Econometrica, Vol. 56, No. 1 (January, 1988), 37-61

DIVIDEND INNOVATIONS AND STOCK PRICE VOLATILITY

BY KENNETH D. WEST[1]

A standard efficient markets model states that a stock price equals the expected present discounted value of its dividends, with a constant discount rate. This is shown to imply that the variance of the innovation in the stock price is smaller than that of a stock price forecast made from a subset of the market's information set. The implication follows even if prices and dividends require differencing to induce stationarity. The relation between the variances appears not to hold for some annual U.S. stock market data. The rejection of the model is both quantitatively and statistically significant.

KEYWORDS: Volatility test, efficient markets, stock price, nonstationary, random walk.

1. INTRODUCTION

THE SOURCES OF FLUCTUATIONS in stock prices have long been argued. Some observers have suggested that a major part of the fluctuations results from self fulfilling rumors about potential price fluctuations. In a famous passage, Keynes, for example, described the stock market as a certain type of beauty contest in which judges try to guess the winner of the contest: speculators devote their "intelligence to anticipating what average opinion expects average opinion to be" (1964, p. 136). An examination of practically any modern finance text (e.g., Brealey and Myers (1981)) indicates that the economics profession tends to hold the opposite view. Stock price fluctuations are argued to result solely from changes in the expected present discounted value of dividends.

The subject has received increased attention in recent years because of the volatility tests of Leroy and Porter (1981) and, especially, Shiller (1981a). These tests seem to indicate that stock price fluctuations are too large to result solely from changes in the expected present discounted value (PDV) of dividends. There is, however, some question as to the validity of this conclusion. Marsh and Merton (1986) have objected to the tests' assumption that dividends are stationary around a time trend; Flavin (1983) and Kleidon (1985, 1986) have argued that in small samples the tests are biased toward finding excess volatility.

This paper develops and applies a stock market volatility test that is not subject to these criticisms. The test is based on an inequality on the variance of the innovation in the expected PDV of a given stock's dividend stream, and was first suggested by Blanchard and Watson (1982).[2] The inequality states that if discount rates are constant this variance is smaller when expectations are conditional on

[1] I thank A. Blinder, J. Campbell, G. Chow, S. Fischer, R. Flood, L. P. Hansen, W. Newey, J. Rotemberg, R. Trevor, J. Taylor, the referees, and an editor of this journal for helpful comments, and the National Science Foundation for financial support. Responsibility for remaining errors is my own. This paper was revised while I was a National Fellow at the Hoover Institution. An earlier version of this paper was circulated under the title "Speculative Bubbles and Stock Price Volatility."

[2] While Blanchard and Watson (1982) do suggest examining the inequality that is the focus of this paper, they do not formally establish the validity of the inequality, consider possible nonstationarity of dividends or prices, or test the inequality rigorously. Subsequent to the initial circulation of this paper, however, M. Watson sent me a proof of this inequality that is valid when prices and dividends are stationary.

the market's information set than when expectations are conditional on a smaller information set. It may be shown that this implies that the variance of the innovation in a stock price is bounded above by a certain function of the variance of the innovation in the corresponding dividend.

The paper checks whether the bound is satisfied by some long term annual data on the Standard and Poor 500 and the Dow Jones indices. It is not. The estimated variance of the stock price innovation is about four to twenty times its theoretical upper bound. The violation of the inequality is in all cases highly statistically significant.

It is to be emphasized that the inequality is valid even when prices and dividends are an integrated ARIMA process with infinite variances, and that the empirical work allows for such nonstationarity. In addition, the test procedure does not require calculation of a perfect foresight price; this price appears to be central to the small sample biases that are argued by Flavin (1983), Kleidon (1985, 1986), and Marsh and Merton (1986) to plague the Shiller (1981a) volatility test. The paper nonetheless performs some small Monte Carlo experiments to check whether under certain simple circumstances small sample bias in this paper's test procedure is likely to explain the results of the test. The answer is no.

While one of the purposes of this paper is to apply a volatility test with a relatively weak set of maintained statistical assumptions, that is not its only aim. It also considers the consistency of some of the test's maintained economic assumptions with the data, to help determine which among these should be relaxed, so that the excess price volatility might be explained. To that end, the paper uses a battery of formal diagnostic tests on the regressions that must be estimated to calculate the inequality. The test results are in general quite consistent with the test's maintained hypotheses of rational expectations and, perhaps surprisingly, of a constant rate for discounting future dividends. Some additional, less formal analysis, which considers further the constant discount rate hypothesis, does not suggest that the excessive price variability results solely from variation in discount rates.

The evidence, then, does not suggest that the excess volatility is caused by a simple failure of the rational expectations or constant discount rate assumptions. This suggests the possibility that the volatility is due either to rational bubbles (e.g., Blanchard and Watson (1982)) or nearly rational "fads" (e.g., Summers (1986)), whose profit opportunities (if any) are difficult to detect. The paper does not, however, attempt to make a case for bubbles, fads or, for that matter, any other factor, as the explanation of the excess volatility. Instead what is emphasized are two empirical regularities that seem to characterize the data studied here. The first is that prices appear to be too variable to be set as the expected PDV of dividends, with a constant discount rate; this holds even if prices and dividends are nonstationary. The second is that it is difficult to attribute the excess variability to variations in discount rates. Reconciliation of these two points is a task left for future research.

Before turning to the details of the subject at hand, two final introductory remarks seem worth making. The first is that the inequality established here may be of general interest in that it could be used to test other infinite horizon present value models. Possible examples include testing whether consumption is too variable to be consistent with the permanent income hypothesis (see Deaton (1985), West (1988)) or whether exchange rates are too variable to be consistent with a standard monetary model (West (1986a)). That the inequality is valid even in a nonstationary environment makes it particularly appealing in these and perhaps other contexts. The second remark concerns the estimation technique. This is in part an application of West's (1986b) result that it is not always necessary to difference regressions on nonstationary variables, to obtain asymptotically normal parameter estimates. The key requirement is that the nonstationary variables have a drift. Since this is plausible for not only stock prices but for many other macroeconomic variables as well, the estimation technique applied in this paper may be of general interest.

The plan of the paper is as follows. Section 2 establishes the basic inequality. Section 3 explains how the inequality may be used to test a rational expectations, constant discount rate stock price model. Section 4 presents formal econometric results. Section 5 considers informally whether small sample bias or discount rate variation are likely to explain the Section 4 results. Section 6 has conclusions. The Appendix has some econometric and algebraic details.

2. THE BASIC INEQUALITY

The following proposition is the basis of this paper.

PROPOSITION 1. *Let I_t be the linear space spanned by the current and past values of a finite number of random variables, with I_t a subset of I_{t+1} for all t. It is assumed that after s differences, all random variables in I_t jointly follow a covariance stationary ARMA (q, r) process, for some finite s, q, $r \geq 0$. This s'th difference is assumed without loss of generality to have zero mean. All variables are assumed to be identically zero for $t \leq q$.*

Let d_t be one of these variables. Let H_t be a subset of I_t consisting of the space spanned by current and past values of some subset of the variables in I_t, including at a minimum current and past values of d_t. Let b be a positive constant, $0 \leq b < 1$. Let $P(\cdot\,|\,\cdot)$ denote linear projections, calculated for $s > 0$ as in Hansen and Sargent (1980). Let

$$x_{tI} = \lim_{k\to\infty} P\left(\sum_0^k b^j d_{t+j} \,\middle|\, I_t\right), \qquad x_{tH} = \lim_{k\to\infty} P\left(\sum_0^k b^j d_{t+j} \,\middle|\, H_t\right).$$

(All summations in this section run over j.) Let E denote mathematical expectations. Then

$$(1) \qquad E[x_{tH} - P(x_{tH}\,|\,H_{t-1})]^2 \geq E[x_{tI} - P(x_{tI}\,|\,I_{t-1})]^2.$$

PROOF:[3] Since d_t is in I_t,

$$x_{tI} = d_t + bP\left(\sum_0^\infty b^j d_{t+j+1} \mid I_t\right) \tag{2}$$

$$= d_t + bx_{t+1,I} - be_{t+1},$$

$$e_{t+1} = x_{t+1,I} - P\left(\sum_0^\infty b^j d_{t+j+1} \mid I_t\right) = x_{t+1,I} - P(x_{t+1,I} \mid I_t).$$

Equation (2) may be rewritten as

$$x_{tI} - d_t = bx_{t+1,I} - be_{t+1}.$$

Recursive substitution for $x_{t+1,I}$, then for $x_{t+2,I}$, etc., yields

$$x_{tI} - \sum_0^{k-1} b^j d_{t+j} = b^k x_{t+k,I} - \sum_1^k b^j e_{t+j}. \tag{3}$$

The assumptions of the proposition insure that as $k \to \infty$, $b^k x_{t+k,I} \to 0$ in mean square. Consider first the ARIMA $(q, s, 0)$ case. The formulas in Hansen and Sargent (1980) state that x_{tI} is a finite distributed lag on the variables in I_t. Since these variables started up at a finite date in the past, and some arithmetic difference of each variable has a finite variance, the rate of growth of the variance of each variable, and therefore of the variance of x_{tI} as well, is some power of t. Since $\lim_{k\to\infty} b^{2k}(t+k)^n = 0$ for any fixed $n \geqslant 0$, $\lim_{k\to\infty} \operatorname{var}(b^k x_{t+k,I}) = 0$. The argument for the ARIMA (q, s, r) case is implied by Hansen and Sargent (1980) since for $j > r$, $P(d_{t+j} \mid I_t)$ is determined by a difference equation that depends only on s and the autoregressive parameters.

