates by including a biased, but independent estimate.

Table 3 presents some Monte Carlo estimates that address this issue. They use the estimates of the variance of k differences that are applied to the data, including small sample corrections described in the next section. We generated artificial data by a variety of the MA(1) models presented above, varying both the permanent/transitory components of each series and the correlation of each series' innovations, while maintaining the assumption of cointegration.

The rows marked '$\mathrm{E}(\tilde{\Psi})$' give the expected value of $1/k$ times the variance and covariance of k differences, which confirms the bias calculations above. The rows marked 'variance–covariance/correlation of $\tilde{\Psi}$' and 'standard deviation' give what they say. To calculate the row marked 'variance and std. dev. of minimum variance linear combination' we calculated the minimum variance linear combination of the three estimates.[5] Note that the variance–covariance matrices are all nearly singular: the correlations are typically 0.98 or 0.99. The minimum variance combinations produce variances almost identical to the individual variances.

[5] Let V denote the 3×1 matrix $[V_1 \ V_2 \ V_3]'$,

$$V = \left[k^{-1}\mathrm{var}(x_t - x_{t-k}) \quad k^{-1}\mathrm{var}(w_t - w_{t-k}) \quad k^{-1}\mathrm{cov}(x_t - x_{t-k})(w_t - w_{t-k})\right]'.$$

Each of the V_i is a possible estimate of the common random walk component, ψ. Let the bias of V be μ and let its variance–covariance matrix be Ω, i.e.,

$$\mathrm{E}(V) = \mu + \psi, \qquad \mathrm{E}((V - (\mu + \psi))(V - (\mu + \psi))') = \Omega.$$

Consider estimating ψ by taking linear combinations of V, $\gamma' V$. The sum of the γ_i is 1, which we can write $\gamma'\underline{1} = 1$, with $\underline{1}$ a vector of 1's. The bias, variance and mean squared error of the new estimate is

$$\mathrm{bias}(\gamma' V) = \gamma'\mu, \quad \mathrm{var}(\gamma' V) = \gamma'\Omega\gamma, \quad \mathrm{MSE}(\gamma' V) = \gamma'(\mu\mu' + \Omega)\gamma.$$

The γ that minimizes mean squared error is

$$\gamma = \left(\underline{1}'(\mu\mu' + \Omega)^{-1}\underline{1}\right)^{-1}(\mu\mu' + \Omega)^{-1}\underline{1},$$

while the γ that minimizes variance is

$$\gamma = \left(\underline{1}'\Omega^{-1}\underline{1}\right)^{-1}\Omega^{-1}\underline{1}.$$

The minimized mean squared error is therefore

$$\left(\underline{1}'(\mu\mu' + \Omega)^{-1}\underline{1}\right)^{-1},$$

while the minimized variance is

$$\left(\underline{1}'\Omega^{-1}\underline{1}\right)^{-1}.$$

We used these formulas to evaluate the optimal combination of V to use, and the costs of using just one (the least biased) and ignoring the others.

We do not report the minimum mean squared error calculations in table 3. We did not constrain linear combinations to be positive, so these calculations basically exploited the bias and ignored the nearly singular variance–covariance matrix. (If one estimate has bias = 1 and another has bias = 2, an 'optimal combination' is -1 times the first + 0.5 times the second.) Since we don't know the bias in applications, these calculations are not interesting.

Table 3

Monte Carlo variance–covariance of $1/k$ variance–covariance of k differences Ψ; 500 trials; $k = 10$.[a]

Model 1: $\begin{cases}(1-L)x_t = (1-0.5L)\varepsilon_t \\ (1-L)w_t = \nu_t\end{cases}$		$\sigma_\varepsilon = 2$, $\sigma_\nu = 1$	$\rho_{-1} = \rho_1 = 0$ $\rho_0 = 1$	
		$\Psi(11)$	$\Psi(12)$	$\Psi(22)$
$\mathrm{E}(\tilde{\Psi})$		1.401	1.103	1.003
Variance–covariance/ correlation of $\tilde{\Psi}$	$\Psi(11)$	0.166	0.156	0.152
	$\Psi(12)$	0.992	0.148	0.146
	$\Psi(22)$	0.985	0.999	0.144
Standard deviation	$\Psi(11)$	0.407	0.395	0.390
	$\Psi(12)$		0.385	0.382
	$\Psi(22)$			0.379
Weights to min. variance[b]		−3.20	6.00	−1.80
Variance and std. dev. of min. variance linear combination		0.130	0.360	

Model 2: $\begin{cases}(1-L)x_t = (1-0.5L)\varepsilon_t \\ (1-L)w_t = (1-0.1L)\nu_t\end{cases}$		$\sigma_\varepsilon = 2$, $\sigma_\nu = 1.11$	$\rho_{-1} = \rho_1 = 0$ $\rho_0 = 1$	
		$\Psi(11)$	$\Psi(12)$	$\Psi(22)$
$\mathrm{E}(\tilde{\Psi})$		1.383	1.114	1.005
Variance–covariance/ correlation of $\tilde{\Psi}$	$\Psi(11)$	0.152	0.143	0.139
	$\Psi(12)$	0.994	0.136	0.132
	$\Psi(22)$	0.986	0.999	0.129
Standard deviation	$\Psi(11)$	0.390	0.378	0.372
	$\Psi(12)$		0.368	0.363
	$\Psi(22)$			0.360
Weights to min. variance[b]		−1.12	6.40	−4.28
Variance and std. dev. of min. variance linear combination		0.11	0.33	

We conclude from these simulations that the most important issue in choosing an optimal combination is the small-k bias discussed above. The extra information available from adding an even more biased estimate is not worth the extra bias. In turn, the small-k bias evaluations of table 2 suggest that the series closest to a random walk is generally the least biased. Our conclusion, which we will apply to the data, is: *When one series has an*

Table 3 (continued)

Model 3: $\begin{cases}(1-L)x_t=\varepsilon_t & \sigma_\varepsilon=1\\(1-L)w_t=\nu_t & \sigma_\nu=1\end{cases}$		$\rho_{-1}=1$ $\rho_1=0$ $\rho_0=0$		
		$\Psi(11)$	$\Psi(12)$	$\Psi(22)$
$E(\tilde{\Psi})$		0.982	0.882	0.983
Variance–covariance/ correlation of $\tilde{\Psi}$	$\Psi(11)$	0.141	0.140	0.142
	$\Psi(12)$	0.999	0.138	0.140
	$\Psi(22)$	0.998	0.999	0.143
Standard deviation	$\Psi(11)$	0.376	0.374	0.377
	$\Psi(12)$		0.372	0.375
	$\Psi(22)$			0.380
Weights to min. variance[b]		−1.11	6.40	−4.28
Variance and std. dev. of min. variance linear combination		0.141	0.341	

Model 4: $\begin{cases}(1-L)x_t=(1-0.5L)\varepsilon_t & \sigma_\varepsilon=2\\(1-L)w_t=(1+0.2L)\nu_t & \sigma_\nu=0.83\end{cases}$		$\rho_{-1}=\rho_1=0$ $\rho_0=1$		
		$\Psi(11)$	$\Psi(12)$	$\Psi(22)$
$E(\tilde{\Psi})$		1.437	1.082	1.002
Variance–covariance/ correlation of $\tilde{\Psi}$	$\Psi(11)$	0.177	0.162	0.159
	$\Psi(12)$	0.988	0.152	0.150
	$\Psi(22)$	0.981	0.999	0.148
Standard deviation	$\Psi(11)$	0.421	0.403	0.399
	$\Psi(12)$		0.390	0.387
	$\Psi(22)$			0.384
Weights to min. variance[b]		−2.69	6.54	−2.85
Variance and std. dev. of min. variance linear combination		0.130	0.360	

[a]All models are MA(1)'s in first differences, cointegrated with cointegrating vector $[1\ \ -1]'$, and with variance of random walk component = 1. The form is given in eqs. (3.14) and (3.15). The sampling uncertainty (500 trials) is about 0.01 in all the above experiments.

