NBER WORKING PAPER SERIES

DIVIDEND INNOVATIONS AND STOCK PRICE VOLATILITY

Kenneth D. West

Working Paper No. 1833

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1986

A previous version of this paper was circulated under the title "Speculative Bubbles in Stock Price Volatility." The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1833
February 1986

Dividend Innovations and Stock Price Volatility

ABSTRACT

This paper establishes an inequality that may be used to test the null hypothesis that a stock price equals the expected present discounted value of its dividend stream, with a constant discount rate. The inequality states that if this hypothesis is true, the variance of the innovation in the stock price is bounded above by a certain function of the variance in the innovation in the dividend. The bound is valid even if prices and dividends are nonstationary.

The inequality is used to test the null hypothesis, for some long term annual U.S. stock price data. The null is decisively rejected, with the stock price innovation variance exceeding its theoretical upper bound by a factor of as much as twenty. The rejection is highly significant statistically. Regression diagnostics and some informal analysis suggest that the results are more consistent with there being speculative bubbles in the U.S. stock market than with a failure of the rational expectations or constant discount rate hypothesis.

Kenneth D. West
Woodrow Wilson School
Princeton University
Princeton, NJ 08544

The source of fluctuations in stock prices has long been argued. Some observers have suggested that a major part of the fluctuations result 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" [15, p136]. An examination of practically any modern finance text (e.g., Brealey and Myers [5]) 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 [21] and, especially, Shiller [33]. 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 [24,25] have objected to the tests' assumption that dividends are stationary around a time trend; Flavin [8] and Kleidon [20] 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 [4].[2] The inequality states that if discount rates are constant this variance is larger when expectations are conditional on just the set of current and past dividends than when expectations are conditional on a larger 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 S and P 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 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 [8] and Kleidon [20] to plague the Shiller [33] volatility test. The paper nonetheless performs a small Monte Carlo experiment to check whether under certain simple circumstances the small sample bias in this paper's test procedure could 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 suprisingly, 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 test maintains essentially only one additional assumption, which is a transversality condition that puts an upper bound on expected growth in stock prices. If this condition is false, the excess volatility might be due to speculative bubbles of the sort considered by, for example, Blanchard and Watson [4].

The evidence, then, is more consistent with a failure of the transversality condition than of the rational expectations or constant discount rate assumptions. The paper does not, however, attempt to make a detailed case for bubbles, 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 a rational expectations, constant discount rate model appears to characterize these data remarkably well. Reconciliation of these two points is a task left for future research.

Before turning to the details of the subject at hand, a final introductory remark seems worth making. 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 (a subject considered in Deaton [6]), or whether exchange rates are too variable to be consistent with a standard monetary model (West [40]). That the inequality is valid even in a nonstationary environment makes it particularly appealing in these and perhaps other contexts.

The plan of the paper is as follows. Section II establishes the basic inequality. Section III explains how the inequality may be used to test a rational expectations, constant discount rate stock price model. Section IV presents formal econometric results. Section V considers informally whether small sample bias or discount rate variation are likely to explain the section IV results. Section VI has conclusions. An appendix has econometric details.

II. The Basic Inequality

The following proposition is the basis of this paper.

Proposition 1: Let I_t be an information set consisting of the space spanned by the current and past values of a finite number of random variables. After suitable differencing, the random variables are assumed to be covariance stationary, and, without loss of generality, to have zero mean. 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, $0 \leq b < 1$. Define $x_t = \sum_0^\infty b^j d_{t+j}$. (The preceding and all other summations in this section run over j.) Suppose that $x_{tI} = Ex_t|I_t$ and $x_{tH} = Ex_t|H_t$ both exist. Then

(1) $$E(x_{tH} - Ex_{tH}|H_{t-1})^2 \geq E(x_{tI} - Ex_{tI}|I_{t-1})^2$$

Proof:[3] Write

(2) $$x_t = d_t + bx_{t+1}$$

Project (2) onto I_t to obtain

(3) $$\begin{aligned} x_{tI} &= d_t + bEx_{t+1}|I_t \\ &= d_t + bx_{t+1,I} - be_{t+1} \\ e_{t+1} &= x_{t+1,I} - Ex_{t+1}|I_t = x_{t+1,I} - Ex_{t+1,I}|I_t \end{aligned}$$

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

$$(4) \qquad x_{tI} = \sum_0^\infty b^j d_{t+j} - \sum_1^\infty b^j e_{t+j} = x_t - \sum_1^\infty b^j e_{t+j}$$

By a similar argument, involving projections onto H_t,

$$(5) \qquad \begin{aligned} x_{tH} &= x_t - \sum_1^\infty b^j f_{t+j} \\ f_{t+j} &= x_{t+j,H} - Ex_{t+j,H}|H_{t+j-1} \end{aligned}$$

Now, since $\text{var}(x_t - x_{tI})$ and $\text{var}(x_t - x_{tH})$ are finite (see below), we have

$$(6) \qquad \begin{aligned} \text{var}(x_t - x_{tH}) &= \text{var}(x_t - x_{tI} + x_{tI} - x_{tH}) = \text{var}(x_t - x_{tI}) + \text{var}(x_{tI} - x_{tH}) \\ &\geq \text{var}(x_t - x_{tI}) \end{aligned}$$

The second equality follows since $x_t - x_{tI} = x_t - Ex_t|I_t$ is uncorrelated with anything in I_t, including, in particular, $x_{tI} - x_{tH}$. The assumptions of the proposition insure that e_t and f_t have zero mean, constant variance and are serially uncorrelated. So $\text{var}(x_t - x_{tH}) = b^2(1-b^2)^{-1}Ef_t^2$, $\text{var}(x_t - x_{tI}) = b^2(1-b^2)^{-1}Ee_t^2$, and (6) implies $E(x_{tH} - Ex_{tH}|H_{t-1})^2 \geq E(x_{tI} - Ex_{tI}|I_{t-1})^2$. 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. $E(x_{tI} - Ex_{tI}|I_{t-1})^2$ and $E(x_{tH} - Ex_{tH}|H_{t-1})^2$ 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. Further details on some of the comments may be found in footnote 4. First, (1) holds whenever x_{tI} and x_{tH} are well defined, as they will be if, for example, the variables in I_t and H_t follow a finite parameter ARIMA process. Note that this includes in particular processes with unit AR roots (an example is given below). Second, (1) does not extend immediately if logarithms or logarithmic differences are required to induce stationarity in d_t. If, for example, $\log(d_t)=\log(d_{t-1})+\varepsilon_t$, $\varepsilon_t \sim N(0,\sigma^2)$, it may be shown that $E(x_{tH}-Ex_{tH}|H_{t-1})^2$ does not even exist for $H_t=\{d_{t-j}\}$. 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 $\sum_0^\infty b^j d_{t+j}$. 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.[4]

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

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

with $|\phi|\leq 1$, 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).

$$(8)\qquad Ed_{t+j}|H_t = \phi^j d_t$$

$$\Longrightarrow\ x_{tH} \equiv E\sum_0^\infty b^j d_{t+j}|H_t = (1-b\phi)^{-1}d_t$$

$$\Longrightarrow\ E(x_{tH}-Ex_{tH}|H_{t-1})^2 = E[(1-b\phi)^{-1}v_t]^2 = (1-b\phi)^{-2}(\sigma_1^2+\sigma_2^2)$$

$$Ed_t|I_t = d_t$$

$$Ed_{t+j}|I_t = \phi^j d_t + \phi^{j-1}z_t \qquad j>0$$

$$\Longrightarrow\ x_{tI} \equiv E\sum_0^\infty b^j d_{t+j}|I_t = (1-b\phi)^{-1}(d_t+bz_t)$$

$$\Longrightarrow\ E(x_{tI}-Ex_{tI}|I_{t-1})^2 = E[(1-b\phi)^{-1}(\epsilon_{1t}+bz_t)]^2$$

$$= (1-b\phi)^{-2}(\sigma_1^2+b^2\sigma_2^2)$$

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.

