

NBER WORKING PAPER SERIES

INTERPRETING COINTEGRATED MODELS

John Y. Campbell

Robert J. Shiller

Working Paper No. 2568

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 1988

The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Interpreting Cointegrated Models

ABSTRACT

Error-correction models for cointegrated economic variables are commonly interpreted as reflecting partial adjustment of one variable to another. We show that error-correction models may also arise because one variable forecasts another. Reduced-form estimates of error-correction models cannot be used to distinguish these interpretations. In an application, we show that the estimated coefficients in the Marsh-Merton [1987] error-correction model of dividend behavior in the stock market are roughly implied by a near-rational expectations model wherein dividends are persistent and prices are disturbed by some persistent random noise. Their results thus do not demonstrate partial adjustment or "smoothing" by managers, but may reflect little more than the persistence of dividends and the noisiness of prices.

John Y. Campbell  
Woodrow Wilson School  
Princeton University  
Princeton, New Jersey 08544  
609 452-4787

Robert J. Shiller  
Cowles Foundation  
Yale University  
Box 2125 Yale Station  
New Haven, CT 06520  
203 432-3708

## 1. Introduction

The notion of cointegration among variables has introduced a new flexibility into the modelling of economic time series. As defined by Engle and Granger [1987], two variables are cointegrated (of order (1,1)) if each variable individually is stationary in first differences (integrated of order 1), but some linear combination of the variables is stationary in levels (integrated of order 0). The notion of cointegration is a special case of the notion of dynamic aggregation introduced by Aoki [1968,1971]. Many economic variables might plausibly be cointegrated when correctly measured, sometimes in natural or sometimes in log units; examples are consumption and income, short and long term interest rates, and stock prices and dividends<sup>1</sup>.

The concepts of cointegration and error-correction are closely related. An error-correction model for two variables relates the changes in the variables to lagged changes and a lagged linear combination of levels<sup>2</sup>. This type of model was introduced by Phillips [1954] and Sargan [1964] and has been promoted by David Hendry and others in a series of papers (Davidson, Hendry, Srba and Yeo [1978], Hendry and von Ungern-Sternberg [1981], Davidson and Hendry [1981], Hendry and Richard [1983], Hendry [1983,1986]). Engle and Granger [1987], following the work of

---

<sup>1</sup> More generally, a set of variables is cointegrated of order (d,b) if each variable individually is integrated of order d, but at least one linear combination exists which is of order (d-b). Most of the literature focuses on the case d=1, b=1 and we will do the same here.

<sup>2</sup> This definition follows Engle and Granger [1987]. Much of the literature uses a single equation relating the change in one variable to the contemporaneous change in the other, lagged changes in both variables and a lagged linear combination of levels. We discuss the single-equation approach further below.

Granger [1981,1983], show that two variables which are cointegrated of order (1,1) have an error-correction representation. The linear combination of levels which enters the error-correction model is just that combination which is stationary in levels.

There is also a less formal link between cointegration and error-correction. The same kind of story is often used to motivate both these concepts. In the words of Granger [1986], for example,

"At the least sophisticated level of economic theory lies the belief that certain pairs of economic variables should not diverge from each other by too great an extent, at least in the long-run. Thus, such variables may drift apart in the short-run or according to seasonal factors, but if they continue to be too far apart in the long-run, then economic forces, such as a market mechanism or government intervention, will begin to bring them together again".

There is a suggestion here that "economic forces", at least those that we understand, are better at explaining long-run tendencies than short-run wiggles in the series. Economic theory, Granger seems to say, is valid for describing the long-run equilibrium, but random shocks knock the economy away from equilibrium and it moves back only slowly. Why is the adjustment back to equilibrium not instantaneous? He does not say. We suppose that he may intend such things as sticky prices, long-term contracts, costs of adjustment, and other factors that may be difficult to model.

Engle and Granger [1987] introduce some terminology which fits naturally with this account. They describe cointegrated variables as being in equilibrium when the stationary linear combination of their levels is at its unconditional mean (assumed equal to zero for simplicity). Most of the time, this combination of levels is not zero

and the system is out of equilibrium; but because the combination of levels is stationary, there is a tendency for the system to return to equilibrium. Engle and Granger call the stationary combination of levels the "equilibrium error". An error-correction model can then be thought of as a description of the stochastic process by which the economy eliminates or corrects the equilibrium error.

This motivation for error-correction models suggests a world in which economic theory describes the long run rather than the short run, and in which unspecified factors cause the economy to respond slowly to random shocks<sup>3</sup>. The terminology makes it natural to think that the equilibrium error appears in the equations describing changes of economic variables because these variables respond to the error in order to eliminate it. As Davidson and Hendry [1981] put it, there is a "servo-mechanism" which returns the economy to equilibrium.

The purpose of this paper is to make it clear that there is an alternative way in which cointegration may arise, and to illustrate such an alternative interpretation of an error-correction model proposed by Marsh and Merton [1987] to describe the dividend-setting behavior of firms.

Engle and Granger's result, that cointegrated variables have an error-correction representation, can be thought of as a statement about Granger causality: the stationary linear combination of levels must Granger cause the change in at least one of the cointegrated variables. Now it is well

---

<sup>3</sup> Engle and Granger do say that their notion of equilibrium is just "a stationary point characterized by forces which tend to push the economy back toward equilibrium whenever it moves away". Their notion of equilibrium is specific to the error-correction model they define, and has no clear relation to other concepts of equilibrium in economics.

known that Granger causality from a variable  $z_1$  to a variable  $z_2$  can arise for two reasons. The variable  $z_1$  may cause  $z_2$  in the common-language sense, or instead the variable  $z_1$  may anticipate or forecast  $z_2$ . In the former case an intervention which changes the stochastic process for  $z_1$  will change the behavior of  $z_2$ , while in the latter case an intervention which changes the stochastic process for  $z_2$  will change the behavior of  $z_1$ . In Sims' [1977] terminology, while both show causal orderings from  $z_1$  to  $z_2$ , in the former case there is a "structural" ordering from  $z_1$  to  $z_2$ , in the latter case a "structural" ordering from  $z_2$  to  $z_1$ .