[b]These weights are extremely unstable in repetitions of the Monte Carlo experiment, because of the near singularity of the variance–covariance matrix.They typically also imply very large biases of a weighted estimate.

important stationary component and the other is nearly a random walk, use $1/k$ *times the variance of k differences of whichever series is closest to a random walk as the estimate of the permanent component of both series.*

The possible exceptions to this advice are the cases in table 2 in which the covariance line was less biased than the variance of the near-random walk series. However, recognizing such a case depends on knowledge of the correlation at leads and lags of the univariate innovations, ρ_{-1} and ρ_{+1}, and in these cases the bias of the covariance line was only slightly better than the bias of the near-random walk series. Therefore, we judge that these cases are of little practical usefulness. In our applications (tables 4 and 5) the covariances give answers that are quite close to the variance of the near random walk series, so the distinction is of little importance.

3.5. Small sample corrections

We corrected for two sources of small sample bias in estimating the variance–covariance matrix of k differences. First, when we removed an estimated drift term, we used for all k the sample mean of the first differences to estimate the drift term μ, instead of using the mean of k differences at each k. Second, we made a degrees of freedom correction that gives unbiased estimates for the variances when the processes follow pure random walks. With these corrections, the variance estimates are precisely unbiased at all k for a random walk. The covariance estimates suffer a downward bias that depends on the correlation between innovations of the two series at leads and lags. Since we can't correct for this bias without knowledge of these correlations, this observation argues further for our procedure which ignores the covariance estimates.

When we remove a drift term, the estimator we use for $1/k$ times the variance–covariance matrix of k differences is

$$\hat{\psi}_{11} = \frac{T}{k(T-k)(T-k+1)} \sum_{t=k}^{T} \left(x_t - x_{t-k} - \frac{k}{T}(x_T - x_0) \right)^2,$$

$$\hat{\psi}_{12} = \hat{\psi}_{21} = \frac{T}{k(T-k)(T-k+1)}$$

$$\times \sum_{t=k}^{T} \left(x_t - x_{t-k} - \frac{k}{T}(x_T - x_0) \right) \left(w_t - w_{t-k} - \frac{k}{T}(w_T - w_0) \right),$$

$$\hat{\psi}_{22} = \frac{T}{k(T-k)(T-k+1)} \sum_{t=k}^{T} \left(w_t - w_{t-k} - \frac{k}{T}(w_T - w_0) \right)^2, \tag{3.18}$$

where T is the sample size (T differences or $T+1$ levels).

In some cases, like the price/dividend or consumption/income ratios, it is inappropriate to remove a drift term. In these cases, we impose that the drift term μ is zero, and we use the following estimator:

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_{11} & \tilde{\psi}_{12} \\ \tilde{\psi}_{21} & \tilde{\psi}_{22} \end{bmatrix}$$

$$= \frac{1}{k(T-k+1)} \begin{bmatrix} \sum_{t=k}^{T} (x_t - x_{t-k})^2 & \sum_{t=k}^{T} (x_t - x_{t-k})(w_t - w_{t-k}) \\ & \sum_{t=k}^{T} (w_t - w_{t-k})^2 \end{bmatrix}. \tag{3.19}$$

To evaluate the potential bias of these estimators, assume that x_t and w_t are each random walks with no drift – the model of eq. (3.14) with $\theta = \phi = \mu_x = \mu_w = 0$. Now,

$$x_t - x_{t-k} = k\mu_x + \sum_{j=0}^{k-1} (1 + \phi L)\varepsilon_{t-j},$$

and (3.20)

$$x_T - x_0 = T\mu_x + \sum_{j=1}^{T} (1 + \phi L)\varepsilon_j.$$

Analogous expressions hold for w_t. Then the estimator of $1/k$ times the variance of k differences for x_t ($\hat{\psi}_{11}$) is, by substitution,

$$\hat{\psi}_{11} = \frac{T}{k(T-k)(T-k+1)} \sum_{t=k}^{T} \left[\sum_{j=0}^{k-1} \varepsilon_{t-j} - \frac{k}{T} \sum_{j=1}^{T} \varepsilon_j \right]^2. \tag{3.21}$$

Since $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$, the expected value of the term in brackets in (3.21) is

$$\left[k + \frac{k^2}{T} - \frac{2k^2}{T} \right] \sigma_\varepsilon^2,$$

and this term cancels with the fraction outside the sum in (3.21). *The variance estimators are unbiased for a random walk*: $E(\hat{\psi}_{11}) = \sigma_\varepsilon^2$.

For the covariance term, $\hat{\psi}_{12}$, the summation in (3.18) reduces to

$$\sum_{t=k}^{T}\left[\sum_{j=0}^{k-1}\varepsilon_{t-j}-\frac{k}{T}\sum_{j=1}^{T}\varepsilon_j\right]\left[\sum_{j=0}^{k-1}\nu_{t-j}-\frac{k}{T}\sum_{j=1}^{T}\nu_j\right]. \tag{3.22}$$

To get an idea of the behavior of this term, we examine the case in which the correlations ρ_j are nonzero at most up to order 1. Then we can write the expected value of the term in brackets in eq. (3.22) as

$$\begin{aligned}\sigma_\varepsilon\sigma_\nu(T-k+1)\Bigg\{&\left[k\rho_0+\frac{k-1}{k}(\rho_1+\rho_{-1})\right]\\ &+\frac{k^2}{T^2}\left[T\rho_0+(T-1)(\rho_1+\rho_{-1})\right]\\ &-\frac{2k}{T}\left[k\rho_0+\left(1-\frac{1}{k(T-k+1)}(\rho_1+\rho_{-1})\right)\right]\Bigg\}.\end{aligned} \tag{3.23}$$

That gives

$$\begin{aligned}\mathrm{E}(\hat{\psi}_{12})=&\frac{1}{1-\frac{k}{T}}\left(\rho_0+\frac{k-1}{k}(\rho_1+\rho_{-1})\right)\\ &+\frac{k}{T-k}\left[\rho_0+\left(1-\frac{1}{T}\right)(\rho_1+\rho_{-1})\right]\\ &-\frac{2k}{T-k}\left[\rho_0+\left(1-\frac{1}{k(T-k+1)}(\rho_1+\rho_{-1})\right)\right].\end{aligned} \tag{3.24}$$

This estimator is asymptotically unbiased. Taking the limit as $T\to\infty$,

$$\lim_{T\to\infty}\mathrm{E}(\hat{\psi}_{12})=\sigma_\varepsilon\sigma_\nu\left(\rho_0+\frac{k-1}{k}(\rho_1+\rho_{-1})\right). \tag{3.25}$$

Then, as $k\to\infty$, the estimator converges to $\sigma_\varepsilon\sigma_\nu$ ($\Sigma\rho_j=1$ by the assumption of cointegration). However, from (3.24), $\hat{\psi}_{12}$ is still biased for any finite k and T, if ρ_1 or ρ_{-1} are nonzero. The bias is always downward, but its value depends on the structure of the correlations between ε and ν at leads and lags, ρ_1 and ρ_{-1}.

The same conclusions apply to the model with $\theta=\phi=0$ and $\mu_x=\mu_w=0$ (no drift term) estimated by (3.19). The diagonal elements are unbiased for any k and T, while the off-diagonal elements are downward biased in a way that depends on ρ_1 and ρ_{-1}.

