

NBER TECHNICAL WORKING PAPER SERIES

SMART MONEY, NOISE TRADING AND STOCK PRICE BEHAVIOR

John Y. Campbell

Albert S. Kyle

Technical Working Paper No. 71

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue

Cambridge, MA 02138

October 1988

We are grateful to Jim Stock for helpful comments on several drafts of this paper, and to John Montgomery and Ludger Hentschel for able research assistance. We acknowledge support from the National Science Foundation and the John M. Olin Fellowship at the NBER (Campbell). An earlier version of Section 3 of the paper was circulated as Albert S. Kyle, "An Explicit Model of Smart Money and Noise Trading." This research is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

SMART MONEY, NOISE TRADING AND STOCK PRICE BEHAVIOR

ABSTRACT

This paper derives and estimates an equilibrium model of stock price behavior in which exogenous "noise traders" interact with risk-averse "smart money" investors. The model assumes that changes in exponentially detrended dividends and prices are normally distributed, and that smart money investors have constant absolute risk aversion. In equilibrium, the stock price is the present value of expected dividends, discounted at the riskless interest rate, less a constant risk premium, plus a term which is due to noise trading. The model expresses both stock prices and dividends as sums of unobserved components in continuous time.

The model is able to explain the volatility and predictability of U.S. stock returns in the period 1871-1986 in either of two ways. Either the discount rate is 4% or below, and the constant risk premium is large; or the discount rate is 5% or above, and noise trading, correlated with fundamentals, increases the volatility of stock prices. The data are not well able to distinguish between these explanations.

## 1. Introduction

Recently there has been a great deal of interest in explaining the behavior of aggregate stock prices. The simplest model of stock market movements, which was generally accepted at least as an approximation until a few years ago, is that the stock price equals the present value of expected future dividends, discounted at a constant rate. This model attributes stock price movements to news about future dividends, and implies that percentage stock returns are unpredictable.

Recent empirical work has convincingly rejected this model. It appears to be inadequate even as a first approximation: Fama and French [1987] and Campbell and Shiller [1988a,b], for example, show that dividend-price ratios are powerful predictors of percentage stock returns when these returns are measured over periods of several years. Campbell and Shiller [1988b] relate this finding to the earlier literature on "excess volatility" of stock prices (LeRoy and Porter [1981], Shiller [1981]). While that literature encountered some econometric difficulties (Flavin [1983], Kleidon [1986], Marsh and Merton [1986]), recent work which corrects these problems upholds the conclusion that stock returns are too volatile to be explained by the simple present value model with a constant discount rate (Mankiw, Romer and Shapiro [1985], West [1988a,b], Campbell and Shiller [1988a,b]).

If this model cannot account for stock price variation, what can? Some people have suggested that time variation in discount rates is important for stock prices, but simple models of observable discount rate variation have met with little success (Hansen and Singleton [1982,1983], Campbell and Shiller [1988a]). Of course discount rates may change in an

unobservable manner, but this possibility is indistinguishable from the presence of serially correlated noise or measurement error in the stock price.

In this paper we take a somewhat different tack. We explicitly consider the possibility that stock prices are influenced by exogenous serially correlated noise. While this noise may be interpreted in several ways, we present a theoretical derivation in which it arises from the competitive interaction of "noise traders" and "smart money" investors. The exogenous actions of the noise traders are able to influence the stock price because the smart money investors are risk averse. This is in the spirit of Black [1986], and provides an equilibrium foundation for the model of Shiller [1984]. (See also DeLong, Shleifer, Summers and Waldmann [1987], Fama and French [1988], Miller [1977], Poterba and Summers [1987] and Summers [1986]).

Our model has several other important features. First, we assume that after exponential detrending, changes in the levels of dividends and stock prices have constant variance and are normally distributed. A more conventional approach (Kleidon [1986], LeRoy and Parke [1987], Campbell and Shiller [1988a,b]) would assume that changes in log dividends and prices are normal with constant variance. This would imply a constant volatility of percentage returns and percentage dividend changes. Our approach forces the variance of percentage returns and dividend growth rates to increase as the level of prices and dividends falls. This is a phenomenon which has been noted in U.S. stock market data (Black [1976], Nelson [1987], Schwert [1987]), although our modelling assumption probably overstates the effect. Negative prices or dividends are a theoretical possibility in our model,

which must be assumed to break down as prices and dividends approach zero<sup>1</sup>.

Secondly, we assume that the smart money investors have constant absolute risk aversion (exponential utility), as opposed to the constant relative risk aversion (power utility) assumed in much recent work (e.g. Hansen and Singleton [1982,1983]). When stock prices and dividends are normally distributed, this means that the equilibrium stock price discounts expected future dividends at the riskless rate of interest. Risk aversion increases the expected return on stock by subtracting a constant or exogenously time-varying term from the price, rather than by increasing the discount rate. Put another way, investors demand a risk premium per share of stock rather than per dollar invested.

This implication of our model turns out to be helpful in explaining stock price behavior. It means that when stock prices are low, expected percentage returns on stock are high because the volatility of percentage returns is high. This effect can account for at least some of the ability of dividend-price ratios to predict percentage stock returns. Also, the discounting of expected future dividends at a relatively low riskless interest rate helps to explain the high volatility of the stock market<sup>2</sup>.

A third feature of our model is that it expresses the relationship

---

<sup>1</sup> Stock markets in other countries, such as Russia, have lost all their value in the past. Thus a zero value for U.S. stocks is not inherently impossible. Below we calculate the probability of negative dividends within our sample period, conditional on the estimated parameters of our model.

<sup>2</sup> Much of the literature on "excess volatility" (e.g. Campbell and Shiller [1987], LeRoy and Porter [1981], Mankiw, Romer, and Shapiro [1985], Shiller [1981], and West [1988a]) also assumes that dividends and prices follow a linear process in levels. But that literature does not allow for a constant risk premium per share of stock, with discounting at the riskless interest rate.

between dividends and stock prices as an unobserved components model in continuous time. Once we have estimates of the model's parameters, we are able to decompose stock prices into several components which are not directly observed: information about future dividends contained in the history of dividends, other information about future dividends, and noise. This explicit approach to modelling information is in the spirit of a large literature in accounting.

While our model is set in continuous time, we have to estimate it, of course, using discrete-time data. But we are careful to compute the true likelihood function of the data, given the continuous-time model, rather than using an approximate discrete-time model. This approach avoids potential difficulties to do with time aggregation. (For work in a similar spirit, see Christiano, Eichenbaum and Marshall [1987] and Grossman, Melino and Shiller [1987])<sup>3</sup>.

The structure of the paper is as follows. In section 2, we outline the continuous-time econometric framework. Section 3 provides the theoretical justification for it. Section 4 confronts the econometric model with the data, and Section 5 concludes. Technical details are given in several Appendices.

---

<sup>3</sup> Engle and Watson [1985] also estimate an unobserved components model on stock market data, but their model is set in discrete rather than continuous time.

## 2. Stock Prices, Dividends, and Noise

Let  $D^0(t)$  and  $P^0(t)$  denote observed real dividends and stock prices, respectively, and let  $\xi$  denote an exponential growth trend. We will write detrended dividends as  $D(t)$  and detrended prices as  $P(t)$ : thus we have

$$(2.01) \quad D(t) = D^0(t)\exp(-\xi t), \quad P(t) = P^0(t)\exp(-\xi t).$$

As discussed above, we will assume that changes in  $D(t)$  and  $P(t)$  are homoskedastic and normally distributed. We will also assume that dividends and prices need to be first differenced to make them stationary (that is, they are first-order integrated), but that there is a particular linear combination of levels of dividends and prices which is stationary (that is, dividends and prices are cointegrated, sharing a common unit root). As Campbell and Shiller [1987] point out, cointegration follows naturally from the notion that dividends are first-order integrated and stock prices forecast future dividends. Finally, we will model  $D(t)$  and  $P(t)$  as sums of unobserved components, each of which follows a first-order autoregression (perhaps with a unit root) in continuous time.

Let  $r$  denote the time-invariant riskless rate of interest. In our model the particular linear combination of prices and dividends which is stationary is  $D(t) - (r-\xi)P(t)$ . The unconditional mean of this variable is

$$(2.02) \quad E[D(t) - (r-\xi)P(t)] = \lambda,$$

where  $\lambda$  is the unconditional expected excess return per detrended share of stock. Note that the units of  $\lambda$  are (real) dollars per share, not dollars

per dollar invested as would be the case for a percentage return.

This section proposes several specific models for the relationship between stock prices and dividends, all of which allow prices to be decomposed into the sum of "fundamental value"  $V(t)$  and "noise"  $Y(t)$ . We thus write

$$(2.03) \quad P(t) = V(t) - \lambda/(r-\xi) + Y(t).$$

The quantity  $V(t)$  measures the expected present value of future dividends conditional on information available to smart money investors, discounted at rate  $r$  with no risk adjustment. Since  $V(t)$  is calculated by discounting anticipated dividends at the riskless rate  $r$ , it is the price which would prevail if smart money investors were risk neutral. The rational expectation  $V(t)$  incorporates all information about future dividends contained in past dividends as well as contemporaneous non-dividend information available to smart money investors, but not available to economists with a dataset limited to prices and dividends. Because smart money investors are risk averse, the stock price  $P(t)$  deviates from  $V(t)$  by a constant term  $\lambda/(r-\xi)$  (the capitalized value of the unconditional expected excess return per share) and a zero-mean random variable  $Y(t)$  which represents noise. The noise makes expected returns per share fluctuate randomly through time, and in fact does nothing more than this.

The differences in the various models considered involve differences in the relationship between fundamentals  $V(t)$  and dividends  $D(t)$ . If  $V(t)$  incorporates non-dividend information, then increases in  $V(t)$  not related to increases in past dividends should anticipate increases in future



dividends. The models discussed below allow different specifications of the way in which  $V(t)$  anticipates future dividends and different specifications of the univariate dividend process to accommodate both a unit root and some degree of mean reversion in dividends.

The rest of this section describes two models (A and B) of the relationship between  $V(t)$  and  $D(t)$ , discusses a generalization (model C), and describes formally how noise is incorporated into the generalized model.

Model A.

According to model A, the actual dividend is the sum of two components,

$$(2.04) \quad D(t) = D_0(t) + D_1(t),$$

both of which are observed by smart money investors. The first component,  $D_0(t)$ , follows a Brownian motion while the second component,  $D_1(t)$ , follows a continuous-time AR(1) ("Ornstein-Uhlenbeck") process. The two components are independently distributed. Thus we have

$$(2.05) \quad dD_0 = \sigma_0 dz_0, \quad dD_1 = -\alpha_1 D_1 dt + \sigma_1 dz_1.$$

Here  $dz_0(t)$  and  $dz_1(t)$  are zero-mean, unit-variance (i.e. standard) Brownian motions which are independently distributed,  $\sigma_0$  and  $\sigma_1$  are the standard deviations of innovations in the two components of the dividend, and the parameter  $\alpha_1$  measures the speed of mean reversion in the second dividend component. The innovation variance of the dividend itself is just

$$\sigma_0^2 + \sigma_1^2.$$

Expected future dividends conditional on the information set of a smart money investor observing both components are given by

$$(2.06) \quad E(D(t+s) \mid \text{history of } D_0, D_1 \text{ to } t) = D_0(t) + \exp(-\alpha_1 s) D_1(t).$$

The fundamental value of the stock,  $V(t)$ , was defined as the discounted value of expected future dividends, conditional on the information of the smart money investor:

$$V(t) = E_t \int_{s=0}^{\infty} e^{-rs} D^o(t+s) ds = E_t \int_{s=0}^{\infty} e^{-(r-\xi)s} D(t+s) ds.$$

In Model A this becomes

$$(2.07) \quad V(t) = \frac{1}{r - \xi} D_0(t) + \frac{1}{r - \xi + \alpha_1} D_1(t).$$

It is not possible for economists, who observe only the history of the total dividend  $D(t) = D_0(t) + D_1(t)$ , to infer the values of  $D_0(t)$  and  $D_1(t)$  exactly. However it is possible to estimate these values. Let  $\hat{D}_0(t)$  and  $\hat{D}_1(t)$  denote, respectively, the expectations of the permanent and transitory components conditional on the history of the dividend process but not conditional on separate observation of the two components. Since the current dividend is observed, these expectations sum to the current dividend:

$$(2.08) \quad D(t) = \hat{D}_0(t) + \hat{D}_1(t) = D_0(t) + D_1(t).$$

Let  $I(t)$  denote the error on the transitory component. It satisfies

$$(2.09) \quad I(t) = \hat{D}_1(t) - D_1(t) - D_0(t) - \hat{D}_0(t).$$

The values  $\hat{D}_0(t)$  and  $\hat{D}_1(t)$  measure "dividend information". Similarly,  $I(t)$  measures "non-dividend information", i.e., it captures what would be known if both components were observed.

Now the expectations of future dividends  $D(t+s)$  and the fundamental value  $V(t)$ , which were given in (2.06) and (2.07) respectively, can be decomposed into dividend information and non-dividend information as follows:

$$(2.10) \quad E(D(t+s) \mid \text{history of } D_0 \text{ and } D_1 \text{ to } t) \\ = \hat{D}_0(t) + \exp(-\alpha_1 s) \hat{D}_1(t) + [1 - \exp(-\alpha_1 s)] I(t),$$

$$(2.11) \quad V(t) = \left[ \frac{1}{r - \xi} \hat{D}_0(t) + \frac{1}{r - \xi + \alpha_1} \hat{D}_1(t) \right] + \\ \left[ \frac{1}{r - \xi} - \frac{1}{r - \xi + \alpha_1} \right] I(t).$$

It can be shown using a Kalman filter argument that  $I(t)$  follows a univariate AR(1) process which cannot be forecast from  $\hat{D}_0(t)$  and  $\hat{D}_1(t)$ .

#### Model B.

Model B is based on the assumption that the fundamental valuation  $V(t)$  can be written as a linear combination of the dividend  $D(t)$  and a non-dividend information variable  $I(t)$  such that the following three properties hold:

1. The vector  $\langle D, I \rangle$  process is linear, i.e., it is a continuous-time first-order vector autoregression.

2. Current and future values of  $I(t)$  are independent of current and past values of  $D(t)$ .

3. The univariate  $D$  process is a Brownian motion, and the univariate  $I$  process is stationary.

Model A does not have the first or third of these properties. In Appendix A it is shown that these three assumptions imply that  $I(t)$  is a univariate AR(1) process. Thus, there exist constants  $\sigma_0$ ,  $\alpha_I$ ,  $\sigma_I$  such that we can write

$$(2.12) \quad dD_0(t) = \sigma_0 dz_0(t),$$

$$(2.13) \quad dI(t) = -\alpha_I I(t) + \sigma_I dz_I(t),$$

where  $dz_0(t)$  and  $dz_I(t)$  are standard Brownian motions, but are not independently distributed.