The assumptions of the proposition also guarantee that e_t has constant, finite variance and is serially uncorrelated. For the ARIMA $(q, s, 0)$ case, this follows directly from inspection of the formula for x_{tI} in Hansen and Sargent (1980). Once again, this argument immediately extends to the ARIMA (q, s, r) case. Therefore, $\sum_1^\infty b^j e_{t+j}$ exists, in the sense that $\lim_{k\to\infty} E(\sum_1^k b^j e_{t+j} - \sum_1^\infty b^j e_{t+j})^2 = 0$, $\operatorname{var}(\sum_1^\infty b^j e_{t+j}) = b^2(1-b^2)^{-1} Ee_t^2$ (Fuller (1976, p. 36)). Equation (3) therefore implies

$$x_{tI} - \sum_0^\infty b^j d_{t+j} = -\sum_1^\infty b^j e_{t+j}. \tag{4}$$

By a similar argument, involving projections onto H_t,

$$x_{tH} - \sum_0^\infty b^j d_{t+j} = -\sum_1^\infty b^j f_{t+j}, \tag{5}$$

$$f_{t+j} = x_{t+j,H} - P(x_{t+j,H} \mid H_{t+j-1}), \qquad \operatorname{var}\left(-\sum_1^\infty b^j f_{t+j}\right) = b^2(1-b^2)^{-1} Ef_t^2.$$

[3] J. Campbell suggested the basic idea of this proof. L. P. Hansen and M. Watson have provided alternative proofs. S. Leroy has pointed out to me that a similar proposition is implied in the stationary case in Leroy and Porter (1981, p. 568). My own, rather tedious, proof may be found in an earlier version of this paper (West (1984)).

Now,

$$\text{(6)} \qquad \operatorname{var}\left(-\sum_1^\infty b^j f_{t+j}\right) = \operatorname{var}\left(x_{tH} - \sum_0^\infty b^j d_{t+j}\right)$$
$$= \operatorname{var}\left(x_{tH} - x_{tI} + x_{tI} - \sum_0^\infty b^j d_{t+j}\right)$$
$$= \operatorname{var}\left(x_{tH} - x_{tI} - \sum_1^\infty b^j e_{t+j}\right)$$
$$= \operatorname{var}(x_{tH} - x_{tI}) + \operatorname{var}\left(-\sum_1^\infty b^j e_{t+j}\right)$$
$$\geqslant \operatorname{var}\left(-\sum_1^\infty b^j e_{t+j}\right).$$

The last equality follows since for $j \geqslant 1$, e_{t+j} is uncorrelated with anything in I_t, including, in particular, $x_{tH} - x_{tI}$. It follows from (6) that $b^2(1-b^2)^{-1}Ef_t^2 \geqslant b^2(1-b^2)^{-1}Ee_t^2 \Rightarrow Ee_t^2 \geqslant Ef_t^2$, i.e., $E[x_{tH} - P(x_{tH} \mid H_{t-1})]^2 \geqslant E[x_{tI} - P(x_{tI} \mid I_{t-1})]^2$. *Q.E.D.*

A verbal restatement of Proposition 1 is as follows. Suppose we are forecasting the present discounted value of d_t, by calculating x_{tI} and x_{tH}. Each period as new data become available we revise our forecast. $E(x_{tI} - P(x_{tI} \mid I_{t-1})]^2$ and $E(x_{tH} - P(x_{tH} \mid H_{t-1})]^2$ are measures of the average size of this period to period revision. Proposition 1 says that with less information the size of the revision tends to be larger. That is, when less information is used, the variance of the innovation in the expected present discounted value of d_t is larger.

It is worth making four comments on the conditions under which (1) is valid. First, if the random variables in I_t are stationary without differencing, Proposition (1) does not require that the variables follow a finite parameter ARMA (q, r) process. The ARIMA assumption is maintained because to my knowledge infinite horizon prediction for nonstationary variables has been developed only for ARIMA processes. Second, (1) may not extend immediately if logarithms or logarithmic differences are required to induce stationarity in d_t, even if linear projections are replaced with mathematical expectations. If, for example, $\log(d_t) = \log(d_{t-1}) + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma^2)$, and H_t is the information set generated by past d_t, it may be shown that $[x_{tH} - E(x_{tH} \mid H_{t-1})]^2$ is proportional to d_{t-1}^2. Third, the inequality need not hold for a finite horizon. That is, it need not hold if we consider the variance of the innovation in the expected PDV of $\sum_0^n b^j d_{t+j}$ instead of $\sum_0^\infty b^j d_{t+j}$. An example is given in footnote 4 of West (1986c). The reason is that terms of the form $b^{n+1}x_{t+n+1,I}$ and $b^{n+1}x_{t+n+1,H}$ are present. See equation (3). Fourth, (1) does not hold for arbitrary subsets of I_t. If, for example, H_t were the empty set, x_{tH} would also be the empty set, and the left-hand side of (1) would be identically zero.

Before developing the implications of (1) for stock price volatility, it may be helpful to work through a simple example. Suppose I_t consists of lags of d_t and of one other variable, z_t. Let H_t consist simply of lags of d_t. Let the bivariate (d_t, z_t) representation be

$$\begin{bmatrix} d_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} d_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{7}$$

with $|\phi| \leqslant 1$, ε_{1t} and ε_{2t} i.i.d., $E\varepsilon_{1t}\varepsilon_{2s} = 0$ for all t, s. Let $E\varepsilon_{1t}^2 = \sigma_1^2$, $E\varepsilon_{2t}^2 = \sigma_2^2$. The univariate representation of d_t clearly is $d_t = \phi d_{t-1} + v_t$, $v_t = \varepsilon_{1t} + z_{t-1} = \varepsilon_{1t} + \varepsilon_{2t-1}$, $Ev_t^2 = \sigma_1^2 + \sigma_2^2$. Let us calculate both sides of (1).

$$\begin{aligned}
(8) \qquad & P(d_{t+j} \mid H_t) = \phi^j d_t \\
& \Rightarrow x_{tH} = P\left(\sum_0^\infty b^j d_{t+j} \,\middle|\, H_t \right) = (1 - b\phi)^{-1} d_t \\
& \Rightarrow E[x_{tH} - P(x_{tH} \mid H_{t-1})]^2 = E[(1-b\phi)^{-1} v_t]^2 = (1-b\phi)^{-2}(\sigma_1^2 + \sigma_2^2). \\
& P(d_t \mid I_t) = d_t. \\
& P(d_{t+j} \mid I_t) = \phi^j d_t + \phi^{j-1} z_t, \qquad j > 0, \\
& \Rightarrow x_{tI} = P\left(\sum_0^\infty b^j d_{t+j} \mid I_t \right) = (1-b\phi)^{-1}(d_t + bz_t) \\
& \Rightarrow E[x_{tI} - P(x_{tI} \mid I_{t-1})]^2 = E[(1-b\phi)^{-1}(\varepsilon_{1t} + bz_t)]^2 \\
& \qquad = (1-b\phi)^{-2}(\sigma_1^2 + b^2\sigma_2^2).
\end{aligned}$$

Since $b^2 < 1$, $\sigma_1^2 + \sigma_2^2 > \sigma_1^2 + b^2\sigma_2^2$, so (1) holds. Observe that (1) holds even when $\phi = 1$ so that d_t is nonstationary.

3. THE MODEL

According to a standard efficient markets model, a stock price is determined by the relationship (9) (Brealey and Myers (1981, pp. 42–45)):

$$p_t = bE[(p_{t+1} + d_{t+1}) \mid I_t], \tag{9}$$

where p_t is the real stock price at the end of period t, b the constant ex-ante real discount rate, $0 < b = 1/(1+r) < 1$, r the constant expected return, E denotes mathematical expectations, d_{t+1} the real dividend paid to the owner of the stock in period $t+1$, and I_t the information set common to traders in period t. I_t is assumed to contain, at a minimum, current and past dividends, and, in general, other variables that are useful in forecasting dividends.

Equation (9) may be solved recursively forward to get

$$p_t = \sum_1^n b^j E(d_{t+j} \mid I_t) + b^n E(p_{t+n} \mid I_t). \tag{10}$$

If the transversality condition

$$(11)\qquad \lim_{n\to\infty} b^n E(p_{t+n}\,|\,I_t)=0$$

holds, then

$$(12)\qquad p_t=\sum_{1}^{\infty} b^j E(d_{t+j}\,|\,I_t).$$

It will be assumed that in forecasts of $\sum_1^\infty b^j d_{t+j}$ from I_{t-k}, for any $k\geqslant 0$, mathematical expectations conditional on the market's information set are the same as linear projections. So x_{tI}, defined in Proposition 1 as the linear projection of $\sum_0^\infty b^j d_{t+j}$ onto a period t set of random variables equals $E(\sum_0^\infty b^j d_{t+j}\,|\,I_t)$. Similarly, the linear projection of x_{tI} onto the market's period $t-1$ set of random variables equals $E(x_{tI}\,|\,I_{t-1})$.

