Cointegration and Tests of Present Value Models

Citation

Published Version

Permanent link

Terms of Use

Share Your Story

Cointegration and Tests of Present Value Models

John Y. Campbell

Princeton University

Robert J. Shiller

Yale University

Application of some advances in econometrics (in the theory of cointegrated vector autoregressive models) enables us to deal effectively with two problems in rational expectations present value models: nonstationarity of time series and incomplete data on information of market participants. With U.S. data, we find some relatively encouraging new results for the rational expectations theory of the term structure and some puzzling results for the present value model of stock prices.

Present value models are among the simplest dynamic stochastic models of economics. A present value model for two variables, y_t and Y_t, states that Y_t is a linear function of the present discounted value of expected future y_t:

$$Y_t = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i E_t y_{t+i} + c, \tag{1}$$

where c, the constant, θ, the coefficient of proportionality, and δ, the discount factor, are parameters that may be known a priori or may need to be estimated. Here and in what follows, E_t denotes mathematical expectation, conditional on the full public information set $\mathbf{I}_t$, which includes y_t and Y_t themselves and in general exceeds the information set $\mathbf{H}_t$ available to the econometrician. Models of this form include the expectations theory for interest rates (Y_t is the long-term yield and y_t the one-period rate), the present value model of stock prices (Y_t is the stock price and y_t the dividend), and, with some modification, the permanent income theory of consumption.[1]

Despite the simplicity of their structure, there is a surprising degree of controversy about the validity of present value models for bonds, stocks, and other economic variables.[2] The controversy seems to be stimulated by three problems that arise in testing equation (1). First, there are several test procedures in the literature: these include single-equation regression tests, tests of cross-equation restrictions on a vector autoregression (VAR), and variance bounds tests. It is not clear how these alternative approaches are related to one another.

Second, a statistical rejection of the model (1) may not have much economic significance. It is entirely possible that the model explains most of the variation in Y_t even if it is rejected at the 5 percent level. Most work on present value models concentrates on statistical testing rather than informal evaluation of the "fit" of the models.

Finally, the variables y_t and Y_t usually require some transformation before the theory of stationary stochastic processes can be applied. One approach is to remove a deterministic linear trend, but this can bias test procedures against the model (1) if in fact y_t and Y_t are nonstationary in levels.[3]

In this paper we develop a test of the present value relation that is valid when the variables are stationary in first differences.[4] Hansen and Sargent (1981a), Mankiw et al. (1985), and West (1986, 1987)

[1] The discounted sum in eq. (1) extends to an infinite horizon. Most of the methods in this paper can be applied to the finite horizon case, at the cost of some additional complexity. Throughout this paper we will treat conditional expectations as equivalent to linear projections on information.

[2] For bonds, see Sargent (1979), Shiller (1979, 1981a, 1987), Hansen and Sargent (1981a), Shiller, Campbell, and Schoenholtz (1983), and Campbell and Shiller (1984). For stocks, see LeRoy and Porter (1981), Shiller (1981b, 1984), Mankiw, Romer, and Shapiro (1985), Scott (1985), Marsh and Merton (1986), and West (1986, 1987).

[3] This point is made for stocks by Kleidon (1986) and Marsh and Merton (1986). Mankiw and Shapiro (1985) present a similar argument for the permanent income theory of consumption.

[4] It might be attractive to model the variables y and Y as stationary in log first differences. However, since the model (1) is linear in levels, a log specification is intractable unless one is willing to focus on a special case (Kleidon 1986) or to approximate the model (Campbell and Shiller 1986).

have also studied this case. We follow Hansen and Sargent and differ from Mankiw et al. and West by using a relatively large information set $\mathbf{H}_t$. We include in $\mathbf{H}_t$ current and lagged values not just of y_t but also of Y_t.

Our choice of information set has several advantages. By including Y_t in the vector stochastic process for analysis, we in effect include *all* relevant information of market participants, even if we econometricians do not observe all their information variables. We can test *all* the implications of the model for the bivariate (y_t, Y_t) process, giving a natural extension of Fama's (1970) notion of a "weak-form" test. We can exploit the recently developed theory of cointegrated processes (Phillips and Durlauf 1986; Phillips and Ouliaris 1986; Engle and Granger 1987; Stock 1987). Our test procedure can be interpreted as a single-equation regression or as a test of restrictions on a VAR. We propose a way to assess the economic significance of deviations from (1), comparing the forecast of the present value of future y_t embodied in Y_t with an unrestricted VAR forecast. Because the information set $\mathbf{H}_t$ includes Y_t, the two forecasts should be equal if the model is true.

We examine the present value models for bonds and stocks, while a companion piece by one of us (Campbell 1987) studies the permanent income theory of consumption. The paper is organized as follows. Section I discusses alternative tests of the present value relation when y_t and Y_t are stationary in first differences rather than levels. Section II is an introduction to the literature on cointegration, summarizing the results we use in testing the present value model. Section III applies the method to data on bonds and stocks. Section IV presents conclusions.

I. Alternative Tests of the Present Value Relation

One straightforward way to test the model (1) is to use it to restrict the behavior of the variable $\xi_t \equiv Y_t - (1/\delta)[Y_{t-1} - \theta(1 - \delta)y_{t-1}]$. Substitution from (1) shows that

$$\xi_t = Y_t - E_{t-1}Y_t + c\left(1 - \frac{1}{\delta}\right). \tag{2}$$

Apart from a constant, ξ_t is the true innovation at time t in Y_t (i.e., the innovation with respect to the full market information set $\mathbf{I}_t$). The model has the striking implication that this innovation is observable when only Y_t, Y_{t-1}, y_{t-1}, and the parameters c, θ, and δ are known.[5] In

---

[5] The variable ξ_t can also be written as a constant plus the true innovation in the expected present value of all future y_t. We note, however, that in general the model does *not* identify the true innovation in y_t itself.

the applications of this paper, ξ_t has the economic interpretation of an asset return. In the term structure it is the excess return on long bonds over short bills, while in the stock market it is the excess return on stocks over a constant mean, multiplied by the stock price.

Since the right-hand side of (2), adjusted for a constant, is orthogonal to all elements of the information set $\mathbf{I}_{t-1}$, one can test the present value relation by regressing ξ_t on variables in this set and testing that the coefficients are jointly zero. This approach is standard in the literature and seems attractively simple. However, there are some econometric pitfalls and issues of interpretation that need careful handling.

First, the regressors used to predict ξ_t must be stationary if conventional asymptotic distribution theory is to apply. Of course, there are many stationary elements of $\mathbf{I}_{t-1}$, but one may want to choose variables that summarize the joint history of y_t and Y_t. It is not clear how the stationarity requirement can be reconciled with this objective if y_t and Y_t are themselves nonstationary.

Second, while (1) implies (2), the reverse is not true. Equation (2) is consistent with a more general form of (1) that includes a "rational bubble," a random variable b_t satisfying $b_t = \delta E_t b_{t+1}$. Recently there has been considerable interest in testing (1) against the alternative that Y_t is influenced by a rational bubble (Blanchard and Watson 1982; Hamilton and Whiteman 1985; Quah 1986; West 1987).

Third, it is not clear what are the implications for Y_t of nonzero coefficients in a regression of ξ_t on information. Predictability of returns has consequences for asset price behavior, and one may want to calculate these explicitly.

Further insight into these issues can be gained by defining a new variable $S_t \equiv Y_t - \theta y_t$. We will refer to S_t as the "spread." In the case of the term structure, it is just the spread between long- and short-term interest rates; for stocks, it is the difference between the stock price and a multiple of dividends. The spread can also be written as a linear combination of the variables ΔY_t, Δy_t, and ξ_t: $S_t = [1/(1 - \delta)]\Delta Y_t - \theta \Delta y_t - [\delta/(1 - \delta)]\xi_t$.

The present value model (1) implies two alternative interpretations of the spread. Subtracting θy_t from both sides of equation (1) and rearranging, one obtains

$$S_t = E_t S_t^* + c, \tag{3}$$

where

$$S_t^* = \theta \sum_{i=1}^{\infty} \delta^i \Delta y_{t+i},$$

and

$$S_t = \left(\frac{\delta}{1 - \delta}\right)E_t\Delta Y_{t+1} + c. \qquad (4)$$

Equation (3) says that the spread is a constant plus the optimal forecast of S_t^*, a weighted average of future changes in y; equation (4) says that the spread is linear in the optimal forecast of the change in Y.

Equation (4) can be used in an alternative test of the present value model, in which one regresses ΔY_t on a constant, S_{t-1}, and other variables. The coefficient on S_{t-1} should be $(1 - \delta)/\delta$, and the coefficients on the other variables should be zero. This regression is just a linear transformation of the regression that has ξ_t as the dependent variable, and it yields the same test statistic.