It is also shown in Appendix A that any bivariate process satisfying the assumptions of model B also satisfies the following:

$$(2.14) \quad \begin{bmatrix} dD_0 \\ dI \end{bmatrix} = \begin{bmatrix} 0 & \alpha_I \\ 0 & -\alpha_I \end{bmatrix} \begin{bmatrix} D \\ I \end{bmatrix} dt + \begin{bmatrix} \sigma_0 & 0 \\ -\rho_I \sigma_0 & [2\rho_I - \rho_I^2]^{1/2} \sigma_0 \end{bmatrix} \begin{bmatrix} dz_0^* \\ dz_I^* \end{bmatrix}$$

Here  $dz_0^*(t)$  and  $dz_I^*(t)$  are standard Brownian motions which are independently distributed. The covariance between  $dD_0$  and  $dI$  is now captured by the off-diagonal elements of the matrices in (2.14). (See (2.17) and (2.18) below for an explicit statement of the relation between  $dz_0$ ,  $dz_I$ ,  $dz_0^*$  and  $dz_I^*$ .)

For the purposes of discussion equation (2.14) can be abbreviated

$$(2.14)' \quad dy = Aydt + Cdz.$$

The matrices A and C contain three exogenous parameters:  $\alpha_I$ ,  $\sigma_0$ , and  $\rho_I$ . The parameters  $\alpha_I$  and  $\sigma_0$  are the same as those describing the univariate processes followed by dividends and prices in (2.12) and (2.13) above. The parameter  $\rho_I$  is proportional to  $\sigma_I^2$ :

$$(2.15) \quad \rho_I = \sigma_I^2 / 2\sigma_0^2.$$

The parameter  $\rho_I$  is the regression coefficient obtained when  $dI$  is regressed on  $-dD_0$ . It satisfies the constraint

$$(2.16) \quad 0 \leq \rho_I \leq 2.$$

This inequality, together with the restrictions in (2.14), places a "variance bound" on innovations in  $V(t)$  relative to innovations in dividends. It is generally true, as West [1988a] has shown, that extra information reduces the innovation variance of an asset price. The form of (2.14) ensures that the innovation variance of  $V(t)$  is no greater than it would be in the absence of superior information  $I(t)$  (i.e., it is no greater than the innovation variance of  $D(t)/(r-\xi)$ ). The restriction (2.16) ensures that the covariance of  $V(t)$  and  $D(t)$  is no greater in absolute value than it would be in the absence of superior information.

Since it takes eight parameters to specify two arbitrary  $2 \times 2$  matrices

A and C, our three-parameter model evidently incorporates five restrictions relative to a model in which A and C are completely unrestricted. A precise derivation of these five restrictions is given in Appendix A, but a brief intuitive discussion is given here. One restriction follows from our assumption that  $D(t)$  is not mean-reverting (is a Brownian motion), two of the restrictions are "normalizing conventions" and two are substantive implications of the assumption that the history of  $D$  cannot predict future  $I$ . In the  $2 \times 2$  matrix A, the upper left corner is zero because the Brownian motion  $D$  is not mean-reverting. The lower left corner is zero because of the assumption that dividends cannot predict non-dividend information. The parameter  $\alpha_I$  in the lower right corner is the univariate autoregressive parameter for  $I(t)$ . The fact that  $\alpha_I$  appears again in the upper right corner is a normalizing convention which (as described in Appendix A) scales the units in which non-dividend information  $I(t)$  is measured.

In the  $2 \times 2$  matrix C, the zero in the upper right corner is a normalizing convention based on the fact that it takes only three parameters to specify the relevant covariance structure implicit in C. The bottom row of C incorporates a restriction on the covariance of innovations in  $I(t)$  necessary to make it impossible to use past dividends to forecast future values of  $I(t)$ . The term  $-\rho_I \sigma_0$  in the lower left gives a negative correlation between changes in dividends and non-dividend information because when anticipated changes occur, the future changes expected from the same information are reduced. Note that  $\sigma_I^2$  is the sum of the squares of the elements in the bottom row of C.

Since the specification in (2.15) involves three parameters equivalent

to the three parameters describing the two univariate processes, model B is the most parsimonious way to model non-dividend information such that both dividends and non-dividend information follow AR(1) processes, with a unit root in the dividend.

It is clear from comparing (2.12) and (2.13) with (2.14) that the various Brownian motions must satisfy

$$(2.17) \quad \sigma_0 dz_0 = \alpha_1 I dt + \sigma_0 dz_0^*$$

$$(2.18) \quad \sigma_1 dz_1 = -\rho_1 \sigma_0 dz_0^* + [2\rho_1 - \rho_1^2]^{1/2} \sigma_0 dz_1^*$$

Equation (2.18) merely states that the sum of two Brownian motions is a Brownian motion, and that the variance is consistent with (2.15). The fact that the right-hand-side of (2.17) is a Brownian motion is one of those surprising properties of Brownian motions which follows from the fact that  $I$  is not anticipated by past changes in  $D$  (see Davis [1977]).

It is an implication of the discussion in Appendix A that

$$(2.19) \quad E\{D(t+s) \mid \text{history of } D, I \text{ to } t\} = D_0(t) + [1 - \exp(-\alpha_1 s)] I(t).$$

The expected discounted value of the flow of dividends is therefore

$$(2.20) \quad V(t) = \frac{1}{r - \xi} D_0(t) + \left[ \frac{1}{r - \xi} - \frac{1}{r - \xi + \alpha_1} \right] I(t).$$

### Differences between Models A and B.

Consider first the manner in which non-dividend information affects  $V(t)$ . In both models, the non-dividend information term  $I(t)$  follows an AR(1) process (unforecastable from dividends) which captures an expectation that dividends will drift, at a rate proportional to  $I(t)$ , to a level eventually  $I(t)$  units higher than that forecast from dividends alone (see (2.10) and (2.19)). In this sense, non-dividend information behaves similarly in both models.

Consider next differences in the univariate dividend process. In model B, dividends follow a random walk. In model A, dividends are the sum of a random-walk term and a mean-reverting term. Thus model A, but not model B, can accommodate some degree of mean reversion even when a unit root is present in the dividend process. In model A, however, the degree of mean reversion in the dividend process is linked inflexibly to the amount of non-dividend information incorporated into  $V(t)$ .

### Generalizations and Model C.

In order to make it possible to measure independently the presence of both non-dividend information and partial mean reversion in dividends, something more general than either model A or model B is needed. One generalization, which combines both models, allows the dividend to be the sum of two independent components, one of which is mean-reverting and each of which is forecast by stock market participants in the manner of model B. There are then two dividend components ( $D_0$  and  $D_1$ ) and two information terms ( $I_0$  and  $I_1$ ). Specifying such a model requires seven parameters. Non-dividend information arises from observing both components, as well as from observing separate signals about each component. Model B is obtained



as a special case when  $D_1$  and  $I_1$  are zeroed out, and model A is obtained as a special case when  $I_0$  and  $I_1$  are zeroed out.

Since this model contains more parameters than the data would make it possible to estimate accurately, model C is defined as the special case in which there is no special non-dividend information about the mean-reverting component of dividends. We then have two dividend components  $D_0(t)$  and  $D_1(t)$ , together with extra information  $I(t) = I_0(t)$  about the random walk component. The model has five parameters.

In comparison with model B, model C makes it possible for dividends to have a mean-reverting component, and in comparison with model A, the importance of this component is not tied inflexibly to the amount of non-dividend information in prices. In model C, the fundamental value  $V(t)$  is defined by

$$(2.21) \quad V(t) = \frac{1}{r - \xi} D_0(t) + \left[ \frac{1}{r - \xi} - \frac{1}{r - \xi + \alpha_1} \right] I(t) + \frac{1}{r - \xi + \alpha_1} D_1(t).$$

The coefficients multiplying  $I(t)$  and  $D_1(t)$  are determined by separate parameters  $\alpha_I$  and  $\alpha_1$ .

#### Specification of Model C with Noise.

The noise process  $Y(t)$  is modelled in the same way as the components of fundamental value, as a continuous-time AR(1) process. This implies that although prices are continuously perturbed by noise, the perturbed prices are continuously pulled back towards their fundamental value at a rate proportional to how far they are from this fundamental. Returns tend to be high when prices are below fundamentals and low when prices are above fundamentals. If the noise process were to have a unit root, there would

be no tendency for prices to return to their fundamental value, but this would not be the same as a rational speculative bubble because there is no compounding of returns on the noise component of prices.

The innovations in the noise term are allowed to be either independent of innovations in the components of fundamental value  $V(t)$ , or correlated with them. This makes it possible to model explicitly the hypothesis that prices "overreact" to new information, as well as the hypothesis that price noise is independent of fundamentals.

Including noise, there are four underlying components in model C:  $D_0$ ,  $I$ ,  $D_1$ , and  $Y$ . The observables  $D$  and  $P$  are the linear combinations

$$(2.22) \quad D = D_0 + D_1,$$

$$P = -\lambda/(r-\xi) + V + Y = -\lambda/(r-\xi) + \pi_0 D_0 + \pi_I I + \pi_1 D_1 + Y,$$

where the coefficients  $\pi_0$ ,  $\pi_I$ , and  $\pi_Y$  are defined by the expression (2.21) for  $V$ . That is,  $\pi_0 = 1/(r-\xi)$ ,  $\pi_I = 1/(r-\xi) - 1/(r-\xi+\alpha_I)$ , and  $\pi_1 = 1/(r-\xi+\alpha_1)$ . The four components satisfy the following vector process:

$$(2.23) \quad \begin{bmatrix} dD_0 \\ dI \\ dD_1 \\ dY \end{bmatrix} = \begin{bmatrix} 0 & \alpha_I & 0 & 0 \\ 0 & -\alpha_I & 0 & 0 \\ 0 & 0 & -\alpha_1 & 0 \\ 0 & 0 & 0 & -\alpha_Y \end{bmatrix} \begin{bmatrix} D_0 \\ I \\ D_1 \\ Y \end{bmatrix} dt + \begin{bmatrix} \sigma_0 & 0 & 0 & 0 \\ -\rho_I \sigma_0 & (2\rho_I - \rho_I^2)^{1/2} \sigma_0 & 0 & 0 \\ 0 & 0 & \sigma_1 & 0 \\ \theta_0 \gamma_0 & \theta_I \gamma_I & \theta_1 \gamma_1 & \theta_Y \gamma_Y \end{bmatrix} \begin{bmatrix} dz_0^* \\ dz_I^* \\ dz_1^* \\ dz_Y^* \end{bmatrix}$$

The parameters  $\theta_0$ ,  $\theta_I$ ,  $\theta_1$  and  $\theta_Y$  are endogenous constants defined by

$$(2.24) \quad \theta_0 = (\pi_0 - \pi_I \rho_I) \sigma_0, \quad \theta_I = \pi_I (2\rho_I - \rho_I^2)^{1/2} \sigma_0, \quad \theta_1 = \pi_1 \sigma_1,$$

$$\theta_Y = (\theta_0^2 + \theta_I^2 + \theta_1^2)^{1/2}.$$

The reduced form specified in (2.21)-(2.24) is the basic structure examined empirically below. It is an unobserved components model with various restrictions imposed. Note that there are ten parameters, five describing the "fundamentals" ( $\alpha_I, \alpha_1, \sigma_0, \rho_I, \sigma_1$ ) and five describing noise ( $\alpha_Y, \gamma_0, \gamma_I, \gamma_1, \gamma_Y$ ).

The three parameters  $\gamma_0, \gamma_I$  and  $\gamma_1$  measure three kinds of overreaction, and the parameter  $\gamma_Y$  measures the intensity of noise which is independent from fundamentals. The scaling parameters  $\theta_0, \theta_1$ , and  $\theta_I$  are defined so that price innovations can be written

(2.25)

$$dp - Edp = \theta_0(1 + \gamma_0)dz_0^* + \theta_I(1 + \gamma_I)dz_I^* + \theta_1(1 + \gamma_1)dz_1^* + \theta_Y\gamma_Y dz_Y^*.$$

Thus, the overreaction parameters  $\gamma_0, \gamma_I$  and  $\gamma_1$  measure the multiple by which prices overreact to fundamentals. For example,  $\gamma_0 = 2$  implies that in response to an innovation in the permanent dividend component  $dz_0$ , prices go up three times as much as the fundamental valuation  $V(t)$ . If one imposes the constraint  $\gamma_0 = \gamma_I = \gamma_1$  (as we shall do in our empirical work below), this says that prices overreact to innovations in  $V(t)$  with the same degree of intensity, regardless of which component of  $V(t)$  induces the innovation. The parameter  $\theta_Y$  scales the independent noise parameter  $\gamma_Y$  so that it measures the standard deviation of innovations in independent noise

as a multiple of the standard deviation of innovations in  $V(t)$ . Thus,  $\gamma_Y = 2$  implies that the innovations in independent noise have a standard deviation which is twice as great (and a variance four times as great) as the innovations in fundamental value.

The reduced form (2.23) can be abbreviated as  $dy = Aydt + Cdz$ . The first two zeros in the third rows and columns of both  $A$  and  $C$  make the dividend component  $D_0$  independent from  $D_1$ . The zeros in the fourth column of  $A$  say that fundamentals are not affected by noise. The zeros in the fourth row of  $A$  are a modelling assumption which says that noise does not react systematically to past levels of the fundamental components. All elements above the diagonal in the matrix  $C$  are zero because they are not necessary to specify the relevant covariance matrix.

Within model  $C$  are nested many special cases of particular interest. Model  $A$  is obtained from (2.23) by zeroing out  $I$  (setting  $\rho_I = \sigma_I = \gamma_I = 0$ ;  $\alpha_I$  becomes unidentified) and by dropping the second row and column of all matrices. Similarly, model  $B$  is obtained by zeroing out  $D_1$  (setting  $\sigma_1 = \gamma_1 = 0$ ;  $\alpha_1$  becomes unidentified) and by dropping the third row and column of all matrices. If prices are not affected by noise,  $Y$  can be zeroed out entirely (setting  $\gamma_0 = \gamma_I = \gamma_1 = \gamma_Y = 0$ ;  $\alpha_Y$  becomes unidentified), and the fourth row and column can be dropped. If there is "independent noise" but no overreaction, we zero out only the three overreaction parameters (setting  $\gamma_0 = \gamma_I = \gamma_1 = 0$ ). In our empirical work we shall assume that there is the same degree of overreaction to innovations in  $V(t)$  regardless of the source of the innovation, so we shall constrain the overreaction parameters to be the same (setting  $\gamma_0 = \gamma_I = \gamma_1$ ).

### 3. An Equilibrium Model of Smart Money and Noise Trading

In this section we show that the reduced form given in (2.21)-(2.24) is implied by an equilibrium model of stock market trading with the following assumptions:

1. Smart money investors are infinitely lived, have exponential utility, and form expectations rationally based on observing the  $D_0(t)$ ,  $D_1(t)$ ,  $I(t)$ , and  $P(t)$  processes.

2. Noise traders trade randomly and their aggregate stock position follows an Ornstein-Uhlenbeck process. In the resulting equilibrium, this process is perfectly correlated with the price-noise process  $Y(t)$  and perfectly negatively correlated with the risk premium on stocks.

3. The number of smart money investors and the number of noise traders grows through time at the exogenous rate  $\xi$ .

The discussion in this section is divided into three parts. First, the investment opportunity set implied by the reduced form (2.24) is described. Second, the corresponding optimal portfolio and consumption rules of an infinitely lived smart money investor with exponential utility are calculated. Third, assumptions about noise trading and population growth which make markets clear at all times are obtained.