The motivation for cointegration given above stresses the idea that the equilibrium error causes changes in the variables of the model. We wish to emphasize instead the possibility that the equilibrium error results from agents' forecasts of these changes.

As a concrete example, consider the term structure of interest rates. If the "expectations theory of the term structure" holds -- that is, if expected returns on bonds of all maturities are equal, up to a constant term premium -- then the yield on a long bond can be written approximately as a present value of expected future short-term interest rates<sup>4</sup>. The yield spread, the difference between the long bond yield and the current short-term rate, can be written as a present value of expected future changes in short-term interest rates. It is also proportional to the expected change in the long-term rate over the next period. Campbell and Shiller [1987a] point out that interest rates must

---

<sup>4</sup> See Shiller [1979], Shiller, Campbell and Schoenholtz [1983] or Shiller [1987].

then be cointegrated if short rates are integrated of order one. The vector of short- and long-term interest rates follows an error-correction model in which the yield spread is the equilibrium error.

In this example it is possible that the yield spread actually does have some causal influence on changes in short rates (presumably by influencing the behavior of the Federal Reserve Board), but this is not necessary for the cointegration result. The example does rely crucially on the idea that market participants have information which is relevant for forecasting short-term interest rates, beyond the history of short rates themselves. If this were not true -- for example, if bond market participants relied on a univariate AR(1) model to forecast short rates - - then the long rate would be a deterministic function of current and lagged short rates, and the error-correction model for the term structure would be trivial. The expectations theory of the term structure explains randomness in the relation between short and long term interest rates, not by allowing random error in the equation relating long rates to expected future short rates, but by superior market information which breaks the deterministic link between current and lagged short rates and expected future short rates<sup>5</sup>.

It is clear from this example that cointegration can arise even in a well-organized market with no adjustment costs, where there is no true causal role for the equilibrium error. The essential features of the example are that one variable (the long-term interest rate) reflects

---

<sup>5</sup> Nickell [1985] discusses some univariate forecasting models. He derives deterministic equations relating agents' forecasts to the history of the variable they are forecasting. Salmon [1982] appeals to random optimization error to break the deterministic relationship.

agents' rational expectations of the future of another variable (the short-term interest rate) which follows an integrated process. Agents have more information about the variable they are trying to forecast than is contained in the history of that variable alone. We believe that these features characterize a wide variety of economic situations.

In the next section of the paper we make these points more formally. We first show how an error-correction model can be rewritten as a vector autoregression (VAR). The VAR form of the model is particularly convenient for analyzing forward-looking behavior. We then show formally how cointegration can arise when agents are forecasting and have rational expectations. In particular, we study the implications of a "present value model" which restricts one variable  $Y_t$  to be linear in the rationally expected present value of future realizations of another variable  $y_t$ . This model implies that  $y_t$  and  $Y_t$  are cointegrated if  $y_t$  is stationary in first differences. It also imposes more specific restrictions on the joint time-series behavior of  $y_t$  and  $Y_t$ .

These restrictions can be tested on the VAR which corresponds to the error-correction model for  $y_t$  and  $Y_t$ . It is also possible to test restrictions on single equations. We note, however, that the present value model has no implications for the error-correction equation in which  $\Delta y_t$  (or  $\Delta Y_t$ ) is regressed on  $\Delta Y_t$  (or  $\Delta y_t$ ),  $S_{t-1}$ , and perhaps lagged changes in  $y_t$  and  $Y_t$ .<sup>6</sup>

In the third section of this paper we show how these points apply to Marsh and Merton's [1987] model of dividends and prices on common stock.

---

<sup>6</sup> Nickell [1985] makes a similar point. Davidson, Hendry, Srba and Yeo [1978] and later papers by David Hendry and his co-authors use the single "error-correction equation" described here.



Marsh and Merton estimate a nonlinear error-correction equation which (approximately) relates log dividend changes to log price changes and the lagged log dividend-price ratio. They interpret their results as evidence for a structural model of managers' dividend-setting behavior; the structural model involves both partial adjustment by managers and rational valuation of firms by the stock market.

Applying the analysis of cointegrated models sketched above, we argue that the Marsh-Merton equation is evidence neither for partial adjustment nor for rational valuation. An error-correction model for dividends and prices should exist whenever there is any forward-looking behavior of stock prices, whether or not managers set dividends by partial adjustment, and whether or not prices are fully rational. The hypothesis of full rationality can be tested jointly with a hypothesis about the behavior of discount rates in a variety of ways (Campbell and Shiller [1987a, 1987b]), but not using the error-correction equation which Marsh and Merton estimate.

To illustrate the point more specifically, we show that very similar results to Marsh and Merton's can be obtained from artificial data for which the log dividend follows a Wiener process and the stock price is a rational forecast "contaminated" by random noise.

## 2. Forward-Looking Cointegrated Models: A VAR Approach

Let us consider a vector  $x_t$  containing two variables  $y_t$  and  $Y_t$ :  $x_t = [y_t \ Y_t]'$ . We assume that  $x_t$  is cointegrated of order (1,1), with an equilibrium error  $S_t$  equal to  $Y_t - \theta y_t$ . The use of the notation  $S_t$  follows Campbell and Shiller [1987a], who call  $S_t$  the "spread" between  $Y_t$  and  $y_t$ . Without loss of generality, we have normalized the spread so that it has a unit coefficient on  $Y_t$ .

The Granger Representation Theorem (Engle and Granger [1987]) implies that  $x_t$  obeys an error-correction model of the form

$$(1) \quad B(L)\Delta x_t = -\lambda S_{t-1} + e_t,$$

where  $B(L)$  is a two-by-two matrix polynomial in the lag operator of order  $q$ ,  $\lambda$  is a column vector with two elements, not both of which are zero, and  $e_t$  is a white noise error term<sup>7</sup>. Here, since at least one element of  $\lambda$  is nonzero, there must be Granger causality from  $S_t$  to  $\Delta x_t$ , as discussed above.

Equation (1) is not a vector autoregression, but it is straightforward to rewrite the system as a VAR for  $\Delta y_t$  and  $S_t$ , or alternatively as a VAR for  $\Delta Y_t$  and  $S_t$ . It is easiest to see how this is done in a first-order error-correction model (where the order  $q$  of  $B(L)$  is one). In this case we can rewrite (1) as

$$(2) \quad \Delta x_t = B\Delta x_{t-1} - \lambda S_{t-1} + e_t.$$

---

<sup>7</sup> Engle and Granger allow  $e_t$  to have an invertible moving average structure. However by increasing the order of  $B(L)$ , one can obtain a representation in which  $e_t$  is arbitrarily close to white noise.