Next, suppose that the univariate representations for x and w are not random walks but MA(1)'s in first differences – model (3.14) with θ and $\phi \neq 0$. From (3.18), the diagonal terms are

$$E(\hat{\psi}_{11}) = \left\{ (1+\phi)^2 - \frac{2\phi}{k}\left[\frac{1+\dfrac{k^2}{T^2}-\dfrac{2k}{T(T-k+1)}}{1-\dfrac{k}{T}} \right] \right\} \sigma_\varepsilon^2. \tag{3.26}$$

The same holds for $\hat{\psi}_{22}$ with θ in the place of ϕ, and σ_ν in place of σ_ε.

As $T \to \infty$, (3.26) converges to $[(1+\phi)^2-(2\phi/k)]\sigma_\varepsilon^2$, which is the variance of k differences of a first differences stationary MA(1) process, and for $k \to \infty$ as well (3.26) converges to the variance of the random walk component, $(1+\phi)^2\sigma_\varepsilon^2$. Hence, $\hat{\psi}_{11}$ is an asymptotically unbiased estimate of $1/k$ times the variance of k differences of the MA(1) process. The same results hold if there is no drift term, and we estimate Ψ by (3.19).

For the off-diagonal (covariance) term, the estimators have expected value

$$E(\hat{\psi}_{12}) = \frac{1}{k}\sigma_\varepsilon\sigma_\nu(1+\phi)(1+\theta)\{(k+\phi\theta)\rho_0 + [(k-1)+2\theta+\phi\theta]\rho_1$$
$$+[(k-1)+2\phi+\theta\phi]\rho_{-1}\}, \tag{3.27}$$

in the case of no drift term. For $k \to \infty$, this expression converges to $\sigma_\varepsilon\sigma_\nu(1+\phi)(1+\theta)$.

Therefore, for MA(1) processes, the diagonal elements are asymptotically unbiased ($T \to \infty$, k finite) estimates of the variance of k differences of the underlying MA(1) processes. They are of course biased estimates of the random walk component as $k < \infty$. The off-diagonal or covariance terms are further biased as before, depending on the values of ρ_{-1}, ρ_1, and the MA(1) parameters θ and ϕ.

4. Application to stock prices and GNP

We will impose the cointegration of prices and dividends and of consumption and income to read the random walk or permanent component of prices (income) from that of dividends (consumption). We can cast the maintained assumption of cointegration either as a result of theory or as a plausible statistical characterization of the actual data.

4.1. Proof that the present value relation implies a stationary dividend/price ratio

It is well known that, if prices are given by the expected present value of dividends, the discount rate is constant and dividends are a random walk, then the dividend/price ratio is a constant. In this section, we generalize this statement: we show that the d/p ratio is stationary even if the discount rate is stochastic, so long as the discount rate and dividend growth are stationary and so long as dividends don't grow 'too fast'. [Hansen, Roberds and Sargent (1987) and Campbell and Shiller (1987) present related derivations.]

Assume that stock prices are generated from their underlying dividends by a present value relation,

$$P_t = \mathrm{E}_t \sum_{j=1}^{\infty} \beta^j \left(\prod_{k=1}^{j} \gamma_{t+k} \right) d_{t+j}, \qquad \beta < 1. \tag{4.1}$$

The discount factors γ_t could be generated from consumer's first-order conditions in complete markets, $\gamma_{t+k} = u'(c_{t+k-1})/u'(c_t)$. In this case (4.1) is

$$P_t = \mathrm{E}_t \sum_{j=1}^{\infty} \beta^j u'(c_{t+j})/u'(c_t) d_{t+j}. \tag{4.2}$$

If we are applying the model to nominal quantities γ_{t+k} also includes the ratio of price levels.

Dividing (4.1) by d_t, we obtain

$$P_t/d_t = \mathrm{E}_t \sum_{j=1}^{\infty} \beta^j \left(\prod_{k=1}^{j} \gamma_{t+k} \right) d_{t+j}/d_t$$

$$= \mathrm{E}_t \sum_{j=1}^{\infty} \beta^j \left(\prod_{k=1}^{j} \gamma_{t+k} \delta_{t+k} \right), \tag{4.3}$$

where

$$\delta_{t+k} = d_{t+k}/d_{t+k-1}.$$

Next, we assume that γ_t and δ_t are strongly stationary (the joint distribution of $\{\gamma_t, \gamma_{t+1}, \ldots, \gamma_{t+k}\}$ does not depend on t, and similarly for δ_t), and that they have at least second moments. These assumptions imply that γ_t and δ_t are also weakly stationary [$E(\gamma_t)$ exists and is independent of t, $E(\gamma_t^2)$ exists and is independent of t, $\text{cov}(\gamma_t, \gamma_{t-s})$ exists and depends only on s].

The function (4.3) is time invariant, so P_t/d_t is strongly stationary. To show that it is also weakly stationary, we must show that its second moment exists. Hence, we must show that

$$E\left(E_t \sum_{j=1}^{\infty} \beta^j \left(\prod_{k=1}^{j} \gamma_{t+k}\delta_{t+k} \right)\right)^2 < \infty. \tag{4.4}$$

Hansen and Sargent (1980) show that this is guaranteed if

$$\lim_{j \to \infty} \beta^{j/2} \left(\prod_{k=1}^{j} \gamma_{t+k}\delta_{t+k} \right) = 0 \quad \text{a.s.} \tag{4.5}$$

In turn, the condition (4.5) is guaranteed if (but not only if)

$$\lim_{j \to \infty} \beta^{j/2} \prod_{k=1}^{j} \delta_{t+k} = 0 \quad \text{a.s.,}$$

and

$$\lim_{j \to \infty} \beta^{\varepsilon j} \prod_{k=1}^{j} \gamma_{t+k} = 0 \quad \text{a.s.} \quad \text{for} \quad \varepsilon > 0. \tag{4.6}$$

This condition says that marginal utility doesn't grow asymptotically, and that dividends grow at less than the rate $\beta^{j/2}$.

We have just shown that if the discount factors γ_t (marginal rates of substitution, times inflation if we use nominal quantities) and dividend growth δ_t are strongly and weakly stationary, if they satisfy the bounds on their growth (4.6) and if prices are set by the present value relation (4.1), then the d_t/P_t ratio is also both strongly and weakly stationary.

4.2. Proof that the permanent income hypothesis implies that consumption and income are cointegrated

In the standard version of the permanent income theory [Hall (1978), Flavin (1981), Hansen (1987)] a representative consumer maximizes a quadratic expected utility function

$$\max \mathrm{E}_t\left[-\tfrac{1}{2}\sum_{j=0}^{\infty}\beta^j(c_{t+j}-\bar{c})^2\right], \tag{4.7}$$

subject to the resource constraint

$$A_{t+1}=(1+r)A_t+e_t-c_t,$$

$$\lim_{j\to\infty}\beta^j A_{t+j}=0 \quad \text{a.s.}, \quad \beta=(1+r)^{-1},$$

where e_t denotes labor income (or an endowment), and A_t is wealth at time t. The optimal consumption sequence is

$$c_t=r\left[A_t+\mathrm{E}_t\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j+1}e_{t+j}\right]$$

$$=rA_t+(1-\beta)\mathrm{E}_t\sum_{j=0}^{\infty}\beta^j e_{t+j}. \tag{4.8}$$

Substituting (4.8) back in the resource constraint, we get

$$A_{t+1}-A_t=e_t-(1-\beta)\mathrm{E}_t\sum_{j=0}^{\infty}\beta^j e_{t+j}. \tag{4.9}$$

Output (GNP) includes both labor income and property income, $y_t=rA_t+e_t$. With this definition, (4.8) implies

$$c_t-y_t=(1-\beta)\mathrm{E}_t\sum_{j=0}^{\infty}\beta^j e_{t+j}-e_t. \tag{4.10}$$

Therefore, to show that $x_t=c_t-y_t$ is a stationary process, or that c_t and y_t are cointegrated with cointegrating vector $\alpha=[1\ \ -1]'$, we need to prove that the right-hand side of (4.10) is a stationary process. This is trivially true if e_t is stationary, stationary about a linear trend, or stationary about a geometric trend of order less than β, so that the sum in (4.10) converges.