#### Investment Opportunities.

Traders can invest in two assets: a riskless asset yielding constant real return  $r$  and a risky stock whose returns can be deduced from the system (2.21)-(2.24). In describing stock returns it is useful to make the notational assumption that "stock dividends" are paid continuously at rate  $\xi$ . By normalizing the number of shares at time zero to unity, the number of shares outstanding at time  $t$  becomes  $e^{\xi t}$ . This notational convention

makes the per capita number of shares outstanding constant in an equilibrium with investor population growth at rate  $\xi$ , and makes returns per share stationary, because the price of one share of stock is  $P(t)$ .

To describe investment opportunities in the stock market, let  $M(t)$  denote the undiscounted cumulative cash flow from a zero-wealth portfolio long one share of stock (financed fully by borrowing at the riskless rate of interest). The process  $M(t)$  satisfies the stochastic differential equation

$$(3.01) \quad dM(t) = e^{-\xi t} [D^0(t)dt + dP^0(t) - rP^0(t)dt],$$

where the three terms in brackets are, respectively, the dividends on the stock market, the capital gains on the stock market, and the financing cost of holding the stock market. The term  $e^{-\xi t}$  deflates quantities to per-share units. Note that sales of stock at rate  $\xi$  to keep investment at one share do not generate cash flows which affect the value of the portfolio because the proceeds are used to retire the debt which finances the position.

Define the process  $N(t)$  by

$$(3.02) \quad N(t) = -(r - \xi + \alpha_Y)Y(t).$$

Note that  $N(t)$  is perfectly negatively correlated with  $Y(t)$ . Now we show that  $\langle M(t), N(t) \rangle$  has a simple bivariate representation:

**THEOREM 3.1.** There exists a vector Brownian motion  $\langle z_M(t), z_N(t) \rangle$  with univariate innovation variances normalized to unity and with correlation  $dz_M dz_N = \zeta dt$ , such that the vector process  $\langle M(t), N(t) \rangle$  satisfies

$$(3.03) \quad dM(t) = [\lambda + N(t)]dt + \sigma_M dz_M,$$

$$(3.04) \quad dN(t) = -\alpha_Y N(t)dt + \sigma_N dz_N.$$

The variances  $\sigma_M^2$  and  $\sigma_N^2$  and the correlation  $\zeta$  are given by

$$(3.05) \quad \sigma_M^2 = \left[ \pi_0 \sigma_0 - \rho_I \sigma_0 + \theta_0 \gamma_0 \right]^2 + \left[ (2\rho_I - \rho_I^2)^{1/2} \sigma_0 + \theta_I \gamma_I \right]^2 \\ + (\pi_1 \sigma_1 + \theta_1 \gamma_1)^2 + (\theta_Y \gamma_Y)^2$$

$$\sigma_N^2 = (r - \xi + \alpha_Y)^2 \left[ \theta_0^2 \gamma_0^2 + \theta_I^2 \gamma_I^2 + \theta_1^2 \gamma_1^2 + \theta_Y^2 \gamma_Y^2 \right]$$

$$\zeta = -(r - \xi + \alpha_Y) \left[ (\pi_0 \sigma_0 - \rho_I \sigma_0 + \theta_0 \gamma_0) \theta_0 \gamma_0 + \left[ (2\rho_I - \rho_I^2)^{1/2} \sigma_0 + \theta_I \gamma_I \right] \theta_I \gamma_I \right. \\ \left. + (\pi_1 \sigma_1 + \theta_1 \gamma_1) \theta_1 \gamma_1 + \theta_Y^2 \gamma_Y^2 \right] / \sigma_M \sigma_N.$$

The proof of this theorem is given in Appendix B.

Equations (3.03)-(3.05) describe succinctly the investment opportunity set of an investor, and it is only this structure which is relevant in the rest of this section. Investment opportunities have the following properties.

1. Expected returns on a share of stock (whose price is  $P(t)$ ) are  $rP(t) + \lambda + N(t)$ . Investment opportunities are not constant through time, because the risk premium  $\lambda + N(t)$  on stocks at time  $t$  deviates from its long-run mean  $\lambda$  by a random amount  $N(t)$ . Thus, the random variable  $N(t)$ , which follows an Ornstein-Uhlenbeck process, is a state variable characterizing the investment opportunities available at time  $t$ .

2. The quantity  $\sigma_M^2$  is the variance of returns on a share of stock. When  $\gamma_Y$  is nonzero or  $\gamma_0$ ,  $\gamma_1$  or  $\gamma_I$  are positive, then the variance of stock prices is increased relative to the "fundamental variance" which would prevail if  $\gamma_0 = \gamma_1 = \gamma_I = \gamma_Y = 0$ . In this sense, noise increases the volatility of prices. It is conceptually possible, however, for the volatility of prices to be reduced by virtue of noise if any of  $\gamma_0$ ,  $\gamma_1$  or  $\gamma_I$  are negative, i.e., if prices "under-react" to new information about fundamentals.

3. When  $\gamma_0$ ,  $\gamma_1$  and  $\gamma_I$  are non-negative (and  $\gamma_0$ ,  $\gamma_1$ ,  $\gamma_I$  and  $\gamma_Y$  are not all zero), the parameter  $\zeta$  is negative, making innovations in prices negatively correlated with innovations in expected returns. Intuitively, this occurs because innovations in the noise process  $Y(t)$ , which push prices up temporarily, must be accompanied by lower expected returns in the future in order for prices to drift back towards their fundamental value in the long run. Thus, even when the noise process  $Y(t)$  is independent from the fundamental processes  $D(t)$  and  $I(t)$ , innovations in expected risk premiums are negatively correlated with price innovations.

#### Intertemporal Optimization by a Smart Money Investor.

Let  $X(t)$  denote the number of shares of the speculative asset held by a smart money investor, and let  $C(t)$  denote the investor's continuous



consumption stream. Then the investor's wealth  $W(t)$  satisfies

$$(3.06) \quad dW(t) = rW(t)dt + X(t)dM(t) - C(t)dt \\ = rW(t)dt + X(t)[(\lambda + N(t))dt + \sigma_M dz_M(t)] - C(t)dt.$$

We assume the smart money investor is infinitely lived, has exponential utility, and observes both  $P(t)$  and  $N(t)$ . The value function is

$$(3.07) \quad V(W, N) = \max E_0 \int_{t=0}^{\infty} -e^{-(\rho t + \psi C)} dt,$$

where  $\rho$  is the time preference parameter and  $\psi$  the measure of risk aversion.

This formulation allows a tractable solution to the investor's optimization problem:

**THEOREM 3.2.** The investor's value function is given by

$$(3.08) \quad V(W, N) = \frac{1}{r} e^{-(\psi rW + \phi_0 + \phi_1 N + \phi_2 N^2/2)},$$

where  $\phi_0$ ,  $\phi_1$ , and  $\phi_2$  are the unique constants which solve the equations

$$(3.09) \quad (r/2 + \alpha_Y)\phi_2 + \psi\phi_2^2\sigma_N^2 - (1 - \phi_2\sigma_M\sigma_N)^2/2\sigma_M^2 = 0, \quad \phi_2 > 0$$

$$(r + \alpha_Y - \phi_2\sigma_N^2)\phi_1 - (\lambda - \phi_1\psi\sigma_M\sigma_N)(1 - \phi_2\psi\sigma_M\sigma_N)/2 = 0,$$

$$r\phi_0 = -r + \psi - (\phi_1^2 - \phi_2)\sigma_N^2/2 + (\lambda - \phi_1\psi\sigma_M\sigma_N)^2/2\sigma_M^2.$$

The investor's optimal consumption and portfolio rules are

$$(3.10) \quad C(t) = rW(t) + (1/\psi)[\Phi_0 + \Phi_1 N(t) + \Phi_2 N^2(t)/2],$$

$$(3.11) \quad X(t) = \frac{\lambda + N(t) - [\Phi_1 + \Phi_2 N(t)]\zeta\sigma_M\sigma_N}{\psi r\sigma_M^2}.$$

The proof of this theorem is in Appendix B.

Before proceeding further, it is useful to take a closer look at the investor's optimal portfolio and consumption rules. Because of the exponential utility assumption, the smart money investor's demand for the risky asset does not depend on wealth. (This is one of the features of the model which makes it tractable, since we do not need to analyze the evolution of the smart money investor's wealth.) In the demand function (3.11), the term involving  $\Phi_1$  and  $\Phi_2$  represents what is usually called "hedging demand" (against changes in future investment opportunities). In the presumed case where  $\zeta$  is negative (and because  $\Phi_2$  is positive), the investor responds more elastically to changes in expected returns than if his demand were "myopic". The demand function for a myopic investor who ignored the fact that changes in future expected returns are correlated with returns would be obtained by setting  $\Phi_1 - \Phi_2 = 0$  in the forward-looking demand function (3.11). Intuitively, when price fluctuations have a temporary component due to noise, the infinitely-lived investor is willing to expand demand because he realizes that some of the risk he bears today is reversed in the long run.

Now consider the consumption function (3.10). The exponential utility assumption implies that if the investor obtains a gift which is added to

his wealth, his optimal consumption policy is to invest the gift in a perpetuity and consume the interest payments, while following with the rest of his wealth the consumption and portfolio policy optimal in the absence of the gift. Changes in the state variable  $N(t)$  also affect consumption.

#### Market Clearing with Noise Trading.

For the smart money investor's optimization problem described above to be part of a general equilibrium, it is merely necessary to specify an exogenous noise trading process such that noise traders each period sell what smart money investors buy, and smart money investors and noise traders together hold the aggregate outstanding supply of stock. To do this let us assume that the number of smart money investors and the number of noise traders grow at rate  $\xi$ . Normalizing the initial population of each group to unity, the number of each type of investor at time  $t$  is  $e^{\xi t}$ . This normalization makes the per capita supply of stock equal to one share per smart money investor (or noise trader). Thus, markets clear (and the aggregate supply of stock is held) if the representative noise trader's demand for stock at time  $t$  is  $1 - X(t)$  shares, where  $X(t)$  is the demand of a smart money investor in (3.11). According to this assumption, the holdings of noise traders consist of a constant plus an amount which is a linear function of  $N(t)$  (or  $Y(t)$ ). Since the holdings of smart money investors are increasing in  $N(t)$  (or decreasing in  $Y(t)$ ), the holdings of noise traders are decreasing in  $N(t)$  (or increasing in  $Y(t)$ ). Thus, there is a direct proportionality between price noise  $Y(t)$  and the deviation of holdings by noise traders from their mean.

These assumptions about noise trading and growth in the number of investors give a general equilibrium model which implies (2.21)-(2.24).

#### 4. Econometric Methods and Empirical Results

In this section we estimate the reduced form (2.21)-(2.24) on annual U.S. time series data for real stock prices and dividends. The data are taken from Campbell and Shiller [1987], and are similar to data used in other studies such as Engle and Watson [1985], Mankiw, Romer and Shapiro [1985], Shiller [1981] and West [1988a]. Our real stock price is the Standard and Poors Composite Stock Price Index for January, measured in each year from 1871 to 1986, divided by the January producer price index. (Before 1900 an annual average producer price index is used. Data from before 1926 are from Cowles [1939].) Our real dividend series is the corresponding dividend per share adjusted to index, measured each year from 1871 to 1985, divided by the annual average producer price index. Our econometric methods will take account of the fact that the price data are point-sampled, while the dividend data are time-averaged.

We write the raw discrete-time data as  $D_t^o$  and  $P_t^o$ . We will use the notational convention that  $D_t^o$  is the dividend paid during year  $t$ , and  $P_t^o$  is the stock price at the end of year  $t$ . This convention differs from that in Campbell and Shiller [1987,1988a,1988b] because it is helpful for us to define backward rather than forward time averages. The first step in our analysis is to transform the raw data in the manner of equation (2.01), dividing by an exponential trend  $\exp(\xi t)$ . We choose  $\xi$  to equal the mean dividend growth rate over the sample, 0.013, and write the transformed discrete-time data as  $D_t$  and  $P_t$ . We also normalize  $D_t$  and  $P_t$  so that the sample mean of  $P_t$  equals one.

This transformation has two important effects. First, it removes exponential growth from the ex-ante mean of the data; this effect may be

called "detrending", although it is important to note that the transformation does not force the data to revert to a trend line. Indeed, we will assume that there is a unit root in the detrended series. Secondly, the transformation removes exponential growth from the variance of the data; this may be called "scaling", and is similar to the effect of a log transformation.

#### Preliminary data analysis.

Before we estimate the system (2.21)-(2.24) and restricted versions of it, it is important to confirm that the data are not grossly at variance with this system. As a preliminary therefore, we plot the exponentially detrended price and dividend series in Figure 1 (the dividend is multiplied by 10 in the figure), and we summarize some of the main features of the data in Tables 1 and 2.

Table 1 presents Dickey-Fuller tests of the null hypotheses that the series  $D_t^0$ ,  $P_t^0$ ,  $D_t$  and  $P_t$  have unit roots in their univariate time series representations. For comparison, we also carry out the tests for  $\ln(D_t^0)$  and  $\ln(P_t^0)$ . The test statistics are formed from a regression of the change in the series on a constant, time trend and lagged level. We report both straight Dickey-Fuller test statistics, and statistics adjusted for fourth-order serial correlation as proposed by Phillips [1987] and Phillips and Perron [1986].

The results of the tests are rather mixed. The null hypothesis of a unit root in dividends is not rejected for  $D_t^0$ , but is rejected at the 10% level for  $D_t$ . This rejection is not just due to detrending by the sample mean growth rate of dividends; the null is rejected even more strongly, at the 5% level, for  $\ln(D_t^0)$ . However there is no evidence against the null

hypothesis for stock prices.

Similarly mixed results were obtained in Campbell and Shiller [1987, 1988a]. As in those papers, we proceed to assume that there is in fact a unit root in stock prices and dividends. We do this because in the present context it is the most conservative assumption. A unit root in dividends will tend to reduce the role of noise in explaining stock price volatility, by giving a greater role to movements in expected future dividends (Kleidon [1986], Marsh and Merton [1986]).

Our model assumes not only that prices and dividends are integrated processes, but also that they are cointegrated. At the bottom of Table 1 we estimate the "cointegrating regression" of  $D_t$  on  $P_t$ . As Stock [1987] has shown, the coefficient in this regression converges rapidly (at a rate proportional to the sample size) to the true parameter which defines the cointegrating vector, the unique stationary linear combination of  $D_t$  and  $P_t$ . Engle and Granger [1987] show how one can test the null hypothesis that two series are not cointegrated by running Dickey-Fuller or Augmented Dickey-Fuller tests on the residual from the cointegrating regression, and we also perform these tests. We find fairly strong evidence that  $D_t$  and  $P_t$  are cointegrated.

A striking feature of the cointegrating regression results is the low interest rate which they imply. Having normalized the data so that the sample mean of  $P_t = 1$ , we estimate  $D_t = 0.030 + 0.016P_t$ . Thus about two thirds of the sample mean of  $D_t$  is attributed to the constant term. In our model, the coefficient on  $P_t$  equals  $(r-\xi)$ , the interest rate less the trend growth rate (0.013), so the implied interest rate is 0.029. The constant term equals  $\lambda$ , the unconditional expected excess return per share of stock

demanded by risk-averse smart money investors<sup>4</sup>. It is important to note that the coefficient  $(r-\xi)$  can be so low in this regression, only because  $\lambda$  is high. If we impose  $\lambda = 0$ , we estimate  $(r-\xi) = 0.050$ , the ratio of the sample mean dividend to the sample mean price<sup>5</sup>, implying  $r = 0.063$ .