Equations (3) and (4) help to resolve the issues raised above. If Δy_t is stationary, it follows from (3) that S_t is stationary; (4) then implies that ΔY_t is stationary. Thus one can use S_t and Δy_t, or S_t and ΔY_t, as stationary variables that summarize the bivariate history of y_t and Y_t in a regression test of the model. (The pair Δy_t and ΔY_t is also stationary, but by using these one would lose information on the relative levels of y_t and Y_t.) Our strategy is to work with S_t and Δy_t.

The effect of a "rational bubble" alternative is easily seen using (3) and (4). If a term b_t is added to the right-hand side of equation (1), satisfying $b_t = \delta E_t b_{t+1}$, it appears on the right-hand side of (3) but does not affect equations (2) and (4). The term b_t is explosive by construction, so it causes explosive behavior of S_t by (3), and this is passed through to ΔY_t by (4).[6]

One way to test for the importance of rational bubbles is therefore to test the stationarity of S_t and ΔY_t. This approach has been proposed by Diba and Grossman (1984), among others. As we noted above, S_t can be written as a linear combination of ΔY_t, Δy_t, and ξ_t. Therefore, independent of any model, if three of the variables S_t, ΔY_t, Δy_t, and ξ_t are stationary, the fourth must be also. This linear dependence needs to be taken into account in testing for stationarity.

Finally, (3) and (4) suggest a way to compute the implications for Y_t of predictable ξ_t. Consider estimating a VAR representation for Δy_t and S_t (with their means removed):

$$\begin{bmatrix} \Delta y_t \\ S_t \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ c(L) & d(L) \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ S_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \qquad (5)$$

[6] Quah (1986) gives an example in which b_t satisfies $b_t = \delta E_t b_{t+1}$ but is stationary. However, this example violates the equivalence of conditional expectations and linear projections, which we assume here.

where the polynomials in the lag operator $a(L)$, $b(L)$, $c(L)$, and $d(L)$ are all of order p. This VAR can be used for multiperiod forecasting of Δy_t, and it includes the variable S_t, which, according to (3), is the optimal forecast of the present value of future Δy_t.

To simplify notation, (5) can be stacked into a first-order system

$$
\begin{bmatrix} \Delta y_t \\ \cdot \\ \cdot \\ \Delta y_{t-p+1} \\ S_t \\ \cdot \\ \cdot \\ S_{t-p+1} \end{bmatrix} = \begin{bmatrix} a_1 \ldots a_p \ b_1 \ldots b_p \\ 1 \\ \quad \cdot \\ \qquad 1 \\ c_1 \ldots c_p \ d_1 \ldots d_p \\ \qquad\qquad 1 \\ \qquad\qquad\quad \cdot \\ \qquad\qquad\qquad 1 \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \cdot \\ \cdot \\ \Delta y_{t-p} \\ S_{t-1} \\ \cdot \\ \cdot \\ S_{t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ 0 \\ \cdot \\ 0 \\ u_{2t} \\ 0 \\ \cdot \\ 0 \end{bmatrix} , \qquad (6)
$$

where blank elements are zero. This can be written more succinctly as $\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{v}_t$. The matrix $\mathbf{A}$ is called the companion matrix of the VAR. For all i, $E(\mathbf{z}_{t+i}|\mathbf{H}_t) = \mathbf{A}^i\mathbf{z}_t$, where $\mathbf{H}_t$ is the limited information set containing current and lagged values of y_t and Y_t or, equivalently, of $\mathbf{z}_t$. As elsewhere in the paper, we are taking conditional expectations to be linear projections on information.

We can now discuss the implications of the present value relation for the VAR system. A rather weak implication is that S_t must linearly Granger-cause Δy_t unless S_t is itself an exact linear function of current and lagged Δy_t (which is a stochastic singularity we do not observe in the data; it would require, e.g., that the variance-covariance matrix of u_{1t} and u_{2t}, $\mathbf{\Omega}$, be singular).

The intuitive explanation for this result is that S_t is an optimal forecast of a weighted sum of future values of Δy_t, conditional on agents' full information set. Therefore, S_t will have incremental explanatory power for future Δy_t if agents have information useful for forecasting Δy_t beyond the history of that variable. If agents do not have such information, they form S_t as an exact linear function of current and lagged Δy_t.[7]

The full set of restrictions of the present value model is more demanding. We obtain these restrictions by projecting equation (3) onto the information set $\mathbf{H}_t$, noting that the left-hand side is unchanged because S_t is in $\mathbf{H}_t$ and rewriting as

$$
\mathbf{g}'\mathbf{z}_t = \theta \sum_{i=1}^{\infty} \delta^i \mathbf{h}' \mathbf{A}^i \mathbf{z}_t,
$$

[7] A formal proof is as follows. Suppose that S_t does not Granger-cause Δy_t. Then $E(\Delta y_{t+i}|\mathbf{H}_t) = E(\Delta y_{t+i}|\Delta y_t, \Delta y_{t-1}, \ldots)$ for all i, and from (3), $E(S_t|\mathbf{H}_t) = E(S_t|\Delta y_t, \Delta y_{t-1}, \ldots)$, an exact linear function of current and lagged Δy_t. But because S_t is itself in the information set $\mathbf{H}_t$, $S_t = E(S_t|\mathbf{H}_t)$.

where $\mathbf{g}'$ and $\mathbf{h}'$ are row vectors with $2p$ elements, all of which are zero except for the $p + 1$st element of $\mathbf{g}'$ and the first element of $\mathbf{h}'$, which are unity. If this expression is to hold for general $\mathbf{z}_t$ (i.e., for nonsingular $\mathbf{\Omega}$), it must be the case that

$$\mathbf{g}' = \theta \sum_{i=1}^{\infty} \delta^i \mathbf{h}' \mathbf{A}^i = \theta \mathbf{h}' \delta \mathbf{A} (\mathbf{I} - \delta \mathbf{A})^{-1}. \tag{7}$$

Here the second equality follows by evaluating the infinite sum, noting that it must converge because the variables Δy_t and S_t are stationary under the null.[8]

The restrictions of equation (7) appear to be highly nonlinear cross-equation restrictions of the type described by Hansen and Sargent (1981b) as the "hallmark" of rational expectations models. However, it turns out that (7) can be simplified so that (taking θ and δ as given) its restrictions are linear and easily interpreted. Postmultiplying both sides of (7) by $(\mathbf{I} - \delta \mathbf{A})$, one obtains

$$\mathbf{g}'(\mathbf{I} - \delta \mathbf{A}) = \theta \mathbf{h}' \delta \mathbf{A}. \tag{8}$$

From the structure of the matrix $\mathbf{A}$, the constraints imposed by (8) on individual coefficients are $c_i = -\theta a_i, i = 1, \ldots, p; d_1 = (1/\delta) - \theta b_1;$ and $d_i = -\theta b_i, i = 2, \ldots, p$. By adding $\theta \Delta y_t$ to S_t, one can interpret these restrictions. They state that $\xi_t = S_t - (1/\delta) S_{t-1} + \theta \Delta y_t$ is unpredictable given lagged Δy_t and S_t, which is what equation (2) implies for the information set $\mathbf{H}_t$. In our empirical application, we obtain a Wald test statistic for equation (8) that is numerically identical to the Wald test statistic for a regression of ξ_t on lagged Δy_t and S_t.[9]

The major advantage of the VAR framework is that it can be used to generate alternative measures of the economic importance, not merely the statistical significance, of deviations from the present value relation. To see this more clearly, suppose that the present value model is false so that $E_t \xi_{t+i} \neq 0$ for $i \geq 1$. Then equations (3) and (4) no longer hold. We define the "theoretical spread," S_t', as the optimal forecast, given the information set $\mathbf{H}_t$, of the present value of all future changes in y:

$$S_t' \equiv E(S_t^* | \mathbf{H}_t) = \theta \mathbf{h}' \delta \mathbf{A} (\mathbf{I} - \delta \mathbf{A})^{-1} \mathbf{z}_t. \tag{9}$$

[8] Under an explosive bubble alternative this infinite sum will not converge, and the matrix $(\mathbf{I} - \delta \mathbf{A})$ will be singular.

[9] However, this statistic is not numerically identical to the Wald statistic for a test of eq. (7), even though (7) and (8) are algebraically equivalent restrictions. Nonlinear transformations of restrictions can change the numerical values of Wald statistics and, as Gregory and Veall (1985) point out, can dramatically alter their power. We report Wald statistics for (8) in the tables that summarize our empirical results and Wald statistics for (7) in notes.