III. The Model

According to a standard efficient markets model, a stock price is determined by the arbitrage relationship (9) (Brealey and Myers [5, pp42-45]):

$$(9)\qquad p_t = bE(p_{t+1}+d_{t+1})|I_t$$

where p_t is the 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 information 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

(10) $$p_t = \sum_{1}^{n} b^j E d_{t+j}|I_t + b^n E p_{t+n}|I_t$$

If the transversality condition

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

holds, then

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

Proposition 1 is used to test the model (12) as follows. Note first that since $x_{tI} = E\sum_{0}^{\infty} b^j d_{t+j}|I_t$, (12) implies that $x_{tI}=p_t+d_t$. So $E(x_{tI}-Ex_{tI}|I_{t-1})^2 = E[p_t+d_t-E(p_t+d_t|I_{t-1})]^2$, and, therefore,

(13) $$E(x_{tH}-Ex_{tH}|H_{t-1})^2 \geq E[p_t+d_t-E(p_t+d_t|I_{t-1})]^2$$

The intuitive reason that the model (12) implies (13) is as follows. $E(x_{tH}-Ex_{tH}|H_{t-1})^2$ 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 forecast optimally, the market's forecasts tend to be more precise, i.e., (13) holds.[5]

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 [26] and Hansen and Singleton [18]. Rewrite (9) as

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

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 $\hat{b}^{-2}\hat{\sigma}_u^2$.

Estimation of $E(x_{tH}-Ex_{tH}|H_{t-1})^2$ is slightly more involved. It requires first of all specification of H_t. The simplest possible one is $H_t=\{1,d_{t-j}|j\geq 0\}$, and H_t defined this way is what is used in this paper's empirical work.[6] Choices of H_t that include lags of additional variables might produce sharper results, but would also entail more complex calculations. With $H_t=\{1,d_{t-j}\}$, $E(x_{tH}-Ex_{tH}|H_{t-1})^2$ can be calculated as a function of d_t's univariate ARIMA parameters. Suppose d_t~ARIMA(q,s,0)

$$(15) \qquad \Delta^s d_{t+1} = \mu + \phi_1\Delta^s d_t + \ldots + \phi_q\Delta^s d_{t-q+1} + v_{t+1}$$

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 $x_{tH}=E\Sigma b^j d_{t+j}|H_t=m+\sum_1^{q+s}\delta_i d_{t-i+1}$. The δ_i are complicated functions of b and the ϕ_i. Hansen and Sargent [17] 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 [17] formula for δ_1 to calculate $\delta_1^2\sigma_v^2 = E(x_{tH}-Ex_{tH}|H_{t-1})^2$. To test the null hypothesis that prices are determined according to (12), then, we calculate

$$(16) \qquad \delta_1^2\sigma_v^2 - b^{-2}\sigma_u^2$$

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

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

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 <u>sum</u> 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.[7]

What alternatives might explain a rejection of the null hypothesis that (16) is positive? Three have figured prominently in discussions of related work: expectational irrationality (e.g., Ackley [1]), variation in discount rates (e.g., Leroy [22]) and speculative bubbles (e.g., Blanchard and Watson [4]). Elaboration of the relevant implications for asset price variability of the first two seems unnecessary since these are well known from the work of Shiller [33,34,35]. The speculative bubble alternative is perhaps less familiar, so some discussion seems warranted.

Let us begin by noting that $p_t = E\Sigma b^j d_{t+j}|I_t$ is not the only solution to (1). If the transversality condition (11) fails, there is a family of solutions to (1) (Blanchard and Watson [4], Shiller [32], Taylor [36]). For any c_t that satisfies $Ec_t|I_{t-1} = b^{-1}c_{t-1}$, $p_t = E\Sigma b^j d_{t+j}|I_t + c_t$ is also a solution to (1). c_t is by definition a speculative bubble, an otherwise extraneous event that affects stock prices because everyone expects it to do so. An example of a stochastic process for c_t, similar to one described in Blanchard and Watson [4], is

$$(18) \qquad c_t = \begin{cases} (c_{t-1} - \bar{c})/(\pi_t b) & \text{with probability } \pi_t \\ \bar{c}/[(1-\pi_t)b] & \text{with probability } 1-\pi_t \end{cases}$$

$$0 < \pi_t < 1, \quad \bar{c} > 0$$

According to (18), strictly positive bubbles grow and pop. (See Blanchard and Watson [4] for an argument that negative bubbles are inconsistent with rationality.) In this example, the probability that a bubble grows is π_t, that it collapses is $1-\pi_t$. The bubble may reflect events like sunspots that have no connection with the expected present discounted value of dividends. π_t might then be a random variable uncorrelated with anything in I_t. A more interesting possibility is that the bubble is intimately connected with fundamentals, with π_t dependent on news about fundamentals. A simple example is $\pi_t = 1/2$ for all t, with the bubble popping if and only if the innovation in dividends is negative. If π_t is constant ($\pi_t = \pi$ for all t), each bubble has an expected duration of $(1-\pi)^{-1}$. (π is not an identifiable parameter.) Combination of several bubbles are possible; the growth and collapse of the bubbles may be either tightly or loosely related. See Blanchard and Watson [4] for further examples and discussion.

Suppose that $p_t = E\Sigma b^j d_{t+j}|I_t + c_t$ for some bubble c_t (possibly one not following the stochastic process (18)). Since $p_t+d_t=x_{tI}+c_t$, we have

$$(19) \qquad E[p_t+d_t-E(p_t+d_t|I_{t-1})]^2 = E(x_{tI}-Ex_{tI}|I_{t-1})^2 + 2E(x_{tI}-Ex_{tI}|I_{t-1})(c_t-Ec_t|I_{t-1}) + E(c_t-Ec_t|I_{t-1})^2$$

When a bubble is present, the right hand side of (19) may be larger than $E(x_{tI}-Ex_{tI}|I_{t-1})^2$. It will unambiguously be larger if the innovation in the bubble is positively correlated with the innovation in x_{tI}. This will be the case if, for example, the bubble is connected with fundamentals and reflects a tendency of the market to overreact to news about dividends. This is sometimes argued to be plausible (e.g., Blanchard and Watson [4]).[8]

In the presence of bubbles, then, $b^{-2}\sigma_u^2$ will plausibly be bigger than $E(x_{tI}-Ex_{tI}|I_{t-1})^2$, and, therefore, (16) will be positive. In light of some empirical evidence yet to be presented, it is of particular interest to consider how to distinguish between bubbles on the one hand and expectational irrationality and time varying discount rates on the other as possible explanations of any excess price volatility. Formal econometric tests will help here. Consider, for example, diagnostic tests on the residual to equation (14). As long as (9) is correct--which it will be if expectations are rational and the discount rate is constant, even if there are bubbles--u_{t+1}, the disturbance to (14), is an expectational error. So u_{t+1} should be serially uncorrelated and uncorrelated with anything in I_t, including, in particular, lagged dividends. But if expectations are not rational, u_{t+1} will not in general have these properties. Nor will it if discount rates vary through time (see footnote 12).

Other diagnostics may also help distinguish between bubbles and other alternatives as possible explanations; some of these were calculated and are described in the next section. For the present, the important point to note is that when bubbles are absent, the arbitrage equation (14) and the dividend equation (15) together imply that (16) is positive. The only apparent form of misspecification that leaves (14) and (15) legitimate, but is still consistent with (16) being negative, is speculative bubbles. So an essential part of the strategy used here to distinguish between bubbles and other alternatives as explanations of excessive price variability is to perform diagnostic tests on equations (14) and (15). If these appear to be well specified, a logical inference is that bubbles explain the excess volatility.[9]

Such an inference may of course be incorrect. There may be small sample biases in the diagnostic tests. In addition, one may have a strong theoretical presumption that speculative bubbles are not present, or that the basic model has been misspecified in that, say, discount rates vary through time: it is certainly true that a consensus view on how general are the equilibria that admit bubbles is far from established, and that intertemporal asset pricing theories suggest that discount rates vary in general.[10] It would then be reasonable to give little credence to formal econometric evidence based on asymptotic distributions.