It is known that in finite samples the coefficient in the cointegrating regression is seriously biased downwards (Banerjee, Dolado, Hendry and Smith [1986], Stock [1987]). Indeed, when we reverse the direction of the regression, making  $P_t$  the dependent variable, we estimate  $1/(r-\xi) = 23.7$ , implying  $(r-\xi) = 0.042$  and  $r = 0.055$ . This interest rate is closer to conventional estimates of the discount rate on stock, although it is still considerably less than the mean rate of return on our stock index (0.082).

Since the cointegrating regression does not yield a reliable estimate of the interest rate  $r$ , we estimate our model under several different assumptions about the interest rate. We fix it a priori at 4% and 6%, allowing a free coefficient  $\lambda$ ; we estimate  $r$  as part of our model, but impose  $\lambda = 0$ ; and we allow the data to estimate both  $r$  and  $\lambda$  freely.

In Table 2 we summarize some of the other time-series properties of our data. The table reports the sample standard deviations of  $\Delta D_t$  and  $\Delta P_t$ , and the sample correlations between  $\Delta D_t$ ,  $\Delta P_t$  and their lags. The stock price

---

<sup>4</sup> Recall that because stock prices and dividends are normally distributed, and smart money investors have constant absolute risk aversion, they demand compensation for risk in the form of a constant discount on the stock price (or premium on the dividend), rather than by discounting expected future dividends at a higher interest rate.

<sup>5</sup> Campbell and Shiller [1987] ran a similar cointegrating regression on the raw levels (rather than the detrended levels) of stock prices and dividends. They found a similar constant term and low implied interest rate. In their model, the constant term was supposed to be related to linear trend growth in dividends; but it had the wrong sign for this interpretation. Their model did not allow for any equivalent of our risk adjustment term  $\lambda$ .

series is very noisy relative to the dividend series; the standard deviation of  $\Delta D_t$  is only 0.032 times the standard deviation of  $\Delta P_t$ . This is of course the well-known "volatility" of stock returns. We shall see that our model can fit this feature of the data only by combining a highly persistent dividend process with a very low interest rate, or by attaching considerable importance to noise in the stock price.

The sample correlations of the series are also of interest. The first autocorrelation of  $\Delta D_t$  is 0.17, and the next four autocorrelations are all negative. Since  $D_t$  is a time-averaged series, one would expect the first autocorrelation of  $\Delta D_t$  to be positive. Indeed, if the continuous-time process  $D(t)$  were a random walk, the first autocorrelation of  $\Delta D_t$  would be 0.25 and the higher autocorrelations would all be zero (Working [1960]). The data therefore suggest that there is some mean-reverting component in the dividend (which is presumably also responsible for the rejection of the unit root hypothesis for dividends in Table 1).

The autocorrelations of the change in the stock price are generally smaller in absolute value (although the second autocorrelation is large at -0.22), and more erratic; there is less evidence of mean-reversion in the first few autocorrelations of this series. Finally, there is a surprisingly low contemporaneous correlation of 0.08 between  $\Delta D_t$  and  $\Delta P_t$ , but a high correlation of 0.57 between  $\Delta D_t$  and  $\Delta P_{t-1}$ . The correlation between  $\Delta D_t$  and  $\Delta P_{t-2}$  is also quite high at 0.21. We shall see that our model has some trouble matching these cross-correlations; "superior information"  $I(t)$  is one element which helps it to do so<sup>6</sup>.

---

<sup>6</sup> In interpreting these numbers, it is important to keep in mind the timing of the data. The correlation of 0.08 is between the price change from the end of one year to the end of the next, and the dividend change



### Estimating our model.

The next step in our analysis is to estimate the reduced form model (2.21)-(2.24) from our discrete-time detrended data. The method we use to do this is discussed in Appendix C. There we start from a continuous-time vector first-order process, of which our model is a special case. We show that a stacked vector of point-sampled and time-averaged transformations of the continuous-time variables follows a discrete-time vector AR(1), and we show how the discrete-time transition and variance-covariance matrices are related to the underlying continuous-time parameters.

Of course, we do not observe point-sampled and time-averaged transformations of all the variables in our model. Instead, we observe one point-sampled variable,  $P = -\lambda/(r-\xi) + \pi_0 D_0 + \pi_1 I + \pi_1 D_1 + Y$ , and one time-averaged variable,  $D = D_0 + D_1$ . However we can still estimate the system by using a Kalman filter to construct a likelihood function for any set of parameter values<sup>7</sup>.

Our approach is somewhat different from that of Grossman, Melino and Shiller [1987], who also estimate a continuous-time model from discrete-time data. Grossman, Melino and Shiller start with a mix of point-sampled

---

from one year to the next. The dividend is time-averaged over the period preceding the price measurement.

<sup>7</sup> The Kalman filter is described in Harvey [1981] and elsewhere. Our application is standard, but two details are worth mentioning. First, our state vector is nonstationary so we initialize the filter using the variance-covariance matrix of the system when it is perturbed slightly to make it stationary. We then drop the first element of the vector which makes up the likelihood function (that is, we condition our estimation on the first observation). Secondly, we can drop two of the time-averaged state variables, the time-average of  $I$  and the time-average of  $Y$ , from the state vector in estimation. This is because they are neither observed, nor do they play any role in forecasting the state of the system if the point-sampled  $I$  and  $Y$  are included.

and time-averaged data, but they time-average their point-sampled data and work with a vector of time-averaged variables alone. This vector follows a discrete-time ARMA(1,1) process, which can be technically difficult to estimate. (For further details, see Melino [1985]). Harvey and Stock [1985,1986] have also developed estimation methods for continuous-time models like ours.

#### Model A.

In Table 3 we report the maximized log likelihoods for several different variants of Model A. Each row of the table corresponds to a different assumption about the interest rate  $r$ . As discussed above, we estimate the model fixing  $r$  equal to 4% and 6%, with a free  $r$  but imposing  $\lambda$  equal to zero, and with a free  $r$  and  $\lambda$ . In the latter two cases we report the estimated  $r$  in parentheses below the log likelihood for the model.

Each column of the table corresponds to a different assumption about the role of noise. In the first column, we assume that there is no noise at all ( $\gamma_Y = \gamma_0 = \gamma_1 = 0$ ). In the second column we allow for noise which is independent of fundamentals (free  $\gamma_Y$ ,  $\gamma_0 = \gamma_1 = 0$ ). In the third column we allow for "overreaction" noise which is correlated with fundamentals, but no independent noise ( $\gamma_Y = 0$ , free  $\gamma_0 = \gamma_1$ ). In the last column we allow for both types of noise (free  $\gamma_Y$ , free  $\gamma_0 = \gamma_1$ ). Throughout we impose the restriction that the stock price overreacts to innovations in fundamentals in the same way, regardless of the source of the innovation ( $\gamma_0 = \gamma_1$ )<sup>8</sup>.

In Table 3 it is clear that exogenous serially correlated noise helps Model A to fit the data. But it is more helpful for some interest rate

---

<sup>8</sup> We experimented with relaxing this restriction, and found that it could never be rejected. In some cases our maximum likelihood algorithm had difficulty estimating the two parameters  $\gamma_0$  and  $\gamma_1$  separately.

specifications than others. When a higher interest rate is imposed (directly by fixing  $r = 6\%$  or indirectly by fixing  $\lambda = 0$ ), the noise columns have dramatically higher likelihoods than the no-noise column. But when a low interest rate of  $4\%$  is imposed, the difference in likelihood is much smaller. Finally, when  $r$  and  $\lambda$  are estimated freely, low estimates of  $r$  are obtained (in the range  $2.9-3.4\%$ ), and the role of noise is smaller still. (Even in this column, however, "overreaction" noise is significant at the  $4.7\%$  level if one uses a  $\chi^2$  test with 2 degrees of freedom.)<sup>9</sup>

The reason for this pattern of results is that the data show an important mean-reverting component in the dividend, but a volatile stock price. The model can fit these characteristics of the data without giving an important role to noise, only if the interest rate is very low. And a low interest rate is consistent with the mean relative levels of  $P_t$  and  $D_t$ , only if  $\lambda$  is fairly large and positive.

Table 4 gives more details for two of the models estimated in Table 3, those with free  $r$  and  $\lambda$ , and either no noise or full noise. (In fact the full noise model converged almost to the overreaction model.) The table gives parameter estimates with asymptotic standard errors, and the implied values of  $\pi_0$  and  $\pi_1$ . Below this is a "normalized innovations variance-covariance matrix". This is the implied variance-covariance matrix of innovations in  $P$ ,  $Y$ ,  $V$ ,  $\pi_0 D_0$ ,  $\pi_1 D_1$ , and  $\pi_1 I$ , where each element has been divided by the innovations variance of  $V$ . Thus the [1,1] element is the ratio of the variance of stock price innovations to the variance of

---

<sup>9</sup> A test of a "no-noise" null against an alternative with noise runs into the difficulty that one of the parameters under the alternative,  $\alpha_Y$ , is unidentified under the null. This causes well-known problems with statistical inference. However a  $\chi^2(2)$  test should be conservative.

innovations in fundamentals, i.e., the excess volatility of stock price innovations. The [4,4] element is the ratio of the innovations variance of the random walk dividend component in fundamentals, to the total innovations variance of fundamentals. That is, it measures the contribution of the random walk dividend term to the innovations variance of fundamentals; similarly, the [5,5] element measures the contribution of the transitory dividend term.

Finally, the table reports some of the implied moments of the observable data, so that one can see which features of the data are well fit by the model and which are not. The table gives the implied standard deviations of  $\Delta D_t$  and  $\Delta P_t$ , and the correlations of  $\Delta D_t$ ,  $\Delta P_t$  and their lags. These can be compared directly with the sample moments given in Table 2.

The first part of Table 4 summarizes the model with free  $r$  and  $\lambda$ , and no noise. In explaining dividend movements, the parameter estimates give an important role to the transitory component of the dividend. But this component does not have much effect on the stock price because it dies out quickly so its expected present value is small. (Thus  $\pi_0 = 60.4$  while  $\pi_1 = 2.2$ ). The innovations variance of the stock price exactly equals the innovations variance of fundamentals (since there is no noise), and almost exactly equals the innovations variance of the random walk dividend component.

This model fits some moments of the data quite well, notably the relative standard deviations of  $\Delta D_t$  and  $\Delta P_t$  (by picking a low interest rate) and the autocorrelations of  $\Delta D_t$ . But it fits other moments poorly; the implied correlation of  $\Delta D_t$  and  $\Delta P_t$  is 0.30, equal to the implied correlation of  $\Delta D_t$  and  $\Delta P_{t-1}$ , whereas in the data these correlations are

0.08 and 0.57 respectively. In the data stock prices anticipate dividend changes but are not highly correlated with them contemporaneously; Model A without noise has no good way to fit this.

The second part of the table gives details for the model with noise. The parameter  $\gamma_0$  is estimated at 1.1, implying that the stock price reacts twice as much as it should do to a dividend innovation. Accordingly the innovations variance of the stock price is four times the innovations variance of fundamentals. But the noise is highly transitory;  $\alpha_Y$  is estimated at 5.6, implying that the noise disappears almost completely within a year. (The fraction of the noise remaining after a year is  $\exp(-\alpha_Y) = 0.36\%$ .) The effect of this type of noise is to reduce the correlation of  $\Delta D_t$  and  $\Delta P_t$  to 0.20, and to increase the correlation of  $\Delta D_t$  and  $\Delta P_{t-1}$  to 0.35.

#### Model B.

Table 5 reports the maximized log likelihoods for alternative versions of Model B. The results are similar to those in Table 3, in that noise makes a bigger difference to the likelihood when a relatively high interest rate is imposed. However in Table 5 the no-noise models can always be rejected in favor of the noise models at very high levels of confidence. In the final row, for example, with a free  $r$  and  $\lambda$ , the no-noise model can be rejected against the full noise model at the 0.04% level with a  $\chi^2$  test with 3 degrees of freedom.

Table 6 gives detailed results for the versions of Model B with free  $r$  and  $\lambda$ , and either no noise or full noise. In the model with no noise, the stock price innovations variance of course equals the fundamentals innovation variance. But the innovations variance of  $\pi_0 D_0$  is 16% higher

because the presence of superior information reduces the innovations variance of fundamentals in the manner analyzed by West [1988a].

Model B, with or without noise, is unable to capture the negative higher-order autocorrelations of dividends. The implied first-order autocorrelation is 0.25, and all higher-order autocorrelations are zero. (As Working [1960] showed, this must always be true for a time-averaged Brownian motion.) Without noise, Model B also has difficulty with the correlations of  $\Delta D_t$  and  $\Delta P_t$ , and  $\Delta D_t$  and  $\Delta P_{t-1}$  (implied equal to 0.25 and 0.39 respectively). When noise is added to the model, it is able to fit these correlations much more accurately (0.17 and 0.57 respectively). It does so by lowering the interest rate estimate from 0.047 to 0.037, and estimating that the stock price underreacts to dividend innovations. This is a rather anomalous result which disappears when one fixes the interest rate or allows for a transitory component in the dividend.

#### Model C.

Finally, in Tables 7 and 8 we give empirical results for our most general specification, Model C. Once again there is a critical interaction between the level of the interest rate and the role of noise. With a low interest rate of 4% or below, quite high likelihoods can be achieved without noise. With an interest rate of 5% or 6%, however, noise is essential to the fit of the model. When the interest rate  $r$  and the parameter  $\lambda$  are freely estimated, a low interest rate of 3.5% is chosen in the no-noise specification and a high interest rate of 9.6% is chosen in the specifications which allow for overreaction noise. The no-noise model can be rejected against the overreaction model at the 10.3% level using a  $\chi^2$  test with 2 degrees of freedom.

In Table 8 we present detailed results for three specifications. These have a free  $r$  and  $\lambda$  with no noise ( $r$  is estimated at 3.5%), a free  $r$  and  $\lambda$  with full noise ( $r$  is estimated at 9.6% and the noise is estimated to be pure overreaction), and a fixed  $r$  of 6% with full noise (again the noise is estimated to be pure overreaction). The last specification is included to illustrate the effect of the interest rate on the quantitative importance of noise.

The specifications in Table 8 all imply roughly the same decomposition of the dividend  $D$  into a random walk component  $D_0$  and a stationary component  $D_1$ . This decomposition is illustrated in Figure 2 for the model with a free  $r$  and  $\lambda$  and full noise. The major movements in  $D$  are matched by movements in  $D_0$ , but there are some high-frequency movements in  $D$  which are attributed to temporary fluctuations in  $D_1$ . The latter have little effect on stock prices, since  $\pi_1$  is small relative to  $\pi_0$ .

The decomposition of the stock price into fundamental value  $V$  and noise  $Y$  is much more sensitive to model specification. The noise always has the effect of amplifying movements in fundamental value, but its variability depends on the interest rate assumed. In Figure 3 we plot  $V$  and  $P$  for the free  $r$  and  $\lambda$  noise model with  $r = 9.6\%$ ; in Figure 4 we plot  $V$  and  $P$  for the  $r = 6\%$  noise model. In the  $r = 9.6\%$  model the parameter  $\gamma_0$  is estimated at 3.3, implying that the stock price moves 4.3 times too much in response to an innovation in  $V$ , and that its innovations variance is 18.5 times that of fundamentals. The noise is also quite persistent; the parameter  $\alpha_Y$  is estimated at only 0.05, which implies that 95% of the noise remains after 1 year. In the  $r = 6\%$  model the parameter  $\gamma_0$  is estimated at 1.4, implying that the stock price moves 2.4 times too much in response to an innovation

in  $V$ , and that its innovations variance is 5.8 times that of fundamentals. The noise is again persistent, with an  $\alpha_Y$  estimate of 0.03 implying that 97% of the noise remains after 1 year.