Proposition 1 is used to test the model (12) as follows. Since $x_{tI}=E(\sum_0^\infty b^j d_{t+j}\,|\,I_t)$, (12) implies that $x_{tI}=p_t+d_t$. So $E[x_{tI}-E(x_{tI}\,|\,I_{t-1})]^2=E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2$, and, therefore,

$$(13)\qquad E[x_{tH}-P(x_{tH}\,|\,H_{t-1})]^2\geqslant E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2.$$

The intuitive reason that the model (12) implies (13) is as follows. $E(x_{tH}-P(x_{tH}\,|\,H_{t-1})]^2$ is by definition a measure of the average size of the innovation in the expected present discounted value (PDV) of dividends, when expectations are conditional on H_t. According to (12), price adjusts unexpectedly only in response to news about dividends. $E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2$ is a measure of the average size of the innovation in the expected PDV of dividends, with expectations conditional on the market's information set I_t. Since the market is presumed to use the variables in I_t to forecast optimally, the market's forecasts tend to be more precise, i.e., (13) holds.[4]

To make (13) operational, both sides of it must be calculated. Consider first $E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2$. A consistent estimate of this is easily obtained by estimating (9) with the instrumental variables method of McCallum (1976) and Hansen and Singleton (1982). Rewrite (9) as

$$(14)\qquad \begin{aligned} p_t &= b(p_{t+1}+d_{t+1})-b[p_{t+1}+d_{t+1}-E(p_{t+1}+d_{t+1}\,|\,I_t)]\\ &= b(p_{t+1}+d_{t+1})+u_{t+1},\\ \sigma_u^2 &= b^2E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2. \end{aligned}$$

Equation (14) can be estimated by instrumental variables, using as instruments variables known at time t. An estimate of $E[p_t+d_t-E(p_t+d_t\,|\,I_{t-1})]^2$ is then obtainable as $\hat{b}^{-2}\hat{\sigma}_u^2$.

[4] Note that inequality (13) holds even for the class of dividend and price processes studied by Marsh and Merton (1986), as long as arithmetic differences suffice to induce stationarity. This is because the March and Merton (1984) model implies that $H_t=I_t$. When $H_t=I_t$, inequality (13) holds trivially, as a strict equality. See the discussion in West (1984).

Estimation of $E[x_{tH} - P(x_{tH} \mid H_{t-1})]^2$ is slightly more involved. It requires first of all specification of H_t. The simplest possible one is $H_t = \{1, d_{t-j} \mid j \geq 0\}$, and H_t defined this way is what is used in this paper's empirical work.[5] Choices of H_t that include lags of additional variables might produce sharper results, but would also entail more complex calculations. With $H_t = \{1, d_{t-j} \mid j \geq 0\}$, $E[x_{tH} - P(x_{tH} \mid H_{t-1})]^2$ can be calculated as a function of d_t's univariate ARIMA parameters. Suppose $d_t \sim$ ARIMA $(q, s, 0)$,

$$\Delta^s d_{t+1} = \mu + \phi_1 \Delta^s d_t + \cdots + \phi_q \Delta^s d_{t-q+1} + v_{t+1}, \tag{15}$$

where $\Delta^s = (1-L)^s$, L the lag operator. (A moving average component to d_t is assumed absent for notational and computational simplicity.) Then $x_{tH} = P(\sum b^j d_{t+j} \mid H_t) = m + \sum_1^{q+s} \delta_i d_{t-i+1}$. The δ_i are complicated functions of b and the ϕ_i. Hansen and Sargent (1980) provide explicit formulas for the δ_i. In particular, given b and the ARIMA parameters of d_t, one can use the Hansen and Sargent (1980) formula for δ_1 to calculate $\delta_1^2 \sigma_v^2 = E[x_{tH} - E(x_{tH} \mid H_{t-1})]^2$. To test the null hypothesis that prices are determined according to (12), then we calculate

$$\delta_1^2 \sigma_v^2 - b^{-2} \sigma_u^2 \tag{16}$$

and test H_0: $\delta_1^2 \sigma_v^2 - b^{-2} \sigma_u^2 \geq 0$. If the estimate of (16) is negative (that is, the implications of (12) for the innovation variances are not borne out by the data), a convenient way to quantify the extent of the failure of the model (12) is to calculate

$$-100(\delta_1^2 \sigma_v^2 - b^{-2} \sigma_u^2)/(b^{-2} \sigma_u^2). \tag{17}$$

When (16) is negative, (17) yields a number between 0 and 100. I will refer to this somewhat loosely as the percentage of the variance of the innovation in p_t that is excessive. This is of course somewhat imprecise in that $b^{-2} \sigma_u^2$ is the variance of the innovation in the *sum* of dividends and prices. But given that price innovations are much larger than dividend innovations (see the empirical results below), this terminology does not seem misleading.[6]

What alternatives might explain a rejection of the null hypothesis that (16) is positive? Three have figured prominently in discussions of related work: gross expectational irrationality, of the sort that systematically leads to profit opportunities (e.g., Ackley (1983)); variation in discount rates (e.g., Leroy (1984)); and rational or nearly rational bubbles or fads (e.g., Blanchard and Watson (1982), Summers (1986)), whose profit opportunities (if any) are very difficult to detect. In light of some empirical evidence yet to be presented, it is of interest to note that diagnostic tests on the estimates of (14) and (15) can help to distinguish between bubbles and fads on the one hand, and gross expectational irrationality

[5] Proposition 1 assumed that variables had zero mean. If not, H_t and I_t must be expanded to include suitable deterministic terms. In the annual data used here, a constant is the only relevant such term.

[6] In fact, in some empirical work the variable that is here called d_{t+1} is assumed known at time t and thus has an innovation of zero when forecast at time t (Shiller (1981a), Leroy and Porter (1981)).

and time varying discount rates on the other, as possible explanations of any excess price volatility. Technically, when rational bubbles are absent (i.e., the transversality condition (11) holds), equation (14) and the dividend equation (15) together imply that (16) is positive. But when rational bubbles are present, (14) and (15) need not imply that (16) is positive (West (1986c)). So bubbles provide a logical explanation of any excess price volatility if (14) and (15) appear to be well specified. More generally, since it may be difficult to detect small departures from the rational bubble alternative in a given finite sample, evidence that (14) and (15) appear legitimate, despite excess price volatility, is consistent as well with nearly rational bubbles or fads of the sort considered in Summers (1986). So an essential part of the strategy used here to distinguish between rational or nearly rational bubbles or fads versus other alternatives as explanations of excess price variability is to perform diagnostic tests on equations (14) and (15). The greater the extent to which these two equations appear to be well specified, the more persuasive is the inference that bubbles or fads explain the excess volatility.[7]

4. EMPIRICAL EVIDENCE

A. *Data and Estimation Technique*

The data used were those used by Shiller (1981a) in his study of stock price volatility, and were supplied by him. There were two data sets, both containing annual aggregate price and dividend data. One had the Standard and Poor 500 for 1871-1980 (p_t is price in January divided by producer price index (1979 = 100), d_{t+1} is the sum of dividends from that same January to the following December, deflated by the average of that year's producer price index). The other data set was a modified Dow Jones index, 1928-1978 (p_t, d_{t+1} as above). See Shiller (1981a) for a discussion of the data.

The following aspects of estimation will be discussed in turn: (i) selection of the lag length q of the dividend process, (ii) estimation of (14), (15), and (16), (iii) calculation of the variance-covariance matrix of the parameters estimated, and (iv) diagnostic tests performed.

(i) It was assumed that the univariate d_t process required at most one difference to induce stationarity. That is, in equation (15), $s=0$ (the original series was used) or $s=1$ (first difference of original series used). No other values of s were tried.

For both the differenced and undifferenced versions of each data set's dividend process, two values of the lag length q were used. One was arbitrarily selected

[7] Unless, of course, one has a theoretical presumption that bubbles are not present: a consensus view on how general are the equilibria that admit bubbles is far from established. For a general equilibrium model that allows bubbles, see Tirole (1985). For an argument that bubbles are inconsistent with rationality, see Diba and Grossman (1985). For discussions on the use of volatility tests versus other techniques in studying present value models, see Hamilton and Whiteman (1984), Hansen and Sargent (1981), and Shiller (1981b). See West (1987) on the interpretation of the Summers (1986) alternative as a nearly rational bubble.

as $q=4$. The other was the q selected by the information criterion of Hannan and Quinn (1979). This criterion chooses the value of q that minimizes a certain function of the estimated parameters. Conditional on q being no greater than some fixed upper bound, which I set to 4, the correct q will be chosen asymptotically if the process truly has a finite order autoregressive representation.[8]

Thus, for each data set up to four sets of parameter estimates were calculated: $q=4$, where $q=$ lag length selected by the information criterion, for differenced and undifferenced data. In one case (Dow Jones, differenced), the Hannan and Quinn (1979) criterion chose $q=4$. So only three sets of parameters were calculated for the Dow Jones.

(ii) Calculation of (16) required estimation of the bivariate system consisting of equations (14) and (15). Equation (14) was estimated by Hansen's (1982) and Hansen and Singleton's (1982) two-step, two-stage least squares. The first step obtained the optimal instrumental variables estimator. The $q+1$ instruments used were the variables on the right hand side of (15), i.e., a constant term and q lags of $\Delta^s d_t$ ($s=0$ or $s=1$). Equation (15) was estimated by OLS, with the covariance matrix of the parameter estimates adjusted for conditional heteroskedasticity as described in (iii).

With $\Delta^s d_t \sim AR(q)$, the δ_1 parameter in the formula (16) is $[(1-b)^s \Phi(b)]^{-1}$, $\Phi(b)=1-\sum_1^q b^j \phi_j$ (Hansen and Sargent (1980)). Thus, formula (16) was calculated as $[(1-\hat{b})^s(1-\sum_1^q \hat{b}^j \hat{\phi}_j)]^{-1}\hat{\sigma}_v^2 - \hat{b}^{-2}\hat{\sigma}_u^2$.