Suppose then that e_t is a first difference stationary process with representation:

$$(1-L)e_t=\mu+A(L)\varepsilon_t, \tag{4.11}$$

where ε_t is a white noise process; by the Beveridge–Nelson decomposition [eq.

(2.2)], we can write $e_t = z_t + v_t$, where z_t is a random walk representing the stochastic trend component of e_t,

$$(1-L)z_t = \mu + A(1)\varepsilon_t, \tag{4.12}$$

and v_t is stationary. Using the decomposition $e_t = z_t + v_t$,

$$\mathrm{E}_t \sum_{j=0}^{\infty} \beta^j e_{t+j} = \mathrm{E}_t\left(\sum_{j=0}^{\infty} \beta^j z_{t+j}\right) + \mathrm{E}_t\left(\sum_{j=0}^{\infty} \beta^j v_{t+j}\right). \tag{4.13}$$

Now,

$$\begin{aligned}\mathrm{E}_t \sum_{j=0}^{\infty} \beta^j z_{t+j} &= z_t + \mathrm{E}_t \sum_{j=1}^{\infty} \beta^j \left(z_t + j\mu + A(1)\sum_{k=1}^{j} \varepsilon_{t+k}\right) \\ &= \frac{1}{1-\beta} z_t + \frac{\mu\beta}{(1-\beta)^2},\end{aligned} \tag{4.14}$$

and from (2.2)

$$\mathrm{E}_t \sum_{j=0}^{\infty} \beta^j v_{t+j} = -\sum_{j=0}^{\infty} \beta^j \sum_{i \geq j} A_i^* \varepsilon_{t+j-i}. \tag{4.15}$$

Substituting (4.14) and (4.15) into (4.13),

$$\mathrm{E}_t \sum_{j=0}^{\infty} \beta^j e_{t+j} = \frac{1}{1-\beta} z_t + \frac{\mu\beta}{(1-\beta)^2} - \sum_{j=0}^{\infty} \beta^j \sum_{i \geq j} A_i^* \varepsilon_{t+j-i}, \tag{4.16}$$

and substituting this into (4.10), note that the random walk terms z_t cancel, so

$$c_t - y_t = \frac{\mu\beta}{(1-\beta)} - (1-\beta)\sum_{j=0}^{\infty} \beta^j \sum_{i \geq j} A_i^* \varepsilon_{t+j-i} - v_t. \tag{4.17}$$

Therefore, the right-hand side of (4.10) or (4.17) is stationary.

Two warnings are in order about this result. First, it holds for total consumption, including the flow of services provided by durable goods. We use data on components of consumption, e.g., nondurable and services. If the income elasticity of the various components isn't one, the ratio of components to GNP can drift up and down even though the ratio of total consumption to GNP is steady. In this case, the components are cointegrated with GNP with a cointegrating factor $\alpha_1 \neq 1$.

Second, this model of consumption cannot easily accommodate the observed growth in consumption and income. With growing income, it has the easily refutable prediction that savings are predominantly negative. Savings are equal to the expected *decline* in future income, so they are negative if income is expected to *grow*. The model also predicts that the difference of the *levels* of (per capita) income and consumption are stationary, not the more intuitively plausible *ratio* of consumption to income, or the difference of their logs.

A model that delivers cointegration of the logs of consumption and GNP is presented in King, Plosser, Stock and Watson (1987). It is a one-sector growth model driven by a production function shock that follows a logarithmic random walk. Their representative agent maximizes

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t(\theta_c \log(C_t) + v(L_t)), \tag{4.18}$$

and faces a production function

$$\begin{aligned} Y_t &= \lambda_t K_t^{1-\alpha} N_t^{\alpha}, \\ K_{t+1} &= (1-\delta) K_t + I_t, \\ \log(\lambda_t) &= \mu + \log(\lambda_{t-1}) + \eta_t, \qquad \eta_t \text{ i.i.d.} \end{aligned} \tag{4.19}$$

They show that the optimal decision rules for this economy include

$$\begin{aligned} \log(Y_t) &= (1-\alpha)\log(k_t) + \alpha \log(N(k_t)) + (1/\alpha)\log(\lambda_t), \\ \log(C_t) &= \log(C(k_t)) + (1/\alpha)\log(\lambda_t), \end{aligned} \tag{4.20}$$

where $k_t = K_t/\lambda_t^{1/\alpha}$ is stationary. (4.20) shows that $\log(C_t)$ and $\log(Y_t)$ have a common trend $(1/\alpha)\log(\lambda_t)$, and so are cointegrated with cointegrating vector $[1\ -1]'$ or $\log(Y_t) - \log(C_t)$ is stationary.[6]

4.3. Results for stock prices and dividends

4.3.1. Data

The price is the value-weighted New York Stock Exchange portfolio index constructed by the Center for Research in Security Prices; dividends are

[6]Note that the permanent income model generates unit roots in consumption and GNP even if the forcing process e_t is stationary. The King–Plosser–Stock–Watson model requires that the forcing process have a unit root.

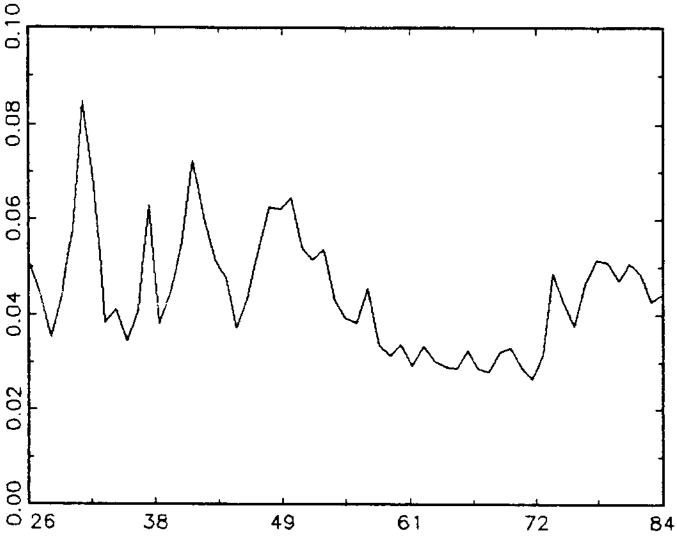


Fig. 3. Dividend/price ratio.

constructed by comparing the with and without dividend portfolios,

$$\operatorname{div}(t) = P(t)(\mathrm{e}^{r(t)} - \mathrm{e}^{rx(t)}),$$

where $P(t)$ is the price, $r(t)$ is the return on the VWNYSE, and $rx(t)$ is the return on the price-only VWNYSE. This technique is taken from Fama and French (1987). We use annual data to avoid the seasonal in dividends. Our dividend series is the total dividends for the year, and our price series is the December value.

4.3.2. Variance ratios for the dividend/price ratio

First, we checked the statistical assumption that the dividend/price ratio is stationary. Fig. 3 presents the dividend/price ratio. (Given the way we constructed our data, the units are arbitrary.) The first block of table 4 presents the variance of k differences and variance ratios (variance of k differences / variance of first differences) for log(dividends) – log(price). Typical point estimates of the variance ratio are 0.16 ($\pm$0.16) at $k = 30$ and 0.31 ($\pm$0.08) at $k = 10$.

Unfortunately the hypotheses of stationarity vs. unit root are adjacent, so any test for stationarity has no power against the alternative of a stationary series plus a small random walk component, in a finite sample. These variance ratios could as easily be generated by a stationary series with moderate persistence as by a small random walk plus another less persistent stationary series. A formal statistical test of stationarity is therefore inappropriate. [Cochrane (1987) discusses this point in detail.]