Now define a matrix M such that  $M\Delta x_t = [\Delta y_t \ \Delta S_t]'$ . The first row of M is just [1 0] and the second row is [- $\theta$  1]. Then from (2) we have

$$(3) \quad \begin{bmatrix} \Delta y_t \\ \Delta S_t \end{bmatrix} = MBM^{-1} \begin{bmatrix} \Delta y_{t-1} \\ \Delta S_{t-1} \end{bmatrix} - MA S_{t-1} + Me_t.$$

This system can be rearranged so that it describes levels rather than differences of  $S_t$ . The result is a second-order VAR for the vector

$[\Delta y_t \ S_t]'$ :

$$(4) \quad G(L) \begin{bmatrix} \Delta y_t \\ S_t \end{bmatrix} = u_t,$$

where the matrix polynomial  $G(L)$  has order  $p=2$ . There are eight coefficients in  $G(L)$ , but by inspection of equation (3) two of these, the coefficients on  $\Delta y_{t-2}$ , are equal to zero. The six remaining coefficients are linear transformations of the six coefficients in (3). We write the coefficient of the  $i$ 'th variable in  $[\Delta y_t \ S_t]'$ , on the  $j$ 'th variable lagged  $k$  times, as  $\gamma_{ijk}$ .

This analysis generalizes in an obvious way to the case where the order  $q$  of the original error-correction model is greater than one. A  $q$ 'th-order error-correction model gives a  $p$ 'th-order VAR, with  $p=q+1$  and two zero coefficients on  $\Delta y_{t-p}$ . To obtain a VAR for  $\Delta Y_t$  and  $S_t$ , one proceeds as above, premultiplying (2) by a matrix N with first row [0 1] and second row [- $\theta$  1]. Again the resulting VAR has two zero restrictions.

The VAR representation in equation (4) is particularly useful when one wishes to analyze a forward-looking model. This is because there is a mechanical formula which gives the optimal k-period-ahead forecast of the variables in the VAR, conditional on an information set  $H_t$  which includes only current and lagged values of these variables (that is, current and lagged values of  $y_t$  and  $Y_t$ ). We define a new vector  $z_t = [\Delta y_t \dots \Delta y_{t-p+1} \ S_t \dots S_{t-p+1}]'$ . Then  $z_t$  follows a vector AR(1) process,  $z_t = Az_{t-1} + v_t$ , where A, the "companion matrix" of the VAR, has the form

$$(5) \ A = \begin{bmatrix} \gamma_{111} & \dots & \gamma_{11p} & \gamma_{121} & \dots & \gamma_{12p} \\ 1 & & & & & \\ & & 1 & & & \\ \gamma_{211} & \dots & \gamma_{21p} & \gamma_{221} & \dots & \gamma_{22p} \\ & & & 1 & & \\ & & & & & 1 \end{bmatrix}.$$

The multi-period forecasting formula is just

$$(6) \ E[z_{t+k} | H_t] = A^k z_t.$$

We now show how agents' forecasting can generate a forward-looking cointegrated model. We use the VAR representation above to analyze this model. Consider the following relationship between the variables  $y_t$  and  $Y_t$ :

$$(7) \ Y_t = \theta(1-\delta) \sum_{i=0}^{\infty} \delta^i E_t y_{t+i} + c.$$

Campbell and Shiller [1987a] call this a "present value model". It

states that  $Y_t$  is a linear function of the present value of expected future  $y_t$ , where expectations  $E_t$  are taken conditional on the information set  $I_t$  of economic agents<sup>8</sup>. We assume that  $H_t$ , the information set defined above consisting of  $y_t$ ,  $Y_t$  and their lags, is a subset of  $I_t$ .

Equation (7) has three parameters:  $\delta$ , the discount factor in the present value formula;  $\theta$ , the slope coefficient; and  $c$ , a constant term. There is no error in (7), so it is what Hansen and Sargent [1981] call an "exact linear rational expectations model". Below we discuss the implications of allowing white noise, or more generally a stationary error term to enter the equation.

A relation of the form (7) can arise in several different ways. If we take  $Y_t$  to be a consol yield and  $y_t$  to be a short-term interest rate, then (7) holds to a linear approximation if expected returns on bonds of all maturities differ only by a constant. In this case  $\delta$  is a parameter of linearization,  $\theta = 1$  and  $c$  is interpreted as a term premium. If we interpret  $Y_t$  as a stock price and  $y_t$  as the corresponding dividend, then (7) holds if the discount rate on the stock is constant through time. In this case  $\theta = \delta/(1-\delta)$  and  $c = 0$ . Both these examples are discussed in Campbell and Shiller [1987a]. Campbell and Shiller [1987b] show that (7) can also be derived as a linear approximation if we interpret  $Y_t$  as a log stock price and  $y_t$  as a log dividend, adjusted if necessary for a time-varying discount rate on stock, and this interpretation will be used in our discussion of the Marsh-Merton model below.

The present value model also describes the relation between

---

<sup>8</sup> Here and elsewhere we take conditional expectations to be equivalent to linear projections on information.

consumption and income under the permanent income hypothesis. If  $Y_t$  is the excess of consumption over capital income and  $y_t$  is labor income, then (7) follows from the model of Hall [1978] and Flavin [1981]. Campbell [1987] discusses this case.

More generally, equation (7) is the solution to a linear-quadratic intertemporal optimization problem where  $Y_t$  is a control variable and  $y_t$  is a target variable. Nickell [1985] shows that agents minimizing a loss function which is linear in the expected present value of squared changes in the control and squared deviations of the control from the target will set  $Y_t$  according to (7), with a zero constant term.