It is beyond the scope of this paper to consider these points in great detail. Section V does, however, analyze informally some of them.

IV. Empirical Results.

A. Data and Estimation Technique.

The data used were those used by Shiller [33] in his study of stock price volatility, and were graciously 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 = price in January divided by producer price index (1979 = 100), d_{t+1} = 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 [33] for a discussion of the data.

The following aspects of estimation are discussed in turn: (i) selection of the dividend process's lag length q, (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 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 lag length q were used. One was arbitrarily selected as $q=4$. The other was the q selected by the information criterion of Hannan and Quinn [14]. This criterion chooses the value of q that minimizes a certain function of the estimated parameters, and asymptotically chooses the correct q if the process truly has a finite order autoregressive representation.[11]

Thus, for each data set up to four sets of parameter estimates were calculated: q=4, q=lag length selected by the information criterion, for differenced and undifferenced series. In one case (Dow Jones, differenced), the Hannan and Quinn [14] criterion chose q=4. So only three sets of parameters were calculated for Dow Jones.

(ii) Calculation of (16) required estimation of the bivariate system consisting equations (14) and (15). Equation (14) was estimated by Hansen's [15] and Hansen and Singleton's [18] two-step two-stage least squares. The first step was standard two stage least squares. The second 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 heteroscedasticity as described in (iii).

With $\Delta^s d_t \sim AR(q)$, the δ_1 parameter in equation

(16) is $[(1-b)^s \Phi(b)]^{-1}$, $\Phi(b) = 1 - \sum_1^q b^i \phi_i$, (Hansen and Sargent [17]).

Thus, equation (16) was calculated as $[(1-\hat{b})^s(1 - \sum_{i=1}^q \hat{b}^i \hat{\phi}_i)]^{-2}\hat{\sigma}_v^2 - \hat{b}^{-2}\hat{\sigma}_u^2$.

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

$$(20) \qquad \hat{\sigma}_u^2 = (T-s)^{-1}\sum_{t=1}^{T-s}\hat{u}_{t+1}^2$$

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

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 $\hat{\theta} = (\hat{b},\hat{\mu}_1,\hat{\phi}_1,\ldots,\hat{\phi}_q,\hat{\sigma}_u^2,\hat{\sigma}_v^2)$. $\hat{\theta}$ is asymptotically normal with an asymptotic covariance V (see the appendix and (iii) below). Let $f(\theta)$ be equation (16) above. The standard error on the estimate of equation (16) was calculated as $[(\partial f/\partial\theta)V(\partial f/\partial\theta)']^{1/2}$ (See Rao [29, pp. 385-86].) 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 [15] and White and Domowitz [42] so that the estimate would be consistent for an <u>arbitrary</u> 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 $j \geq 0$. The Newey and West [27] procedure was used to insure that V was positive definite. 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 to 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}. This was done as suggested in Pagan and Hall [28, pp. 191, 170]. 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 [10, 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 [18]). Let Z_t be the ((q+1) x 1) vector of instruments and $\hat{b}$ the estimate of b. The orthogonality test is computed as:

$$(21) \quad \left(\sum_{t=1}^{T-s} Z_t'[p_t-\hat{b}(p_{t+1}+d_{t+1})]\right)(T\hat{S}_z)^{-1}\left(\sum_{t=1}^{T-s} Z_t[p_t-\hat{b}(p_{t+1}+d_{t+1})]\right)$$

$\hat{S}_z$ is an estimate of $E(Z_t u_{t+1})(Z_t u_{t+1})'$ and was calculated as $T^{-1}(\Sigma Z_t Z_t' \tilde{u}_{t+1}^2)$, $\tilde{u}_t$ the 2SLS residual to (14). The statistic (21) is asymptotically distributed as a chi-squared random variable with q degrees of freedom. As explained in a footnote, this test in general has the power to detect failures of the model (12) such as expectational irrationality and variation in discount rates.[12]

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 IA and IB. The results in Table IA strongly suggest that the basic arbitrage equation (1) is a sensible one. The entries in column (4) allow comfortable acceptance of 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. In addition, the equation (19) test for instrument-residual orthogonality also allows easy acceptance of the null hypothesis of no correlation between the instruments and the residuals. A possible exception is the Standard and Poor's data set, undifferenced, lag length = 2. See column (5). The generally successful results in column (5) are perhaps especially noteworthy since failures of rational expectations models to pass this test are quite common (e.g., Hansen and Singleton [18], West [38]).

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 percent for the Standard and Poor's index (1872-1981) and 7.4 per cent for the Dow Jones index (1929-1979). The estimates of the discount rate therefore are reasonable. The plausibility of the estimates of the discount rate provide special reassurance that the specification of the arbitrage equation (1) is an attractive one, since rational expectations models often fail to estimate ex ante real rates either sensibly or precisely (e.g., Blanchard [3], Rotemberg [30], Sargent [31]). 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 Standard and Poor's, 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 Standard and Poor's 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 quite attractive, with the possible exception of the Standard and Poor's data set with dividends undifferenced. Let us now turn to the estimates for the dividend process, reported in Table IB. 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 exception was the estimate of the first order serial correlation coefficient $\hat{\rho}$ for the Standard and Poor's index, undifferenced, lag length q=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 [14] 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 Standard and Poor's 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, equation (16) was 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 percent (column (8) of Table II).

The residual price fluctuation might reflect 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 is not particularly helpful in discriminating among these possibilities. It is worth noting, however, that for the other specifications, the results of the diagnostic tests were more consistent with the residual volatility being due to speculative bubbles than to a misspecification of the arbitrage or dividend equations.[13]

V.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 (1) below considers whether under certain simple circumstances small sample bias is likely to account for the finding of excess variability. Part (2) studies whether under equally simple circumstances low small sample power of the equation (21) test of instrument residual volatility is likely to explain the generally favorable results of the diagnostic tests.

(1)It is important to consider whether small sample bias explains the finding of excess variability, in light of the evidence in Kleidon [20] and Marsh and Merton [25] suggesting that if prices and dividends are nonstationary, the Shiller [33] 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 [20] and Marsh and Merton [25] was assumed. A Monte Carlo experiment was performed, assuming: (a)dividends follow a random walk, $\Delta d_t=\mu+v_t$, and (b)$H_t=I_t$, so that equations (16) and (17) are zero.

In this 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 IA. For each of 1000 samples, the following was done: A vector of 100 independent normal shocks was drawn, $(v_1,\ldots,v_{100})$. Dividends and prices were calculated as $\Delta d_t=.0373+v_t$; $d_t=d_0+\Sigma\Delta d_s$ $(d_0=1.3)$; $p_t=\Sigma(.9413)^j Ed_{t+j}|I_t = m+\delta_1 d_t$, $m=(.0373)*(.9413)/(1-.9413)$, $\delta_1=.9413/(1-.9413)$. $\hat{\mu}$ 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, equation (17), the percentage of price variability that is excessive, was calculated from the estimated parameters.

Table IIIA presents the empirical distribution of equation (17). Ideally, the median value of this statistic would be zero, with half the samples yielding a positive value to (17). Instead it is 15.0, and about two thirds of the samples produced a positive value. So there is a bias towards a finding of excess variability. The bias is not, however, particularly marked, and fewer than 10 percent of the simulated regressions produced the extreme values of the sort found in all of the Table II specifications.