It is apparent from Table 8 that Model C, in any of its variants, is much better able than Models A or B to fit the autocorrelations of dividend changes and the correlations of dividend changes with current and lagged price changes. In comparing the theoretical autocorrelations with the actual ones in Table 2, the only major phenomenon which is not fit by the model is the large negative second autocorrelation of the stock price change.

#### How likely are negative dividends and stock prices?

Table 8 can also be used to gain insight into one possible problem with our approach. As we noted above, it is theoretically possible in our model for dividends and stock prices to become negative. This is not particularly bothersome if the parameter estimates imply that negative dividends are unlikely to occur. (After all, the normal distribution was originally proposed to describe the distribution of human heights.) But if negative dividends have a substantial probability, this reduces the plausibility of our results.

To study this issue, we consider the parameter estimates reported in Table 8 for the model with free  $r$  and  $\lambda$  and no noise. (Other models estimated in Table 8 are similar.) The random walk component of the dividend has a 1-year conditional variance of  $\sigma_0^2$ , and a  $t$ -year conditional variance of  $\sigma_0^2 t$ . The stationary component of the dividend has a 1-year conditional variance of  $\sigma_1^2$ , and a  $t$ -year conditional variance which approaches  $\sigma_1^2 / 2\alpha_1$  as  $t$  increases. For simplicity, we will ignore this



term, which becomes asymptotically negligible relative to the conditional variance of the random walk component.

As  $t$  increases, of course, the conditional variance of the random walk component of the dividend grows without limit and the probability of observing a negative dividend approaches unity. But over 115 years (the length of our sample), the conditional variance is 0.0017 and the conditional standard deviation is 0.0411. The probability that the random walk is negative after 115 years, given the initial dividend value of 0.0400, is 0.16. The probability that it is negative at some point during the 115 years is twice the probability that it is negative at the end (Ingersoll [1987], p.353). Thus our parameter estimates do imply a one-third probability of negative dividends at some point during the sample period.

How robust are the results to changes in specification?

All of the models estimated so far have a sample period 1871-1986, and a fixed trend growth rate  $\xi = 0.013$ . As a check on the robustness of our results, we estimated Model C over subsamples 1871-1925 and 1926-1986 with  $\xi = 0.013$ , and over 1871-1986 with  $\xi$  set to 0.01, 0.015 and 0.02.

When Model C is estimated over the first and second halves of the sample separately, we find qualitatively similar results in both subperiods but stronger evidence for noise in the 1926-1986 period. Thus in 1871-1925 Model C with a free  $r$  and  $\lambda$  and no noise estimates  $r = 4.1\%$ ; with full noise, the  $r$  estimate rises to 8.6%, but the difference in log likelihood is only 0.16. In 1926-1986 the no-noise  $r$  estimate is 3.2%, while the full-noise  $r$  estimate is 6.1%. The difference in log likelihood is now 2.1. In both subsamples the models with free  $r$  and  $\lambda$  and full noise imply

that stock prices move from 2.5 to 3 times too much in response to an innovation in fundamentals.

When we vary the trend growth rate  $\xi$ , we again have qualitatively similar results. The evidence for noise is weakest when we fix  $\xi$  at a low value of 0.01, and allow a free value of  $r$  and  $\lambda$ . The interest rate is then estimated at 2.9%, and adding noise increases the log likelihood by only 0.04. But with a fixed interest rate of 6%, there is strong evidence for "overreaction" noise regardless of the trend growth rate assumed.

#### Some tests for misspecification

As a final check on our results, we examined the normalized forecast errors from the Kalman filter estimation of our most general model in Table 8. If our model is well specified, these errors should be homoskedastic and independently and normally distributed (with zero mean and unit variance). We test the errors for skewness, excess kurtosis, and serial correlation, all of which should be zero if the model is well specified.

These tests do yield some evidence of specification error. The normalized forecast error for the dividend has skewness of -0.08 (with standard error 0.23) and excess kurtosis of 1.00 (with standard error 0.46). The normalized forecast error for the stock price has skewness of -0.52 and excess kurtosis of 0.88, with the same standard errors as above. Thus both dividend and price errors have excessively fat tails, and price errors are also negatively skewed.

In addition, we find that the price forecast error has a significant negative second autocorrelation of -0.21. This should not be surprising, since we noted above that none of our models were able to fit the second autocorrelation of the stock price change.

## 5. Conclusion

In this paper we have tried to account for the predictability and volatility of stock returns in two ways. First, we suppose that the exponentially detrended levels of stock prices and dividends are normally distributed with constant variance, and that utility-maximizing investors have constant absolute risk aversion. This implies that the percentage stock return required by utility-maximizing investors as a reward for bearing risk declines as the stock market rises, and that in equilibrium stock prices discount dividends at a relatively low riskless rate so they will react strongly to news about dividends.

Secondly, we suppose that stock prices are influenced by the presence of some investors who do not maximize utility but instead trade exogenously. These "noise traders" can affect stock prices because the utility-maximizing or "smart money" investors are risk-averse.

Both aspects of our approach are helpful in explaining U.S. stock market movements over the last century. It turns out that the importance of noise depends sensitively on the interest rate assumed. If one believes that the stock market discounts dividends at a very low rate (roughly, 4% or below), then one can account for stock price movements fairly well without appealing to noise. On the other hand if one believes that the discount rate is 5% or above, one must also believe that noise is extremely important in moving the stock market. The data on dividends and prices are not well able to discriminate between these two views.

The type of noise which appears to be empirically important is highly correlated with fundamental value. We have called this "overreaction", since it makes the stock price respond more to news about fundamentals than

it otherwise would do. One way to think about this overreaction is that it represents the rational behavior of investors whose absolute risk aversion is declining with wealth rather than constant. In our model, where absolute risk is constant, a rise in the stock market which increases investor wealth will stimulate the demand for stocks by investors with declining absolute risk aversion. These investors could be responsible for the observed overreaction of stock prices to dividends<sup>10</sup>.

We conclude with two points which need to be kept in mind when interpreting our results. First, our assumption that detrended stock price and dividend changes are homoskedastic and normally distributed is not literally accurate. Specification tests reveal some evidence of nonnormality in the distribution of our forecast errors. We believe, however, that the model is still of interest as an approximation to the true process which generates the data.

Secondly, we note that even if one believes in a low interest rate and a relatively small role for noise, this does not rehabilitate the view that the stock price equals the expected present value of future dividends and that percentage stock returns are unpredictable. A low interest rate is only consistent with the data if investors demand a large fixed discount from the expected present value of future dividends, a discount which makes percentage stock returns predictable.

---

<sup>10</sup> In our model investors with declining absolute risk aversion follow "portfolio insurance" strategies, increasing their demand for risky assets when their wealth increases, i.e., when the prices of risky assets rise. It has been argued that investment strategies of this sort contributed to the October 1987 stock market crash and the volatility of the stock market in general.

## Bibliography

Banerjee, A., J.J. Dolado, D.F. Hendry and G.W. Smith, "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence", Oxford Bulletin of Economics and Statistics 48:253-277, August 1986.

Bergstrom, A.R., "Continuous Time Stochastic Models and Issues of Aggregation over Time", Chapter 20 in Zvi Griliches and Michael D. Intriligator eds., Handbook of Econometrics Vol. 2, Amsterdam: North-Holland, 1984.

Black, Fischer, "Studies of Stock Price Volatility Changes", Proceedings of the Business and Economic Statistics Section, American Statistical Association, 177-181, 1976.

Black, Fischer, "Noise", Journal of Finance 41:529-543, July 1986.

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

Campbell, John Y. and Robert J. Shiller [1988a], "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", forthcoming Review of Financial Studies, 1988.

Campbell, John Y. and Robert J. Shiller [1988b], "Stock Prices, Earnings and Expected Dividends", Journal of Finance 43:661-676, July 1988.

Christiano, Lawrence J., Martin Eichenbaum and David Marshall, "The Permanent Income Hypothesis Revisited", unpublished paper, 1987.

Cowles, Alfred, Common Stock Indexes, second edition, 1939.

Davis, M.H.A., Linear Estimation and Stochastic Control, London: Chapman and Hall, 1977.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers and Robert J. Waldmann, "The Economic Consequences of Noise Traders", unpublished paper, revised October 1987.

Dickey, David A. and Wayne A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica 49:1057-1072, 1981.

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

Engle, Robert F. and Mark Watson, "Applications of Kalman Filtering in Econometrics", paper presented at World Congress of the Econometric Society, August 1985.

Fama, Eugene F. and Kenneth R. French, "Dividend Yields and Expected Stock Returns", unpublished paper, University of Chicago, July 1987.

Fama, Eugene F. and Kenneth R. French, "Permanent and Transitory Components of Stock Prices", Journal of Political Economy 96:246-273, April 1988.

Flavin, Marjorie A., "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence", Journal of Political Economy 91:929-956, December 1983.

Fuller, Wayne A., Introduction to Statistical Time Series, Wiley: New York, 1976.

Grossman, Sanford J., Angelo Melino and Robert J. Shiller, "Estimating the Continuous Time Consumption Based Asset Pricing Model", Journal of Business and Economic Statistics, 5:315-327, July 1987.

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", Econometrica 50:1269-1268, September 1982.

Hansen, Lars Peter and Kenneth J. Singleton, "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns", Journal of Political Economy 91:249-265, April 1983.

Harvey, Andrew C., Time Series Models, Oxford: Philip Allan, 1981.

Harvey, Andrew C. and James H. Stock, "The Estimation of Higher-Order Continuous Time Autoregressive Models", Econometric Theory 1:97-112, 1985.

Harvey, Andrew C. and James H. Stock, "Estimating Integrated Higher Order Continuous Time Autoregressions", unpublished paper, June 1986.

Ingersoll, Jonathan E., Jr., Theory of Financial Decision Making, Totowa, NJ: Rowman and Littlefield, 1987.

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

LeRoy, Stephen F. and William R. Parke, "Stock Price Volatility: A Test Based on the Geometric Random Walk", unpublished paper, University of California Santa Barbara, August 1987.

LeRoy, Stephen F. and Richard D. Porter, "The Present-Value Relation: Tests Based on Implied Variance Bounds", Econometrica 49:555-574, May 1981.

Mankiw, N. Gregory, David Romer and Matthew D. Shapiro, "An Unbiased Reexamination of Stock Market Volatility", Journal of Finance 40:677-687, July 1985.

Marsh, Terry A. and Robert C. Merton, "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices", American Economic Review 76:483-498, June 1986.

Melino, Angelo, "Estimation of Unit Averaged Diffusion Processes", Working Paper No. 8502, Department of Economics, University of Toronto, March 1985.

Merton, Robert C., "Optimum Consumption and Portfolio Rules in a Continuous Time Model", Journal of Economic Theory 373-413, December 1971.

Miller, Edward M., "Risk, Uncertainty, and Divergence of Opinion", Journal of Finance 32:1151-1168, September 1977.

Nelson, Daniel B., "The Time Series Behavior of Stock Market Volatility and Returns", unpublished paper, MIT, September 1987.

Perron, Pierre, "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach", unpublished paper, Universite de Montreal, 1986.

Phillips, Peter C.B., "Time Series Regression with a Unit Root", Econometrica 55:277-301, March 1987.

Phillips, Peter C.B. and Pierre Perron, "Testing for Unit Roots in Time Series Regression", Cowles Foundation Discussion Paper, June 1986.

Poterba, James M. and Lawrence H. Summers, "Mean Reversion in Stock Prices: Evidence and Implications", unpublished paper, Harvard University, revised March 1987.

Priestley, Maurice B., Spectral Analysis and Time Series, London: Academic Press, 1982.

Schwert, G. William, "Why Does Stock Market Volatility Change Over Time?", unpublished paper, University of Rochester, September 1987.

Shiller, Robert J., "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", American Economic Review 71:421-436, June 1981.

Shiller, Robert J., "Stock Prices and Social Dynamics", Brookings Papers on Economic Activity 2:457-498, 1984.

Stock, James H., "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", Econometrica 55:1035-1056, September 1987.

Summers, Lawrence H., "Does the Stock Market Rationally Reflect Fundamental Values?", Journal of Finance 41:591-601, July 1986.

West, Kenneth D. [1988a], "Dividend Innovations and Stock Price Volatility", Econometrica 56:37-61, January 1988.

West, Kenneth D. [1988b], "Bubbles, Fads and Stock Price Volatility Tests: A Partial View", Journal of Finance 43:639-656, July 1988.

Wilson, Jack W., and Charles P. Jones, "A Comparison of Annual Common Stock Returns: 1871-1925 with 1926-85," Journal of Business 60, April 1987.

Working, Holbrook, "Note on the Correlation of First Differences of Averages in a Random Chain", Econometrica 28:916-918, October 1960.



## APPENDIX A

In this appendix we first summarize (without proof) some results about linear continuous-time stochastic processes. For proofs and extensions, the reader is referred to Davis [1977] and Bergstrom [1984]. Then we derive the structure of Model B given in section 2 of the paper.

Forecasting Observable Continuous-Time Processes. Let  $y(t)$  be a continuous vector stochastic process which satisfies the linear stochastic differential equation

$$(A.01) \quad dy(t) = Ay(t)dt + Cdz(t),$$

where  $y$  is an  $n$ -vector,  $A$  is an  $n \times n$  matrix,  $C$  is an  $n \times k$  matrix and  $z(t)$  is a  $k$ -dimensional normalized Brownian motion. The initial observation  $y(0)$  is normally distributed and independent from the  $z(t)$  process. The process  $y(t)$  can be represented in integral form by

$$(A.02) \quad y(t+s) = e^{As}y(t) + \int_{\tau=0}^s e^{A(s-\tau)}C dz(t+\tau),$$

Here, the exponential notation  $e^{At}$  denotes the  $n \times n$  "transition" matrix  $I + tA + t^2A^2/2! + t^3A^3/3! + \dots$ . The matrix  $e^{At}$  is positive definite, it commutes with  $A$ , and its derivative with respect to  $t$  is  $Ae^{At}$ . It follows from (A.02) that the conditional variance of  $y(t+s)$  given  $y(t)$  can be written

$$(A.03) \quad \text{var} \left\{ y(t+s) \mid y(t) \right\} = \int_{\tau=0}^s e^{A(s-\tau)} CC' e^{A'(s-\tau)} d\tau.$$

Let  $Q(t)$  denote the unconditional variance of  $y(t)$  given by

$$(A.04) \quad Q(t) = \text{var}\{y(t)\} = e^{At} \text{var}\{y(0)\} e^{A't} + \text{var}\{y(t) \mid y(0)\}.$$

The matrix  $Q(t)$  satisfies the differential equation

$$(A.05) \quad \frac{dQ}{dt} = AQ(t) + Q(t)A' + CC'.$$

This result can be derived intuitively by writing

$$(A.06) \quad \begin{aligned} Q + dQ &= \text{var}\{y + dy\} = \text{var}\{(I + Adt)y + Cdz\} \\ &= (I + Adt)Q(I + Adt)' + CC'dt = Q + (AQ + QA' + CC')dt + AQA'dt^2, \end{aligned}$$

then subtracting  $Q$  from both sides, throwing away the  $dt^2$ -term, and dividing by  $dt$ . (A more theoretically precise derivation is obtained by differentiating (A.03)).