The innovation variances $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$ were calculated from the moments of the residuals of the regressions, with a degree of freedom correction used for $\hat{\sigma}_v^2$:

$$\text{(18)} \qquad \hat{\sigma}_u^2 = (T-s)^{-1} \sum_{t=1}^{T-s} \hat{u}_{t+1}^2,$$

$$\hat{\sigma}_v^2 = (T-s-q-1)^{-1} \sum_{t=1}^{T-s} \hat{v}_{t+1}^2.$$

T is the number of observations; $T=110$ for the Standard and Poor's index, $T=51$ for the Dow Jones index.

The parameter vector estimated was thus $\hat{\theta}=(\hat{b}, \hat{\mu}, \hat{\phi}_1, \ldots, \hat{\phi}_q, \hat{\sigma}_v^2, \hat{\sigma}_u^2)$. $\hat{\theta}$ is asymptotically normal with an asymptotic covariance matrix V (see the Appendix and (iii) below).[9] Let $f(\theta)$ be formula (16) above. The standard error on the

[8] The Hannan-Quinn procedure selects the q that minimizes

$$\ln \hat{\sigma}_v^2 + T^{-1} 2qk \ln \ln T,$$

for $q<Q$ for some fixed Q, with $k>1$. I set $Q=4$, $k=1.001$. The choice of $k=1.001$ was made because Hannan and Quinn (1979, p. 194) seem to suggest a value very close to one is appropriate for sample sizes such as those used in this paper.

[9] A referee has suggested that I emphasize that West (1986b), the references for the asymptotic distribution of parameter estimates for differenced specifications, requires $E(\Delta d_t + \Delta p_t) \neq 0$. While this certainly seems reasonable a priori given that the data are from a stock market in a growing economy, the upward drift in the data is not particularly well reflected in formal statistical tests. See, for example, the insignificant constant terms in all the differenced specifications in Table IB. It is reassuring, then, that the Monte Carlo simulations in West (1986b), which assume data as noisy as those used here, suggest that the asymptotic normal approximation can be useful even with the sample sizes used here.

estimate of (16) was calculated as $[(\partial f/\partial\theta)V(\partial f/\partial\theta)']$. The derivatives of f were calculated analytically.

(iii) The estimate of V, the variance covariance matrix of $\hat{\theta}$, was calculated by the methods of Hansen (1982), Newey and West (1987), and West (1986c), so that the estimate would be consistent for an *arbitrary* ARMA process for u_t and v_t. This is necessary because, for example, the correlation between u_t and v_{t+j} may in principle be nonzero for all $j \geqslant 0$. The Newey and West (1987) procedure was used to insure that V was positive semidefinite. Details may be found in the Appendix. It suffices to note here that the procedure for calculating the standard error on (16) properly accounts for the uncertainty in the estimates of both the regression parameters and the variances of the residuals.

(iv) The final item discussed before results are presented is diagnostic tests on equations (14) and (15). Four diagnostic checks were performed.

The first checked for serial correlation in the residuals to the equations, using a pair of tests. As noted above, u_{t+1}, the disturbance to equation (14), is an expectational error. If expectations are rational, then u_{t+1} will be serially uncorrelated. Equation (15)'s disturbance v_{t+1} should also be serially uncorrelated, since v_{t+1} is the innovation in the dividend process.

The first of the pair of serial correlation tests checked for first order serial correlation in u_{t+1} and v_{t+1}. The calculation of the standard errors for this is described in the Appendix. The second of the pair of serial correlation tests, performed only for (15), calculated the Box–Pierce Q statistic for the residuals. This statistic of course simultaneously tests for first and higher order serial correlation. See Granger and Newbold (1977, p. 93).

The second of the four diagnostic checks was performed only on equation (14). This was a test of instrument-residual orthogonality, basically checking whether the residual to (14) is uncorrelated with lagged dividends (Hansen and Singleton (1982)). Let Z_t be the $(q+1)\times 1$ vector of instruments and $\hat{b}$ the estimate of b. The orthogonality test is computed as:

$$\left\{\sum_{t=1}^{T-s} Z_t'[p_t-\hat{b}(p_{t+1}+d_{t+1})]\right\}(T\hat{S}_z)\left\{\sum_{t=1}^{T-s} Z_t[p_t-\hat{b}(p_{t+1}+d_{t+1})]\right\}. \tag{19}$$

$\hat{S}_z$ is an estimate of $E(Z_tu_{t+1})(Z_tu_{t+1})'$ and was calculated as $T^{-1}(\sum Z_tZ_t'\bar{u}_{t+1}^2)$, where $\bar{u}_t$ is the 2SLS residual to (14). The statistic (19) is asymptotically distributed as a chi-squared random variable with q degrees of freedom. This test has the power to detect irrational expectational errors and variations in discount rates that are correlated with dividends.

The third of the four diagnostic checks tested for the stability of the regression coefficients in (14) and (15). Each sample was split in half, a pair of regression estimates was obtained, and equality of the pair was tested. The resulting statistic is asymptotically chi-squared, with one degree of freedom for (14) and $(q+1)$ degrees of freedom for (15). This test clearly has the power to detect shifts in the discount rate, as well as in the dividend process.

The fourth and final diagnostic check performed is implicit in the estimation procedure described above. A variety of specifications for the dividend process were used—differenced and undifferenced, with a variety of lag lengths. Since the results did not prove sensitive to the specification of the dividend process, the likelihood is relatively small that changes in the specification of the dividend process will affect the results.

B. *Empirical Results*

Regression results for (14) and (15) are reported in Tables I-A and I-B. The results in Table I-A suggest that the basic arbitrage equation (1) is a sensible one. The entries in column (4) do not reject the null hypothesis of no serial correlation in u_{t+1}, the disturbance to equation (14). The test statistic in all cases is far from significant at the .05 level. The equation (19) test for instrument-residual orthogonality also does not reject the null hypothesis of no correlation between the instruments and the residuals at the .05 level, for any specification. See column (5).[10]

TABLE 1-A: EQUATION 14
REGRESSION RESULTS

Data Set	(1) Differenced	(2)[c] q	(3)[c] b	(4)[c] ρ	(5)[c] H/sig	(6)[c] Stability/sig
S and P						
1873-1980	no	2[a]	.9311 (.0186)[b]	.0695 (.0766)	5.50/.064	4.55/.033
1874-1980	yes	2[a]	.9413 (.0170)	.0670 (.0974)	2.87/.238	.33/.566
1875-1980	no	4	.9315 (.0158)	.0661 (.0754)	6.96/.138	3.69/.055
1876-1980	yes	4	.9449 (.0136)	.0671 (.0984)	3.15/.533	.28/.594
Modified Dow Jones						
1931-1978	no	3[a]	.9402 (.0301)	−.1040 (.0806)	5.42/.144	1.56/.211
1933-1978	yes	4[a]	.9379 (.0188)	−.1182 (.0752)	5.20/.267	2.02/.154
1932-1978	no	4	.9271 (.0253)	−.1112 (.1493)	6.08/.108	.49/.483

See notes to Table I-B.

[10] Some results of Flood, Hodrick, and Kaplan (1986) should, however, be noted. They apply the test of instrument-residual orthogonality to these data using three lags of d_t/p_t as instruments, and estimating (14) in the form $1 = b(p_{t+1} + d_{t+1})/p_t + \text{error}$. They report $\chi^2(3)$ test statistics with significance levels of .03 for the S and P and .08 for the Dow Jones. This suggests some mild evidence against the model. They also report stronger rejections using some indirect tests of the constant expected return model. See Section 5B for further discussion.

TABLE I-B: EQUATION 15 REGRESSION RESULTS

Data Set	(1) Differenced	(2)[c] q	(3)[c] μ	(4)[c] ϕ_1	(5)[c] ϕ_2	(6)[c] ϕ_3	(7)[c] ϕ_4	(8)[c] ρ	(9)[c] Q/sig	(10)[c] Stability/sig
S and P										
1873–1980	no	2[a]	.168 (.084)[b]	1.196 (.114)	−.238 (.103)			.045 (.025)	36.87/.181	12.93/.005
1874–1980	yes	2[a]	.034 (.029)	.262 (.118)	−.214 (.071)			.002 (.023)	22.79/.824	2.71/.438
1875–1980	no	4	.150 (.080)	1.247 (.116)	−.480 (.093)	.227 (.113)	−.029 (.066)	.001 (.010)	21.39/.875	33.49/.000
1876–1980	yes	4	.036 (.031)	.264 (.115)	−.230 (.094)	.026 (.080)	−.006 (.153)	.001 (.011)	23.98/.773	4.34/.501
Modified Dow Jones										
1931–1978	no	3[a]	1.945 (1.037)	1.265 (.112)	−.664 (.108)	.333 (.098)		.002 (.054)	4.05/1.000	7.53/.111
1933–1978	yes	4[a]	.275 (.405)	.302 (.119)	−.351 (.133)	.051 (.093)	.050 (.176)	−.024 (.067)	9.77/.939	8.06/.153
1932–1978	no	4	1.925 (1.900)	1.263 (.111)	−.662 (.208)	.330 (.209)	.004 (.134)	.005 (.022)	4.06/1.000	10.22/0.69

[a] Lag length q chosen by Hannan and Quinn (1979) procedure.
[b] Asymptotic standard errors in parentheses.
[c] Symbols: q is lag length of dividend autoregression (15); parameters b, μ, ϕ_i are defined in equations (9) and (15); ρ is the first order serial correlation coefficient of disturbance; H is the statistic in equation (19), $H \sim \chi^2(q)$; "stability" is test for stability of coefficients, as described in text, distributed $\chi^2(1)$ in Table I-A and $\chi^2(q+1)$ in Table I-B; Q is Box-Pierce Q statistic, $Q \sim \chi^2(30)$ for S and P, $Q \sim \chi^2(18)$ for Dow Jones. For the "H", "stability", and "Q" columns, "sig" refers to the probability of seeing the statistic under the null hypothesis.