We then have, ignoring constant terms,

$$S_t - S_t' = \sum_{i=1}^{\infty} \delta^i E(\xi_{t+i}|\mathbf{H}_t) \tag{10}$$

and

$$S_t - \left(\frac{\delta}{1-\delta}\right)E(\Delta Y_{t+1}|\mathbf{H}_t) = \left(\frac{1}{1-\delta}\right)E(\xi_{t+1}|\mathbf{H}_t). \tag{11}$$

Equations (10) and (11) measure deviations from the model in two different ways. The metric of equation (11) is the difference between S_t and the optimal forecast, given the information set $\mathbf{H}_t$, of the one-period change in Y. Equation (11) shows that this difference is large if excess returns are predictable one period in advance.

The metric of equation (10) is the difference between S_t and the theoretical spread, which is large if the present value of all future excess returns is predictable. By this measure, a large deviation from the model requires not only that movements in ξ be predictable one period in advance but that they be predictable many periods in advance. Loosely speaking, predictable excess returns must be persistent as well as variable.[10]

We use the VAR framework not only to conduct statistical tests of the present value relation but also to evaluate its failures using the metric of equation (10). We display time-series plots of the spread S_t and the theoretical spread S_t', the unrestricted VAR forecast of the present value of future changes in y. If the present value model is true, these variables should differ only because of sampling error. Large observed differences in the time-series movements of the two variables imply (subject to sampling error) economically important deviations from the model.

The VAR framework can also be used to test the present value model against more specific alternatives. Volatility tests, for example, are designed to test against the alternative that Y_t or some transformation of it "moves too much."

We present two different volatility tests. The first is just a test that the ratio $\text{var}(S_t)/\text{var}(S_t')$ is unity. This ratio, together with its standard error, can be computed from the VAR system. Under the present value model, the ratio should be one but would be larger than one if the spread is too volatile relative to information about future y. A statistic that complements this is the correlation between S_t and S_t'

---

[10] The terminology of our earlier paper (Campbell and Shiller 1984) may be helpful in understanding (10) and (11). The right-hand side of (11) is proportional to what we called the one-period "holding premium," and the right-hand side of (10) is what we called the "rolling premium."

since if the variance ratio and correlation both equal one, then S_t must equal S_t' and the model is satisfied.[11]

We obtain a second volatility test, following West (1987), as follows. Let us define ξ_t' as θ times the innovation from $t-1$ to t in the expected present value of Δy, conditional on the VAR information set:

$$\xi_t' \equiv \theta \sum_{i=0}^{\infty} \delta^i [E(\Delta y_{t+i}|\mathbf{H}_t) - E(\Delta y_{t+i}|\mathbf{H}_{t-1})]$$
$$= S_t' - \left(\frac{1}{\delta}\right)S_{t-1}' + \theta\Delta y_t. \tag{12}$$

Under the present value model, $\xi_t' = \xi_t$ since $S_t' = S_t$. We construct the ratio var(ξ_t)/var(ξ_t'), again with standard error.[12] The model implies that this ratio should be one, while the notion that stock prices are too volatile suggests that it will be greater than one. We call the first of our variance ratios the "levels variance ratio" and the second the "innovations variance ratio."

The fact that a linear combination S_t of y_t and Y_t is stationary in its level, even though y_t and Y_t are individually stationary only in first differences, turns out to be important for understanding present value models. In the language of time-series analysis, the vector $\mathbf{x}_t = (y_t\ Y_t)'$ is cointegrated. Cointegrated vectors have a number of important properties, which we now discuss.

II. Properties of Cointegrated Vectors

In this section we summarize the theory of cointegrated processes and show how it applies to present value models.

DEFINITION (Engle and Granger 1987). A vector $\mathbf{x}_t$ is said to be cointegrated of order (d, b), denoted $\mathbf{x}_t$ CI(d, b), if (i) all components of $\mathbf{x}_t$ are integrated of order d (stationary in dth differences) and (ii) there exists at least one vector $\boldsymbol{\alpha}$ ($\neq 0$) such that $\boldsymbol{\alpha}'\mathbf{x}_t$ is integrated of order $d - b$, $b > 0$.

When y_t is stationary in first differences, the vector $\mathbf{x}_t = (y_t\ Y_t)'$ is CI$(1, 1)$ if the present value model holds. The CI$(1, 1)$ case is the one

[11] We compute the levels variance ratio and correlation from the sample moments of S_t and S_t'. We report numerical standard errors that are conditional on the sample moments of $\mathbf{z}_t$ and take account of sampling error only in the coefficients of the estimated VAR.

[12] We use the estimated variance-covariance matrix of the VAR to compute the innovations variance ratio. The standard error takes account of sampling error in this matrix as well as in the VAR coefficients.

that has been studied almost exclusively in the theoretical literature, and the results that follow apply to it.

Cointegrated systems of order (1, 1) have two unusual properties. These concern the existence of well-behaved vector time-series representations for the cointegrated variables and the estimation of unknown elements of the vector α. Both properties turn out to be relevant for testing present value models.

The first important property of a cointegrated vector is that the vector moving average (VMA) representation of the first difference $\Delta \mathbf{x}_t$ is noninvertible. Equivalently, the spectral density matrix of $\Delta \mathbf{x}_t$ is singular at zero frequency. This singularity is what "holds together" the elements of $\mathbf{x}_t$ so that a linear combination is stationary.

More formally, write $\Delta \mathbf{x}_t = \mathbf{K}(L)\boldsymbol{\epsilon}_t = \mathbf{I}\boldsymbol{\epsilon}_t + \mathbf{K}_1 \boldsymbol{\epsilon}_{t-1} + \ldots$. The matrix $\mathbf{M} = \mathbf{K}(1)\mathbf{K}(1)'$, where $\mathbf{K}(1) = \mathbf{I} + \mathbf{K}_1 + \mathbf{K}_2 + \ldots$, is the spectral density matrix of $\Delta \mathbf{x}_t$ at zero frequency. Now if the variance of $\alpha'\mathbf{x}_t$ exists, it will be given by

$$\text{var}(\alpha'\mathbf{x}_t) = \sum_{t=0}^{\infty} \alpha' \mathbf{C}_t \mathbf{V} \mathbf{C}_t' \alpha,$$

where $\mathbf{V}$ is the variance-covariance matrix of $\boldsymbol{\epsilon}_t$ and $\mathbf{C}_t = \mathbf{I} + \mathbf{K}_1 + \ldots + \mathbf{K}_t$. Ignoring the degenerate case in which $\mathbf{V}$ is singular, the summation above converges only if $\alpha'\mathbf{C}_t$ converges to zero. But the limit of $\mathbf{C}_t$ as $t \to \infty$ is $\mathbf{K}(1)$, so for convergence we must have $\alpha'\mathbf{K}(1) = 0$, which requires $\mathbf{K}(1)$, and hence $\mathbf{M}$, to be singular.

It follows from this that if an economic theory imposes cointegration on a set of nonstationary variables, simple first differencing of all the variables can lead to econometric problems. Noninvertibility of the VMA destroys the usual argument for using a finite VAR representation, that a finite VAR can approximate the true VMA arbitrarily well. Intuitively, the problem arises because a cointegrated system has fewer unit roots than variables, so first differencing all the variables amounts to overdifferencing the system.[13]

Fortunately, there is a simple solution to the difficulty, which is to include $\alpha'\mathbf{x}_t$ in a VAR along with a subset of the elements of $\Delta \mathbf{x}_t$. An equation that relates the change in an element of $\mathbf{x}_t$ to its own lags and lags of $\alpha'\mathbf{x}_t$ is called an error-correction model for that element of $\mathbf{x}_t$. The VAR proposed in the previous section to test present value models is an error-correction model for y_t, along with an equation describing the evolution of $\alpha'\mathbf{x}_t$.

---

[13] Shiller (1981*b*) and Melino (1983) criticized Sargent (1979) on this ground (and on the ground that he failed to test the implications of the model for the relative levels of y_t and Y_t). Baillie, Lippens, and McMahon (1983) also overdifferenced their system. Hansen and Sargent (1981*a*) corrected the problems with Sargent's procedure.

The second major result from the theory of cointegration concerns the "cointegrating vector" α. In a present value model, α is unique up to a scalar normalization and is proportional to $(-\theta \ 1)'$. Stock (1987) and Phillips and Ouliaris (1986) prove that a variety of methods provide estimates that converge to the true parameter at a rate proportional to the sample size T (rather than $\sqrt{T}$ as in ordinary cases). The reason for this is that, asymptotically, all linear combinations of the elements of $\mathbf{x}_t$ other than $\alpha'\mathbf{x}_t$ have infinite variance.