If the process  $y(t)$  has the stationarity property that  $Q(t)$  does not vary with  $t$ , then the time invariant matrix  $Q = Q(t)$  solves the equation

$$(A.07) \quad AQ + QA' + CC' = 0.$$

In this case one can evaluate explicitly the integral in (A.03), yielding

$$(A.08) \quad \text{var}\{y(t+s) \mid y(t)\} = Q - e^{As} Q e^{A's}.$$

Model B. Model B is based upon three assumptions.

The first assumption states that  $D_0$  and  $I$  follow a linear process. This means that for appropriate constants  $A_{ij}$  and  $C_{ij}$  we can write

$$(A.09) \quad d \begin{bmatrix} D_0 \\ I \end{bmatrix} = \begin{bmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{bmatrix} \begin{bmatrix} D_0 \\ I \end{bmatrix} dt + \begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix} \begin{bmatrix} dz_0^* \\ dz_I^* \end{bmatrix},$$

or, more briefly,  $dy = Aydt + Cdz$ . Here,  $dz_0^*$  and  $dz_I^*$  are i.i.d. unit-variance Brownian motions.

The second assumption says that the history of  $D_0$  cannot forecast the future of  $I$ . This means

$$(A.10) \quad E\{I(t+s) | D_0[-\infty, t]\} = I \text{ for all } s \geq 0.$$

The third assumption says that the univariate  $D_0$  process is a Brownian motion and the univariate  $I$  process is stationary.

Here we show that these three assumptions imply that the univariate  $I$  process is AR(1) and the vector  $D_0, I$  process satisfies the constrained version of (A.09) given by (2.14) in the text.

Begin by writing  $\exp(As)$  as

$$(A.11) \quad \exp(As) = \begin{bmatrix} B_{00}(s) & B_{01}(s) \\ B_{10}(s) & B_{11}(s) \end{bmatrix}.$$

It follows from (A.02) that

$$(A.12) \quad E\left\{D_0(t+s) \middle| D_0[-\infty, t], I[-\infty, t]\right\} = B_{00}(s) D_0(t) + B_{01}(s) I(t)$$

$$(A.13) \quad E\left\{I(t+s) \middle| D_0[-\infty, t], I[-\infty, t]\right\} = B_{10}(s) D_0(t) + B_{11}(s) I(t).$$

Taking conditional expectations of both sides of (A.13) with respect to the information set  $D_0[-\infty, t]$  and applying the assumption (A.10) that  $D_0$  does not anticipate  $I$  to eliminate the  $I(t)$  term on the right-hand side yields

$$(A.14) \quad E\left\{I(t+s) \middle| D_0[-\infty, t]\right\} = B_{10}(s) D_0(t).$$

But the assumption that  $D_0$  does not anticipate  $I$  in (A.10) implies that the right-hand side of this equation is zero. Since  $D_0(t)$  is not identically zero, we must have

$$(A.15) \quad B_{10}(s) = 0 \text{ for all } s \geq 0.$$

Now expand the derivative  $d/ds \exp(As) = A \exp(As)$  and use

$$B_{10}(s) = d/ds B_{10}(s) = 0 \text{ to obtain}$$

$$(A.16) \quad \begin{bmatrix} d/ds B_{00}(s) & d/ds B_{01}(s) \\ 0 & d/ds B_{11}(s) \end{bmatrix} = \begin{bmatrix} A_{00}B_{00}(s) & A_{00}B_{01}(s) + A_{01}B_{11}(s) \\ A_{10}B_{00}(s) & A_{10}B_{01}(s) + A_{11}B_{11}(s) \end{bmatrix}.$$

Equating terms in the lower left corner yields  $A_{10}B_{00}(s) = 0$  for all  $s \geq 0$ . Thus, either  $A_{10} = 0$ , or  $B_{00}(s) = 0$  for all  $s$ . Since the exponent of any matrix is nonsingular,  $B_{00}(s)$  cannot be zero for any  $s$ , since this,

combined with  $B_{10}(s) = 0$  as shown above, would make  $\exp(As)$  singular for some  $s$ . We conclude  $A_{10} = 0$ .

Solving the three differential equations in (A.16) subject to the initial conditions  $B_{00}(0) = B_{11}(0) = 1$  and  $B_{01}(0) = 0$  yields (with  $A_{10} = 0$ )

$$(A.17) \quad B_{00}(s) = \exp(A_{00}s) \qquad (A.18) \quad B_{11}(s) = \exp(A_{11}s)$$

$$(A.19) \quad B_{01}(s) = \begin{cases} \frac{\exp(A_{00}s) - \exp(A_{11}s)}{A_{11} - A_{00}} A_{01} & \text{if } A_{11} \neq A_{00}, \\ 0 & \text{if } A_{11} = A_{00}. \end{cases}$$

This result can also be obtained from a power series expansion of  $\exp(As)$  with  $A_{10} = 0$ . The three initial conditions are obtained by setting  $s = 0$  in (A.12) and (A.13).

Now take conditional expectations of both sides of (A.12) with respect to  $D_0[-\infty, t]$ , take conditional expectations of both sides of (A.13) with respect to  $I[-\infty, t]$ , and apply (A.10), (A.15), (A.17), and (A.18) to obtain

$$(A.20) \quad E\left\{D_0(t+s) \mid D_0[-\infty, t]\right\} = \exp(A_{00}s) D_0(t),$$

$$(A.21) \quad E\left\{I(t+s) \mid I[-\infty, t]\right\} = \exp(A_{11}s) I(t).$$

Our third assumption -- that  $D_0(t)$  is a Brownian motion and  $I(t)$  is stationary -- implies that  $\exp(A_{00}s)$  on the right side of (A.20) equals one and  $\exp(A_{11}s)$  on the right side of (A.21) does not explode as  $s \rightarrow \infty$ . It follows that

$$(A.22) \quad A_{00} = 0, \quad A_{11} = -\alpha_I,$$

where  $\alpha_I$  is a positive constant. Furthermore, the stationarity of  $I(t)$  together with (A.21) imply that  $I(t)$  is a univariate continuous-time AR(1) with "mean reversion parameter"  $\alpha_I$ .

We have now pinned down all of the elements of  $A$  except  $A_{01}$ . To obtain a value for  $A_{01}$ , observe that the assumptions underlying model B imply nothing about the units in terms of which  $I$  is scaled. If  $I$  is rescaled by multiplying by some constant  $K_0$ , the vector  $y$  is changed to  $Ky$ , where  $K$  is the matrix

$$(A.23) \quad K = \begin{bmatrix} 1 & 0 \\ 0 & K_0 \end{bmatrix}.$$

Since  $Ky$  satisfies  $d(Ky) = KAK^{-1}(Ky)dt + KCdz$ , the rescaling changes  $A$  to  $KAK^{-1}$ , where

$$(A.24) \quad KAK^{-1} = \begin{bmatrix} 0 & A_{01}/K_0 \\ 0 & -\alpha_I \end{bmatrix}.$$

It is apparent that rescaling  $I$  changes only the upper right element of this matrix. Thus, the initial choice of  $A_{01}$  is equivalent to a choice of units in which  $I$  is scaled. To simplify formulas, the scaling convention adopted here is  $A_{01} = \alpha_I$ . This, combined with previous results, yields

$$(A.25) \quad A = \begin{bmatrix} 0 & \alpha_I \\ 0 & -\alpha_I \end{bmatrix}, \quad \exp(As) = \begin{bmatrix} 1 & 1 - \exp(-\alpha_I s) \\ 0 & \exp(-\alpha_I s) \end{bmatrix}$$

We have thus characterized the matrix A completely. Equations (A.12) and (A.26) become

$$(A.26) \quad E\left\{D_0(t+s) \mid D_0[-\infty, t], I_0[-\infty, t]\right\} = D_0(t) + \left(1 - e^{-\alpha_I s}\right) I(t)$$

$$E\left\{I(t+s) \mid D_0[-\infty, t], I_0[-\infty, t]\right\} = e^{-\alpha_I s} I(t)$$

Note since  $e^{-\alpha_I s} \rightarrow 0$  as  $s \rightarrow \infty$ , we have scaled the units of I so that I(t) measures how much  $D_0(t)$  is expected to increase in the distant future.

Now consider the 2x2 matrix C. Since it takes only three scalars to specify a bivariate covariance structure, let the upper right element of C be zero by convention, i.e.,  $C_{01} = 0$ .

Let Q(t) denote the unconditional covariance matrix of [D(t), I(t)]. Our second assumption that D does not forecast I implies that D(t) and I(t) are independently distributed. Thus, the matrix Q(t) is diagonal.

Now consider the diagonal elements of Q(t). Following (2.12) and (2.13) in the text, let  $\sigma_0^2$  denote the innovations variance of  $D_0$  and let  $\sigma_I^2$  denote the innovations variance of I. Since I is stationary, its unconditional variance is the scalar constant  $\sigma_I^2/2\alpha_I$  (which can be obtained by solving the scalar version of (A.07) for Q). Since D follows a Brownian motion, its unconditional variance is  $Q_{00}(0) + \sigma_0^2 t$  (obtained from the scalar version of (A.05)). Thus, the matrix Q(t) is given by

$$(A.27) \quad Q(t) = \begin{bmatrix} Q_{00}(0) + \sigma_0^2 t & 0 \\ 0 & \sigma_I^2 / 2\alpha_I \end{bmatrix} .$$

Now recall that (A.05) implies  $dQ/dt = AQ(t) + Q(t)A' + CC'$ . Using the specific expression for A in (A.25) and imposing the constraint  $C_{01} = 0$  on C, this becomes

$$(A.28) \quad \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} C_{00}^2 & \sigma_I^2/2 + C_{00}C_{10} \\ \sigma_I^2/2 + C_{00}C_{10} & -\sigma_I^2 + C_{10}^2 + C_{11}^2 \end{bmatrix} .$$

Since these matrices are symmetric, we have three equations in the three unknowns  $C_{00}$ ,  $C_{01}$ ,  $C_{11}$ .

If we define  $\rho_I = \sigma_I^2 / 2\sigma_0^2$ , the solution can be written

$$(A.29) \quad C_{00} = \sigma_0, \quad C_{10} = -\rho_I \sigma_0, \quad C_{11} = \left[ 2\rho_I - \rho_I^2 \right]^{1/2} \sigma_0 .$$

This solution adopts the innocuous convention  $C_{00} > 0$ ; any solution requires  $0 < \rho_I \leq 2$ . Thus, the general linear formulation (A.09) becomes the same as (2.14) in the text.

It remains to show that the constrained system (2.14) actually satisfies all three assumptions. The only assumption not already verified is that the history of D cannot forecast future I. To prove this, it suffices (see Davis) to show that  $I(t+s)$  is uncorrelated with  $D(t)$  for all  $s \geq 0$ , and this follows from (A.14) and (A.15).



## APPENDIX B

This Appendix contains proofs of Theorem 3.1 and Theorem 3.2.

Proof of Theorem 3.1. Using (2.01), equation (3.01) can be written

$$(B.01) \quad dM(t) = D(t)dt + \xi P(t)dt + dP(t) - rP(t)dt$$

This says the excess return on a fully levered portfolio long one share of stock consists of a cash dividend, plus a stock dividend, plus a capital gain, minus a financing cost. Recall from (2.03) that  $P(t)$  can be expressed as the sum of a fundamental value, a noise term, and a risk discount:  $P(t) = V(t) - \lambda/(r-\xi) + Y(t)$ . Thus (B.01) can be written

$$(B.02) \quad dM(t) = [D(t) - (r-\xi)V(t)]dt + dV(t) + dY(t) - (r-\xi)Y(t)dt + \lambda dt.$$

Of the five terms on the right side, the first two terms,  $[D(t) - (r-\xi)V(t)]dt + dV(t)$ , give the returns which would prevail if stocks were valued in a risk neutral manner; the next two terms,  $dY(t) - (r-\xi)Y(t)dt$ , adjust returns for noise; and the last term  $\lambda dt$  is an expected risk premium. Since the first two terms represent risk-neutral returns, they must have a martingale property. To verify this, it can be shown that when the first two terms are expanded using  $V(t) = \pi_0 D_0 + \pi_1 D_1 + \pi_I D_I$  from (2.21) and (2.22), then the reduced form (2.23) implies

$$(B.03) \quad [D(t) - (r-\xi)V(t)]dt + dV(t) \\ = (\pi_0 - \rho_1)\sigma_0 dz_0^* + (2\rho_1 - \rho_1^2)^{1/2}\sigma_0 dz_1^* + \pi_1 \sigma_1 dz_1^*$$

with all dt-terms involving  $\pi_0$ ,  $\pi_1$ , and  $\pi_I$  cancelling. A similar expansion for the two terms involving  $Y(t)$  yields

$$(B.04) \quad dY(t) - (r-\xi)Y(t)dt - \\ - (r-\xi+\alpha_Y)Y(t)dt + \theta_0\gamma_0 dz_0^* + \theta_I\gamma_I dz_I^* + \theta_1\gamma_1 dz_1^* + \theta_Y\gamma_Y dz_Y^* .$$

Thus,  $dM(t)$  satisfies

$$(B.05) \quad dM = \left[ \lambda - (r-\xi+\alpha_Y)Y \right] dt + (\pi_0\sigma_0 - \rho_I\sigma_0 + \theta_0\gamma_0) dz_0^* \\ + \left[ (2\rho_I - \rho_I^2)^{1/2} \sigma_0 + \theta_I\gamma_I \right] dz_I^* + (\pi_1\sigma_1 + \theta_1\gamma_1) dz_1^* + \theta_Y\gamma_Y dz_Y^* .$$

Now define the process  $N(t)$  by

$$(B.06) \quad N(t) = -(r - \xi + \alpha_Y) Y(t) .$$

The process  $N(t)$ , which is perfectly negatively correlated with  $Y(t)$ , satisfies

$$(B.07) \quad dN(t) = -\alpha_Y N(t) - (r-\xi-\alpha_Y)(\theta_0\gamma_0 dz_0^* + \theta_I\gamma_I dz_I^* + \theta_1\gamma_1 dz_1^* + \theta_Y\gamma_Y dz_Y^*) .$$

Equations (B.05) and (B.07) imply equations (3.03)-(3.05).

Proof of Theorem 3.2. This kind of optimization problem is considered by Merton [1971], and our proof follows his analysis.

Begin with the Bellman equation

$$(B.08) \quad 0 = \max_{C, X} \left[ -e^{-\psi C} - \alpha_Y V + [rW + X(\lambda + N) - C]V_W \right. \\ \left. + \sigma_M^2 X^2 V_{WW}/2 - \alpha_Y N V_N + \sigma_N^2 V_{NN}/2 + \zeta \sigma_M \sigma_N X V_{WN} \right]$$

Make the conjecture that the value function  $V(\cdot)$  is given by (3.08).