An important aspect of Campbell and Shiller [1987a] is that the present value model implies cointegration between  $Y_t$  and  $y_t$  if  $y_t$  is first-order integrated. Specifically, the spread  $S_t = Y_t - \theta y_t$  is stationary. To see this, note that the sum of the weights in the first term on the right hand side of (7) is  $\theta$ . Therefore (7) implies

$$(8) \quad S_t = \theta(1-\delta) \sum_{i=0}^{\infty} \delta^i (E_t y_{t+i} - y_t) + c$$

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

where the second equality follows by writing  $(y_{t+i} - y_t) = \Delta y_{t+1} + \dots + \Delta y_{t+i}$ , and collecting terms. Equation (8) shows that  $S_t$  is stationary when  $\Delta y_t$  is stationary.

The present value model also implies a linear relation between  $S_t$  and the expectation of the one-period change in  $Y_t$ . We have

$$(9) \quad E_t \Delta Y_{t+1} = ((1-\delta)/\delta) S_t + ((1-\delta)/\delta) c.$$

This shows that  $\Delta Y_t$  is stationary when  $S_t$  is stationary.

Since  $y_t$  and  $Y_t$  are cointegrated, we know that VAR representations exist for  $\Delta y_t$  and  $S_t$ , or  $\Delta Y_t$  and  $S_t$ . The present value model puts cross-equation restrictions on the parameters of these VARs. Before stating the restrictions formally, we note the following proposition about the pattern of Granger causality between  $S_t$ ,  $\Delta y_t$  and  $\Delta Y_t$ :

Proposition. If the cointegrated present value model holds, then either  $S_t$  is an exact linear function of current and lagged  $\Delta y_t$  or  $S_t$  Granger causes  $\Delta y_t$  and  $\Delta Y_t$ .

The intuition behind this statement is straightforward, and was given in the introduction. Either market participants have extra information relevant for forecasting  $\Delta y_{t+1}$ , beyond the history of this variable, or they do not. If they do not have superior information, then they form  $S_t$  as a linear function of their relevant information  $\Delta y_t, \dots, \Delta y_{t-p}$ . If they do have superior information, then  $S_t$  incorporates it and will Granger cause  $\Delta y_t$ . A similar argument can be made for the relation between  $S_t$  and  $\Delta Y_t$ .

The formal proof of this proposition involves a projection argument. Consider projecting equation (8) onto the information set  $H_t$  which includes current and lagged values of  $y_t$  and  $Y_t$ . The left hand side is unchanged because  $S_t$  is in  $H_t$ . The right hand side becomes an expected present value, where expectations are now taken conditional on  $H_t$  rather than agents' information set  $I_t$ .

Suppose that  $S_t$  does not Granger cause  $\Delta y_t$ . Then the expected value

of  $\Delta y_{t+1}$ , conditional on  $H_t$ , is an exact linear function of current and lagged values of  $\Delta y_t$ . It follows that the right hand side of the projected equation (8) is an exact linear function, so  $S_t$  itself must be an exact linear function of current and lagged values of  $\Delta y_t$ , which is what we needed to prove. A similar argument can be made for  $S_t$  and  $\Delta Y_t$  by projecting equation (9) onto the information set  $H_t$ .

In practical applications we do not observe exact linear relationships between cointegrated variables. The proposition then implies that there must be Granger causality from the spread or equilibrium error  $S_t$  to both cointegrated variables, if the system is to obey the present value model.

In some circumstances one might want to add a stationary error term to equation (7). This could represent a time-varying term premium, the influence of noise traders on stock prices, or transitory consumption, depending on the application. Once an error term appears in (7), the proposition above no longer holds. The error term will make the relationship between  $Y_t$  and  $y_t$  stochastic, even if agents have no information relevant for forecasting  $y_t$  beyond the history of  $y_t$  itself. However  $S_t$  will Granger cause  $Y_t$  and  $y_t$  whenever agents have superior information, and we regard this as the normal situation.

The present value model (7) has much tighter implications than those stated in the proposition above. These can be stated conveniently using the companion form of the VAR for  $\Delta y_t$  and  $S_t$ . Recall that  $z_t$  is defined as the vector  $[\Delta y_t \dots \Delta y_{t-p+1} \ S_t \dots S_{t-p+1}]'$ , and that it follows a vector AR(1) process,  $z_t = Az_{t-1} + v_t$ , where the companion matrix  $A$  is given in equation (5). Let us define  $g'$  to be a row vector which picks out  $S_t$  from the vector  $z_t$ :  $S_t = g'z_t$ . (The vector  $g'$  has all elements



zero except the  $p+1$ 'st, which is unity). Let us define  $h'$  similarly, as a row vector which picks out  $\Delta y_t$ :  $\Delta y_t = h'z_t$ . Then projecting (8) onto the VAR information set  $H_t$ , and using equation (6), we have

$$(10) \quad S_t = g'z_t = S'_t$$

$$S'_t = \theta \sum_{i=1}^{\infty} \delta^i h'A^i z_t = \theta h'\delta A(I-\delta A)^{-1} z_t.$$

This equation says that the spread must equal the unrestricted forecast of the present value of future  $\Delta y_t$  from the VAR, evaluated using the multi-period forecasting formula (5). We call this unrestricted forecast  $S'_t$ <sup>9</sup>.

Equation (10) can be used to evaluate the present value model in several different ways. A standard formal approach would be to test the restrictions on the coefficients of the VAR,  $g' = \theta h'\delta A(I-\delta A)^{-1}$ , which can be rewritten in linear form if one postmultiplies by  $(I-\delta A)$ :

$$(11) \quad g'(I-\delta A) = \theta h'\delta A.$$

Using the structure of the matrix  $A$ , one can write out the implications of (11) for individual coefficients:  $\gamma_{21k} = -\theta\gamma_{11k}$ ,  $k=1\dots p$ ;  $\gamma_{221} = 1/\delta - \theta\gamma_{121}$ ; and  $\gamma_{22k} = -\theta\gamma_{12k}$ ,  $k=2\dots p$ . These linear restrictions operate across the two equations of the VAR; thus the equations describing  $S_t$  and  $\Delta y_t$  are each unrestricted when taken in isolation. The restrictions have

---

<sup>9</sup> The infinite weighted sum of powers of  $A$  must converge because the variables  $S_t$  and  $\Delta y_t$  are stationary. Equation (9) gives the closed-form solution for this sum.

a simple intuition in most applications. In the term structure, for example, the restrictions are that excess returns on long bonds over short bonds are unpredictable given current and lagged values of  $y_t$  and  $Y_t$ .