Most important, the discount rate b is estimated plausibly and extremely precisely in all regressions. See column (3). The implied annual real interest rates are about six to seven per cent. These rates are quite near the arithmetic means for ex post returns: 8.1 per cent for the S and P index (1872–1981) and 7.4 per cent for the Dow Jones index (1929–1979). The estimates of the discount rate therefore are reasonable. Moreover, there is little evidence that the rate was different in the two halves of either sample. As indicated in column (6), the null hypothesis of equality cannot be rejected at the five per cent level for any specification except the S and P, undifferenced, $q=2$. In addition, no evidence against the constancy of the discount rate may be found in a comparison of the two halves' mean ex post returns. For the S and P index, these were (in per cent) 8.09 (1872–1926) versus 8.12 (1927–1981); for the Dow Jones the figures are 7.87 (1929–1954) versus 6.92 (1955–1979).

In general, then, the specification of the arbitrage equation (14) seems acceptable, with the possible exception of the S and P data set with dividends undifferenced. Let us now turn to the estimates for the dividend process, reported in Table I-B. Once again, the entries in columns (8) and (9) allow comfortable acceptance of the null hypothesis of no serial correlation in the disturbance to equation (15). With one exception, both test statistics in all regressions are far from significant. The only possible exception was the estimate of the first order serial correlation coefficient $\hat{\rho}$ for the S and P index, undifferenced, lag length $=2$. Note, however, that this regression's Q statistic in column (9) comfortably accepts the null hypothesis of no serial correlation. Overall, then, no serial correlation to the residual to (15) is apparent. Also, the estimates of most regression coefficients are fairly precise, at least when the lag length q was chosen by the Hannan and Quinn (1979) procedure. Finally, the null hypothesis that the parameters of the dividend process are the same in the two halves of each sample cannot be rejected for any specification except the Standard and Poor's, undifferenced. See column (10). Overall, then, the specification of the dividend process seems quite acceptable, again with the possible exception of the S and P data set, undifferenced.

The null hypothesis that price is the expected present discounted value of dividends, with a constant discount rate, does not, however, appear acceptable, for any specification. As may be seen from column (7) in Table II, formula (16) is always negative, and significantly so. The asymptotic z-stat (ratio of parameter to asymptotic standard error) was always larger than 2.5. This means that the column (7) entries are always significant at the one-half per cent level, using a one-tailed test. The null hypothesis may therefore be rejected at traditional significance levels. Furthermore, the fraction of the variance of the price innovation that is excessive is substantial, about 80 to 95 per cent (column (8) of Table II).

The residual price fluctuation might reflect grossly irrational reaction to news about dividends, variation in discount rates, or some combination of these and other factors. For the S and P undifferenced specifications, the econometric evidence perhaps is not particularly helpful in discriminating among the various possibilities. It is worth noting, however, that for the other specifications, the

TABLE II

TEST STASTICS

Data Set	(1) Differenced	(2) q	(3)[c] $\hat{b}$	(4)[c] $\hat{\delta}_1$	(5)[d] σ_u^2	(6)[d] $\hat{\sigma}_v^2$	(7)[d] Eqn. (16)	(8) Eqn. (17)
S and P	No	2[a]	.9311	10.82	215.2	.1501	−230.66	92.92
			(.0186)[b]	(3.47)	(79.0)	(.0543)	(87.10)	
	Yes	2[a]	.9413	18.06	214.1	.1485	−193.22	79.95
			(.0170)	(6.22)	(80.2)	(.0523)	(71.07)	
	No	4	.9315	10.76	219.4	.1502	−235.51	93.12
			(.0158)	(3.10)	(73.4)	(.0510)	(90.12)	
	Yes	4	.9449	18.45	218.2	.1538	−192.05	78.58
			(.0136)	(5.64)	(81.1)	(.0511)	(73.63)	
Modified Dow-Jones	No	3[a]	.9402	8.28	19,653	9.980	−21,545	96.92
			(.0301)	(2.85)	(5,836)	(3.383)	(5,978)	
	Yes	4[a]	.9379	15.78	19,342	9.014	−19,740	89.79
			(.0188)	(8.13)	(5,871)	(2.655)	(5,852)	
	No	4	.9271	7.55	19,228	10.453	−21,777	97.34
			(.0253)	(3.12)	(3,912)	(2.427)	(4,309)	

[a] Lag length q chosen by Hannan-Quinn (1979) criterion.
[b] Asymptotic standard errors in parentheses.
[c] Symbols: q = lag length in dividend regression; b defined in equation (9); δ_1 defined above equation (18); σ_u^2 and σ_v^2 defined in equation (18).
[d] Units for columns (5)-(7) are 1979 dollars squared. For the S and P, $P_{1979} = 99.71$, $d_{1979} = 5.65$; for the Dow-Jones, $P_{1979} = 468.94$, $d_{1978} = 30.91$.

results of the diagnostic tests were more consistent with the residual volatility being due to bubbles or fads whose profit opportunities are difficult to detect, than to a misspecification of the arbitrage or dividend equations.[11]

5. SOME ADDITIONAL ANALYSIS

This section considers the possibilities that the previous section's results are due to (A) small sample bias, or (B) variation in discount rates. It is to be emphasized that the analysis is informal, and the conclusions are far from definitive. The goal here is simply to gather some evidence on whether either possibility explains the results; a complete, rigorous econometric examination of either possibility would require a separate paper.

A. *Small Sample Bias*

This section uses two small Monte Carlo experiments to get a feel for the importance of two types of bias. Part (a) below considers whether under certain simple circumstances small sample bias is likely to account for the finding of excess variability. Part (b) studies whether under equally simple circumstances low small sample power of the equation (19) test of instrument residual volatility is likely to explain the generally favorable results of the diagnostic tests.

a. *Bias in Estimate of Excess Volatility*

It is important to consider whether small sample bias explains the finding of excess variability, in light of the evidence in Kleidon (1985, 1986) and Marsh and Merton (1986) suggesting that if prices and dividends are nonstationary, the Shiller (1981a) variance bounds test is strongly biased towards finding excess variability. To see whether there is a similar bias in the present paper's test, an environment similar to that in Kleidon (1985, 1986) and Marsh and Merton (1986) was assumed. Two Monte Carlo experiments were performed. The first assumed that dividends follow a random walk, $\Delta d_t = \mu + v_t$, the second that dividends follow a lognormal random walk, $\Delta(\log d_t) = f + w_t$. In both experiments, it was assumed that only lagged dividends were used to forecast future dividends, so that $H_t = I_t$.

[11] This seems an appropriate place to give the results of another test of this model, suggested to me by a referee. Equation (6) states that $\operatorname{var}(x_{tH} - \sum_0^\infty b^j d_{t+j}) - \operatorname{var}(x_{tI} - \sum_0^\infty b^j d_{t+j}) - \operatorname{var}(x_{tH} - x_{tI}) = 0$. So, under the null hypothesis that $x_{tI} = p_t + d_t$,

$$\delta_1^2\sigma_v^2 - b^{-2}\sigma_u^2 - b^{-2}(1-b^2)\operatorname{var}[p_t + d_t - (m + \textstyle\sum_1^{q+s} \delta_i d_{t-i+1})] = 0.$$

The formulas for $m, \delta_1, \ldots, \delta_{q+s}$, which are needed to calculate x_{tH} under the null, may be found in West (1987).

I tested this equality constraint for all seven specifications, with the number of lags used in the calculation of the matrix $\hat{S}$ (defined in the Appendix) set to 11. The z-statistics for the seven specifications, presented in the same order as in Table II, were: 1.88, 2.07, 1.71, 2.23, 1.85, 2.17, 1.71. Thus this suggests some mild evidence against the null hypothesis.

The basic reason for the relatively low statistics was a very noisy estimate of $\operatorname{var}[p_t + d_t - (m + \sum_1^{q+s} \delta_i d_{t-i+1})]$. This was insignificantly different from zero at the five per cent level, for all seven specifications.

In the first experiment, μ and σ_v^2 were matched to the S and P sample values of the mean and variance of Δd_t, $\mu = .0373$, $\sigma_v^2 = .1574$. b was set to .9413, the value estimated in line 2 of Table I-A. For each of 1000 samples, the following was done: A vector of 100 independent normal shocks was drawn, $(v_1, \ldots, v_{100})$, using the IMSL routine GGNPM. Dividends and prices were calculated as $\Delta d_t = .0373 + v_t$; $d_t = d_0 + \sum \Delta d_s$ $(d_0 = 1.3)$; $p_t = \sum (.9413)^j Ed_{t+j} \mid I_t = m + \delta_1 d_t$, $m = (.0373)*(.9413)/(1-.9413)$, $\delta_1 = .9413/(1-.9413)$. $\hat{\mu}$ and $\hat{\sigma}_v^2$ were then estimated by an OLS regression of Δd_t on a constant, $\hat{b}$ and $\hat{\sigma}_u^2$ by an instrumental variables regression of equation (14), with a constant as the only instrument. Finally, formula (17), the percentage of price variability that is excessive, was calculated from the estimated parameters. Since $H_t = I_t$, the population value of (17) is zero.