Then the Bellman equation becomes (upon dividing through by  $V$ )

$$(B.09) \quad 0 = \max_{C, X} \left[ -\frac{1}{V} e^{-\psi C} - \psi r(rW - C) + [(\Phi_1 + \Phi_2 N)^2 - \Phi_2] \sigma_N^2 / 2 - \psi \right. \\ \left. + \alpha_Y (\Phi_1 + \Phi_2 N) N - [\lambda + N - (\Phi_1 + \Phi_2 N) \zeta \sigma_M \sigma_N] \psi r X + (\psi r X)^2 \sigma_M^2 / 2 \right]$$

Notice how the maximization problems for  $C$  and  $X$  can be solved separately. The first order condition for  $C$  is

$$(B.10) \quad e^{-\psi C} = -rV$$

Taking logs of both sides and substituting from (3.08) yields the consumption function

$$(B.11) \quad \psi C = -\log(-rV) = \psi rW + \Phi_0 + \Phi_1 N + \Phi_2 N^2 / 2$$

We thus obtain

$$(B.12) \quad \max \left[ -\frac{1}{V} e^{-\psi C} - \psi r(rW - C) \right] - r(1 + \Phi_0 + \Phi_1 N + \Phi_2 N^2/2) .$$

The first order condition for X generates the demand function (3.11).

Plugging (B.12) and (3.11) into the Bellman equation (B.09) yields

$$(B.14) \quad 0 = r(1 + \Phi_0 + \Phi_1 N + \Phi_2 N^2/2) - \psi + \alpha_Y (\Phi_1 + \Phi_2 N) N + \\ [(\Phi_1 + \Phi_2 N)^2 - \Phi_2] \sigma_M^2 / 2 - [\lambda + N - (\Phi_1 + \Phi_2 N) \sigma_M \sigma_N]^2 / \sigma_M^2 .$$

The desired equations (3.09) are now obtained by equating coefficients on the  $N^2$ -term, the  $N$ -term, and the constant term. The first equation of (3.09) is quadratic in  $\Phi_2$  with a positive and a negative root, but the positive root is economically relevant, because it generates a value function with higher expected utility. Values of  $\Phi_0$  and  $\Phi_1$  are obtained from the other two equations of (3.09) which are linear. As discussed by Merton [1971], the values of  $\Phi_0$ ,  $\Phi_1$ , and  $\Phi_2$  actually characterize a solution to the smart-money investor's problem.

APPENDIX C

Discrete Processes from Point Observations and Time Averages. Let  $y_n^P$  denote the discrete process obtained by taking point observations of the continuous process defined in Appendix A, equation (A.01), at evenly spaced points  $t_0, t_1, \dots, t_n, \dots$ , with  $\Delta t = t_n - t_{n-1}$ . Then  $y_n^P$  satisfies the stochastic difference equation

$$(C.01) \quad y_{n+1}^P = e^{A\Delta t} y_n^P + u_{n+1}^P,$$

where  $u_n^P$  is a normally and independently distributed sequence of innovations (independent from  $y_n^P, y_{n-1}^P, \dots$ ) given (from (A.02)) by

$$(C.02) \quad u_{n+1}^P = \int_{\tau=0}^{\Delta t} e^{A(\Delta t-\tau)} C dz(t_n+\tau).$$

We have from (A.03)

$$(C.03) \quad \text{var} \left\{ u_{n+1}^P \right\} = \int_{\tau=0}^{\Delta t} e^{A(\Delta t-\tau)} C C' e^{A'(\Delta t-\tau)} d\tau,$$

and when  $Q(t)$  is time invariant, we have from (A.08)

$$(C.04) \quad \text{var} \left\{ u_{n+1}^P \right\} = Q - e^{A\Delta t} Q e^{A'\Delta t}.$$

Now let  $y_n^A$  be the discrete process obtained by taking continuously compounded time averages of the continuous process  $y(t)$  over the interval  $[t_{n-1}, t_n]$ , i.e.,

$$(C.05) \quad y_n^A = \int_{r=0}^{\Delta t} e^{rA} y_{n-1+r} dr .$$

An explicit evaluation of this integral yields

$$(C.06) \quad y_n^A = \int_{r',-0}^{\Delta t} e^{r'r} \left[ e^{Ar'} y_{n-1}^P + \int_{r=0}^{r'} e^{A(r'-r)} C dz(t_n+r) \right] dr' \\ = (rI+A)^{-1} \left[ e^{(rI+A)\Delta t} - I \right] y_{n-1}^P + (rI+A)^{-1} \int_{r=0}^{\Delta t} \left[ e^{(rI+A)\Delta t} - e^{rI\tau} \right] C dz$$

Thus,  $y_{n+1}^A$  satisfies the linear stochastic difference equation

$$(C.07) \quad y_{n+1}^A = (rI+A)^{-1} \left[ e^{(rI+A)\Delta t} - I \right] y_n^P + u_n^A .$$

The innovations  $u_n^A$  are serially independent and are also independent from past values of both the point process  $y_n^P$  and the time averages  $y_n^A$ . This makes  $y_n^P, y_n^A$  jointly a vector AR1 process, even though the time averages  $y_n^A$  alone are not. We have

$$(C.08) \quad \text{var} \left\{ u_{n+1}^A \right\} \\ = (rI+A)^{-1} \left[ \int_{r=0}^{\Delta t} \left[ e^{(rI+A)\Delta t} - e^{rI\tau} \right] C C' \left[ e^{(rI+A')\Delta t - A'\tau} - e^{rI\tau} \right] dr \right] (rI+A')^{-1}$$

$$(C.09) \quad \text{cov} \left\{ u_n^A, u_n^P \right\} = (rI+A)^{-1} \int_{r=0}^{\Delta t} \left[ e^{(rI+A)\Delta t} - e^{rI\tau} \right] C C' e^{A'(\Delta t - \tau)} dr .$$

Evaluation of these expressions is a messy exercise with the following

solution when  $Q(t)$  is time invariant and  $rI+A$  and  $rI-A$  are both invertible: Define

$$(C.10) \quad S = e^{(rI+A)\Delta t}, \quad T = e^{2rI\Delta t}$$

$$M_1 = TS - SQS', \quad M_2 = CC'(T - S')(rI - A)^{-1}, \quad M_3 = \frac{1}{2r}(T - I)CC'$$

Then we have

$$(C.11) \quad \text{var}\left\{u_n^A\right\} = (rI + A)^{-1} \left[ M_1 - M_2 - M_2' + M_3 \right] (rI + A')^{-1}$$

$$(C.12) \quad \text{cov}\left\{u_n^A, u_n^P\right\} = (rI + A)^{-1} e^{-r\Delta t} \left[ M_1 - M_2 \right]$$

If  $Q(t)$  is not time invariant, then the above formulas are valid when  $M_1$  is replaced by

$$(C.13) \quad M_1^* = e^{(rI+A)\Delta t} \left[ \int_{r=0}^{\Delta t} e^{Ar} CC' e^{A'r} dr \right] e^{(rI+A')\Delta t}$$

The stacked vector of discrete points and time averages thus satisfies

$$(C.14) \quad \begin{bmatrix} y_{n+1}^P \\ y_{n+1}^A \end{bmatrix} = \begin{bmatrix} e^{A\Delta t} & 0 \\ (rI+A)^{-1} \left[ e^{(rI+A)\Delta t} - I \right] & 0 \end{bmatrix} \begin{bmatrix} y_n^P \\ y_n^A \end{bmatrix} + \begin{bmatrix} u_{n+1}^P \\ u_{n+1}^A \end{bmatrix}$$

$$(C.15) \quad \text{var} \begin{bmatrix} u_{n+1}^P \\ u_{n+1}^A \end{bmatrix} = \begin{bmatrix} \text{var}\{u_{n+1}^P\} & \text{cov}\{u_{n+1}^P, u_{n+1}^A\} \\ \text{cov}\{u_{n+1}^P, u_{n+1}^A\}' & \text{var}\{u_{n+1}^A\} \end{bmatrix}$$

The terms in (C.15) are explicitly evaluated in (C.03), (C.11), and (C.12).

**TABLE 1**  
**UNIVARIATE TESTS FOR UNIT ROOTS**  
**AND TESTS FOR COINTEGRATION**

Variable	Test Statistics	
	$t_{\alpha}$	$Zt_{\alpha}$
$D_t^0$	-2.77	-3.08
$P_t^0$	-2.37	-2.51
$D_t$	-3.40 (10%)	-3.70 (5%)
$P_t$	-2.68	-2.80
$\ln(D_t^0)$	-3.43 (5%)	-3.69 (2.5%)
$\ln(P_t^0)$	-2.68	-2.85

Cointegrating regression:

$$D_t = 0.030 + 0.016 P_t + u_t. \text{ Implied value of } r = 0.029.$$

Engle-Granger [1987] tests for cointegration:

Dickey-Fuller 4.68 (1%), Augmented Dickey-Fuller 3.56 (5%).

Other estimates of r:

Reverse cointegrating regression: implied value of  $r = 0.055$ .

(Mean of  $D_t$ )/(Mean of  $P_t$ ) = 0.050: implied value of  $r = 0.063$ .

Mean stock return = 0.082.



TABLE 2

## TIME-SERIES PROPERTIES OF THE DATA

$$\sigma(\Delta D_t) = 0.006, \sigma(\Delta P_t) = 0.174, \sigma(\Delta D_t)/\sigma(\Delta P_t) = 0.032$$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.08	$\Delta P_t$	0.08	1.00
$\Delta D_{t-1}$	0.17	-0.12	$\Delta P_{t-1}$	0.57	0.04
$\Delta D_{t-2}$	-0.14	0.03	$\Delta P_{t-2}$	0.21	-0.22
$\Delta D_{t-3}$	-0.09	-0.11	$\Delta P_{t-3}$	0.04	0.09
$\Delta D_{t-4}$	-0.15	0.02	$\Delta P_{t-4}$	-0.05	0.07
$\Delta D_{t-5}$	-0.04	0.08	$\Delta P_{t-5}$	-0.04	-0.06

TABLE 3  
MAPPING THE LIKELIHOOD FUNCTION, MODEL A

Interest rate assumption	No noise	Independent noise	Overreaction	Full noise
4%	485.15	490.66	491.41	491.71
6%	451.22	488.35	490.94	491.35
Free r, $\lambda=0$ (estimated r)	474.11 (1.5%)	488.66 (5.6%)	491.15 (5.6%)	491.46 (5.7%)
Free r and $\lambda$ (estimated r)	491.36 (3.0%)	491.83 (3.1%)	494.40 (3.4%)	494.40 (3.4%)

TABLE 4

IMPLICATIONS OF MODEL A ESTIMATES

1) Free r and  $\lambda$ , no noise. Log likelihood = 491.35.

$r$	= 0.029 (0.004)	$\alpha_Y$	= -----
$\lambda$	= 0.029 (0.004)	$\gamma_Y$	= -----
$\sigma_0$	= 0.003 (0.001)	$\gamma_0$	= -----
$\alpha_I$	= -----		
$\rho_I$	= -----	$\pi_0$	= 60.357
$\alpha_1$	= 0.441 (0.126)	$\pi_I$	= -----
$\sigma_1$	= 0.007 (0.001)	$\pi_1$	= 2.185

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0^D$	$\pi_1^D$	$\pi_I$
1.00	0.00	1.00	0.99	0.01	0.00
	0.00	0.00	0.00	0.00	0.00
		1.00	0.99	0.01	0.00
			0.99	0.00	0.00
				0.01	0.00
					0.00

$$\sigma(\Delta D_t) = 0.005, \sigma(\Delta P_t) = 0.174$$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.30	$\Delta P_t$	0.30	1.00
$\Delta D_{t-1}$	0.08	-0.01	$\Delta P_{t-1}$	0.30	-0.00
$\Delta D_{t-2}$	-0.16	-0.01	$\Delta P_{t-2}$	-0.01	-0.00
$\Delta D_{t-3}$	-0.10	-0.01	$\Delta P_{t-3}$	-0.01	-0.00
$\Delta D_{t-4}$	-0.07	-0.00	$\Delta P_{t-4}$	-0.01	-0.00
$\Delta D_{t-5}$	-0.04	-0.00	$\Delta P_{t-5}$	-0.00	-0.00

2) Free r and  $\lambda$ , full noise. Log likelihood = 494.40.

$r = 0.034$  (0.005)     $\alpha_Y = 5.624$  (20.9)  
 $\lambda = 0.025$  (0.005)     $\gamma_Y = 0.001$  (\*)  
 $\sigma_0 = 0.003$  (0.001)     $\gamma_0 = 1.054$  (2.767)  
 $\alpha_I = \text{-----}$   
 $\rho_I = \text{-----}$              $\pi_0 = 48.534$   
 $\alpha_1 = 0.374$  (0.141)     $\pi_I = \text{-----}$   
 $\sigma_1 = 0.006$  (0.001)     $\pi_1 = 2.535$

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0 D_0$	$\pi_1 D_1$	$\pi_I I$
4.22	2.16	2.05	2.03	0.03	0.00
	1.11	1.05	1.04	0.01	0.00
		1.00	0.99	0.01	0.00
			0.99	0.00	0.00
				0.01	0.00
					0.00

$\sigma(\Delta D_t) = 0.005$ ,  $\sigma(\Delta P_t) = 0.179$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.20	$\Delta P_t$	0.20	1.00
$\Delta D_{t-1}$	0.11	-0.03	$\Delta P_{t-1}$	0.35	-0.18
$\Delta D_{t-2}$	-0.14	-0.01	$\Delta P_{t-2}$	-0.01	-0.00
$\Delta D_{t-3}$	-0.10	-0.01	$\Delta P_{t-3}$	-0.01	-0.00
$\Delta D_{t-4}$	-0.07	-0.00	$\Delta P_{t-4}$	-0.01	-0.00
$\Delta D_{t-5}$	-0.05	-0.00	$\Delta P_{t-5}$	-0.00	-0.00

(\*): parameter converged almost to the boundary of the admissible region.

TABLE 5  
MAPPING THE LIKELIHOOD FUNCTION, MODEL B

Interest rate assumption	No noise	Independent noise	Overreaction	Full noise
4%	487.65	489.46	495.86	500.73
6%	481.64	499.57	492.87	499.82
Free r, $\lambda=0$ (estimated r)	486.25 (5.4%)	499.84 (5.7%)	492.28 (5.4%)	499.87 (5.8%)
Free r and $\lambda$ (estimated r)	491.74 (4.7%)	499.96 (5.5%)	-----a (-----)	500.77 (3.7%)

Notes:

a. This model failed to converge.

TABLE 6

IMPLICATIONS OF MODEL B ESTIMATES

1) Free r and  $\lambda$ , no noise. Log likelihood = 491.74.