It can be hard to interpret a statistical rejection of cross-equation restrictions, so Campbell and Shiller [1987a] propose an alternative approach. They construct  $S_t^i$ , and compare its historical behavior with that of  $S_t$ . If the two variables move closely together, there is some truth to the present value model even if the cross-equation restrictions (11) can be rejected at conventional significance levels.

One might want to test the present value model (7) without using the VAR approach developed above. There are at least two ways to do this. First, one could form a variable  $\xi_t = Y_t - (1/\delta)[Y_{t-1} - \delta(1-\delta)y_{t-1}]$ . Substitution from (7) shows that

$$(12) \quad \xi_t = Y_t - E_{t-1}Y_t + c(1-1/\delta).$$

The variable  $\xi_t$  can be interpreted as an excess return in applications to the stock and bond markets. Equation (12) says that  $\xi_t$  is unpredictable given information known in advance; a regression of  $\xi_t$  on lagged variables should give zero coefficients apart from a constant term<sup>10</sup>.

Alternatively, one could use equation (9) to restrict the equation of the error-correction model (1) which describes  $\Delta Y_t$ . In the regression of  $\Delta Y_t$  on  $S_{t-1}$  and lagged values of  $\Delta y$  and  $\Delta Y$ , the coefficient on  $S_{t-1}$

---

<sup>10</sup> This single-equation approach is closely related to a test of (11) on the VAR. In fact, one can obtain numerically identical test statistics using (11) and (12).

should be  $(1-\delta)/\delta$ , and all other coefficients should be zero. In the term structure, for example, the yield spread should be the optimal forecast of the change in the long-term interest rate over the following period.

We have seen that there are valid single-equation regression tests of the present value model. But it is important to note that one cannot test the model by regressing  $\Delta y_t$  (or  $\Delta Y_t$ ) on  $\Delta Y_t$  (or  $\Delta y_t$ ),  $S_{t-1}$  and lags. This type of regression is used by David Hendry and his co-authors, and Marsh and Merton's equation can be interpreted in this way. But the present value model has no implications for this regression.

It is easy to see why this is so. We have seen from equation (9) that the regression of  $\Delta Y_t$  on  $S_{t-1}$  and other lagged variables is restricted by the model. But our analysis of equation (11) showed that the regression of  $\Delta y_t$  on  $S_{t-1}$  and lags is not restricted, and the correlation between  $\Delta Y_t$  and  $\Delta y_t$  is not restricted. Therefore when both  $\Delta y_t$  and  $\Delta Y_t$  appear in the regression equation (one as the dependent variable and one as a regressor), the coefficients of the equation are unrestricted by the present value model.

### 3. The Marsh-Merton Model of Dividend Behavior in the Stock Market

Terry Marsh and Robert Merton [1987] have set forth what they describe as an error-correction model of the dividend-setting behavior of firms' managers. Their argument is provocative, in raising the issue of how dividends are set by managers in the context of rational expectations models, and suggesting that price may well drive dividends rather than the other way around as commonly assumed. They hypothesized that firms' managers have a "target ratio" of dividends per share to price per share, and adjust dividends to price changes in such a way as to bring dividends gradually back to this target ratio. This same hypothesis can be described in another way. Markets are efficient, they assumed, so that price per share is proportional to a long run or "permanent" earnings per share as optimally forecasted by the market. They described their hypothesis as that managers have a target payout ratio of dividends to permanent earnings and adjust dividends when permanent earnings change, so as to bring dividends back into alignment with the target ratio to earnings. Their model is thus an updated version of John Lintner's [1956] partial adjustment model of dividends to earnings, the change being essentially the substitution of permanent earnings for earnings.

They estimated a nonlinear error-correction model using annual data on New York Stock Exchange stocks from the Center for Research in Security Prices, 1926-81. They took  $P_t$  to be the end of year price of a value-weighted portfolio, and  $D_t$  to be the total of dividends paid over the year (summing the monthly dividend series for the year). The (ordinary least squares) estimated model (with standard errors in parentheses) that

they gave was:<sup>11</sup>

$$\begin{aligned} \log(D_{t+1}) - \log(D_t) + D_t/P_{t-1} = & -0.101 + 0.437\log((P_t + D_t)/P_{t-1}) \\ & (0.157) \quad (0.064) \\ (13) \quad & - 0.042\log(D_t/P_{t-1}) + u_t \\ & (0.050) \end{aligned}$$

$$\bar{R}^2 = 0.47$$

$$D. W. = 1.53$$

They reported roughly similar results when generalized least squares was performed, or when a data series using the Standard and Poor Composite Index 1928-80 was used.

Their interpretation of this estimated regression was similar to that made of other estimated error correction models in the literature: solely as representing the sluggish response to errors of economic agents. They noted that the point estimate of the coefficient of  $\log((P_t + D_t)/P_{t-1})$  is "substantial in magnitude and highly significant." They concluded that "this finding is consistent with the hypothesis that the market price is a good indicator of permanent earnings and that managers systematically change dividends in response to an unanticipated change in permanent earnings." The inference is that managers are highly rational forecasters who behave as if they trust the efficient market to forecast their own earnings. They noted also that this estimated coefficient was "significantly less than one," and concluded that "this finding is also consistent with Lintner's stylized fact that managers smooth dividends responding to a partial adjustment fashion to an unanticipated change in

---

<sup>11</sup>Terry Marsh and Robert Merton, "Dividend Behavior for the Aggregate Stock Market," Journal of Business, 1987, p. 17.

permanent earnings."<sup>12</sup> They noted that the coefficient of  $\log(D_t/P_{t-1})$  is negative, which is "consistent with the hypothesis that this ratio converges to a long-run stationary distribution," and that "the point estimates for the speed of adjustment are, however, rather small, which at best suggests that a substantial period of time is required for the dividend-price ratio to converge to its steady-state distribution."<sup>13</sup> They described these results as reassuring us that "the data tend to support such 'superrational' forecasting behavior by managers even in the relatively unimportant area of dividend policy."<sup>14</sup>

But ought we to take their results as suggesting that managers behave this way? In effect, they have discovered a correlation of dividend changes with lagged price changes. But as we have discussed above, a correlation of dividend with lagged price can come about not only because dividend responds to price, but because price responds to information about future dividends. This, in fact, has been the customary interpretation, before Marsh and Merton, of the correlation of prices with dividends, and we should not dismiss this old interpretation

---

<sup>12</sup> ibid. p. 18.