The practical implication is that an unknown element of α may be estimated in a first-stage regression and then treated as known in second-stage procedures, whose asymptotic standard errors will still be correct. This is extremely useful in carrying out the VAR tests of the previous section. In the case of stock prices, for example, the present value model constrains $\theta = \delta/(1 - \delta)$, so one can estimate the discount factor from a preliminary regression and then treat it as known in testing the model.

Two types of preliminary regression have been proposed for estimating the unknown parameter θ. The first, called the cointegrating regression by Engle and Granger (1987), is just a regression of Y_t on y_t. The second is an "error-correction" regression of Δy_t or ΔY_t on lagged changes in and levels of y_t and Y_t. In the first case, one estimates θ as the coefficient on y_t, while in the second case one takes the ratio of the coefficient on lagged y_t to that on lagged Y_t.

One might argue that use of the error-correction regression is preferable because it accounts more fully for the short-run dynamics of Y_t and y_t. However, it has an important disadvantage. For any cointegrated vector with two elements, there are two possible error-correction regressions, one for Δy_t and one for ΔY_t. Cointegration alone does not rule out that, in one of these regressions, lagged Y_t and y_t have zero coefficients in the population, so that the coefficient ratio fails to identify the desired parameter.[14] Of course, under the present value model the error-correction equation for Δy_t has nonzero coefficients (because $\alpha'\mathbf{x}_t$ Granger-causes Δy_t), but this is not implied by all plausible alternatives. Accordingly, we rely primarily on the cointegrating regression to identify θ.

One may want to conduct a formal statistical test of the null hypothesis that $\mathbf{x}_t$ is not cointegrated. This turns out to pose some difficult statistical problems. If a candidate for the cointegrating vector α is available, the null hypothesis is that $\alpha'\mathbf{x}_t$ is nonstationary, and one can use a modified Dickey-Fuller (1981) test, regressing the change in $\alpha'\mathbf{x}_t$ on a constant and a single lagged level. The t-statistics

---

[14] Cointegration does rule out that the coefficients are zero in both error-correction regressions.

and F-statistic are corrected for serial correlation in the equation residual as proposed by Phillips and Perron (1986) and Phillips (1987) and then compared with significance levels computed numerically by Dickey and Fuller. If the statistics are sufficiently high, the null hypothesis is rejected.

If the cointegrating vector is not known but must be estimated from a cointegrating regression, the Dickey-Fuller significance levels are no longer appropriate. Engle and Granger (1987) analyze a variety of tests that use the residual from the cointegrating regression, an estimate of $\alpha' \mathbf{x}_t$. We report two of their test statistics, one based on the Dickey-Fuller regression and one that augments that regression with four lagged dependent variables. Engle and Granger provide significance levels for these tests, based on a Monte Carlo study.[15]

Phillips and Ouliaris (1986) propose an alternative test procedure for the null hypothesis of no cointegration. Their method involves computing the matrix $\mathbf{M}$, the spectral density matrix at zero frequency, nonparametrically. As discussed above, this matrix will be nonsingular under the null and singular under the alternative of cointegration. Unlike the Engle-Granger procedures, their test statistics have a distribution that is asymptotically free of nuisance parameters. They applied their methods to our data, and we note their results below.

III. Testing the Model in Bond and Stock Markets

In this section we apply the methods developed above to test present value models for bonds and stocks. The model for bonds, usually referred to as the "expectations theory of the term structure," is a special case of equation (1) in which the parameters θ and δ are known a priori (θ equals one, and δ is a parameter of linearization), while the constant c is a liquidity premium unrestricted by the model.[16]

We test the present value model for bonds on a monthly U.S. Treasury 20-year yield series, available from 1959 to 1983 from Salomon Brothers' *Analytical Record of Yields and Yield Spreads*. The short rate used is a 1-month Treasury bill rate, obtained from the *Treasury Bulletin* for dates prior to 1982 and from the *Wall Street Journal* thereafter.[17] These data were previously studied in Campbell and Shiller

[15] The Monte Carlo results are based on 10,000 replications of 100 observations of independent random walks, with four lagged residual changes included in the test.

[16] The linearization required to write the expectations theory in this form is explained in Shiller (1979) and Shiller et al. (1983).

[17] The *Treasury Bulletin* and *Wall Street Journal* data are consistent with one another at dates when they are both available.

(1984); Shiller et al. (1983) worked with very similar data. We present empirical results both for the full sample 1959:1–1983:10 and for a short sample ending in 1978:8, which is more likely to correspond to a single interest rate regime.[18]

The present value model for stocks is a special case of equation (1) in which θ is known to equal $\delta/(1-\delta)$. The model restricts the constant c to be zero. The discount factor δ is not known a priori but can be inferred by estimating the cointegrating vector for stock prices and dividends; a consistent estimate is also provided by the sample mean return on stocks.[19]

One difficulty with this formulation for stocks is that Y_t and y_t are not measured contemporaneously. The term Y_t is a beginning-of-period stock price, and y_t is paid sometime within period t. Literal application of the methods outlined in Section I would require us to assume that y_t is known to the market at the start of period t; but, as pointed out by West (1987) and others, this might lead us to a spurious rejection of the model if in fact y_t is known only at the start of period $t+1$. Intuitively, it is not hard to "predict" excess returns using ex post information. In order to avoid this problem, we modify the procedures of Section I by constructing a variable $SL_t \equiv Y_t - \theta y_{t-1}$. We use this variable in our tests and alter the cross-equation restrictions appropriately. The dependent variables in the VAR are now SL_t and Δy_{t-1}, both of which are in the information set at the start of time t but not at the start of time $t-1$ under our conservative assumption about the market's information.[20] Since $SL_t = S_t + \theta \Delta y_t$, it is of course stationary if S_t and Δy_t are.

We tested the model for stocks using time-series data for real annual prices and dividends on a broad stock index from 1871 to 1986. The term Y_t is the Standard and Poor's composite stock price index for January, divided by the January producer price index scaled so that the 1967 producer price index equals 100. (Before 1900 an annual average producer price index was used.) The nominal dividend series is, starting in 1926, dividends per share adjusted to index, four-quarter total, for the Standard and Poor's composite index. The nominal dividend before 1926 was taken from Cowles (1939), who ex-

---

[18] For both samples, the parameter of linearization δ is set equal to $1/(1+R)$, with R at 0.0587/12 (the mean 20-year bond rate in the short sample, expressed at a monthly rate). Our subsequent empirical results are conditional on a fixed value of δ.

[19] The sample mean return converges to the population mean only at rate $\sqrt{T}$ and therefore should not strictly be taken as known in second-stage procedures. However, we ignore this problem in our empirical work.

[20] Engle and Watson (1985) did some regressions similar to ours, using a similar data set on stock prices and dividends. They used the variable S_t rather than SL_t. Their results differ from ours in that they found no evidence of Granger causality from S_t to Δy_t, but they did not reject the present value model more strongly than we do.

TABLE 1

Unit Root Tests (Test Statistic $Zt\alpha$)

Variable	With Trend	Without Trend
	A. In the Term Structure	
Sample 1959–78:		
y_t	-2.78	-1.72
Y_t	-2.76	$-.46$
Δy_t	-17.40 (1%)	-17.44 (1%)
ΔY_t	-15.30 (1%)	-15.32 (1%)
S_t	-3.15 (10%)	-3.08 (5%)
ξ_t	-15.22 (1%)	-15.25 (1%)
Sample 1959–83:		
y_t	-3.83 (2.5%)	-2.32
Y_t	-2.51	$-.50$
Δy_t	-17.05 (1%)	-17.08 (1%)
ΔY_t	-15.27 (1%)	-15.29 (1%)
S_t	-4.77 (1%)	-4.67 (1%)
ξ_t	-15.18 (1%)	-15.19 (1%)
	B. In the Stock Market	
y_t	-2.88	-1.28
Y_t	-2.19	-1.53
Δy_t	-8.40 (1%)	-8.44 (1%)
ΔY_t	-9.91 (1%)	-9.96 (1%)
$\theta = 31.092$:		
SL_t	-4.35 (1%)	-4.31 (1%)
ξ_t	-9.93 (1%)	-9.99 (1%)
$\theta = 12.195$:		
SL_t	-2.68	-2.15
ξ_t	-9.76 (1%)	-9.69 (1%)

Note.—Test statistics for a variable X_t are based on the t-statistics on α in the regression $\Delta X_t = \mu + \beta t + \alpha X_{t-1}$ (with trend) or the regression $\Delta X_t = \mu + \alpha X_{t-1}$ (without trend) The t-statistic is corrected for serial correlation in the equation residual in the manner proposed by Phillips and Perron (1986) and Phillips (1987) Significance levels are with trend 10%, -3.12; 5%, -3.41, 2.5%, -3.66; 1%, -3.96; without trend 10%, -2.57, 5%, -2.86; 2.5%, -3.12, 1%, -3.43

tended the Standard and Poor's series back in time.[21] Finally, y_t is the nominal dividend series, divided by the annual average producer price index scaled so that the 1967 producer price index equals 100.