The point estimates tell us that it is at least not a bad approximation to treat the d/P ratio as a stationary variable. If there is a random walk component, it is quite small, and so would have little effect on statistics in this sample, though it would dominate asymptotically.

The standard errors in this block of table 4, and in all the other tables, are calculated from the Monte Carlo values reported on table 1, scaled by the variance at $k = 30$ (a point estimate of the variance of the underlying random walk component) and by k/T. This is a good approximation for large k, where little is left in the variance of k differences but the random walk component. It is a worse approximation for small k, because it ignores the stationary components.

4.3.3. Variance and covariance of price and dividends

Fig. 1 and the second block of table 4 present $1/k$ times the variance of k differences of log price and log dividends, and $1/k$ times the covariance of k differences.

When we impose the hypothesis that the d/p ratio is stationary, we can read the limiting value of the variance of k differences of price from the dividends line. The discussion above argues that this is in fact the best way to estimate the variance of the one common random walk component using both variances and the covariance.

Hence, we can calculate the ratio of the innovation variance of the random walk component of prices to the variance of first differences of prices (the 'variance ratio') by dividing $1/k$ times the variance of k differences of dividends by the variance of first differences of price. These calculations are presented in table 4, along with the variance ratio of prices calculated from the variance of k differences of price. Some resulting estimates of the variance ratios of prices are 0.21 at $k = 30$ and 0.38 at $k = 10$. These are 1/3 to 1/2 of the values estimated from the price line alone.

Table 4 also presents results for the postwar period. The great volatility of stock prices through WWII contributes most of the low univariate variance ratios. The variance ratio of prices alone declines to 0.75 at a seven-year horizon and then rises to 1.12 at a twenty-year horizon. However, the variance ratio calculated using dividends is about 0.3 at all horizons.

The standard errors are constructed as in the first block, by scaling the Monte Carlo results by k/T and by the variance of dividends at $k = 30$. It is tempting to note that 1 is more than 2 standard errors above the point estimate of the variance ratio, so we finally reject the hypothesis of a pure

Table 4

Results for stock prices and dividends.[a]

	Log dividend − log price													
K	1	2	3	4	5	6	7	8	9	10	15	20	25	30
	$10^5 \times 1/k \times$ variance of k differences													
Var	3.89	3.85	3.02	2.40	1.85	1.60	1.48	1.46	1.32	1.19	1.32	0.97	0.70	0.63
S.e.	0.11	0.14	0.17	0.20	0.22	0.24	0.26	0.28	0.30	0.32	0.41	0.49	0.56	0.64
	Ratio to one-year variance													
Ratio	1.00	0.99	0.78	0.62	0.48	0.41	0.38	0.37	0.34	0.31	0.34	0.25	0.18	0.16
S.e.	0.03	0.04	0.04	0.05	0.06	0.06	0.07	0.07	0.08	0.08	0.10	0.12	0.14	0.16

	Log price, log dividend, 1926–1984													
K	1	2	3	4	5	6	7	8	9	10	15	20	25	30
	$100 \times 1/k \times$ variance of k differences													
Var p	4.34	4.91	4.52	4.26	3.88	3.31	2.71	2.58	2.56	2.75	3.34	3.27	2.94	2.83
Cov p, d	0.90	1.47	1.80	2.01	2.08	1.90	1.56	1.40	1.38	1.49	1.57	1.57	1.40	1.34
Var d	1.43	2.02	2.27	2.34	2.31	2.26	2.10	1.91	1.77	1.67	1.57	1.29	0.92	0.90
S.e.	0.14	0.19	0.24	0.28	0.33	0.37	0.40	0.43	0.46	0.49	0.63	0.74	0.80	0.94
	Ratio of var div to one-year var p													
Ratio	0.33	0.46	0.52	0.54	0.53	0.52	0.48	0.44	0.41	0.38	0.36	0.30	0.21	0.21
S.e.	0.03	0.04	0.06	0.07	0.08	0.09	0.09	0.10	0.11	0.11	0.14	0.17	0.18	0.22
	Ratio of var p to one-year var p													
Ratio	1.00	1.13	1.04	0.98	0.89	0.76	0.62	0.59	0.59	0.63	0.77	0.75	0.68	0.65
S.e.	0.10	0.14	0.17	0.20	0.24	0.27	0.29	0.31	0.33	0.36	0.45	0.53	0.58	0.68

	Log price, log dividend, 1950–1984											
K	1	2	3	4	5	6	7	8	9	10	15	20
	$100 \times 1/k \times$ variance of k differences											
Var p	2.62	2.41	1.66	1.49	1.88	2.11	1.97	2.10	2.24	2.36	2.81	2.93
Cov p, d	0.44	0.44	0.39	0.39	0.48	0.50	0.42	0.37	0.35	0.29	0.20	0.30
Var d	0.32	0.42	0.53	0.62	0.71	0.79	0.82	0.82	0.79	0.74	0.55	0.79
S.e.	0.12	0.17	0.21	0.25	0.29	0.33	0.35	0.38	0.41	0.43	0.55	0.65
	Ratio of var div to one-year var p											
Ratio	0.12	0.16	0.20	0.24	0.27	0.30	0.31	0.31	0.30	0.28	0.21	0.30
S.e.	0.05	0.06	0.08	0.10	0.11	0.12	0.13	0.14	0.16	0.17	0.21	0.25
	Ratio of var p to one-year var p											
Ratio	1.00	0.92	0.63	0.57	0.72	0.81	0.75	0.80	0.86	0.90	1.07	1.12
S.e.	0.17	0.24	0.30	0.35	0.41	0.46	0.50	0.54	0.57	0.61	0.78	0.92

[a]Standard errors are calculated from Monte Carlos, scaled by k/T and the variance of the implied random walk component at $k = 30$.

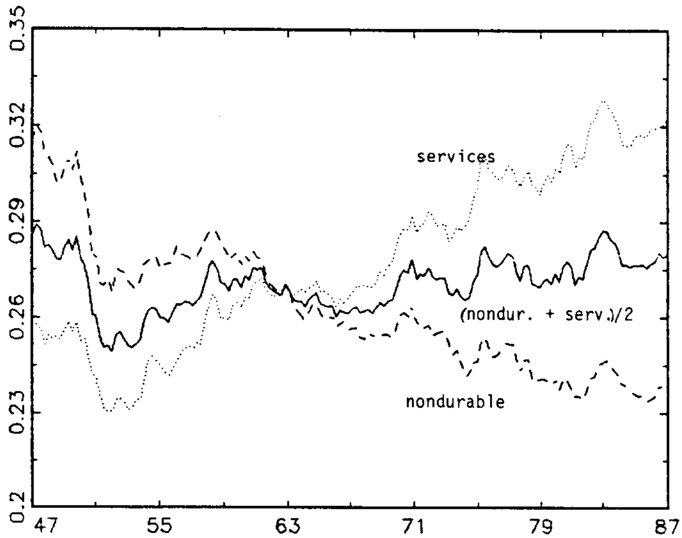


Fig. 4. Consumption/GNP ratio.

random walk at conventional significance levels. However, the standard errors scale with the size of the random walk component, so if the variance ratio was 1, the standard errors would be proportionately larger, and equal to the values given on table 1. Comparing that table with table 4, we can't formally reject the hypothesis of a pure random walk at the conventional 5% level, because the standard errors are so much larger for a pure random walk.

4.4. *Results for consumption and income*

4.4.1. *Data*

We used quarterly seasonally adjusted data on consumption of nondurables, consumption of services, and the sum of nondurable and services consumption together with GNP, all in 1982 dollars, from 1947:1 to 1986:4. Our proximate source was the Citibase-mini data set for PC's.