<sup>13</sup> ibid. p. 19. The steady-state behavior of this equation can be considered only by adding another equation to the model:  $P_{t+1} = P_t/\delta - D_{t+1} + v_{t+1}$ , which follows from the present value relation (equation (7) above applied to prices and dividends as discussed above) where  $v_{t+1}$  is the new information about dividends that appeared between  $t$  and  $t+1$ . For estimated values of parameters, the estimated system does tend to converge gradually to a steady state for most reasonable initial conditions. If equation shocks cause dividends to become too large, the system will crash (arguments of log functions become negative) but this does not normally seem to occur. Simulation of this system confirms that it tends to converge on a steady state even if the coefficient of  $\log(D_t/P_{t-1})$  is small and positive.

<sup>14</sup> ibid., p. 21.

lightly.

In order to understand Marsh and Merton's results in more detail, we would like to be able to relate their equation to a present value model for stock prices and dividends. It is critical to note at this point just what the data alignment is.  $P_t$  is end of year price, and  $D_t$  is total for year dividends. There is thus a time overlap between the log-dividend change on the left-hand side of (13) with the variables on the right-hand side, and we shall argue that the nature of this time overlap is critical to interpreting their results.

Note that our present value model (7) included the current-year dividend in the price, which should therefore be a beginning-of-year price. Accordingly we set  $\log(D_t) = y_t$  and  $\log(P_{t-1}) = Y_t$ . In order to obtain a log-linear equation, we drop the  $D_t/P_{t-1}$  term on the left-hand side and the  $D_t$  in the numerator in the second term on the right-hand side. Given the stochastic properties of the data, neither of these omissions will have much impact on the estimated coefficients of the model.<sup>15</sup> The resulting equation is

$$(14) \quad \Delta y_{t+1} = a + b\Delta Y_{t+1} - cS_t + u_t,$$

where  $S_t = \log(P_{t-1}) - \log(D_t) = Y_t - y_t$ ,  $b = .437$  and  $c = -.042$ .

This equation is similar to Marsh and Merton's, but is easier to

---

<sup>15</sup> More formally, a linearization of (13) is  $\Delta y_{t+1} \approx \text{constant} + 0.437(\exp(w)/(\exp(w)+\exp(x))\Delta Y_{t+1} - (0.437\exp(x)/(\exp(w)+\exp(x)) - .042 - \exp(x))S_t + u_t$ . Choosing points of linearization  $\exp(w) = 1.05$  and  $\exp(x) = .05$ , the equation becomes  $\Delta y_{t+1} \approx \text{constant} + 0.417\Delta Y_{t+1} + 0.072S_t + u_t$ . The coefficients are not substantially different between (13) and the linearized model, hence we expect roughly the same coefficients if (14) is estimated instead of (13).

relate to our previous analysis. Two conclusions follow immediately from that discussion. First, if  $S_t = \log(P_{t-1}) - \log(D_t)$  has explanatory power for  $\Delta y_{t+1} = \Delta \log(D_{t+1})$ , this could just as easily be due to forecasting as to a structural causal relationship. Secondly, the log-linear present value model cannot be tested using (13) or (14) for the reasons given at the end of the previous section. One cannot use the Marsh-Merton equation to judge whether the stock market forecasts dividends rationally.

While these points hold very generally, it is worthwhile also to consider in the Marsh-Merton model the issue of data alignment and time overlap in a more precise way. Consider a world in which dividends flow continuously. Then, operating in continuous time, and disregarding for purposes of approximation the distinction between logs of averages and averages of logs, we may take:

$$(15) \quad \log(D_t) - y_t = \int_{t-1}^t \bar{d}_s ds$$

where  $\bar{d}_s$  is the instantaneous flow of log dividends at time  $s$ . Consider a very simple assumption about the instantaneous flow of log dividends: that these are a Wiener process, where the variance of the one-period change is  $\zeta^2$ .<sup>16</sup> Then, if the information set available to the market for forecasting is lagged dividends, i.e., there is no superior information, and taking a continuous time version of (7) above where  $\delta = \exp(-r)$ , the log price at time  $t$  is proportional to the instantaneous flow of log dividend  $\bar{d}_t$ ; it is given by  $\log(P_t) - Y_{t+1} = \bar{d}_t - \log(r \cdot \zeta^2 / 2)$ . Thus, (14)

---

<sup>16</sup> A discrete time version of this model has been proposed by Kleidon [1986].



can be rewritten:

$$(16) \int_t^{t+1} (\dot{d}_s - \dot{d}_{s-1}) ds = a + b(\dot{d}_t - \dot{d}_{t-1}) + c \left( \int_t^{t+1} \dot{d}_{s-1} ds - \dot{d}_{t-1} \right) + u_t.$$

The theoretical regression coefficients for this regression are  $a = 0$ ,  $b = 1.00$  and  $c = -1.00$ , and the theoretical  $R^2$  is 0.5. With these coefficients, the  $\dot{d}_{t-1}$  terms on the right-hand side of (16) drop out and the right hand side becomes  $\dot{d}_t - \int(t,t+1)\dot{d}_{s-1} ds$ . It is easy to see why these coefficients arise: The dependent variable in (16) equals  $g_t + h_t$  where  $g_t = \int(t,t+1)\dot{d}_s ds - \dot{d}_t$  and  $h_t$  equals  $\dot{d}_t - \int(t,t+1)\dot{d}_{s-1} ds$ . Now  $g_t$  is determined only by innovations after  $t$ ,  $h_t$  by innovations before  $t$ , so the best possible forecast of  $g_t + h_t$  at time  $t$  is  $h_t$ . By symmetry,  $g_t$  and  $h_t$  have the same variance, hence the  $R^2$  is 0.5.