As shown in table 1, parts A and B, we ran unit root tests on our raw data and the various linear combinations discussed in Section I. This is an important preliminary because our approach is appropriate only if y_t is integrated of order one. We present test statistics that are based on the t-statistic on the lagged level in a Dickey-Fuller regres-

[21] The dividend data differ slightly from those used in Shiller (1981b), Mankiw et al. (1985), West (1987), and others. It has recently come to our attention that the second (1939) edition of Cowles's book contains some corrections to the dividend series presented in the original 1938 edition, and these corrections have been incorporated here.

sion, corrected for fourth-order serial correlation as proposed by Phillips and Perron (1986) and Phillips (1987).[22] We ran the Dickey-Fuller regression with and without a time trend; the former is appropriate when the alternative hypothesis is that the series is stationary around a trend, the latter when the alternative is that the series is stationary around a fixed mean.

The results in part A of table 1 are generally supportive of the view that short- and long-term interest rates are cointegrated, with the cointegrating vector equal to $(-1\ 1)$ as implied by the expectations theory. Over the short sample 1959–78, one cannot reject the hypothesis that short and long rates have a unit root at even the 10 percent level; however, there is strong evidence that *changes* in interest rates are stationary. The hypothesis that the long-short spread has a unit root is rejected at the 10 percent level when a trend is estimated and at the 5 percent level when the trend is excluded from the regression. Finally, the excess return ξ_t also appears stationary; this, together with the results for Δy_t and ΔY_t, is indirect evidence for stationarity of the spread because of the linear dependence discussed in Section I.

Results are fairly similar over the full sample 1959–83. There is even stronger evidence that the spread is stationary, and the unit root hypothesis for short rates can be rejected unless a trend in interest rates is ruled out on a priori grounds.[23]

In part B of the table, we repeated these tests for the stock market data. Once again y_t and Y_t appear to be integrated of order one. In the stock market, the parameter θ is not determined by the present value model as it is in the term structure. Therefore, we must compute SL_t and ξ_t using estimates of θ obtained from the data. Strictly speaking, this invalidates the Phillips-Perron tests for SL_t and ξ_t, but we report the statistics as data description.

Table 2 gives details of alternative estimation procedures for θ. The cointegrating regression estimates θ at 31.092; the corresponding real discount rate (the reciprocal of θ) is 3.2 percent, which is lower than the average dividend-price ratio and considerably lower than the sample mean return of 8.2 percent.[24] The error-correction regression

[22] The results are qualitatively unchanged by looking at other statistics from the Dickey-Fuller regression or by varying the order of the serial correlation correction between one and 10.

[23] The results in table 1, pt. A, are more favorable to the hypothesis of cointegration between long and short rates than are the results reported by Phillips and Ouliaris (1986). They reject the null hypothesis of no cointegration at only the 15 percent level (their table 6). However, their procedure does not impose the cointegrating vector a priori, and this may involve a loss of power.

[24] The estimate of θ that corresponds to the sample mean return is 12.195. The higher estimate in the cointegrating regression is associated with a negative constant

TABLE 2

ESTIMATION OF THE COINTEGRATING VECTOR AND TEST FOR COINTEGRATION
IN THE STOCK MARKET

	R^2	Estimate of θ	Implied Discount Rate
1. $Y_t = -12.979 + 31.092y_t$	.842	31.092	3.2%
2 $\Delta y_t = 101 + .165\Delta y_{t-1} + .010\Delta Y_t$			
$-.157y_{t-1} + .004Y_t$	.373	37.021	2.7%
3. Sample mean return $= 8.2\%$	...	12.195	8.2%

Tests of no cointegration Engle and Granger (1987) ξ_2 statistic for eq (1) residual, 3 58, significance levels: 10%, 3 03, 5%, 3 37, 1%, 4 07 Engle and Granger (1987) ξ_3 statistic for eq (1) residual, 2 64 , significance levels: 10%, 2 84, 5%, 3 17, 1%, 3 77

delivers a fairly similar estimate of θ, 37.021 with an implied real discount rate of 2.7 percent. We proceed to construct SL_t using discount rates of 8.2 percent and 3.2 percent as a check on the robustness of our methods.

Engle and Granger's tests for no cointegration, based on the residual from the cointegrating regression, give mixed results: the ξ_2 statistic rejects at the 5 percent level, while the ξ_3 statistic narrowly fails to reject at the 10 percent level. The Phillips-Perron tests in part B of table 1 are also mixed. Both SL_t and ξ_t appear to be stationary when the 3.2 percent discount rate is used, but at an 8.2 percent discount rate the tests fail to reject the unit root null for SL_t even though they reject for Δy_t, ΔY_t, and ξ_t. There seems to be some evidence for cointegration between stock prices and dividends, but it is weaker than the evidence for cointegration in the term structure.[25]

The results in table 1 do not suggest that a "rational bubble" is present in the term structure or the stock market since a bubble would cause both ΔY_t and S_t to be nonstationary. Accordingly, we interpret the test statistics below in terms of predictable excess returns.

In table 3, part A, we report summary statistics for a VAR test of the expectations theory of the term structure. The VAR includes Δy_t

term; under the present value model, the constant should be proportional to the unconditional mean change in dividends, so it should be positive rather than negative. An estimated discount rate lower than the mean dividend-price ratio is consistent with the model only if dividends are expected to decline through time, the historical rise being due to sampling error.

[25] Phillips and Ouliaris (1986) did not reject the null hypothesis of no cointegration between stock prices and dividends at even the 25 percent level (their table 6). Campbell and Shiller (1986) report unit root tests for log dividends, log prices, and the log dividend-price ratio. There is some evidence for trend stationarity of log dividends, no evidence against the unit root null for log prices, and strong evidence for stationarity of the dividend-price ratio.

TABLE 3

TESTS OF PRESENT VALUE MODEL

A. In the Term Structure

Sample 1959–78:
  Akaike criterion selects 11-lag VAR
  Δy equation $R^2 = .216$; S Granger-causes Δy at 0.01% level
  S equation $R^2 = .877$; Δy Granger-causes S at 0.3% level
  Test of present value model: $\chi^2(22) = 83.02$; P-value $< 0.005\%$
  Summary statistics:

$E(\Delta y)$	=	.016	$\sigma(S)$	=	1.060
$E(S)$	=	1.144	var(S)/var(S')	=	.987 (.360)
$E(S')$	=	.016	corr(S, S')	=	.978 (.011)
$\sigma(\Delta y)$	=	.442	var(ξ)/var(ξ')	=	1.160 (1.146)

Sample 1959–83:
  Akaike criterion selects six-lag VAR
  Δy equation $R^2 = .171$; S Granger-causes Δy at 0.3% level
  S equation $R^2 = .772$; Δy Granger-causes S at 1.3% level
  Test of present value model: $\chi^2(12) = 35.63$; P-value $= 0.03\%$
  Summary statistics:

$E(\Delta y)$	=	.021	$\sigma(S)$	=	1.320
$E(S)$	=	1.138	var(S)/var(S')	=	3.394 (3.948)
$E(S')$	=	.021	corr(S, S')	=	.956 (.098)
$\sigma(\Delta y)$	=	.793	var(ξ)/var(ξ')	=	.502 (.506)

B. In the Stock Market

Sample 1871–1986:
$\theta = 12.195$ (8.2% discount rate): Akaike criterion selects four-lag VAR
  Δy equation $R^2 = .400$; SL Granger-causes Δy at $< 0.001\%$ level
  SL equation $R^2 = .837$; Δy Granger-causes SL at 63.3% level
  Test of present value model with mean restriction: $\chi^2(9) = 15.74$; P-value $= 7.2\%$
  Test of present value model without mean restriction: $\chi^2(8) = 15.72$; P-value $= 4.7\%$
  Summary statistics:

$E(\Delta y)$	=	.017	$\sigma(SL)$	=	15.51
$E(SL)$	=	16.07	var(SL)/var(SL')	=	67.22 (86.04)
$E(SL')$	=	2.563	corr(SL, SL')	=	−.459 (.801)
$\sigma(\Delta y)$	=	.168	var(ξ)/var(ξ')	=	11.27 (4.49)

$\theta = 31.092$ (3.2% discount rate): Akaike criterion selects two-lag VAR
  Δy equation $R^2 = .378$; SL Granger-causes Δy at $< 0.001\%$ level
  SL equation $R^2 = .516$; Δy Granger-causes SL at 1.8% level
  Test of present value model with mean restriction: $\chi^2(5) = 14.90$; P-value $= 1.1\%$
  Test of present value model without mean restriction: $\chi^2(4) = 5.75$; P-value $= 21.8\%$
  Summary statistics:

$E(\Delta y)$	=	.017	$\sigma(SL)$	=	9.937
$E(SL)$	=	−12.52	var(SL)/var(SL')	=	4.786 (5.380)
$E(SL')$	=	16.66	corr(SL, SL')	=	.911 (.207)
$\sigma(\Delta y)$	=	.167	var(ξ)/var(ξ')	=	1.414 (.441)

and S_t as variables, and the number of lags is chosen by the Akaike information criterion (AIC).[26] White's (1984) heteroscedasticity-consistent covariance matrix estimator is used in constructing standard errors and test statistics. The VARs are estimated for the short sample 1959–78 and the full sample 1959–83; they have 11 and six lags, respectively.