4.4.2. *Consumption / income ratios*

Fig. 4 presents consumption/income ratios for each of our definitions of consumption. The nondurable c/y ratio drifted down over the sample, the services c/y ratio drifted up, while the nondurable + services has little visible trend. Table 5, part A presents $1/k$ times the variance of k differences of log GNP − log consumption. (No drift term is removed in the calculations.) Results for the consumption/GNP ratio are the same to 0.02.

Here, as in all the $1/k$ times variance of k differences of variance ratios in table 5, the variance rises out to a year, and then declines. We suspect that this represents a spurious smoothness of quarterly data due to seasonal adjustment. Therefore, we present variance ratios using the one-year mark in the denominator, as well as the one-quarter mark.[7] As with the dividend/price ratio, table 5 suggests small point estimates of a random walk component in the c/y ratio.

4.4.3. *Variance and covariance of income and consumption*

Table 5, part A also presents $1/k$ times the variance and covariance of k differences of GNP and nondurable + services consumption. Fig. 2 presents the corresponding graph. As above, we constructed variance ratios using $1/k \times$ the variance of k differences of consumption dividend by the one-year and the one-quarter variance of income. Table 5, part A also presents the corresponding variance ratios calculated using income alone for comparison.

The ratio of consumption to yearly income changes is 0.31 at 10 years and 0.42 at 15 years, while the ratio to quarterly changes is 0.58 at 10 years and 0.79 at 15 years.[8] These estimates are roughly 1/2 to 2/3 the corresponding univariate estimates of 0.60 at 10 years and 0.59 at 15 years using one-year changes in the denominator, and 1.14 at 10 years and 1.11 at 15 years using quarterly changes in the denominator.

4.4.4. *Other definitions of the series*

We also tried using nondurable and services consumption separately. We clearly cannot model nondurable or services consumption as series that are cointegrated with GNP with cointegrating vector $[1\ -1]'$. Hence, we estimated a cointegrating vector $[1\ -\alpha_1]'$ by OLS, and then we took the variance ratio of $\log(y) - \alpha_1 \log(c)$. These are presented in table 5, parts B and C. (The estimated cointegrating vector for nondurable + services consumption was 0.97. Imposing $\alpha_1 = 1$ as above slightly raised the variance ratios, from 0.24 to 0.28 and from 0.18 to 0.21 at the fifteen-year horizon when one-quarter and one-year variances are used in the denominator, respectively.)

[7] As far as we can tell, this accounts for most of the difference between Cochrane's (1986) and Campbell and Mankiw's (1986, 1987) results for postwar data: Cochrane used annual and Campbell and Mankiw used quarterly data, so Cochrane had a larger denominator (one year) than Campbell and Mankiw (one quarter).

[8] We scaled the standard errors based on the fifteen-year estimate of the random walk component. If we had scaled based on the ten-year estimate, the standard errors would have been proportionally smaller.

Table 5

Results for consumption and GNP.

(A) Nondurable + services consumption and GNP

Log $y - \log c$													
k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
$10^4 \times 1/k$ var of k differences													
Var	1.05	1.18	1.41	1.34	1.35	1.26	1.00	0.83	0.66	0.56	0.47	0.36	0.29
S.e.	0.03	0.04	0.05	0.08	0.10	0.12	0.14	0.15	0.16	0.18	0.19	0.20	0.25
Ratio to one-quarter variance													
Var ratio	1.00	1.12	1.34	1.27	1.28	1.19	0.95	0.79	0.62	0.54	0.44	0.34	0.28
S.e.	0.02	0.04	0.05	0.07	0.09	0.11	0.13	0.14	0.16	0.17	0.18	0.19	0.24
Ratio to one-year variance													
Var ratio	0.75	0.84	1.00	0.95	0.96	0.89	0.71	0.59	0.47	0.40	0.33	0.26	0.21
S.e.	0.02	0.03	0.04	0.06	0.07	0.09	0.10	0.11	0.12	0.13	0.14	0.14	0.18

Log y, $\log c$													
k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
$10^4 \times 1/k$ var of k differences													
Var y	1.22	1.69	2.29	2.40	2.33	2.15	1.94	1.84	1.69	1.63	1.51	1.39	1.36
Cov c, y	0.25	0.45	0.66	0.76	0.70	0.65	0.68	0.72	0.75	0.77	0.77	0.78	0.88
Var c	0.35	0.40	0.47	0.53	0.53	0.55	0.59	0.61	0.63	0.66	0.68	0.71	0.97
S.e.	0.08	0.14	0.17	0.25	0.33	0.40	0.44	0.49	0.54	0.59	0.63	0.67	0.83
Ratio of var c to one-year var y													
Var ratio	0.15	0.17	0.20	0.23	0.23	0.24	0.26	0.27	0.28	0.29	0.30	0.31	0.42
S.e.	0.04	0.06	0.08	0.11	0.14	0.17	0.19	0.21	0.24	0.26	0.27	0.29	0.36
Ratio of var c to one-quarter var y													
Var ratio	0.29	0.33	0.38	0.43	0.43	0.45	0.48	0.50	0.52	0.54	0.56	0.58	0.79
S.e.	0.07	0.11	0.14	0.21	0.27	0.32	0.36	0.40	0.44	0.48	0.51	0.55	0.68
Ratio of var y to one-year var y													
Var ratio	0.53	0.74	1.00	1.05	1.02	0.94	0.85	0.80	0.74	0.71	0.66	0.60	0.59
S.e.	0.05	0.08	0.11	0.16	0.20	0.24	0.27	0.30	0.33	0.36	0.38	0.41	0.51
Ratio of var y to one-quarter var y													
Var ratio	1.00	1.39	1.88	1.97	1.91	1.76	1.59	1.51	1.39	1.34	1.24	1.14	1.11
S.e.	0.10	0.16	0.20	0.29	0.37	0.46	0.51	0.57	0.62	0.68	0.72	0.77	0.96

Parts B and C of table 5 also report $1/k \times$ the variance and covariance of k differences of GNP and nondurable and services consumption, and the variance ratios as calculated above.

For nondurable or services taken alone, the calculations of variance ratios using the variance of consumption/variance of income impose the stability of c/y or the cointegrating vector $[1\ -1]'$, contrary to fact. If we take the trends in these ratios seriously, we should modify our estimates of the variance of the

Table 5 (continued)

(B) Nondurable consumption and GNP

Log $y - \alpha_1 \log c$ ($\alpha_1 = 1.24$)

k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
						$10^4 \times 1/k$ var of k differences							
Var	1.59	1.51	1.61	1.34	1.33	1.25	0.96	0.74	0.57	0.46	0.36	0.24	0.22
S.e.	0.02	0.03	0.04	0.06	0.07	0.08	0.09	0.11	0.12	0.13	0.13	0.14	0.18
						Ratio to one-quarter variance							
Var ratio	1.00	0.95	1.02	0.85	0.84	0.79	0.61	0.47	0.36	0.29	0.22	0.15	0.14
S.e.	0.01	0.02	0.03	0.04	0.05	0.05	0.06	0.07	0.07	0.08	0.08	0.09	0.11
						Ratio to one-year variance							
Var ratio	0.98	0.94	1.00	0.83	0.82	0.78	0.60	0.46	0.35	0.28	0.22	0.15	0.14
S.e.	0.02	0.04	0.04	0.06	0.08	0.10	0.11	0.13	0.14	0.15	0.16	0.17	0.21