Table III-A presents the empirical distribution of the estimates of formula (17). Ideally, the median value of this distribution would be zero, with half the samples yielding a positive value to (17). Instead, there appears to be a very slight bias towards finding excess variability, with 53 per cent of the estimates being positive. The bias is not, however, particularly marked, and fewer than 5 per cent of the simulated regressions produced the extreme values of the sort found in all of the Table II specifications.

The second experiment assumed that log differences are required to induce stationarity, as in the Monte Carlo evidence in Kleidon (1986). It was noted earlier that the proof of Proposition 1 assumes that arithmetic differences suffice to induce stationarity. Since this is not true in the present Monte Carlo experiment, it is not clear what value (if any) formula (17) will converge to as the sample size grows.[12] The aim of the experiment, then, is not to evaluate the small sample divergence of estimates of (17) from a population value, but to see if this form of nonstationarity is likely to account for the large values found in column (8) of Table II.

The experiment assumed that Δ (log d_t) $= f + w_t$, with f and σ_w^2 matched to the S and P sample values for the mean and variance of $\Delta(\log d_t)$, $f = .013$, $\sigma_w^2 = .016$. b was again set to .9413. The log d_t data were generated by the obvious analogue to the procedure described above for the first experiment, with $d_t = \exp[\log(d_t)]$ and $p_t = \delta d_t$, $\delta = \exp(f + \sigma_w^2/2)/[b^{-1} - \exp(f + \sigma_w^2/2)] = 24.82$. This experiment used a different seed than did the first to initiate the generation of the random w_t. The parameters needed to calculate (17) were estimated exactly as in the first

TABLE III-A

MONTE CARLO DISTRIBUTION OF FORMULA (17) FOR ARITHMETIC RANDOM WALK

Percentile	1	5	50	53
Formula (17)	100.0	31.1	1.7	0.0

[12] In this model one can, however, place a theoretical lower bound on the variance of the innovation in the first difference of log dividends. I tested this in West (1987) and, once again, found that this variance is so small that it is unlikely that a lognormal random walk model generates the data.

experiment. So, for example, arithmetic first differences of d_t were regressed on a constant, even though logarithmic first differences were in fact required to induce stationarity.

Table III-B presents the results of the experiment. This time, over four fifths of the estimates of (17) were positive, suggesting a tendency to find excess volatility. Once again, however, fewer than 5 per cent of the simulated regressions produced the extreme values found in all the Table II specifications. This indicates that it is unlikely that the basic results of the empirical work are attributable to the small sample effects of the misspecification considered in this experiment. More generally, in conjunction with the other Monte Carlo experiment and the empirical results in the previous section, this indicates that the apparent inconsistency of the simple efficient markets model with the S and P and Dow Jones data is unlikely to result from the nonstationarity that is central to Kleidon's (1986) critique of Shiller (1981a).

That the estimates for the artificial data rarely display the extreme Table II excess variability suggests more strongly than might be immediately apparent that small sample bias does not explain the Table II results. This is because Table III-A, and possibly Table III-B as well, contain worst case figures, since they are based on simulations in which $H_t = I_t$. Proposition 1 implies that for any given b and univariate Δd_t process, σ_u^2 will be smaller when I_t contains additional variables useful in forecasting d_t than when $I_t = H_t$. This suggests that when I_t contains these variables estimates of σ_u^2 and of formula (17) will be smaller as well. But a simulation with such variables in I_t does not seem worth undertaking because even under worst case circumstances assumed here, there is little to suggest that small sample bias explains the extreme excess variability reported in Table II.

b. *Bias in Test of Instrument-Residual Orthogonality*

It is possible that the diagnostic tests reported basically favorable results because the tests have low power against some interesting alternatives; see Summers (1986), for example, on tests for serial correlation. It is particularly difficult to consider this comprehensively, even if only one of the diagnostic tests is analyzed. This is because Monte Carlo experiments here are potentially quite burdensome computationally. This will be true if p_t or d_t are generated nonlinearly under the alternative, as will be the case, for example, in most formulations of the Lucas (1978) asset pricing model.

TABLE III-B

MONTE CARLO DISTRIBUTION OF FORMULA (17) FOR LOGNORMAL RANDOM WALK

Percentile	1	5	50	83
Formula (17)	81.0	73.9	41.1	0.0

So this section has a relatively modest aim, of using a single diagnostic test and a single, simple form of misspecification, to suggest whether the data and sample size are such that the diagnostic tests are unlikely to detect plausible misspecifications. The test that is used is the equation (19) test of instrument residual orthogonality. The misspecification that is assumed is that expectations are static rather than rational, $Ed_{t+j} \mid I_t = d_t$. In such a case, the disturbance to the arbitrage equation (14) is $-b(\Delta p_{t+1} + \Delta d_{t+1})$. So the test must pick up a correlation between $\Delta p_{t+1} + \Delta d_{t+1}$ on the one hand and lagged Δd_t (the instruments, assuming a differenced specification) on the other. Note that there are variations in (mathematically) expected returns.

Under this alternative, $p_t = [b/(1-b)]d_t$; $b = .9413$ was again assumed. Dividends were assumed to be generated by an ARIMA (2, 1, 0) process, with the parameters given by line (2) of Table I-B. The following was done 1000 times. A vector of 100 independent normal disturbances was generated, with the variance of the disturbances equal to that reported in line (2), column (6) of Table II, and with a different random number seed than those used in the other experiments. One hundred Δd_t's, and then one hundred d_t's and p_t's, were computed, with initial conditions matching the initial values of the S and P ($\Delta d_{-1} = .16$, $\Delta d_0 = .11$, $d_0 = 1.61$). $\hat{b}$ was then estimated by two-step, 2SLS, with a constant, Δd_t, and Δd_{t-1} as instruments. Finally, the equation (19) statistic was calculated.

The distribution of this statistic, which is a $\chi^2(2)$ random variable under the null, is reported in Table III-C. In over three fourths of the samples, the statistic was above 5.99, the ninety-five per cent level for a $\chi^2(2)$ random variable. In over nine tenths of the samples, the statistic was over 2.87, the value reported in line (2), column (5), in Table I-A.

Against this alternative, then, the test of instrument residual orthogonality appears to have reasonable power. Whether this applies to other alternatives or to the other diagnostic tests performed is uncertain. But the limited amount of evidence presented here at any rate does not suggest that the favorable results of the diagnostic tests result solely from low power of the tests.

B. *Variation in Discount Rates*

One possible explanation for the excess variability found in Section 4 is that discount rates are time varying, so that the error in equation (14) reflects not only news about dividends but also about discount rates (or, equivalently, expected returns). Special consideration of the plausibility of this variation as an explanation seems warranted, given theoretical work such as Lucas (1978) and

TABLE III-C
MONTE CARLO DISTRIBUTION OF EQUATION (19)

Percentile	5	10	50	77	95
Equation (19)	∞6.57	14.83	8.86	5.99	2.89

empirical evidence such as in Shiller (1984) and Flood, Hodrick, and Kaplan (1986).

This will be done in two separate exercises. The first (part (a) below) assumes as in, e.g., Hansen and Singleton (1982) that a consumption based asset pricing model determines expected returns, with the representative consumer's utility function displaying constant relative risk aversion. For small values of the coefficient of relative risk aversion, this permits exact calculation of formula (17), the percentage excess variability. The second (part (b) below) does not model expected returns parametrically but instead uses Shiller's (1981a) linearized version of a completely general model. This permits calculation of a lower bound to how large a standard deviation in expected returns is required to explain the excess variability reported in Table II.

a. *Consumption Based Asset Pricing Model*

Consider the class of models (e.g., Hansen and Singleton (1982)) in which the first order condition for the return on a stock is $E\{\{\beta C_{t+1}/C_t)^{-\alpha}[(p_{t+1}+d_{t+1})/p_t]\}|I_t\}=1$, where β, $0<\beta<1$, is the representative consumer's subjective discount rate, C_t is his real consumption, α his coefficient of relative risk aversion, with E, d_t, p_t, and I_t defined as above. This may be rearranged as

$$\tilde{p}_t=\beta E[(\tilde{p}_{t+1}+\tilde{d}_{t+1})|I_t], \tag{20}$$

$$\tilde{p}_t=p_tC_t^{-\alpha}, \quad \tilde{d}_t=d_tC_t^{-\alpha}.$$

Equation (20) is of the same form as equation (9). R. Flood has pointed out to me that if $\tilde{d}_t$ is stationary, perhaps after one or more differences are taken, the statistics computed in the constant discount rate case can be computed in this model as well. Repetition of the entire procedure is beyond the scope of this paper (and, in light of the results about to be presented, seems pointless). Instead, I will focus on obtaining a point estimate of formula (17), the percentage excess variability, for various imposed values of β and α.

The C_t variable used in these calculations was the Grossman and Shiller (1981) annual figure on real, per Capita consumption of nondurables and services, 1889-1978. $\tilde{d}_t$ and $\tilde{p}_t$ were calculated using the S and P data for various values of α. A simple plot of $\tilde{d}_t$ suggested that $\tilde{d}_t$ in neither levels nor first nor higher differences is stationary for α much bigger than one. The problem is that for big α, $\tilde{d}_t$ displays a marked secular decline, because annual C_t growth was slightly higher than annual d_t growth.