In both sample periods the lagged variables have a fair degree of explanatory power for the change in short rates. The R^2 for the Δy_t equation is 21.6 percent in the short sample and 17.1 percent in the full sample. This argues against the view of Mankiw and Miron (1986) that short-rate changes are essentially unpredictable in the postwar period in the United States. Furthermore, there is strong evidence that spreads Granger-cause short-rate changes, as they should do if the expectations theory is true. The hypothesis of no Granger causality can be rejected at the 0.01 percent level for the short sample and the 0.3 percent level for the full sample.

A formal test of the expectations theory restrictions in equation (8) rejects very strongly. The null that excess returns on long bonds are unpredictable can be rejected at less than the 0.005 percent level in the short sample and at the 0.03 percent level in the full sample. The R^2 values for excess returns are 26.3 percent and 16.7 percent, respectively.[27] In the corresponding regression (4), which has the change in the long rate as its dependent variable, the coefficient on the spread has the wrong sign (-0.020 in the short sample and -0.039 in the full sample).[28]

Despite these negative results, the summary statistics in table 3, part A, suggest that there is an important element of truth to the expectations theory of the term structure. The spread does seem to move very closely with the theoretical spread, the unrestricted forecast of the present value of future short-rate changes. In both sample periods the variance of the spread is insignificantly different from the variance of the theoretical spread (i.e., our "levels variance ratio" does not reject), and the two variables have similar innovation variances and an extremely high correlation. In the 1959–78 period the correlation between the actual and theoretical spreads is 0.978 with a standard error of 0.011, while in the 1959–83 period it is 0.956 with a

---

[26] That is, we pick the number of lags to minimize ($-\ln$ likelihood + number of parameters) in the VAR. Sawa (1978) has argued that the AIC tends to choose models of higher order than the true model but states that the bias is negligible when $p < T/10$, as it is here. The test statistics in table 3 are not highly sensitive to small changes in the lag length of the VAR system.

[27] Nonlinear Wald tests of eq. (7) reject at significance levels of less than 0.005 percent in the short sample and 8.4 percent in the full sample.

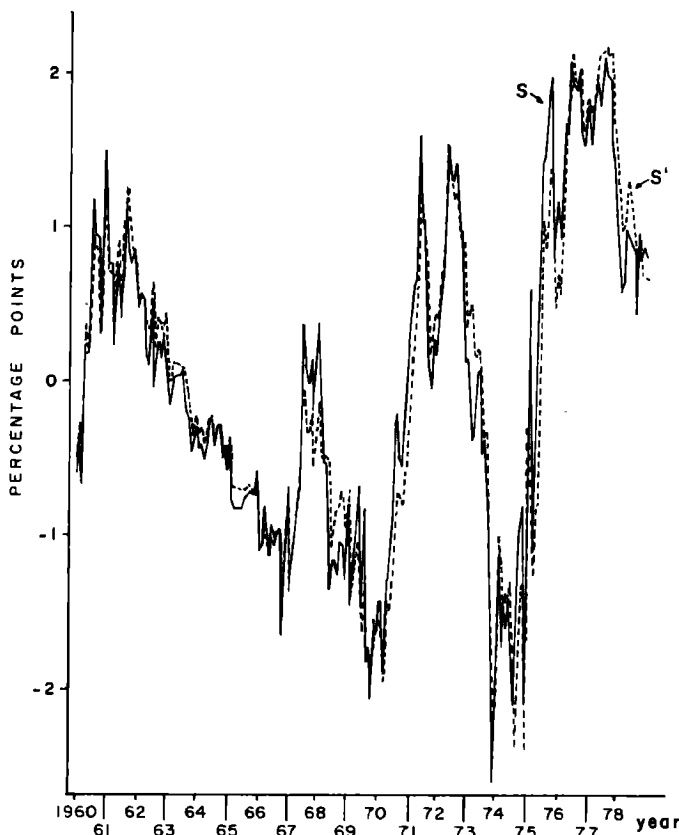[28] This is consistent with the results of Shiller et al. (1983).

FIG. 1.—Term structure: deviations from means of long-short spread S_t and theoretical spread S'_t.

standard error of 0.098. Figure 1 illustrates the comovement of S_t and S'_t in the short sample.[29]

What this suggests is that tests of predictability of returns are highly sensitive to deviations from the expectations theory—so sensitive, in fact, that they may obscure some of the merits of the theory. An example illustrates the point. Suppose long and short rates differ from the expectations theory in the following manner: $S_t = S'_t + w_t$,

[29] The high correlation of these variables in postwar U.S. data might also have been inferred from results in Modigliani and Shiller (1973) (see particularly their fig. 6). Despite the evidence reported in Modigliani and Shiller and in the present paper, one of us (Shiller 1979) presented evidence suggesting that long-term interest rates are too volatile to accord with the expectations theory. By contrast with Modigliani and Shiller and the present paper, Shiller (1979) assumed that *levels* of short rates are stationary, an assumption more clearly appropriate for prewar data sets.

where w_t is serially uncorrelated noise. As Campbell and Shiller (1984) point out, excess bond returns will be predicted by S_t, and a regression of ΔY_{t+1} on S_t may find that the coefficient has the opposite sign from that predicted by (4), even if the variance of w_t is quite small. However, a regression of S_t^* on S_t will find that the coefficient has the same sign as predicted by (3), and downward bias caused by w_t will be small if the variance of w_t is small. Moreover, the variance ratios $\text{var}(S_t)/\text{var}(S_t')$ and $\text{var}(\xi_t)/\text{var}(\xi_t')$ may not be much greater than one. In this example the spread predicts short-rate movements almost correctly, even though it badly misforecasts long-rate movements. Deviations from the present value model are transitory rather than persistent, so the metric of equation (10) reveals the strengths of the expectations theory that are obscured by the metric of equation (11).[30]

In part B of table 3, we repeated the exercises above for stock prices and dividends. We worked with one sample period but two discount rates. The Akaike criterion selected a four-lag representation for the data when the sample mean discount rate 8.2 percent was used and a two-lag representation when the cointegrating regression discount rate 3.2 percent was used.

The VAR estimates suggest that dividend changes are rather highly predictable; the R^2 values for the equations that explain them are around 40 percent. There is very strong evidence that price-dividend spreads Granger-cause dividend changes, which is what one would expect if there is any truth to the present value model for stock prices.

We conducted two formal tests of the model. The first restricted the mean of the price-dividend difference, while the second left the mean unconstrained and restricted only the dynamics of the variable. (In the case of the term structure, the mean spread is always unconstrained because we allowed a constant risk premium.)

The results of these tests include some statistical rejections at conventional significance levels, but they are not nearly as strong as the rejections in the term structure. The pattern of results is sensitive to the choice of discount rate. When the sample mean return is used, the mean restriction on SL_t is satisfied almost exactly. Therefore, the test of only the dynamic restrictions in equation (8) rejects more strongly, at the 4.7 percent level as compared with the 7.2 percent level for the full set of restrictions. When the discount rate from the cointegrating regression is used, the complete set of restrictions is rejected at the 1.1 percent level while the significance level for the dynamic restrictions is

---

[30] We do not claim that this example is literally correct for our data. The model $S = S' + w$ can be tested, for any MA(q) process for w, by regressing ξ on information known $q + 2$ periods earlier. We found that this test rejected the model for q up to 8 using the bond data for 1959–78.

only 21.8 percent.[31] For both discount rates, a regression of ΔY_{t+1} on SL_t gives a coefficient estimate with a negative sign rather than the positive sign implied by the present value model.[32]

These tests are "portmanteau" tests of the present value model against an unspecified alternative. We also present variance ratios in order to test against the specific alternative that stock prices "move too much" in levels or innovations. The point estimate of the levels variance ratio $\text{var}(SL_t)/\text{var}(SL_t')$ is dramatically different from unity, at 67.22, when the sample mean discount rate is used. Not surprisingly, the variance ratio is smaller when future dividend changes are discounted at the lower rate estimated by the cointegrating regression, but it is still considerable at 4.79. However, the asymptotic standard errors on these ratios are huge, and one cannot reject the hypothesis that both of them equal unity.