That these theoretical regression coefficients will apply approximately to the Marsh-Merton regression (13) (for plausible parameter values if log dividends were a Wiener process and prices were set in accordance with rational expectations) can be confirmed with a Monte-Carlo experiment. In each iteration, a monthly Gaussian random walk was generated where the standard deviation  $\zeta$  of the monthly change was .03. This was exponentiated to create a monthly lognormal dividend series, and the monthly series was averaged over 12 periods to create an annual series of dividends. Price  $P_t$  was  $D_t^*/(r - \zeta^2/2)$  where  $r$ , the monthly interest rate, was .007 and  $D_t^*$  was the dividend in the last month of the year. The variables in equation (13) were generated, and the regression run for 54 observations (as with Marsh and Merton). With 1000 iterations, the average coefficient of  $\log((D_t + P_t)/P_{t-1})$  was 1.076 (with a standard deviation across iterations of 0.213), the average coefficient

of  $\log(D_t/P_{t-1})$  was -1.014 (with a standard deviation of 0.327), the average  $\bar{R}^2$  was 0.432 (with standard deviation 0.098) and the average Durbin-Watson statistic was 1.762 (with a standard deviation of 0.073).<sup>17</sup>

Now these regression coefficients have the same signs as the coefficients in the Marsh-Merton regression. This suggests that the Marsh-Merton results are related to the data overlap. But the estimated coefficients are much larger in absolute value than are the coefficients that Marsh and Merton estimated. We can easily fix this by adding a little noise to price, so that by standard errors in variables results, there is a downward bias to the coefficients.

We have indeed argued in earlier papers that there is likely to be a 'noise' or 'fads' component in stock prices (Shiller [1981,1984], Campbell and Kyle [1986]). If such a noise component resembles, say, an AR-1 process independent of dividends and with a high autoregressive parameter, then this component may substantially increase the variance of prices without making returns on stocks very forecastable. The noise story thus generates a 'near rational expectations' model in which there are no dramatic profit opportunities to people who understand the noise, but in which stock prices behavior may be dominated by the noise. That stock prices appear to have such a noise component has been claimed by

---

<sup>17</sup>Of course, the standard errors for the averages can be derived by dividing the standard deviations by the square root of 1000. We also experimented with an AR-1 process for the log dividend, instead of a Wiener process. Monte-Carlo experiments analogous to the one just described tended to produce a coefficient of  $\log((P_t+D_t)/P_{t-1})$  greater than one and of  $\log(D_t/P_{t-1})$  as negative and substantially less than one in absolute value. This example does not accurately mimic actual stock prices, since if the autoregressive coefficient is even slightly less than one it implies that prices are substantially less volatile than dividends.

DeBondt and Thaler [1985], Fama and French [1986], and Poterba and Summers [1987]. For our present purposes, this particular noise story is very helpful, in that it contributes a much bigger error term (relative to the true value) to the third term in equation (13) (which has the level of stock prices in it) than in the second term of equation (13) (which has the change in stock prices in it). Thus, this noise term could greatly reduce absolute value of the coefficient  $c$  and reduce the coefficient  $b$  to a lesser extent. In considering this possibility, bear in mind that if the third term in (16) is omitted from the regression, then the theoretical coefficient  $b$  of the second term is .5, roughly the value obtained by Marsh and Merton, and the theoretical  $R^2$  is 0.375. If a large enough error term is added to the third term in (16), then standard errors in variables results imply that its coefficient should become only slightly less than zero, the coefficient  $b$  to be reduced to near 0.5, and the  $R^2$  around 0.375. This would bring us to the vicinity of the Marsh-Merton results.

We confirmed that this may explain the general nature of the Marsh-Merton results (13) using a Monte-Carlo procedure. For each iteration of the Monte Carlo we generated, as before, a Gaussian random walk  $w_t$  for monthly log dividends, where the standard deviation of the error term was  $\zeta = .03$ . We generated another independent first-order autoregressive process  $x_t$  (with initial value not zero but a drawing from the unconditional distribution of the process) with autoregressive parameter  $g = .97$  for annual data and standard deviation of the error term as .05. The monthly dividend was  $\exp(w_t)$ , and the annual dividend series  $D_t$  was taken as the sum for each year of the 12 monthly dividends. Price was  $P_t$

$-\exp(x_t)D_t^*/(r - \zeta^2/2)$ , where  $D_t^*$  is the dividend for the last month of the year, and the monthly interest rate was 0.7%. These parameters were chosen as roughly plausible values for actual data. The AR-1 process has an unconditional standard deviation of  $.05/(1-g^2)^{.5} = .206$ , which means that price typically runs  $\pm 20\%$  of the rational expectations value  $D_t^*/(r - \zeta^2/2)$ , and occasionally deviates by 50% or more from the rational expectations value. Such a standard deviation is well within the bounds claimed by Black [1986] and others. With these values of parameters and a sample size of 54, for 1000 iterations the average estimated constant term was -0.015 (with a standard deviation across iterations of 0.288), the average estimated coefficient of  $\log((D_t + P_t)/P_{t-1})$  was 0.424 (standard deviation 0.116), the average coefficient of  $\log(D_t/P_{t-1})$  was -.024 (standard deviation 0.110), the average  $\bar{R}^2$  was 0.270 (standard deviation 0.100) and the average Durbin-Watson statistic was 2.088 (standard deviation 0.186). The regression coefficients are remarkably close to the Marsh-Merton figures. The  $\bar{R}^2$  and Durbin Watson statistic are less closely matched (the Monte Carlo values are a couple of standard deviations away from the Marsh-Merton values). The model that generated these coefficients was extremely simple, and we would not expect a perfect correspondence of parameters. We think these results suggest a very different interpretation of the Marsh-Merton results.<sup>18</sup>

---

<sup>18</sup> If we drop the standard deviation of the innovation in the AR-1 from .05 to .03, then in 1000 iterations the average coefficient of  $\log((P_t + D_t)/P_{t-1})$  rises to 0.546, the average coefficient of  $\log(D_t/P_{t-1})$  moves to -0.166, the average  $\bar{R}^2$  becomes 0.308, and the average Durbin-Watson statistic becomes 2.123.