I nonetheless calculated (17), the percentage excess variability, for a wide range of α, just in case $\tilde{d}_t$ really is stationary for large α. This was done for $\beta=.95$ and $\beta=.98$, with very similar results. In all cases the lag length of the $\tilde{d}_t$ autoregression was set to four. Table IV-A contains the figures that resulted for some of the α, with $\beta=.98$. Since (17) was not only positive but large, the price and dividend data are as inconsistent with the model implied by (20) as they are with the constant expected return model assumed in Sections 3 and 4. There is

TABLE IV-A
PERCENTAGE EXCESS PRICE VARIABILITY

α	.5	1	2	3	10	25	50
Formula (17)	96.5	97.5	80.9	88.4	99.6	100.0	100.0

therefore no evidence supporting the hypothesis that the excess variability displayed in Table II is explained solely by the sort of variation in expected returns predicted by this asset pricing model.[13]

Since $\tilde{d}_t$ does not appear stationary for α much bigger than unity, it is equally true that Table IV-A contains no evidence against the hypothesis that the Table II excess variability is explained by variation in expected returns associated with a coefficient of risk aversion greater than, say, one. Table IV-A does, however, suggest if the model of expected returns assumed here is correct, that the Table II excess variability is unlikely to be due to variation in expected returns associated with a coefficient of relative risk aversion of less than, say, one.

b. *Linearized Model*

Let us now consider a general model that does not parameterize expected returns, linearized as in Shiller (1981a) to make the analysis tractable. Let r_{t+j} be the one period return expected by the market at period $t+j$, assumed covariance stationary. Suppose $p_t = E\{\{\sum_{j=1}^{\infty} [\prod_{i=1}^{j} (1+r_{t+i-1})^{-1}]d_{t+j}\} \mid I_t\}$. Let us linearize the quantity in braces around $\bar{r}$ and $\bar{d}$. $\bar{r}$ is the mean of r_t; selection of $\bar{d}$ is discussed below. Define $b=(1+\bar{r})^{-1}$, $a=-\bar{d}/\bar{r}$. Then (Shiller (1981a)) $p_t \approx E\{\{\sum_{j=1}^{\infty} b^j[a(r_{t+j-1}-\bar{r})+d_{t+j}]\} \mid I_t\}$. Let $u_{t+1}=p_t-b(p_{t+1}+d_{t+1})$. Proposition 1 may be used to show that in this linearized model

$$(21) \qquad \delta_1^2\sigma_v^2 - b^{-2}\sigma_u^2 \geqslant -[a^2+(1-b^2)^{-1}a^2]\sigma_r^2 - [2(1-b^2)^{-1/2}a\delta_1\sigma_v]\sigma_r,$$

where σ_r is the standard deviation of r_t, and δ_1 and σ_v are as defined in formula (16). The algebra to derive (21) is in the Appendix.

The left-hand side of (21) is precisely the quantity studied in Sections 3 and 4. If this is positive, as it will be in the model (12), $\sigma_r = 0$ would of course satisfy the inequality. The empirical estimates of (16), in Table II, column (7), however, were negative; the minimum return variability needed to explain the Table II results is given by the positive σ_r that makes (21) hold with equality.

This lower bound σ_r was calculated for all seven of the specifications. σ_u^2, σ_v^2, δ_1, and b were set equal to the estimated values reported in Table II. When dividends were assumed stationary, $\bar{d}$ was set equal to mean dividends, $\bar{d} = T^{-1}\sum d_t$. When dividends were assumed nonstationary, $\bar{d}$ was set equal to average

[13] Note that the entries in the table are not a monotonic function of α. To make sure that the entries were representative, I calculated the percentage excess variability for α in steps of 0.1 from 0 to 3.0, in steps of 1.0 from 3.0 to 10.0, and in steps of 5.0 from 10.0 to 50.0. The results were quite similar to those reported in the table. The lowest percentage happened to occur at $\alpha = 2.0$.

TABLE IV-B

MINIMUM σ_r NEEDED TO EXPLAIN EXCESS VARIABILITY

Data Set	S&P	S&P	S&P	S&P	DJ	DJ	DJ
Differenced	no	yes	no	yes	no	yes	no
Lags	2	2	4	4	3	4	4
σ_r	.146	.222	.146	.201	.127	.176	.169

expected discounted dividends, $\bar{d}=(1-b)\sum_{t=1}^{\infty} b^{t-1}E_0d_t$, where: $E_0d_t = E_0d_0 + tE\Delta d_t$, $E_0d_0 = d_0$, d_0 the level of dividends at the beginning of the sample, and $E\Delta d_t$ calculated as $T^{-1}\sum \Delta d_t$. The parameter $\underset{\sim}{a}$ was in all cases set to $-\bar{d}/\bar{r}$, with $\bar{r}$ defined implicitly by $(1+\bar{r})^{-1}=b$.

The resulting lower bound values may be found in Table IV-B. They are rather large. None of the estimates are less than .12. With $\sigma_r = .12$ and $\bar{r} = .07$, a two standard deviation confidence interval for the (real) expected return is about -17 per cent to $+31$ per cent. This would seem to be an implausibly large range.

In the linearized model considered here, then, Table IV-B suggests that variations in ex ante discount rates do not plausibly explain the excess variability of stock prices. How well this conclusion applied to any given nonlinear model of course depends on how well the linear model approximates the nonlinear one. An example in Shiller (1981a) suggests that if dividends are stationary the approximation can be quite good, even when changes in expected returns are larger than are typically considered reasonable. It is of course debatable that the approximation makes any sense, let alone is very accurate, if dividends are nonstationary. But the results here can in any case be said not to lend support to the hypothesis that the excess price variability reported in Table II is solely due to variation in expected returns.

6. CONCLUSIONS

This paper has derived and applied a stock price volatility test. The test required neither of two strong assumptions required by the Shiller (1981a) volatility test: that prices and dividends have finite variance, and that a satisfactory approximation to a perfect foresight price can be calculated from a finite data series.

The test indicated that stock prices are too volatile to be the expected present discounted value of dividends, with a constant discount rate. Among the explanations for the test results are that discount rates vary and that there are rational or nearly rational bubbles or fads. The possibility that the excess volatility is caused by discount rate fluctuations has been considered in detail, with largely negative results. The possibility that the excess volatility is due to bubbles has received little direct attention. But since this alternative is consistent with the econometric diagnostics, it seems worthy of further investigation.

A detailed case for bubbles, or, for that matter, any other factor as the explanation of the excess volatility is, however, beyond the scope of this paper.

A challenging task for future research is to make such a case, explaining the apparently excessive price volatility.

Woodrow Wilson School, Princeton University, Princeton, NJ 08544, U.S.A.

Manuscript received December, 1984; final revision received March, 1987.

APPENDIX

A. CALCULATION OF THE VARIANCE-COVARIANCE MATRIX

This describes the calculation of the variance-covariance matrix of the parameter vector $\theta = (b, \phi, \sigma_u^2, \sigma_v^2) = (b, \mu, \phi_1, \ldots, \phi_q, \sigma_u^2, \sigma_v^2)$.

Let $Z_t = (1, \Delta^s d_t, \ldots, \Delta^s d_{t-q+1})'$ be the $(q+1)\times 1$ vector of instruments, $s=0$ or $s=1$, $n_{t+1} = (d_{t+1}+p_{t+1})$ be the right-hand side variable in (14). One way of describing the estimation technique is to note that $\hat{\theta}$ was chosen to satisfy the orthogonality condition

$$0 = T^{-1}\sum h_t(\hat{\theta}) = \begin{bmatrix} T^{-1}(T^{-2s}\sum n_{t+1}Z_t')(\hat{S}_z)^{-1}\sum Z_t(p_t - n_{t+1}\hat{b}) \\ T^{-1}\sum Z_t(\Delta^s d_{t+1} - Z_t'\hat{\phi}) \\ \hat{\sigma}_u^2 - T^{-1}\sum (p_t - n_{t+1}\hat{b})^2 \\ \hat{\sigma}_v^2 - T^{-1}\sum(\Delta^s d_{t+1} - Z_t'\hat{\phi})^2 \end{bmatrix}.$$

(The degrees of freedom corrections in $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ are suppressed for notational simplicity.) The summations in the orthogonality condition run over t, from 1 to $T-s$. $\hat{S}_z$ is an estimate of $EZ_tZ_t'u_{t+1}^2$, calculated as described below equation (19). Thus $\hat{b}$ is estimated by two-step, 2SLS, $\hat{\phi}$ by OLS, $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$ from moments of the residuals.