That the Table IIIA distribution is only slightly biased suggests more strongly than might be immediately apparent that small sample bias does not explain the Table II results. For Table IIIA contains worst case figures, since it is 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 equation (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 excess variability reported in Table II.

(2) It is possible that the diagnostic tests reported basically favorable results because the tests have low power. 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 [23] asset pricing model.

So this section has a relatively modest aim, of using a single diagnostic test and a single, simple form of misspecfication, 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 (21) test of instrument residual orthogonality. The misspecification that is assumed is that expectations are static rather than rational, $Ed_{t+j}|I_t = d_t$. 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. That the results of a simulation for this alternative might produce representative results is perhaps suggested by the fact that for just about any alternative, the residual is some function, possibly nonlinear, of expected and/or actual prices, dividends, and possibly, other variables. (See footnote 12.) So to have power against plausible alternatives, the test will basically have to be able to pick up a correlation between prices and dividends on the one hand and lagged Δd_t on the other, and, again, this is exactly what it must do to have power against the static expectations alternative.

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 IB. 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. 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 (21) statistic was calculated.

The distribution of this statistic, which is a $\chi^2(2)$ random variable under the null, is reported in Table IIIB. In about fourth fifths of the cases, 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 cases, the statistic was over 2.87, the value reported in line (2), column (5), in Table IA.

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 IV 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). The diagnostic tests performed in section IV do not seem to suggest such variation, and the section VA(2) analysis just completed does not seem to indicate that the results of these tests are easily dismissed. Further consideration of the plausibility of this variation as an explanation seems warranted nonetheless, given theoretical work such as Lucas [23] and empirical evidence such as in Shiller [35].

This will be done in two separate exercises. The first (part (1) below) assumes as in, e.g., Hansen and Singleton [18], 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 equation (17), the percentage excess variability. The second (part (2) below) does not model expected returns parametrically but instead uses Shiller's [33] 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.

(1) Consider the class of models (e.g., Hansen and Singleton [18]) 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

$$(22) \qquad \tilde{p}_t = \beta E(\tilde{p}_{t+1}+\tilde{d}_{t+1})|I_t$$
$$\tilde{p}_t=p_tC_t^{-\alpha}, \ \tilde{d}_t=d_tC_t^{-\alpha}$$

Equation (22) is of the same form as equation (9). So 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.[13] 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 equation (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

[11] 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 or 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; this is unsuprising given that annual C_t growth was nearly 2 per cent per year, d_t growth slightly above one per cent.

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 β=.95 and β=.98, with very similar results. In all cases the lag length of the $\tilde{d}_t$ autoregression was set to four. Table IVA contains the figures that resulted for some of the α, with β=.98. As may be seen, there is 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.[15]

Since $\tilde{d}_t$ does not appear stationary for α much bigger than unity, it is equally true that Table IVA 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 IVA 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.

(2)Let us now consider a general model that does not parameterize expected returns, linearized as in Shiller [33] 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 $p_t = E\{\sum_{j=1}^{\infty}[\prod_{i=1}^{j}(1+r_{t+i-1})^{-1}]d_{t+j}\}|I_t$. 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 [33]), $p_t \approx E\{\sum_{j=1}^{\infty} b^j [a(r_{t+j-1}-\bar{r})+d_{t+j}]\}|I_t$. 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

$$(23) \qquad \delta_1^2\sigma_v^2 - b^{-2}\sigma_u^2 \geq -[a^2+(1-b^2)^{-1}a^2]\sigma_r^2 - [2(1-b^2)^{-1/2}a\delta_1\sigma_v]\sigma_r$$

where σ_r is the standard deviation of r_t, and δ_1 and σ_v are as defined in equation (16). The algebra to derive (23) is in a footnote.[16]

The left hand side of (23) is precisely the quantity studied in sections III and IV. 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 (23) 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, $\bar{d}=T^{-1}\Sigma d_t$. When dividends were assumed nonstationary, $\bar{d}$ was set equal to average expected discounted dividends, $\bar{d}=(1-b)\sum_{t=1}^{\infty} b^{t-1}E_0 d_t$, where: $E_0 d_t=E_0 d_0+tE\Delta d_t$, $E_0 d_0=d_0$, d_0 the level of dividends at the beginning of the sample, and $E\Delta d_t$ calculated as $T^{-1}\Sigma\Delta d_t$. The parameter 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 IVB. 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 percent to +31 percent. This would seem to be an implausibly large range.

In the linearized model considered here, then, variation in ex ante discount rates do not plausibly explain the excess variability of stock prices. How well this conclusion applies to any given nonlinear model of course depends on how well the linear model approximates the nonlinear one. An example in Shiller [33] 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.

VI. Conclusions

This paper has derived and applied a stock price volatility test. The test required neither of two strong assumptions required by the Shiller [33] 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. Possible explanations for the test results include that expectations are not rational, that discount rates vary and that there are speculative bubbles. The econometric diagnostics and the informal analysis were notably more consistent with the bubble explanation than with the other two.

A detailed case for bubbles, or, for that matter, any other factor as the explanation of the excess volatililty is, however, beyond the scope of this paper. A challenging task for future research is to make such a case, reconciling the apparently excessive price volatility with the apparently good performance of a rational expectations, constant discount rate specification.

FOOTNOTES

1. I thank A. Blinder, J. Campbell, G. Chow, S. Fischer, R. Flood, L.P. Hansen, W. Newey, J. Rotemberg, R. Trevor, and J. Taylor for helpful comments and discussions, and the National Science Foundation for partial financial support. Responsibility for remaining errors is my own. This paper was revised while I was a National Fellow at the Hoover Institution.

2. While Blanchard and Watson [4] 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.

3. I thank J. Campbell for this proof. I also thank L.P. Hansen and M. Watson for providing alternative proofs. S. Leroy has suggested to me that a similar proposition is implied in Leroy and Porter [21,p568]. My own, rather tedious, proof may be found in an earlier version of this paper [41].

4. Elaboration on the first three comments: (1)I believe that x_{tI} and x_{tH} are always well defined, given the assumptions of Proposition 1. The statement in the text specifies finite parameter ARIMA models because to my knowledge the theory for prediction of linear processes, which is well developed for stationary variables, has been extended to nonstationary variables only for such models. See Hansen and Sargent [17] for the ARIMA (q,s,0) case. It follows from Hansen and Sargent [17] that x_{tI} and x_{tH} exist for the ARIMA

(q,s,r) case. This is because after r periods, the expectations follow the same difference equation in each case. So if the discounted sum for an arbitrary ARIMA (q,s,0) process converges, so does that for an arbitrary ARIMA (q,s,r) process. (Strictly speaking, this statement does not hold for a process with a unit MA root and an infinite past, which may be the case for a stationary ARMA(q,r) process, since such a process does not have a convergent autoregressive representation. See Granger and Newbold [10,pp142-145].)

(2)In this case, $x_{tH}=kd_t=kd_{t-1}\exp(\varepsilon_t)$ for a certain constant k (Kleidon [20]). So $Ex_{tH}|H_{t-1} = kd_{t-1}E\exp(\varepsilon_t)$, and $x_{tH}-Ex_{tH}|H_{t-1}$ is proportional to d_{t-1}. An interesting project for future research is to develop an analogue of equation (1) that holds when logs or log differences are required to induce stationarity.

(3)An example: Let n=1, so $x_t=d_t+bd_{t+1}$. Suppose d_t~MA(1), $d_t=v_t+\theta v_{t-1}$, $-1<\theta<1$; $H_t=\{d_{t-j}|j\geq 0\}=\{v_{t-j}|j\geq 0\}$; $I_t=\{d_{t-j},v_{t-j+1}|j\geq 0\}=\{v_{t-j+1}|j\geq 0\}$. Then $x_{tH}-Ex_{tH}|H_{t-1} = (1+b\theta)v_t$, $x_{tI}-Ex_{tI}|I_{t-1} = bv_{t+1}$. Inequality (1) will be violated if, for example, b=.9 and $\theta<-.2$.