Log y, log c

k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
						$10^4 \times 1/k$ var of k differences							
Var y	1.22	1.69	2.29	2.40	2.33	2.15	1.94	1.84	1.69	1.63	1.51	1.39	1.36
Cov y, c	0.30	0.55	0.81	0.98	0.89	0.82	0.84	0.86	0.87	0.91	0.90	0.89	0.97
Var c	0.73	0.78	0.89	0.93	0.84	0.83	0.82	0.78	0.77	0.80	0.81	0.78	0.97
S.e.	0.08	0.14	0.17	0.25	0.33	0.40	0.44	0.49	0.54	0.59	0.63	0.67	0.83
						Ratio of var c to one-year var y							
Var ratio	0.32	0.34	0.39	0.40	0.37	0.36	0.36	0.34	0.34	0.35	0.35	0.34	0.42
S.e.	0.04	0.06	0.08	0.11	0.14	0.17	0.19	0.21	0.24	0.26	0.27	0.29	0.36
						Ratio of var c to one-quarter var y							
Var ratio	0.60	0.64	0.73	0.76	0.69	0.68	0.67	0.64	0.63	0.66	0.66	0.64	0.79
S.e.	0.07	0.11	0.14	0.21	0.27	0.33	0.36	0.40	0.44	0.48	0.52	0.55	0.68
						$(1.24)^2 \times$ ratio of var c to one-year var y							
Var ratio	0.49	0.52	0.60	0.62	0.57	0.56	0.55	0.53	0.52	0.54	0.55	0.53	0.65
S.e.	0.06	0.09	0.12	0.17	0.22	0.27	0.30	0.33	0.36	0.40	0.42	0.45	0.56
						$(1.24)^2 \times$ ratio of var c to one-quarter var y							
Var ratio	0.93	0.99	1.13	1.18	1.07	1.05	1.04	0.99	0.98	1.02	1.03	0.99	1.23
S.e.	0.11	0.18	0.22	0.32	0.41	0.50	0.56	0.62	0.69	0.75	0.80	0.85	1.06

random walk component in GNP by the square of the cointegrating factor, according to eq. (2.12). Table 5, parts B and C include these estimates of the variance ratio of GNP. Since the nondurable/GNP ratio trends down, its cointegrating factor α_1 is $1.2 > 1$, so this raises the estimate of the variance ratio of GNP; since the services/GNP ratio trends up, its cointegrating factor is $\alpha_1 = 0.8 < 1$ and including this correction lowers the estimate of the variance ratio in GNP. With a few exceptions, these results are similar to the results in part A.

Table 5 (continued)

	(C) Services consumption and GNP												
	Log $y - \alpha_1 \log c$ ($\alpha_1 = 0.80$)												
k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
	$10^4 \times 1/k$ var of k differences												
Var	1.07	1.33	1.63	1.64	1.61	1.46	1.18	1.00	0.79	0.70	0.59	0.44	0.29
S.e.	0.03	0.04	0.06	0.08	0.09	0.11	0.12	0.14	0.15	0.16	0.17	0.18	0.24
	Ratio to one-quarter variance												
Var ratio	1.00	1.24	1.52	1.53	1.50	1.36	1.10	0.93	0.74	0.66	0.55	0.42	0.27
S.e.	0.02	0.04	0.05	0.07	0.09	0.10	0.11	0.13	0.14	0.15	0.16	0.17	0.22
	Ratio to one-year variance												
Var ratio	0.66	0.82	1.00	1.01	0.99	0.90	0.73	0.61	0.48	0.43	0.36	0.27	0.18
S.e.	0.02	0.03	0.03	0.05	0.06	0.07	0.08	0.08	0.09	0.10	0.11	0.11	0.14
	Log y, log c												
k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
	$10^4 \times 1/k$ var of k differences												
Var y	1.22	1.69	2.29	2.40	2.33	2.15	1.94	1.84	1.69	1.63	1.51	1.39	1.36
Cov y, c	0.22	0.37	0.55	0.59	0.57	0.54	0.60	0.66	0.71	0.72	0.73	0.77	0.94
Var c	0.32	0.37	0.38	0.39	0.46	0.51	0.60	0.67	0.74	0.78	0.83	0.91	1.29
S.e.	0.11	0.18	0.23	0.34	0.43	0.53	0.59	0.66	0.72	0.78	0.84	0.89	1.11
	Ratio of var c to one-year var y												
Var ratio	0.14	0.16	0.17	0.17	0.20	0.22	0.26	0.29	0.32	0.34	0.36	0.40	0.56
S.e.	0.05	0.08	0.10	0.15	0.19	0.23	0.26	0.29	0.31	0.34	0.36	0.39	0.48
	Ratio of var c to one-quarter var y												
Var ratio	0.26	0.30	0.31	0.32	0.38	0.42	0.49	0.55	0.61	0.64	0.68	0.75	1.06
S.e.	0.09	0.15	0.19	0.28	0.36	0.43	0.49	0.54	0.59	0.64	0.69	0.73	0.91
	$(0.8)^2 \times$ ratio of var c to one-year var y												
Var ratio	0.09	0.10	0.11	0.11	0.13	0.14	0.17	0.19	0.21	0.22	0.23	0.25	0.36
S.e.	0.03	0.05	0.06	0.09	0.12	0.15	0.16	0.18	0.20	0.22	0.23	0.25	0.31
	$(0.8)^2 \times$ ratio of var c to one-quarter var y												
Var ratio	0.17	0.19	0.20	0.20	0.24	0.27	0.32	0.35	0.39	0.41	0.43	0.48	0.68
S.e.	0.06	0.10	0.12	0.18	0.23	0.28	0.31	0.34	0.38	0.41	0.44	0.47	0.58

We are not sure whether to take these cointegrating factors seriously. A cointegrating factor $\alpha > 1$ for a component of consumption implies that the ratio of that component to GNP will tend to 0 as income rises, while a cointegrating factor $\alpha < 1$ implies that the consumption/GNP ratio tends to ∞. Only $\alpha = 1$ does not have either property. Also, if the logs of two components have cointegrating factors with GNP different from 1, the log of their sum is not cointegrated with GNP. (If the levels were cointegrated with GNP with $\alpha \neq 1$, the sum can be cointegrated with GNP.) A second way of

Table 5 (continued)

(D) Per capita nondurable + services consumption and GNP

$y - \alpha_1 c$ ($\alpha_1 = 1.72$)

k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
	$100 \times 1/k \times$ var of k differences												
Var	1.21	1.36	1.55	1.44	1.32	1.14	0.87	0.72	0.58	0.53	0.46	0.34	0.25
S.e.	0.02	0.04	0.05	0.07	0.08	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.20
	Ratio to one-quarter variance												
Var ratio	1.00	1.12	1.27	1.18	1.09	0.94	0.72	0.59	0.48	0.43	0.38	0.28	0.21
S.e.	0.02	0.03	0.04	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.12	0.13	0.17
	Ratio to one-year variance												
Var ratio	0.79	0.88	1.00	0.93	0.86	0.74	0.56	0.47	0.38	0.34	0.30	0.22	0.16
S.e.	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.08	0.09	0.10	0.10	0.13