#### 4. Conclusion

In describing human behavior, we must always seek some simplifications, since a complete and accurate description of behavior would be vastly complicated. The interpretation commonly given to reduced-form error correction models, as reflecting costly or sluggish adjustment back to equilibrium, may be viewed sympathetically in this light. In the Marsh and Merton example, managers do not literally set dividends in response to price changes just as Marsh and Merton assert, but perhaps there is an element of truth to this characterization. Managers do know the price of their stock, and this enters in incalculable ways into the decision process that leads to their decision on dividend payouts. If their behavior warrants, we may say that they are behaving roughly as if they were obeying a structural error-correction model.

We should be careful however not to read too much into such 'as if' interpretations of results, because people may behave simultaneously as if they were doing many very different things. We have seen in our dividend example that an alternative structural model, a near-rational expectations model, also implies the same sort of reduced-form error-correction behavior. The near-rational expectations model, we think, is quite as plausible as the sluggish adjustment model offered by Marsh and Merton.

Of course, there is other evidence that can be brought to bear to distinguish these two models and to refine them further. Relevant evidence includes surveys of managers (Lintner [1956]), direct tests of stock market pricing models (Campbell and Shiller [1987a, 1987b]) and many

others), and evidence on the univariate time series behavior of dividends and prices. In other specific examples, yet other kinds of evidence may be adduced. But in the absence of additional evidence, one should regard the structural interpretation of an estimated error-correction model as highly tentative at best.

## Bibliography

Aoki, M., Aggregation in Optimization Methods for Large-Scale Systems, D. Wismer, ed., McGraw-Hill, 1971.

Aoki, M., "Control of Large-Scale Dynamic Systems by Aggregation," IEEE T-AC, AC-13, p. 246-53, June 1968, also reprinted in Distributed Control, R. Larson et al. (eds), IEEE, Computer Society, 1978.

Black, Fisher, "Noise," Journal of Finance, 41:529-43, July 1986.

Campbell, John Y., "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis", forthcoming Econometrica, 1987.

Campbell, John Y. and Albert S. Kyle, "Smart Money, Noise Trading and Stock Price Behavior", unpublished paper, Princeton University, September 1986.

Campbell, John Y., and Robert J. Shiller, "A Simple Account of the Behavior of Long-Term Interest Rates," American Economic Review Papers and Proceedings, 74:44-8, May 1984.

Campbell, John Y. and Robert J. Shiller [1987a], "Cointegration and Tests of Present Value Models", Journal of Political Economy 95:1062-1088, October 1987.

Campbell, John Y. and Robert J. Shiller [1987b], "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", unpublished paper, revised May 1987.

Davidson, James E.H. and David F. Hendry, "Interpreting Econometric Evidence: The Behavior of Consumers' Expenditure in the U.K.", European Economic Review 16:1777-192, 1981.

Davidson, James E.H., David F. Hendry, Frank Srba and Stephen Yeo, "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom", Economic Journal 88:661-692, December 1978.

DeBondt, Werner, and Richard Thaler, "Does the Stock Market Overreact?" Journal of Finance, 40:793-805, July 1985.

Engle, Robert F. and C.W.J. Granger, "Co-Integration and Error-Correction: Representation, Estimation and Testing", Econometrica 55:251-276, March 1987.

Fama, Eugene F., and Kenneth R. French, "Permanent and Temporary Components of Stock Prices," Center for Research in Security Prices Working Paper No. 178, July 1986.

Flavin, Marjorie A., "The Adjustment of Consumption to Changing Expectations about Future Income", Journal of Political Economy 89:974-1009, October 1981.

Granger, C.W.J., "Some Properties of Time Series Data and Their Use in Econometric Model Specification", Journal of Econometrics 121-130, 1981.

Granger, C.W.J., "Cointegrated Variables and Error-Correcting Models", working paper 83-13, University of California San Diego, 1983.

Granger, C.W.J., "Developments in the Study of Cointegrated Variables", Oxford Bulletin of Economics and Statistics 48:213-228, August 1986.

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

Hansen, Lars P. and Thomas J. Sargent, "Exact Linear Rational Expectations Models: Specification and Estimation", Staff Report No. 71, Federal Reserve Bank of Minneapolis, November 1981.

Hendry, David F., "Econometric Modelling: The 'Consumption Function' in Retrospect", Scottish Journal of Political Economy 30:193-220, November 1983.

Hendry, David F., "Econometric Modelling with Cointegrated Variables: An Overview", Oxford Bulletin of Economics and Statistics 48:201-212, August 1986.

Hendry, David F. and Jean-Francois Richard, "Econometric Analysis of Economic Time Series", International Statistical Review 51:111-164, August 1983.

Kleidon, Allan W., "Variance Bounds Tests and Stock Price Valuation Models", Journal of Political Economy 94:953-1001, October 1986.

Lintner, John, "Distribution of Incomes of Corporations among Dividends, Retained Earnings, and Taxes," American Economic Review 61:97-113, 1956.

Marsh, Terry A., and Robert C. Merton, "Dividend Behavior for the Aggregate Stock Market," Journal of Business, 60:1-40, 1987.

Nickell, Stephen, "Error Correction, Partial Adjustment and All That: An Expository Note", Oxford Bulletin of Economics and Statistics 47:119-129, 1985.

Phillips, A.W., "Stabilization Policy in a Closed Economy", Economic Journal 64:290-323, 1954.



Poterba, James M., and Lawrence H. Summers, "Mean Reversion in Stock Prices: Evidence and Implications," unpublished paper, Massachusetts Institute of Technology, March, 1987.

Salmon, Mark, "Error Correction Mechanisms", Economic Journal 92:615-629, September 1982.

Sargan, J.D., "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in P.E. Hart, G. Mills and J.N. Whittaker eds., Econometric Analysis for National Economic Planning, London: Butterworths, 1964.

Shiller, Robert J., "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure", Journal of Political Economy 87:1190-1219, December 1979.

Shiller, Robert J., "The Term Structure of Interest Rates," in Benjamin Friedman and Frank Hahn, editors, The Handbook of Monetary Economics, North Holland, forthcoming 1987.

Shiller, Robert J., John Y. Campbell and Kermit L. Schoenholtz, "Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates", Brookings Papers on Economic Activity 1:173-217, 1983.

Sims, Christopher A., "Exogeneity and Causal Ordering in Macroeconomic Models", in New Methods in Business Cycle Research: Proceedings from a Conference, Federal Reserve Bank of Minneapolis, 1977.