The reason the proof of (1) cannot be adapted to the finite horizon case is that a term of the form $b^{n+1}x_{t+n+1,I}$ will appear in equation (4), with an analogous term in equation (5). The fact that x_{tI} exists means that $b^{n+1}x_{t+n+1,I}$ is expected to get arbitrarily small for arbitrarily large n, and so can be ignored in the infinite horizon case. But for any finite n, the term cannot be ignored, and the argument in the proof will not apply.

5. To emphasize that inequality (13) holds even when dividends and prices are nonstationary, it is perhaps worth considering the class of dividend and price processes studied by Marsh and Merton [25]. Marsh and Merton argue that both theory and empirical evidence on dividends suggest that dividends are a

distributed lag on prices: $d_t = \sum\theta_i p_{t-i} = \theta(L)p_t$.

They also show that if dividends in fact are such a distributed lag on prices, and if dividends and prices are nonstationary, the basic Shiller [33] volatility test is no longer valid.

The test in this paper is, however, still valid if dividends are a distributed lag on prices and dividends are nonstationary. For it may be shown that $d_t = \theta(L)p_t$ and $p_t = E\Sigma b^i d_{t+i}|I_t$ together imply that $H_t = I_t$ -- only lagged dividends are used to forecast future dividends. When $H_t = I_t$, inequality (13) holds trivially, as a strict equality. (See footnotes 5 and 6 in West [41].) Even when dividends and prices are determined as suggested by Marsh and Merton [25], then, a violation of inequality (13) is evidence against the model (12).

6. 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.

7. 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 [33], Leroy and Porter [21]).

8. Even if there are bubbles, the right hand side of (19) clearly is not gauranteed to be larger, and, in particular, will not be larger in the (implausible) case of a purely deterministic bubble, $c_t = b^{-t}c_0$. A related paper (West [39]) develops and applies a test that is capable of finding such a bubble. The results of that paper are consistent with the results of this paper.

Since the test in West [39] is a test of cross equation restrictions, similar to the tests developed in Sargent [31] and Hansen and Sargent [17], this seems the appropriate place to comment on Hansen and Sargent's [16] point that tests of cross equation restrictions test all the restrictions of a linear rational expectations model, while volatility tests do not. The latter part of this statement is illustrated for the present paper's test by the comments in the preceding paragraph.

There are at least two reasons why stock market volatility tests are valuable nonetheless. The first is that a volatility test may have more power against a particular alternative than a test of cross equation restrictions. In the present context, this is perhaps reflected by the stronger rejection of the null in the present paper than in West [39], for differenced specifications. The second is that if a model is rejected by both tests, characterization of prices as "excessively volatile" may to some economists be a more provocative stimulus to future research than is a characterization of prices as "failing to obey cross equation constraints." That a characterization as "excessively volatile" is provocative to some is perhaps evidenced by the the strong reaction, both favorable and unfavorable, to the Shiller [33] volatility test.

In any case, the West [39] test of cross equation restrictions and the present paper are complementary studies. Those who argue for regression tests instead of volatility tests (see the discussion Hansen and Sargent [16]) are likely to prefer West [39], while those who argue for the converse (see the discussion in Shiller [34]) are likely to prefer the present paper.

9. Standard diagnostic tests will not suffice to find a misspecification of the dividend equation if the sample size is not large enough to infer the

parameters of the true dividend process. This might be the case because a small probability event, which is rationally considered by market participants, has not occurred. The best protection against such a biased sample is obviously to use a large sample, which is what I did.

In addition, it is worth noting that one important example of such a low probability event is allowed in the present framework. Suppose the market is considering a disaster such as nationalization that will set dividends to zero. The probability of disaster is θ. Shiller [34] shows that equation (9) is still valid, with b interpreted as the product of a discount rate and $1-\theta$. It follows that if (12) is true, (13) should hold.

10. For a nice general equilibrium model that allows bubbles, see Tirole [37]; it is perhaps worth noting that in Tirole's deterministic, perfect foresight steady state, asset returns are constant, just as are expected returns in the stochastic environment considered here. For an argument that volatility tests cannot be used to infer the presence of bubbles, see Hamilton and Whiteman [12].

11. The Hannan-Quinn procedure selects the r that minimizes

$$\ln \sigma_v^2 + T^{-1} 2rk \ln \ln T, \quad \sigma_v^2 = T^{-1} \sum v_t^2$$

for $r<R$ for some fixed R, with $k>1$. I set $R=4$, $k=1.001$.

12. Suppose discount rates are time varying. Let b_t be the one period rate from period t to period t+1, $\bar{b}$ the probability limit of the instrumental variables estimate of the discount rate in equation (14), and $n_{t+1} = p_{t+1}+d_{t+1}$. The proper specification of equation (1) is thus $p_t = b_t E n_{t+1} | I_t$. Equation (14) is then

$$p_t = \bar{b}n_{t+1} + (b_t-\bar{b})n_{t+1} + b_t(n_{t+1}-En_{t+1}|I_t) = \bar{b}n_{t+1} + u_{t+1}.$$

In general the instruments (lagged dividends) will be correlated with the residual u_{t+1} since they will be correlated with $(b_t-\bar{b})n_{t+1} = (b_t-\bar{b})(p_{t+1}+d_{t+1})$. The only apparent exceptions are implausible or uninteresting -- e.g., when $b_t-\bar{b}$, the deviation of ex-ante rates from a fixed level, is uncorrelated with both $p_{t+1}+d_{t+1}$ and with lagged dividends (the instruments).

Observe also that in general the residual to the equation above will be serially correlated when discount rates are not constant (i.e., when b_t not equal $\bar{b}_t$ for all t). Thus testing for serial correlation in the residual to (14) checks not only whether expectations are rational but also whether discount rates are constant.

13. This seems an appropriate place to give the results of another test of this model. Equation (6) states that $\text{var}(x_t-x_{tH}) - \text{var}(x_t-x_{tI}) - \text{var}(x_{tI}-x_{tH}) = 0$. So, under the null hypothesis that $x_{tI}=p_t+d_t$,

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

The parameters needed to calculate x_{tH} under the null--m, $\delta_1,\ldots,\delta_{q+s}$--are complicated functions of b,μ, and the ϕ_i. The formula for m may be found in West [39], for the δ_i in Hansen and Sargent [17].

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 $\text{var}[p_t+d_t - (m+\sum_1^{q+s}\delta_i d_{t-i+1})]$. This was insignificantly different from zero at the five per cent level, for all seven specifications. One possible reason for this noisy estimate is that there are bubbles: if so, $\mathrm{var}[p_t+d_t - (m+\sum_1^{q+s}\delta_i d_{t-i+1})]$ is not even finite.

14. I thank R. Flood for pointing this out to me.

15. 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 α=2.0.