y, c

k	1qtr	2qtr	1yr	2	3	4	5	6	7	8	9	10	15
	$100 \times 1/k \times$ var of k differences												
Var y	1.47	1.98	2.65	2.84	2.60	2.18	1.74	1.63	1.55	1.56	1.56	1.43	0.91
Cov y, c	0.18	0.31	0.48	0.60	0.56	0.49	0.43	0.43	0.45	0.48	0.52	0.54	0.46
Var c	0.12	0.15	0.20	0.25	0.25	0.26	0.25	0.24	0.25	0.27	0.30	0.32	0.38
S.e.	0.03	0.05	0.07	0.10	0.13	0.16	0.17	0.19	0.21	0.23	0.25	0.26	0.33
	Ratio of var c to one-year var y												
Var ratio	0.05	0.06	0.08	0.09	0.10	0.10	0.09	0.09	0.09	0.10	0.11	0.12	0.14
S.e.	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.07	0.08	0.09	0.09	0.10	0.12
	Ratio of var c to one-quarter var y												
Var ratio	0.08	0.10	0.14	0.17	0.17	0.18	0.17	0.17	0.17	0.18	0.20	0.22	0.26
S.e.	0.02	0.04	0.05	0.07	0.09	0.11	0.12	0.13	0.14	0.16	0.17	0.18	0.22
	$(1.72)^2 \times$ ratio of var c to one-year var y												
Var ratio	0.14	0.17	0.22	0.28	0.28	0.29	0.28	0.27	0.28	0.30	0.33	0.36	0.43
S.e.	0.04	0.06	0.08	0.11	0.14	0.17	0.19	0.22	0.24	0.26	0.28	0.29	0.37
	$(1.72)^2 \times$ ratio of var c to one-quarter var y												
Var ratio	0.25	0.31	0.40	0.50	0.51	0.52	0.50	0.49	0.51	0.55	0.60	0.65	0.77
S.e.	0.07	0.11	0.14	0.20	0.26	0.31	0.35	0.39	0.43	0.47	0.50	0.53	0.66
	Ratio of var y to one-year var y												
Var ratio	0.55	0.75	1.00	1.07	0.98	0.82	0.66	0.61	0.58	0.59	0.59	0.54	0.34
S.e.	0.03	0.05	0.06	0.09	0.12	0.14	0.16	0.17	0.19	0.21	0.22	0.24	0.29
	Ratio of var y to one-quarter var y												
Var ratio	1.00	1.35	1.81	1.94	1.78	1.49	1.19	1.11	1.06	1.06	1.06	0.98	0.62
S.e.	0.05	0.09	0.11	0.16	0.21	0.25	0.28	0.32	0.35	0.38	0.40	0.43	0.53

stating the same problem is that the 'trends' in consumption/GNP ratios are generated by different income elasticities for the components, but as income $\rightarrow \infty$, all the income elasticities must approach 1 for goods whose shares do not approach zero. On the other hand, it is clearly inappropriate to ignore them in our sample.

We also tried applying our technique to the per capita levels, rather than the log levels, as suggested by a literal interpretation of the permanent income model. Table 5, part D presents our results. First we applied the variance ratio to GNP per capita $-$ α_1 consumption per capita, using consumption of nondurable + services. The variance ratios were around 0.2 at ten and fifteen-year horizons. The high value of the cointegrating factor α_1 comes from the fact that we are looking at the *difference* between output and consumption. Since it is the *ratio* of output to consumption that is steady over time, the difference increases steadily.

The $1/k$ variance of k differences and variance ratios for this case are broadly similar to our results in part A out to the ten-year horizon. At the fifteen-year horizon, the variance of consumption increases and the variance of income declines. However, the values of the variance ratios are still low compared to univariate results.

We regard these latter results as a check that our main results for consumption and GNP presented in table 5, part A are robust to the definitions of the variables.

5. Conclusions

The heart of our technique is: to measure the variance of a permanent component of a series, or its variance ratio, find another series that is nearly a random walk, yet is cointegrated with the original series. Then, use $1/k$ times the variance of k differences of the new series to measure the limiting variance ratio or the variance of the permanent component of the original series. This technique improves the bias, but not the precision, of corresponding univariate variance ratios. (Estimated standard errors are lower when the estimated random walk component is smaller, but are the same for a given size of random walk component.)

Applying this technique to stock prices and GNP, using dividends and consumption as the near random walk variables, we get estimates of the variance ratios or variance of permanent components of stock prices and GNP about half of corresponding univariate estimates – variance ratios of 0.2 to 0.3 for stock prices and 0.3 to 0.4 for GNP were typical, compared to corresponding univariate estimates of 0.65 – 0.75 or more for stocks and 0.6 for GNP.

These are interesting stylized facts. They indicate that stock prices and GNP follow processes that are unlike random walks, by reverting towards a mean at long horizons. The connection of these stylized facts to theory is still tenuous. Most interesting business cycle models are capable of generating a variety of variance ratios with small changes in their assumptions – by changing the process of the forcing variables, for example. Poterba and Summers (1987) show that for some very special cases, the variance ratio of stock prices puts a lower bound on the variance of discount rates, or marginal rates of substitution, and that these seem 'too high'. If these bounds can be generalized, they may provide a sharp statement of the asset pricing puzzle that marginal rates of substitution seem 'too variable'.

References

Beveridge, S. and C.R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle, Journal of Monetary Economics 7, 151–174.

Campbell, J.Y., 1985, Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis, Manuscript (Princeton University, Princeton, NJ).

Campbell, J.Y. and N.G. Mankiw, 1986, Are output fluctuations transitory?, Quarterly Journal of Economics, forthcoming.

Campbell, J.Y. and N.G. Mankiw, 1987, International evidence on the persistence of economic fluctuations, Manuscript (Princeton University, Princeton, NJ).

Campbell, J.Y. and N.G. Mankiw, 1987, Permanent and transitory components in macroeconomic fluctuations, American Economic Review 77, 111–117.

Campbell, J.Y. and R.J. Shiller, 1987, Cointegration and tests of present value models, Journal of Political Economy 95, 1062–1088.

Cochrane, J.H., 1986, How big is the random walk in GNP?, Journal of Political Economy, forthcoming.

Cochrane, J.H., 1987, The application of unit root tests: A critique, Manuscript (University of Chicago, Chicago, IL).

Clark, P.K., 1987, The cyclical component of US economic activity, Quarterly Journal of Economics, forthcoming.

Dickey, D.A. and W.A. Fuller, 1979, Distributions of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427–443.

Dickey, D.A. and W.A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 49, 1057–1072.

Dickey, D.A., W.R. Bell and R.B. Miller, 1986, Unit roots in time series models: Tests and implications, American Statistician 40, 12–26.

Durlauf, S.N. and P.C.B. Phillips, 1986, Trends versus random walks in time series analysis, Cowles Foundation discussion paper no. 788 (Yale University, New Haven, CT).

Engle, R.F. and C.W.J. Granger, 1987, Cointegration and error-correction: Representation, estimation and testing, Econometrica 55, 251–276.

Fama, E.F. and K. French, 1986, Permanent and temporary components of stock prices, Journal of Political Economy, forthcoming.

Fama, E.F. and K. French, 1987, Dividend yields and expected returns, Manuscript (University of Chicago, Chicago, IL).

Flavin, M.A., 1981, The adjustment of consumption to changing expectations about future income, Journal of Political Economy 89, 974–1009.

Hall, R.E., 1978, Stochastic implications of the life-cycle permanent income hypothesis: Theory and evidence, Journal of Political Economy 86, 971–987.

Hansen, L.P., 1987, Calculating asset prices in three example economies, in: T.F. Bewley, ed., Advances in econometrics: Proceedings of the fifth world congress (Cambridge University Press, Cambridge).

Hansen, L.P. and T.J. Sargent, 1980, Formulating and estimating dynamic linear rational expectations models, Journal of Economic Dynamics and Control 2, 7–46.

Hansen, L. P., W. Roberds and T.J. Sargent, 1987, Time series implications of present-value budget balance and of martingale models of consumption or taxes, Manuscript (University of Chicago, Chicago, IL).

King, R., C. Plosser, J. Stock and M. Watson, 1987, Stochastic trends and economic fluctuations, Manuscript (University of Rochester, Rochester, NY).

Koopmans, T.J., 1974, The spectral analysis of times series (Academic Press, New York).

Lo, A.W. and A.C. Mackinlay, 1987, Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies, forthcoming.

Nelson, C.R. and C. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, Journal of Monetary Economics 10, 139–162.

Phillips, P.C.B. and S. Ouliaris, 1986, Testing for cointegration, Cowles Foundation discussion paper no. 809 (Yale University, New Haven, CT).

Poterba, J.M. and L.H. Summers, 1987, Mean reversion in stock prices: Evidence and implications, NBER working paper no. 2343.

Stock, J.H. and M. Watson, 1986, Testing for common trends, Manuscript (Harvard University, Cambridge, MA).

Strang, G., 1976, Linear algebra and its applications (Academic Press, New York).