$r$ = 0.047 (0.002)	$\alpha_Y$ = -----
$\lambda$ = 0.012 (0.003)	$\gamma_Y$ = -----
$\sigma_0$ = 0.007 (0.000)	$\gamma_0$ = -----
$\alpha_I$ = 0.300 (0.088)	
$\rho_I$ = 0.768 (0.135)	$\pi_0$ = 29.715
$\alpha_1$ = -----	$\pi_1$ = 26.718
$\sigma_1$ = -----	$\pi_1$ = -----

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0^D$	$\pi_1^D$	$\pi_1^I$
1.00	0.00	1.00	0.36	0.00	0.64
	0.00	0.00	0.00	0.00	0.00
		1.00	0.36	0.00	0.64
			1.16	0.00	-0.80
				0.00	0.00
					1.44

$\sigma(\Delta D_t) = 0.005, \sigma(\Delta P_t) = 0.181$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.25	$\Delta P_t$	0.25	1.00
$\Delta D_{t-1}$	0.25	-0.00	$\Delta P_{t-1}$	0.39	0.02
$\Delta D_{t-2}$	-0.00	-0.00	$\Delta P_{t-2}$	0.17	0.01
$\Delta D_{t-3}$	-0.00	-0.00	$\Delta P_{t-3}$	0.13	0.01
$\Delta D_{t-4}$	-0.00	-0.00	$\Delta P_{t-4}$	0.10	0.01
$\Delta D_{t-5}$	-0.00	-0.00	$\Delta P_{t-5}$	0.07	0.01

2) Free  $r$  and  $\lambda$ , full noise. Log likelihood = 500.77.

$r = 0.037$  (0.009)     $\alpha_Y = 0.089$  (0.056)  
 $\lambda = 0.021$  (0.009)     $\gamma_Y = 0.315$  (0.172)  
 $\sigma_0 = 0.007$  (0.000)     $\gamma_0 = -0.435$  (0.207)  
 $\alpha_I = 1.103$  (0.336)  
 $\rho_I = 1.126$  (0.214)     $\pi_0 = 41.031$   
 $\alpha_1 = \text{-----}$          $\pi_I = 40.144$   
 $\sigma_1 = \text{-----}$          $\pi_1 = \text{-----}$

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0^D$	$\pi_1^D$	$\pi_I$
0.42	-0.15	0.57	-0.06	0.00	0.63
	0.29	-0.43	0.05	0.00	-0.48
		1.00	-0.11	0.00	1.11
			1.05	0.00	-1.16
				0.00	0.00
					2.26

$\sigma(\Delta D_t) = 0.006$ ,  $\sigma(\Delta P_t) = 0.182$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.17	$\Delta P_t$	0.17	1.00
$\Delta D_{t-1}$	0.25	0.06	$\Delta P_{t-1}$	0.57	0.07
$\Delta D_{t-2}$	-0.00	0.05	$\Delta P_{t-2}$	0.30	0.05
$\Delta D_{t-3}$	-0.00	0.05	$\Delta P_{t-3}$	0.10	0.05
$\Delta D_{t-4}$	-0.00	0.04	$\Delta P_{t-4}$	0.03	0.04
$\Delta D_{t-5}$	-0.00	0.04	$\Delta P_{t-5}$	0.01	0.04

TABLE 7  
MAPPING THE LIKELIHOOD FUNCTION, MODEL C

Interest rate assumption	No noise	Independent noise	Overreaction	Full noise
4%	502.25	502.25 <sup>a</sup>	503.14 <sup>b</sup>	503.14 <sup>b</sup>
6%	481.64 <sup>c</sup>	499.74	504.59 <sup>b</sup>	504.59 <sup>b</sup>
Free r, $\lambda=0$ (estimated r)	486.25 <sup>c</sup> (5.4%)	500.34 (5.4%)	504.38 <sup>b</sup> (5.5%)	504.38 <sup>b</sup> (5.5%)
Free r and $\lambda$ (estimated r)	502.81 (3.5%)	503.13 (3.6%)	505.08 <sup>b</sup> (9.6%)	505.08 <sup>b</sup> (9.6%)

Notes:

- a. This model converged to the no-noise model.
- b. This model converged with  $\rho_I$  at its maximum allowable value of 2. The full-noise model converged to the overreaction model.
- c. This model converged to the equivalent version of model B.





2) Free  $r$  and  $\lambda$ , full noise. Log likelihood = 505.08.

$r = 0.096$  (0.053)     $\alpha_Y = 0.050$  (0.034)

$\lambda = -0.043$  (0.061)     $\gamma_Y = 0.000$  (\*)

$\sigma_0 = 0.003$  (0.001)     $\gamma_0 = 3.302$  (2.737)

$\alpha_I = 1.796$  (0.575)

$\rho_I = 2.000$  (\*)     $\pi_0 = 12.006$

$\alpha_1 = 0.383$  (0.185)     $\pi_I = 11.474$

$\sigma_1 = 0.006$  (0.000)     $\pi_1 = 2.145$

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0 D_0$	$\pi_1 D_1$	$\pi_I I$
18.51	14.20	4.30	-4.24	0.43	8.11
	10.90	3.30	-3.26	0.33	6.22
		1.00	-0.99	0.10	1.89
			1.08	0.00	-2.07
				0.10	0.00
					3.95

$\sigma(\Delta D_t) = 0.005$ ,  $\sigma(\Delta P_t) = 0.171$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.08	$\Delta P_t$	0.08	1.00
$\Delta D_{t-1}$	0.12	-0.03	$\Delta P_{t-1}$	0.53	-0.02
$\Delta D_{t-2}$	-0.12	-0.03	$\Delta P_{t-2}$	0.16	-0.02
$\Delta D_{t-3}$	-0.08	-0.03	$\Delta P_{t-3}$	-0.01	-0.02
$\Delta D_{t-4}$	-0.06	-0.02	$\Delta P_{t-4}$	-0.02	-0.02
$\Delta D_{t-5}$	-0.04	-0.02	$\Delta P_{t-5}$	-0.02	-0.02

(\*): parameter converged to the boundary of the admissible region.

3) r = 6% full noise. Log likelihood = 504.59.

$r = 0.060$  (-----)     $\alpha_Y = 0.034$  (0.035)  
 $\lambda = -0.006$  (0.008)     $\gamma_Y = 0.000$  (\*)  
 $\sigma_0 = 0.003$  (0.001)     $\gamma_0 = 1.402$  (0.356)  
 $\alpha_I = 2.210$  (0.590)  
 $\rho_I = 2.000$  (\*)     $\pi_0 = 21.236$   
 $\alpha_1 = 0.406$  (0.159)     $\pi_I = 20.793$   
 $\sigma_1 = 0.006$  (0.000)     $\hat{\pi}_1 = 2.205$

Normalized innovations variance-covariance matrix:

P	Y	V	$\pi_0^D$	$\pi_1^D$	$\pi_I$
5.77	3.37	2.40	-2.42	0.09	4.73
	1.96	1.40	-1.41	0.05	2.76
		1.00	-1.01	0.04	1.97
			1.05	0.00	-2.06
				0.04	0.00
					4.03

$\sigma(\Delta D_t) = 0.005$ ,  $\sigma(\Delta P_t) = 0.172$

Correlations:

	$\Delta D_t$	$\Delta P_t$		$\Delta D_t$	$\Delta P_t$
$\Delta D_t$	1.00	0.05	$\Delta P_t$	0.05	1.00
$\Delta D_{t-1}$	0.12	-0.02	$\Delta P_{t-1}$	0.53	-0.01
$\Delta D_{t-2}$	-0.13	-0.02	$\Delta P_{t-2}$	0.15	-0.01
$\Delta D_{t-3}$	-0.09	-0.02	$\Delta P_{t-3}$	-0.00	-0.01
$\Delta D_{t-4}$	-0.06	-0.01	$\Delta P_{t-4}$	-0.01	-0.01
$\Delta D_{t-5}$	-0.04	-0.01	$\Delta P_{t-5}$	-0.01	-0.01

(\*): parameter converged to the boundary of the admissible region.

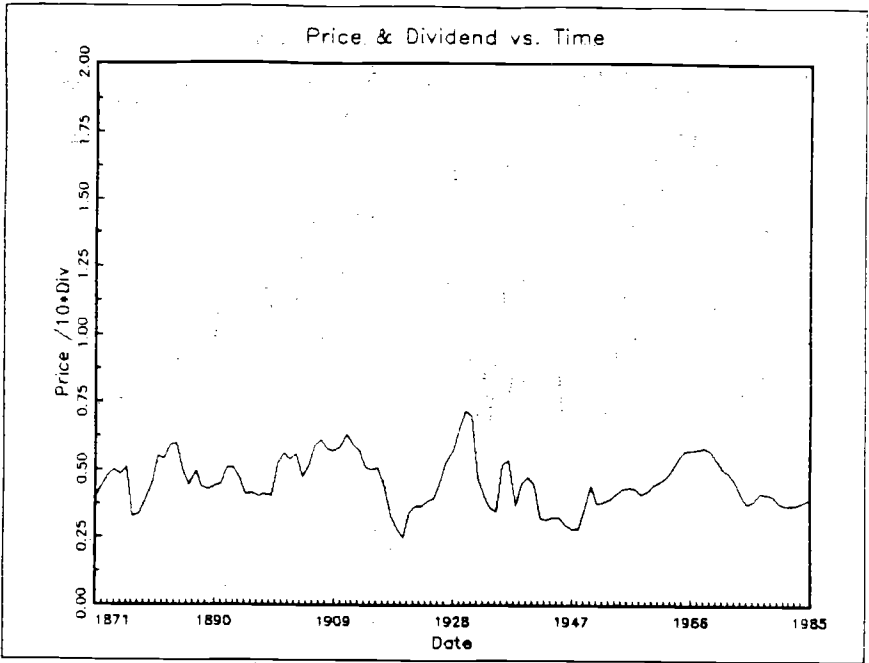


Figure 1

..... price      \_\_\_\_\_ dividend x 10

Note: Stock price and dividend have been exponentially detrended using the sample mean growth rate of the dividend, and then normalized so sample mean price = 1.

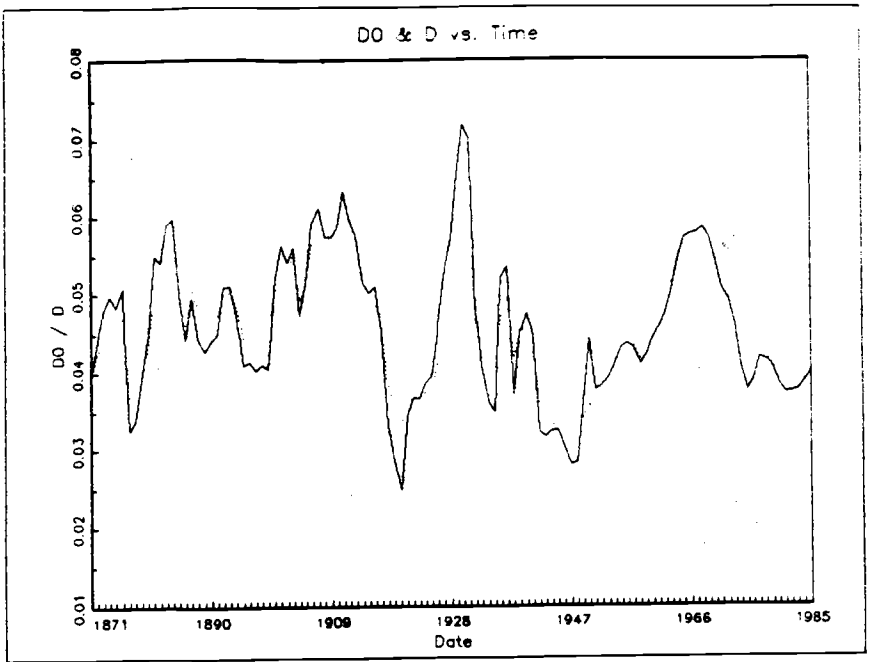


Figure 2

.....  $D_0$       \_\_\_\_\_  $D$

Note: Model estimated has free  $r$  and  $\lambda$ , and full noise.

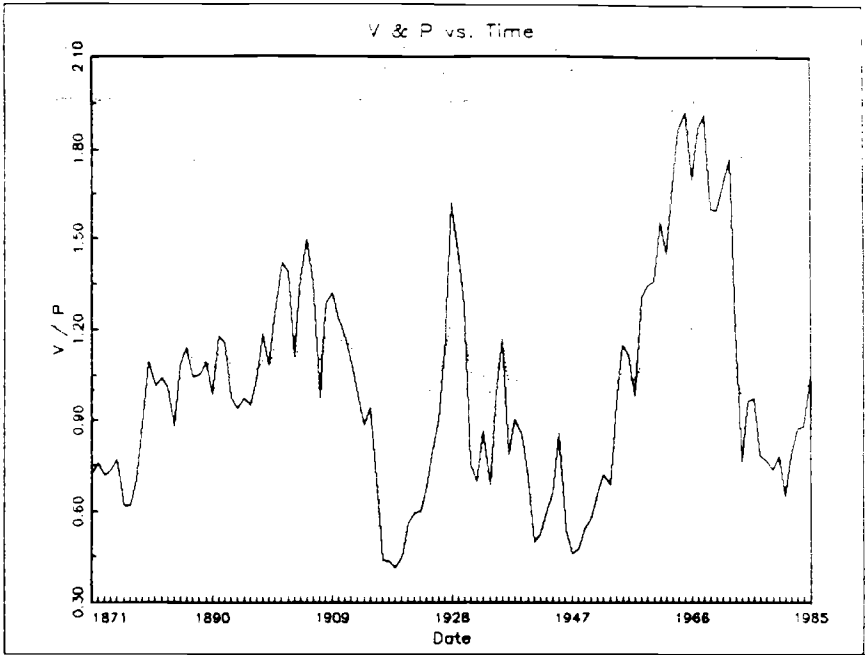


Figure 3

..... V      \_\_\_\_\_ P

Note: Model estimated has free  $r$  and  $\lambda$ , and full noise.

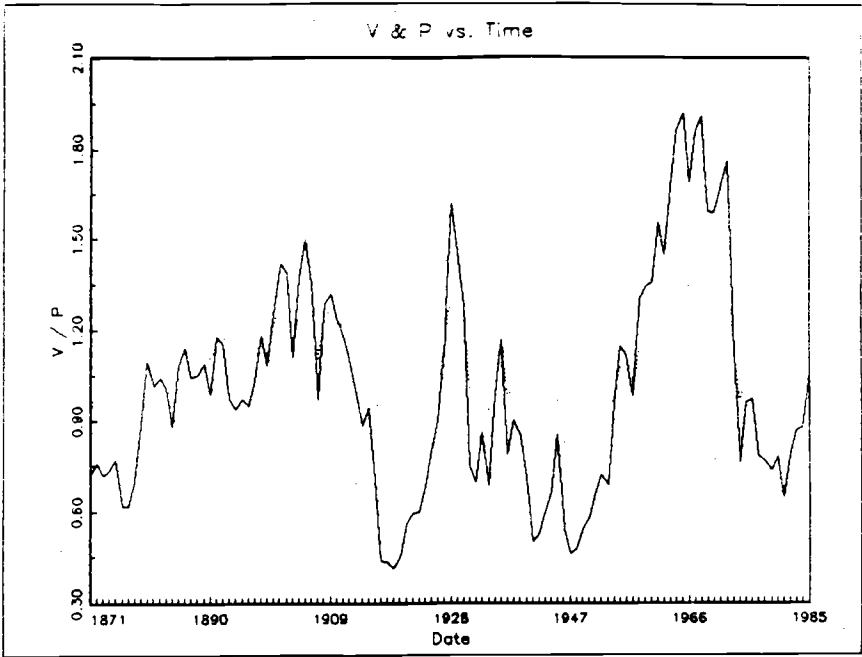


Figure 4

..... V      —— P

Note: Model estimated has  $r = 0.06$ , and full noise.