The innovations variance ratios $\text{var}(\xi_t)/\text{var}(\xi_t')$ are also estimated larger than unity, and here the standard errors are less extreme. In the sample mean discount rate case, one can reject at the 5 percent level the hypothesis that the innovation variance ratio is unity; it is estimated to be 11.27, with a standard error of 4.49. With the lower discount rate, the ratio is estimated at 1.41, with a standard error of 0.44.

Plots of the price-dividend difference and the unrestricted VAR forecast of dividend changes give a visual image of these variance results. At an 8.2 percent discount rate (fig. 2), SL_t and SL_t' are negatively correlated (but there is a very large standard error on the correlation) and the excess volatility of the spread is very dramatic. At a 3.2 percent discount rate (fig. 3), SL_t and SL_t' have a correlation of 0.911 (with standard error 0.207) and the excess volatility is much less dramatic.[33]

To compare our results on volatility with results using earlier methods, we also computed sample values of S_t^* using the terminal condition $S_T^* = S_T$, where T is the last observation in our sample. We computed SL_t^* analogously. Equation (3) implies $\sigma(S_t^*) > \sigma(S_t)$ and $\sigma(SL_t^*) > \sigma(SL_t)$. For the bond data in the period 1959–78, $\sigma(S_t^*) = 1.217$, while $\sigma(S_t) = 1.060$, so the inequality is satisfied. For the stock

---

[31] Nonlinear Wald tests of the dynamic restrictions in the form (7), rather than (8), reject at less than the 0.005 percent level for the 8.2 percent discount rate and at the 7.3 percent level for the 3.2 percent discount rate.

[32] The coefficient is -0.064 for the 8.2 percent discount rate and -0.079 for the 3.2 percent discount rate.

[33] It should be emphasized that excess volatility of the spread SL_t is not quite the same as the excess volatility discussed in Shiller (1981b). That analysis suggested that stock prices should very nearly follow a trend. If that were in fact what was observed, the spread SL_t would be quite volatile because of dividend movements.
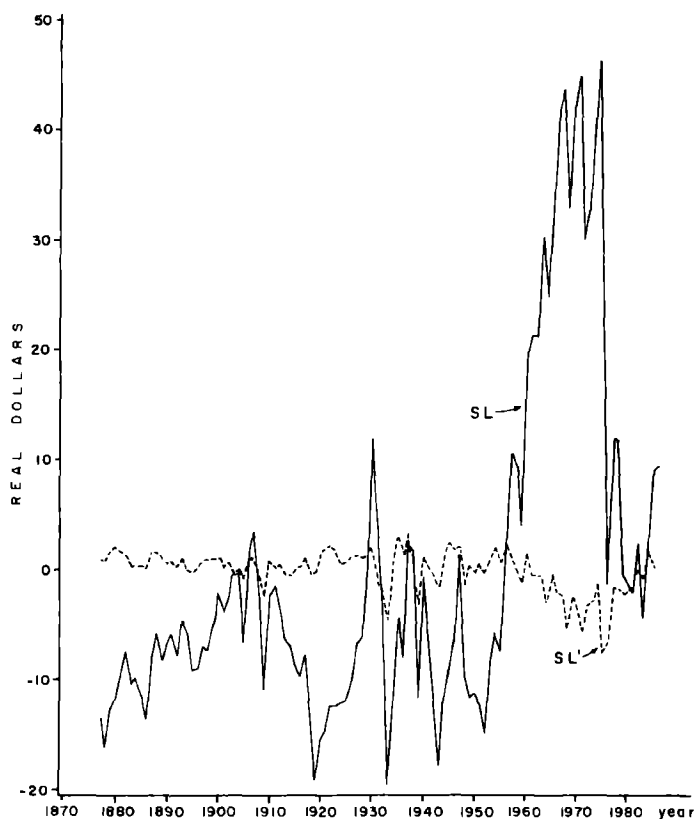
FIG. 2.—Stock market: deviations from means of actual spread (SL_t = Price$_t$ − θ · Dividend$_{t-1}$) and theoretical spread SL_t', θ = 12.195.

data at an 8.2 percent discount rate, $\sigma(SL_t^*)$ = 7.928, while $\sigma(SL_t)$ = 15.506, so the inequality is sharply violated. The inequality is again satisfied by the stock data at a 3.2 percent discount rate, where $\sigma(SL_t^*)$ = 12.888 and $\sigma(SL_t)$ = 9.937.

Following Scott (1985), we also regressed S_t^* on S_t and a constant. If the present value model is true, the coefficient on S_t should be one. The same holds for the corresponding regression with SL_t^* and SL_t. For bonds in 1959–78, we estimated the coefficient at 0.81; for stocks at an 8.2 percent discount rate we estimated it at 0.16, while for stocks at a 3.2 percent discount rate we estimated it at 0.02. Thus the results using S_t^* and SL_t^* generally support the conclusion that the present value model for bonds fits the data comparatively well, whereas the model for stocks has a poor fit even though it cannot be rejected statistically at high levels of confidence.

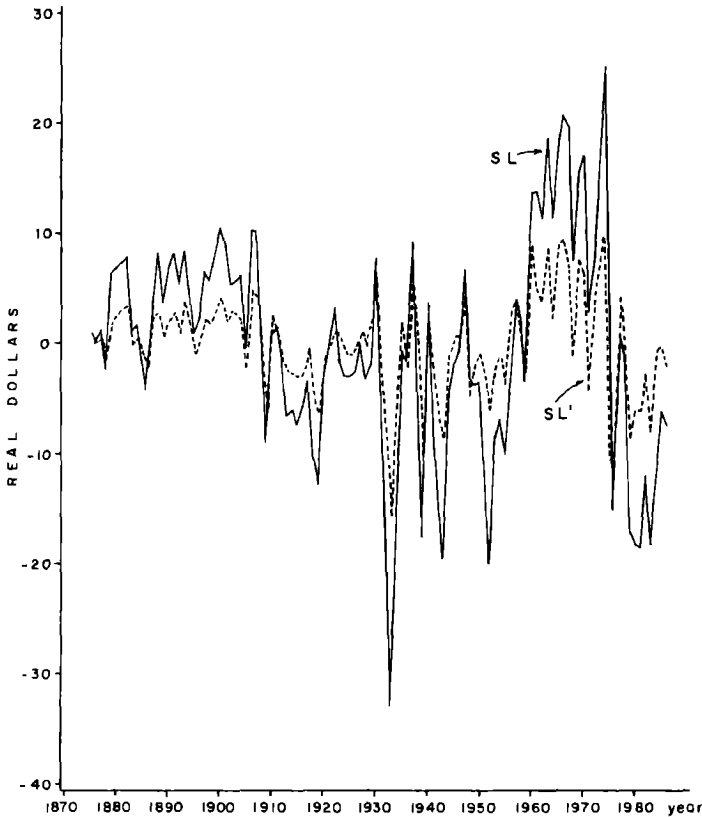We close with a caveat about the plots and summary statistics gener-

FIG. 3.—Stock market: deviations from means of actual spread (SL_t = Price$_t$ − θ · Dividend$_{t-1}$) and theoretical spread SL'_t, θ = 31.092.

ated by the VAR system. The VAR simulation method may be misleading if the wrong value of θ is chosen so that the spread variable is nonstationary. For example, if θ is chosen too large, the movements of S_t are dominated by the movements of $-\theta y_t$. The VAR results are then approximately those one would get if one regressed Δy_t and $-\theta y_t$ on lagged values of these variables. It is well known that in finite samples estimates of autoregressive parameters for nonstationary variables are biased downward, and this problem will afflict the VAR if θ is too large.

In a simple case in which y_t follows an AR(1) process with a unit root and the VAR includes one lag only, one can show that the estimated VAR companion matrix will have first column zero and second column $((1 - \rho)/\theta \quad \rho)'$, where ρ is a downward-biased estimate of the unit root. This companion matrix satisfies the restrictions of equation

(9) almost exactly, whatever the behavior of the variable Y_t. A symptom of this misspecification would be that mean returns would not obey the model, even though the dynamics of returns would appear to satisfy the restrictions.

It is possible that a problem of this sort affects our results for the stock market when we use a low 3.2 percent discount rate corresponding to a high θ of 31.092. The cointegrating regression that generates this θ estimate—a regression of the level of Y on the level of y—is dominated by the enormous postwar hump in stock prices. Since this hump coincided with a much milder hump in real dividends, the regression estimates a coefficient for y that is much larger than the historical average price-dividend ratio. The negative intercept prevents the fitted value from overpredicting Y over the sample period as a whole. As a result, over the bulk of the sample period, the spread SL_t is distinctly negatively correlated with the lagged dividend.[34] The VAR estimates place considerable weight on this earlier part of the sample period because the dividend equation is specified in terms of dividend changes that are more variable before 1946. Thus the high correlation of SL_t and SL'_t may be to some extent spurious.