16.In the linearized model the analogue to equation equation (9) is $p_t=bE[a(r_t-\bar{r})+d_{t+1}+p_{t+1}]|I_t$. Let $y_{t+j}=a(r_{t+j-1}-\bar{r})+d_{t+j}$ and redefine $x_t=\Sigma b^j y_{t+j}$, $x_{tI}=Ex_t|I_t$. (Of course, if expected returns are constant, $r_t=\bar{r}$ for all t, x_t and x_{tI} as defined here reduce to their Proposition 1 counterparts.) The efficient markets model considered in section III implied $x_{tI}=d_t+p_t$; 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}, $x_{tI}-Ex_{tI}|I_{t-1} = d_t+p_t-E(d_t+p_t|I_{t-1})$. Now,

$$(*)\ u_{t+1} = p_t-b(d_{t+1}+p_{t+1}) = [ba(r_t-\bar{r})+bE(p_{t+1}+d_{t+1}|I_t) - b(d_{t+1}+p_{t+1})]$$

$$=b[a(r_t-\bar{r})-(x_{t+1,I}-Ex_{t+1,I}|I_t)]$$

$$\Longrightarrow\ b^{-2}\sigma_u^2 = a^2\sigma_r^2 + E(x_{t+1,I}-Ex_{t+1,I}|I_t)^2$$

$$\Longrightarrow\ E(x_{t+1,I}-Ex_{t+1,I}|I_t)^2 = b^{-2}\sigma_u^2 - a^2\sigma_r^2$$

Now define J_t as the space spanned by a constant and all current and lagged dividends and expected returns, $x_{tJ}=Ex_t|J_t$. Let $x_{tJ}-Ex_{tJ}|J_{t-1} = aw_{1t}+w_{2t}$, where w_{1t} and w_{2t} are the innovations in the expected present discounted values of r_t and d_t. Shiller [35] shows that $\sigma^2_{w_1} \le \sigma^2_r/(1-b^2)$. 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 $\sigma^2_{w_2} \le \delta_1^2\sigma_v^2$, where, as previously, σ_v^2 is the variance of the univariate dividend innovation and δ_1 is defined above equation (16). So

$$
\begin{aligned}
(**) \qquad E(x_{tJ}-Ex_{tJ}|J_{t-1})^2 &= a^2\sigma^2_{w_1} + 2a\sigma_{w_1w_2} + \sigma^2_{w_2} \\
&\le a^2\sigma^2_{w_1} + 2a\sigma_{w_1}\sigma_{w_2} + \sigma^2_{w_2} \\
&\le (1-b^2)^{-1}a^2\sigma^2_r + 2a(1-b^2)^{-1/2}\delta_1\sigma_v\sigma_r + \delta_1^2\sigma_v^2
\end{aligned}
$$

Since J_t is a subset of I_t, Proposition 1 tells us that $E(x_{tI}-Ex_{tI}|I_{t-1})^2 \le E(x_{tJ}-Ex_{tJ}|J_{t-1})^2$. So with a little rearrangement, (*) and (**) together imply equation (23) in the text.

APPENDIX

This describes the calculation of the variance-covariance matrix of the parameter vector $\theta=(b,\phi,\sigma_u^2,\sigma_v^2) = (b,\mu,\phi_1,\ldots,\phi_q,\sigma_u^2,\sigma_v^2)$. It also establishes suitable conditions for the calculation to be appropriate when dividends are assumed nonstationary.

Let $Z_t=(1,\Delta^s d_t,\ldots,\Delta^s d_{t-q+1})'$ be the (q+1) x 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

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

(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. $\hat{S}_z$ is an estimate of $EZ_tZ_t'u_{t+1}^2$, calculated as described below equation (21). 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, $\sqrt{T}(\hat{\theta}-\theta)$ is asymptotically normal with a covariance matrix $V = (\text{plim}T^{-1}\Sigma h_{t\theta})^{-1}S(\text{plim}T^{-1}\Sigma h_{t\theta}')^{-1}$ (Hansen [15], White and Domowitz [42]). $h_{t\theta}$ is the (q+4)x(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 [27] 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: $m \rightarrow \infty$ as $T \rightarrow \infty$ and m is $o(T^{1/4})$; $w(i,m) = i/(m+1)$; $\hat{\Omega}_i = T^{-1}\sum_{t=i+1}^{T} \tilde{h}_t \tilde{h}'_{t-i}$, $\tilde{h}_t = h_t(\tilde{\theta})$, $\tilde{\theta}$ an initial consistent estimate (2SLS and OLS). The weights $w(i,m)$ insure that $\hat{S}$ is positive definite. 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 <u>largest</u> 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.

The conditions in Hansen [15], White and Domowitz [42] and Newey and West [27] unfortunately do not cover the case when n_{t+1} is nonstationary. The formulas just given are, however, still basically applicable, at least under the conditions listed in the assumptions given below. The only difference between the stationary and the nonstationary cases is that a certain term in $h_{t\theta}$ that depends on $\text{plim}T^{-1}\Sigma n_{t+1}u_{t+1}$ is set to zero in the nonstationary case.

The remainder of this appendix sketches the argument necessary to establish the asymptotic distribution of $\hat{\theta}$ in the nonstationary case. A detailed argument is available on request.

Theorem 1 below establishes the asymptotic distribution of $\hat{b}$, Theorem 2 that of $\hat{\sigma}_u^2$, Theorem 3 that of the joint asymptotic distribution of the elements of $\hat{\theta}$. (In light of assumption (a4) below, standard theory applies for $\hat{\phi}$ and $\hat{\sigma}_v^2$.)

<u>Assumptions</u>.

<u>(a1)</u>Let W_t be the $(2q+4)\times 1$ vector $(Z_t u_{t+1}, Z_t v_{t+1}, u_{t+1}^2 - \sigma_u^2, v_{t+1}^2 - \sigma_v^2).'$ Then W_t is (i)ergodic and covariance and fourth order stationary, with (ii)iid innovations, and (iii)a moving average representation whose weights are absolutely summable.

(a2)The innovations in the (y_t, z_t) process are zero for all $t \leq t_0$, for some

$t_0 \leq -q$; n_{t_0+1} is a nonstochastic constant, that, for simplicity is assumed to be zero.

(a3) $E\Delta n_t \neq 0$.

(a4) The $(\Delta n_t, \Delta d_t)$ process is covariance stationary, with a moving average representation whose weights are absolutely summable.

Remark. The heart of the argument is in the lemma, which proves that $T^{-2}\Sigma Z_t n_{t+1}$ converges in probability to a matrix of constants of rank one. Asymptotic normality then follows easily. The convergence in probability is established by showing that $\lim (E_{t_0} T^{-2}\Sigma Z_t n_{t+1})$ is a vector of constants of rank one and $\lim [\mathrm{var}_{t_0}(T^{-2}\Sigma Z_t n_{t+1})]$ is zero. (E_{t_0} and var_{t_0} denote expectations and variances calculated conditional on the history of the (y_t, z_t) process at date t_0; by assumption (a2), this means expectations and variances calculated assuming that all past innovations in the (y_t, z_t) process are zero.) In reading the lemma, it will be helpful to note that (a) by (a2), $E_{t_0} y_t = Ey_t$ and $E_{t_0} z_t = Ez_t$, for all $t \geq t_0$; (b) $\mathrm{var}_{t_0}(.) \leq \mathrm{var}(.)$, where (.) is any function of y_t's and z_t's, $t \geq t_0$, with finite unconditional variance. The unconditional operators $E(.)$ and $\mathrm{var}(.)$ are understood to act as if the y_t and z_t processes have infinite pasts, i.e., these operators do not condition on assumption (a2).

Lemma. $T^{-2}\Sigma Z_t n_{t+1}$ converges in probability to a $(q+1)\times 1$ constant vector of rank 1.

Proof: The first element of $T^{-2}\Sigma Z_t n_{t+1}$ is $T^{-2}\Sigma n_{t+1}$. I will show that this converges in probability to $(1/2)E\Delta n_t$, which is nonzero by (a3). A similar but considerably messier argument can be used to establish that each of the other elements of $T^{-2}\Sigma Z_t n_{t+1}$ converge in probability to a constant.