Since $Eh_t(\theta) = 0$, where θ is the true but unknown parameter vector, it may be shown that under fairly general conditions, $C_T(\hat{\theta} - \theta)$ is asymptotically normal with a covariance matrix $V = (\text{plim } F_T^{-1}\sum h_{t\theta}F_T^{-1})^{-1}S(\text{plim } F_T^{-1}\sum h_{t\theta}'F_T^{-1})^{-1}$ (Hansen (1982), West (1986c)). C_T and F_T are $(q+4)\times(q+4)$ diagonal normalizing matrices, $C_T = F_T = \text{diag}\,(T^{1/2}, \ldots, T^{1/2})$ for undifferenced specifications, $C_T = \text{diag}\,(T^{3/2}, T^{1/2}, \ldots, T^{1/2})$ and $F_T = \text{diag}\,(T, T^{1/2}, \ldots, T^{1/2})$ for differenced specifications. $h_{t\theta}$ is the $(q+4)\times(q+4)$ matrix of derivatives of h_t with respect to θ and $S = Eh_th_t' + \sum_{j=1}^{\infty}[Eh_th_{t-j}' + (Eh_th_{t-j}')']$. $h_{t\theta}$ is straightforward to calculate. Calculation of S is slightly more involved. Newey and West (1986) and West (1986c) show that in general S and thus V are consistently estimated if $\hat{S} = \hat{\Omega}_0 + \sum_{i=1}^{m} w(i, m)(\hat{\Omega}_i + \hat{\Omega}_i')$, where $m\to\infty$ as $T\to\infty$ and m is $o(T^{1/2})$; $w(i, m) = i/(m+1)$; $\hat{\Omega}_i = T^{-1}\sum_{t=1+i}^{T}\tilde{h}_t\tilde{h}_{t-i}'$, $\tilde{h}_t = h_t(\tilde{\theta})$, $\tilde{\theta}$ an initial consistent estimate (2SLS and OLS). The weights $w(i, m)$ insure that $\hat{S}$ is positive semidefinite. In the absence of any theoretical or Monte Carlo evidence on the small sample properties of various choices of m, I tried various values: $m = 3$, 7, or 11. The value of m that led to the *largest* standard error in column (7) of Table II is what is reported in Table II. For all specifications, this turned out to be $m = 11$.

It is easy to show that the $T^{3/2}$ rate for $\hat{b}$ when $s=1$ implies that uncertainty about b can be ignored when calculating the standard error for $\hat{\rho}$ in column (5) of Table I-A. I therefore did so, and used the OLS standard error of the regression of the 2SLS residual on a lagged residual. The standard errors for the undifferenced specifications in Table I-A were calculated according to equation (50) in Pagan and Hall (1983). All standard errors in column (8) of Table I-B were calculated according to Theorem 3 in Pagan and Hall (1983).

B. DERIVATION OF EQUATION (21)

In the linearized model the analogue to equation (9) is $p_t = bE[a(r_t - \bar{r}) + d_{t+1} + p_{t+1} \mid I_t]$. Let $y_{t+j} = a(r_{t+j-1} - \bar{r}) + d_{t+j}$ and redefine $x_{tI} = E(\sum b^j y_{t+j} \mid I_t)$. (Of course, if expected returns are constant, $r_t = \bar{r}$ for all t, x_{tI} as defined here reduces to its Proposition 1 counterparts.) To simplify the argument, it will be assumed throughout this section that linear projections and mathematical expectations are equivalent. The efficient markets model considered in Section 3 implies $x_{tI} = d_t + p_t$;

the one currently under consideration implies $x_{tI} = y_t + p_t = a(r_{t-1} - \bar{r}) + d_t + p_t$. So with r_{t-1} an element of I_{t-1}, $x_{tI} - E(x_{tI} \mid I_{t-1}) = d_t + p_t - E(d_t + p_t \mid I_{t-1})$. Now,

$$
\begin{aligned}
\text{(A.1)} \qquad u_{t+1} &= p_t - b(d_{t+1} + p_{t+1}) = [ba(r_t - \bar{r}) + bE(p_{t+1} + d_{t+1} \mid I_t) - b(d_{t+1} + p_{t+1})] \\
&= b\{a(r_t - \bar{r}) - [x_{t+1,I} - E(x_{t+1,I} \mid I_t)]\} \\
&\Rightarrow b^{-2}\sigma_u^2 = a^2\sigma_r^2 + E[x_{t+1,I} - E(x_{t+1,I} \mid I_t)]^2 \\
&\Rightarrow E[x_{t+1,I} - E(x_{t+1,I} \mid I_t)]^2 = b^{-2}\sigma_u^2 - a^2\sigma_r^2.
\end{aligned}
$$

Now define J_t as the information set determined by a constant and all current and lagged dividends and expected returns, $x_{tJ} = E(\sum b^j y_{t+j} \mid J_t)$. Let $x_{tJ} - E(x_{tJ} \mid J_{t-1}) = aw_{1t} + w_{2t}$, where w_{1t} and w_{2t} are the innovations in the expected present discounted values of r_t and d_t. Shiller (1981a) shows that $\sigma_{w_1}^2 \leq \sigma_r^2/(1-b^2)$. Assume that d_t or Δd_t follows the autoregression (15). Then since H_t is a subset of J_t, Proposition 1 tells us that $\sigma_{w_2}^2 \leq \delta_1^2\sigma_v^2$, where, as previously, σ_v^2 is the variance of the univariate dividend innovation and δ_1 is defined above formula (16). So

$$
\begin{aligned}
\text{(A.2)} \qquad E[x_{tJ} - E(x_{tJ} \mid J_{t-1})]^2 &= a^2\sigma_{w_1}^2 + 2a\sigma_{w_1w_2} + \sigma_{w_2}^2 \\
&\leq a^2\sigma_{w_1}^2 + 2a\sigma_{w_1}\sigma_{w_2} + \sigma_{w_2}^2 \\
&\leq (1-b^2)^{-1}a^2\sigma_r^2 + 2a(1-b^2)^{-1/2}\delta_1\sigma_v\sigma_r + \delta_1^2\sigma_v^2.
\end{aligned}
$$

Since J_t is a subset of I_t, Proposition 1 tells us that $E[x_{tI} - E(x_{tI} \mid I_{t-1})]^2 \leq E[x_{tJ} - E(x_{tJ} \mid J_{t-1})]^2$. With a little rearrangement, (A.1) and (A.2) together imply equation (21) in the text.

REFERENCES

ACKLEY, G. (1983): "Commodities and Capital: Prices and Quantities," *American Economic Review*, 73, 1-16.

BLANCHARD, O., AND M. WATSON (1982): "Bubbles, Rational Expectations and Financial Markets," NBER Working Paper No. 945.

BREALEY, R., AND S. MYERS (1981): *Principles of Corporate Finance*. New York, NY: McGraw Hill.

DIBA, B. T., AND H. I. GROSSMAN (1985): "The Impossibility of Rational Bubbles," NBER Working Paper No. 1615.

DEATON, A. (1985): "Life Cycle Models of Consumption: Is the Evidence Consistent with the Theory?" manuscript, Princeton University.

FLAVIN, M. (1983): "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence," *Journal of Political Economy*, 91, 929-956.

FLOOD, R., R. HODRICK, AND P. KAPLAN (1986): "An Evaluation of Recent Evidence on Stock Market Bubbles," manuscript, Northwestern University.

FULLER, W. A. (1976): *Introduction to Statistical Time Series*. New York: John Wiley and Sons.

GRANGER, C. W. J., AND P. NEWBOLD (1977): *Forecasting Economic Time Series*. New York, NY: Academic Press.

GROSSMAN, S., AND R. SHILLER (1981). "The Determinants of the Variability of Stock Prices," *American Economic Review*, 71, 222-227.

HAMILTON, J., AND C. WHITEMAN (1984): "The Observable Implications of Self Fulfilling Prophecies," manuscript, University of Virginia.

HANNAN, E. J., AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society Series B*, 41, 190-195.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

HANSEN, L. P., AND T. J. SARGENT (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7-46.

——— (1981): "Exact Linear Rational Expectations Models: Specification and Estimation," Federal Reserve Bank of Minneapolis Staff Report 71.

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinar Rational Expectations Models," *Econometrica*, 50, 1269-1286.

KEYNES, J. M. (1964): *The General Theory of Employment, Interest and Money*. New York, NY: Harcourt, Brace, and World.

KLEIDON, A. W. (1985): "Bias in Small Sample Tests of Stock Price Rationality," Stanford University Graduate School of Business Research Paper 819R.

——— (1986): "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 94, 953-1001.
LEROY, S. (1984): "Efficiency and Variability of Asset Prices," *American Economic Review*, 74, 183-187.
LEROY, S. AND R. PORTER (1981): "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, 64, 555-574.
LUCAS, R. E., JR. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429-1445.
MARSH, T. A., AND R. C. MERTON (1986): "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, 76, 483-498.
MCCALLUM, B. (1976): "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," *Econometrica*, 44, 43-52.
NEWEY, W. K., AND K. D. WEST (1987): "A Simple, Positive, Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
PAGAN, A. R., AND A. D. HALL (1983): "Diagnostic Tests as Residual Analysis," *Econometric Reviews*, 2, 159-218.
SHILLER, R. J. (1981a): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421-436.
——— (1981b): The Use of Volatility Measures in Assessing Market Efficiency," *Journal of Finance*, 35, 291-304.
——— (1984): "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity*, 457-498.
SUMMERS, L. H. (1986): "Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance*, 41, 591-600.
TIROLE, J. (1985): "Asset Bubbles and Overlapping Generations," *Econometrica*, 53, 1071-1100.
WEST, K. D. (1984): "Speculative Bubbles and Stock Price Volatility," Princeton University Financial Research Memorandum No. 54.
——— (1987): "A Specification Test for Speculative Bubbles," *Quarterly Journal of Economics*, 102, 553-580.
——— (1986a): "A Standard Monetary Model and the Variability of the Deutschemark-Dollar Exchange Rate," forthcoming, *Journal of International Economics*.
——— (1986b): "Asymptotic Normality, When Regressors Have a Unit Root," Princeton University Woodrow Wilson School Discussion Paper No. 110.
——— (1986c): "Dividend Innovations and Stock Price Volatility," NBER Working Paper No. 1833.
——— (1988): "The Insensitivity of Consumption to News About Income," forthcoming, *Journal of Monetary Economics*.