This view is supported by the results from regressing SL_t^* on SL_t. This is a levels regression that is dominated by the postwar hump in stock prices, and here we find the coefficient to be essentially zero rather than one as required by the model. Further support comes from the fact that we strongly reject the implications of the model for the mean of the data when we impose a 3.2 percent discount rate.

IV. Conclusion

In this paper we have shown how a present value model may be tested when the variables of the model, y_t and Y_t, follow linear stochastic processes that are stationary in first differences rather than in levels. If the present value model is true, a linear combination of the variables—which we call the spread—is stationary. Thus y_t and Y_t are cointegrated. The model implies that the spread is linear in the optimal forecast of the one-period change in Y_t and also in the optimal forecast of the present value of all future changes in y_t. We have shown how to conduct formal Wald tests of these implications.

We have also proposed an informal method for evaluating the "fit" of a present value model. A VAR is used to construct an optimal unrestricted forecast of the present value of future y_t changes, and this is compared with the spread. If the model is true, the unrestricted

---

[34] Over the period 1871–1946, the spread has a correlation of -0.7 with the lagged dividend when θ is set equal to 31.092.

forecast or "theoretical spread" should equal the actual spread. We computed the variances and correlation of the two variables and plotted their historical movements.

We applied our methods to the controversial present value models for stocks and bonds. We found that both models can be rejected statistically at conventional significance levels, with much stronger evidence for bonds. However, in our data set, the spread between long- and short-term interest rates seems to move quite closely with the unrestricted forecast of the present value of future short-rate changes. This can be interpreted as evidence that deviations from the present value model for bonds are transitory. In contrast, our evaluation of the present value model for stocks indicates that the spread between stock prices and dividends moves too much and that deviations from the present value model are quite persistent, although the strength of the evidence for this depends sensitively on the discount rate assumed in the test.

References

Baillie, Richard T.; Lippens, Robert E.; and McMahon, Patrick C. "Testing Rational Expectations and Efficiency in the Foreign Exchange Market." *Econometrica* 51 (May 1983): 553–63.
Blanchard, Olivier J., and Watson, Mark W. "Bubbles, Rational Expectations and Financial Markets." In *Crises in the Economic and Financial Structure: Bubbles, Bursts and Shocks,* edited by Paul Wachtel. Lexington, Mass.: Lexington, 1982.
Campbell, John Y. "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis." *Econometrica* (1987), in press.
Campbell, John Y., and Shiller, Robert J. "A Simple Account of the Behavior of Long-Term Interest Rates." *A.E.R. Papers and Proc.* 74 (May 1984): 44–48.
————. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." Working Paper no. 2100. Cambridge, Mass.: NBER, 1986.
Cowles, Alfred. *Common-Stock Indexes.* 2d ed. Bloomington, Ind.: Principia, 1939.
Diba, Behzad T., and Grossman, Herschel I. "Rational Bubbles in the Price of Gold." Working Paper no. 1300. Cambridge, Mass.: NBER, 1984.
Dickey, David A., and Fuller, Wayne A. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49 (July 1981): 1057–72.
Engle, Robert F., and Granger, Clive W. J. "Cointegration and Error-Correction: Representation, Estimation and Testing." *Econometrica* 55 (March 1987): 251–76.
Engle, Robert F., and Watson, Mark W. "Applications of Kalman Filtering in Econometrics." Paper presented at the World Congress of the Econometric Society, Cambridge, Mass., August 1985.

Fama, Eugene F. "Efficient Capital Markets: A Review of Theory and Empirical Work." *J. Finance* 25 (May 1970): 383–417.

Gregory, Allan W., and Veall, Michael R. "Formulating Wald Tests of Nonlinear Restrictions." *Econometrica* 53 (November 1985): 1465–68.

Hamilton, James D., and Whiteman, Charles H. "The Observable Implications of Self-fulfilling Expectations." *J. Monetary Econ.* 16 (November 1985): 353–73.

Hansen, Lars Peter, and Sargent, Thomas J. "Exact Linear Rational Expectations Models: Specification and Estimation." Staff Report no. 71. Minneapolis: Fed. Reserve Bank, November 1981. (*a*)

———. "Linear Rational Expectations Models for Dynamically Interrelated Variables." In *Rational Expectations and Econometric Practice,* edited by Robert E. Lucas, Jr., and Thomas J. Sargent. Minneapolis: Univ. Minnesota Press, 1981. (*b*)

Kleidon, Allan W. "Variance Bounds Tests and Stock Price Valuation Models." *J.P.E.* 94 (October 1986): 953–1001.

LeRoy, Stephen F., and Porter, Richard D. "The Present-Value Relation: Tests Based on Implied Variance Bounds." *Econometrica* 49 (May 1981): 555–74.

Mankiw, N. Gregory, and Miron, Jeffrey A. "The Changing Behavior of the Term Structure of Interest Rates." *Q.J.E.* 101 (May 1986): 211–28.

Mankiw, N. Gregory; Romer, David; and Shapiro, Matthew D. "An Unbiased Reexamination of Stock Market Volatility." *J. Finance* 40 (July 1985): 677–87.

Mankiw, N. Gregory, and Shapiro, Matthew D. "Trends, Random Walks, and Tests of the Permanent Income Hypothesis." *J. Monetary Econ.* 16 (September 1985): 165–74.

Marsh, Terry A., and Merton, Robert C. "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices." *A.E.R.* 76 (June 1986): 483–98.

Melino, Angelo. "Essays on Estimation and Inference in Linear Rational Expectations Models." Ph.D. dissertation, Harvard Univ., 1983.

Modigliani, Franco, and Shiller, Robert J. "Inflation, Rational Expectations and the Term Structure of Interest Rates." *Economica* 40 (February 1973): 12–43.

Phillips, P. C. B. "Time Series Regression with Unit Roots." *Econometrica* 55 (March 1987): 277–302.

Phillips, P. C. B., and Durlauf, S. N. "Multiple Time Series Regression with Integrated Processes." *Rev. Econ. Studies* 53 (August 1986): 473–95.

Phillips, P. C. B., and Ouliaris, S. "Testing for Cointegration." Discussion Paper no. 809. New Haven, Conn.: Yale Univ., Cowles Found., October 1986.

Phillips, P. C. B., and Perron, P. "Testing for Unit Roots in Time Series Regression." Discussion Paper. New Haven, Conn.: Yale Univ., Cowles Found., June 1986.

Quah, Danny. "Stationary Rational Bubbles in Asset Prices." Manuscript. Cambridge: Massachusetts Inst. Tech., 1986.

Sargent, Thomas J. "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure." *J. Monetary Econ.* 5 (January 1979): 133–43.

Sawa, Takamitsu. "Information Criteria for Discriminating among Alternative Regression Models." *Econometrica* 46 (November 1978): 1273–91.

Scott, Louis O. "The Present Value Model of Stock Prices: Regression Tests and Monte Carlo Results." *Rev. Econ. and Statis.* 67 (November 1985): 599–605.

Shiller, Robert J. "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure." *J.P.E.* 87 (December 1979): 1190–1219.

———. "Alternative Tests of Rational Expectations Models: The Case of the Term Structure." *J. Econometrics* 16 (May 1981): 71–87. (*a*)

———. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *A.E.R.* 71 (June 1981): 421–36. (*b*)

———. "Stock Prices and Social Dynamics." *Brookings Papers Econ. Activity*, no. 2 (1984), pp. 457–98.

———. "Conventional Valuation and the Term Structure of Interest Rates." In *Macroecomonics and Finance: Essays in Honor of Franco Modigliani*, edited by Rudiger Dornbusch, Stanley Fischer, and John Bossons. Cambridge, Mass.: MIT Press, 1987.

Shiller, Robert J.; Campbell, John Y.; and Schoenholtz, Kermit L. "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates." *Brookings Papers Econ. Activity*, no. 1 (1983), pp. 173–217.

Stock, James H. "Asymptotic Properties of Least Squares Estimates of Cointegrating Vectors." *Econometrica* (1987), in press.

West, Kenneth D. "A Specification Test for Speculative Bubbles." Manuscript. Princeton, N.J.: Princeton Univ., 1986.

———. "Dividend Innovations and Stock Price Volatility." *Econometrica* (1987), in press.

White, Halbert. *Asymptotic Theory for Econometricians.* Orlando, Fla.: Academic Press, 1984.