We have $\Sigma n_{t+1} = Tn_1 + \Sigma(n_{t+1} - n_1)$. It is easliy shown that (a2) implies that $\lim [E_{t_0} T^{-2}(Tn_1)] = \lim [\mathrm{var}_{t_0} T^{-2}(Tn_1)] = 0$. Now, $n_{t+1} = \Delta n_{t+1} + \ldots + \Delta n_2 +$

n_1. It follows that

$$\text{(A.1)}\qquad \Sigma(n_{t+1}-n_1) = \Delta n_2 + (\Delta n_2+\Delta n_3) + \ldots + (\Delta n_2+\ldots+\Delta n_{T+1})$$
$$= T\Delta n_2 + (T-1)\Delta n_3 + \ldots + \Delta n_{T+1}$$
$$\Longrightarrow \quad E\Sigma(n_{t+1}-n_1) = [T+(T-1)+\ldots+1]E\Delta n_t$$
$$= [(T^2+T)/2]E\Delta n_t$$
$$\Longrightarrow \quad \lim T^{-2}E\Sigma(n_{t+1}-n_1) = (1/2)E\Delta n_t$$

Let $\gamma(j)$ denote $\mathrm{cov}(\Delta n_t,\Delta n_{t-j})$. As stated in the Remark, to establish that $\lim \mathrm{var}_{t_0}[T^{-2}\Sigma(n_{t+1}-n_1])$ = 0, it suffices to establish that $\lim \mathrm{var}[T^{-2}\Sigma(n_{t+1}-n_1]) = 0$. To show this, note that (A.1) implies

$$\begin{aligned}\mathrm{var}[\Sigma(n_{t+1}-n_1)] &= [T^2+(T-1)^2+\ldots 1^2]\gamma(0)\\ &\quad +2[T(T-1)+(T-1)(T-2)+\ldots+2.1]\gamma(1)\\ &\quad + \ldots + 2[T.1]\gamma(T-1)\\ &\le (\Sigma t^2)[\gamma(0) + 2\textstyle\sum_{j=1}^{T-1}|\gamma(j)|]\\ &\le (\Sigma t^2)[\gamma(0) + 2\textstyle\sum_{j=1}^{\infty}|\gamma(j)|]\end{aligned}$$

Assumption (a4) implies that the right hand side of the above is finite for given T (Hannan [13,p211]). The fact that Σt^2 is of order T^3 now implies that $\lim \mathrm{var}[T^{-2}\Sigma(n_{t+1}-n_1)]$ is zero. So $T^{-2}\Sigma n_{t+1}$ converges in mean square and thus in probability to $(1/2)E\Delta n_t$.

<u>Theorem 1.</u> $T^{3/2}(\hat{b}-b)$ converges in distribution to a $N(0,V_b)$ random variable, $V_b = \mathrm{plim}\,[(T^{-2}\Sigma n_{t+1}Z_t')S_z^{-1}(T^{-2}\Sigma Z_t n_{t+1})]^{-1}$.

<u>Proof:</u> We have

$T^{3/2}(\hat{b}-b) =$

$$[(T^{-2}\Sigma n_{t+1}Z_t')\hat{S}_z^{-1}(T^{-2}\Sigma Z_t n_{t+1})]^{-1}(T^{-2}\Sigma n_{t+1}Z_t')\hat{S}_z^{-1}(T^{-1/2}\Sigma Z_t u_{t+1})$$

Assumption (a1) insures that (a)$T^{-1/2}\Sigma Z_t u_{t+1}$ converges in distribution to a $N(0, S_z)$ random variable, and (b)plim $\hat{S}_z = S_z$. The lemma insures that $T^{-2}\Sigma Z_t n_{t+1}$ converges in probability to a constant vector of rank 1. The theorem now follows.

Theorem 2. $\sqrt{T}(T^{-1}\Sigma\hat{u}_t^2 - \sigma_u^2)$ has the same asymptotic distribution as $\sqrt{T}(T^{-1}\Sigma u_t^2 - \sigma_u^2)$. (Note: summation signs here and in the proof of Theorem 2 run from 2 to T+1.)

Proof: We have

$$\hat{u}_t^2 = (p_{t-1} - n_t\hat{b})^2 = u_t^2 - 2(\hat{b}-b)n_t u_t + (\hat{b}-b)^2 n_t^2$$

$$\begin{aligned} \Longrightarrow \quad \sqrt{T}(T^{-1}\Sigma\hat{u}_t^2 - \sigma_u^2) &= \sqrt{T}(T^{-1}\Sigma u_t^2 - \sigma_u^2) \\ &\quad - 2[T^{3/2}(\hat{b}-b)](T^{-2}\Sigma n_t u_t) \\ &\quad + [T^{3/2}(\hat{b}-b)]^2(T^{-7/2}\Sigma n_t^2) \end{aligned}$$

It may be shown that $T^{-2}\Sigma n_t u_t$ and $T^{-7/2}\Sigma n_t^2$ each converge in probability to zero. For $T^{-2}\Sigma n_t u_t$ this follows because $E\Sigma n_t u_t$ is of order T, var$(\Sigma n_t u_t)$ of order T^3. A similar argument applies to $T^{-7/2}\Sigma n_t^2$.

Since $T^{3/2}(\hat{b}-b)$ has a well defined asymptotic distribtution, this implies that $\sqrt{T}(T^{-1}\Sigma\hat{u}_t^2 - \sigma_u^2)$ has the same asymptotic distribution as $\sqrt{T}(T^{-1}\Sigma u_t^2 - \sigma_u^2)$. It also obviously will imply that $\hat{\sigma}_u^2$ is a consistent estimate of σ_u^2.

Theorem 3.

(a)The normalized parameter vector $[T^{3/2}(\hat{b}-b), \sqrt{T}(\hat{\phi}-\phi), \sqrt{T}(\hat{\sigma}_u^2-\sigma_u^2), \sqrt{T}(\hat{\sigma}_v^2-\sigma_v^2)]$ converges in distribution to a N(0,V) random variable, where: $V = HS_wH'$; $S_w = EW_tW_t' + \sum_{j=1}^{\infty}[EW_tW_{t-j}' + (EW_tW_{t-j}')']$, W_t defined in (a1); H is a block diagonal matrix with plim $[(T^{-2}\Sigma n_{t+1}Z_t')\hat{S}_z^{-1}(T^{-2}\Sigma Z_t n_{t+1})]^{-1}(T^{-2}\Sigma n_{t+1}Z_t')\hat{S}_z^{-1}$ in the upper left hand block, plim $(T^{-1}\Sigma Z_t Z_t')$ in the middle block, and a (2x2) identity matrix in the lower right hand block.

(b)If $\hat{S}_w$, the Newey and West estimate of S_w, is calculated using 2SLS and OLS residuals, $\hat{S}_w$ converges in probability to S_w.

Proof: It may be shown than assumption (a1) is strong enough to insure that cross products of instruments and disturbances, and of instruments and residuals calculated using estimated parameters, are well behaved. So parts (a) and (b) of the theorem both follow from Theorems 1 and 2, given assumption (a1).

Note that the covariance matrix in the nonstationary case, HS_wH', is the same as the covariance matrix in the stationary case, $(\text{plim}T^{-1}\Sigma h_{t\theta})^{-1}S(\text{plim}T^{-1}\Sigma h_{t\theta}')^{-1}$, except that a term depending in part on plim $T^{-1}\Sigma n_{t+1}u_{t+1}$ appears in the latter but not in the former.

REFERENCES

[1] Ackley, G.: "Commodities and Capital: Prices and Quantities," American Economic Review, 73 (1983), pp. 1-16.

[2] Anderson, T.W.: "On The Asymptotic Distributions of Parameters of Stochastic Difference Equations," Annals of Mathematical Statistics, (1959), 30, pp. 676-687.

[3] Blanchard, O.: "The Production and Inventory Behavior of the American Automobile Industry," Journal of Political Economy, 91 (1983), pp. 364-400.

[4] Blanchard, O. and M. Watson: "Bubbles, Rational Expectations and Financial Markets," NBER Working Paper No. 945, (1982).

[5] Brealey, R. and S. Myers: Principles of Corporate Finance, New York: McGraw Hill, 1981.

[6] Deaton, A.: "Life Cycle Models of Consumption: Is the Evidence Consistent with the Theory?", manuscript, Princeton University, 1985.

[7] Diba, B. and H. Grossman: "Rational Asset Price Bubbles," National Bureau of Economic Research Working Paper No. 1059, 1983.

[8] Flavin, M.: "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence," Journal of Political Economy, 91 (1983), pp. 929-956.

[9] Fuller, W.A.: Introduction to Statistical Time Series, New York: John Wiley and Sons, 1976.

[10] Granger, C.W.J. and P. Newbold: Forecasting Economic Time Series, New York: Academic Press, 1977.

[11] Grossman, S. and R. Shiller: "The Determinants of The Variability of Stock Prices," American Economic Review, 71 (1981), pp. 222-227.

[12] Hamilton, J. and C. Whiteman: "The Observable Implications of Self Fulfilling Prophecies," manuscript, University of Virginia, 1984.

[13] Hannan, E.J.: Multiple Time Series, New York: John Wiley and Sons, 1970.

[14] Hannan, E.J. and B.G. Quinn: "The Determination of the Order of an Autoregression," Journal of the Royal Statistical Society Series B, 41 (1979), pp. 190-195.

[15] Hansen, L. P.: "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, 50 (1982), pp. 1029-1054.

[16] Hansen, L. P. and T. J. Sargent: "Exact Linear Rational Expectations Models: Specification and Estimation," Federal Reserve Bank of Minneapolis Staff Report No. 71 (1981).

[17] __________ "Formulating and Estimating Dynamics Linear Rational Expectations Models," Journal of Economic Dynamics and Control, 2 (1980).

[18] Hansen, L. P. and K. Singleton: "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," Econometrica, 50 (1982), pp. 1269-1286.

[19] Keynes, J. M.: The General Theory of Employment, Interest and Money, New York: Harcourt, Brace and World, 1964.

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

[21] Leroy, S. and R. Porter: "The Present Value Relation: Tests Based on Implied Variance Bounds," Econometrica, 49 (1981), pp. 555-574.

[22] Leroy, S.: "Efficiency and the Variability of Asset Prices," American Economic Review, 74 (1984), pp. 183-187.

[23] Lucas, R.E. Jr.: "Asset Prices in an Exchange Economy," Econometrica 46 (1978), pp. 1429-1445.

[24] Marsh, T. A. and R. C. Merton: "Aggregate Dividend Behavior and its Implications for Tests of Stock Market Rationality," Sloan School of Management Working Paper, No. 1475-83, 1983.

[25] __________ "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," Sloan School of Management Working Paper No. 1584-84, 1984.

[26] McCallum, B.: "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," Econometrica, 44 (1976), pp. 43-52.

[27] Newey, W.K. and K.D. West: "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Princeton University Woodrow Wilson School Discussion Paper No. 92, 1985.

[28] Pagan, A.R. and A.D. Hall: "Diagnostic Tests as Residual Analysis," Econometric Review, 2 (1983), pp. 159-218.

[29] Rao, C.R.: Linear Statistical Inference and its Applications, New York: John Wiley and Sons, 1973.

[30] Rotemberg, J.: "Sticky Prices in the United States," Journal of Political Economy, 90 (1982), pp. 1187-1211.

[31] Sargent, T. J.: "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," pp. 463-500 in Lucas, Robert E. and Thomas J. Sargent (eds.) Rational Expectations and Econometric Practice, Minneapolis: University of Minnesota Press, 1981.

[32] Shiller, R. J.: "Rational Expectations and the Dynamic Structure Of Macro Models," Journal of Monetary Economics, 4 (1981) pp. 1-44.

[33] __________ "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," American Economic Review, 71 (1981), pp. 421-436.

[34] __________ "The Use of Volatility Measures in Assessing Market Efficiency," Journal of Finance, XXXVI (1981), pp. 291-301.

[35] __________ "Stock Prices and Social Dynamics," Brookings Papers on Economic Activity, (2) 1984, pp. 457-498.

[36] Taylor, J. B.: "On the Conditions for Uniqueness in the Solution of Rational Expectations Models," Econometrica, 45 (1977), pp. 1377-1385.

[37] Tirole, J.: "Asset Bubbles and Overlapping Generations," Econometrica 53 (1985), pp. 1071-1100.

[38] West, K. D.: "A Variance Bounds Test of the Linear Quadratic Inventory Model," forthcoming, Journal of Political Economy, 1986.

[39] __________ "A Specification Test for Speculative Bubbles," manuscript, Princeton University Woodrow Wilson School Discussion paper No. 97, 1985.

[40] __________ "A Standard Monetary Model and The Variability of the Deutschemark - Dollar Exchange Rate, manuscript, Princeton University, 1985.

[41] __________ "Speculative Bubbles and Stock Price Volatility," Princeton University Financial Research memorandum, No. 54, 1985.

[42] White, H. and I. Domowitz: "Nonlinear Regression with Dependent Observations," Econometrica, 52 (1984), pp. 143-162.

Regression Results

Table IA: Equation 14

Data Set	(1) differenced	(2) q	(3) b	(4) ρ	(5) H/sig	(6) stability/sig
S and P						
1873-1980	no	2^a	.9311 (.0186)	.2346 (.4633)	5.50/.064	4.55/.033
1874-1980	yes	2^a	.9413 (.0170)	-.7140 (.6358)	2.87/.238	.33/.566
1875-1980	no	4	.9315 (.0158)	-.6131 (.4115)	6.96/.138	3.69/.055
1876-1980	yes	4	.9449 (.0136)	-.6132 (.5022)	3.15/.533	.28/.594
Modified Dow Jones						
1931-1978	no	3^a	.9402 (.0301)	-.5135 (.5756)	5.42/.144	1.56/.211
1933-1978	yes	4^a	.9379 (.0188)	-.5131 (.4877)	5.20/.267	2.02/.154
1932-1978	no	4	.9271 (.0253)	-.2027 (.4124)	6.08/.108	.49/.483

See notes to Table IB

Regression Results

Table IB: Equation 15

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Data Set	differenced	q	μ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ρ	Q/sig	stability/sig
S and P 1873-1980	no	2^a	.168 (.084)	1.196 (.114)	-.238 (.103)			.806 (.408)	36.87/.181	12.93/.005
1874-1980	yes	2^a	.034 (.029)	.262 (.118)	-.214 (.071)			.042 (.463)	22.79/.824	2.71/.438
1875-1980	no	4	.150 (.080)	1.247 (.116)	-.480 (.093)	.227 (.113)	-.029 (.066)	1.570 (3.531)	21.39/.875	33.49/.000
1876-1980	yes	4	.036 (.031)	.264 (.115)	-.230 (.094)	.026 (.080)	-.006 (.153)	.149 (1.626)	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)		.011 (.454)	4.05/1.000	7.53/.111
1933-1978	yes	4^a	.275 (.405)	.302 (.119)	-.351 (.133)	.051 (.093)	.050 (.176)	-.198 (2.892)	9.77/.939	8.06/.153
1932-1978	no	4	1.925 (1.900)	1.263 (.111)	-.662 (.208)	.330 (.209)	.004 (.134)	-4.693 (37.187)	4.06/1.000	10.22/.069

Notes to Tables IA, and IB:

(a) Lag length q chosen by Hannan and Quinn (1979) procedure.

(b) Asymptotic standard errors in parentheses.

(c) Symbols: q = lag length of dividend autoregression (15); parameters b, μ, ϕ_i defined in equations (9) and (15); ρ = first order serial correlation coefficient of disturbance; H = statistic in equation (21), $H \sim \chi^2(q)$; "stability" is test for stability of coefficients, as described in text, distributed $\chi^2(1)$ in Table IA and $\chi^2(q+1)$ in Table IB; 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.

Table II
Test Statistics

Data Set	(1) differenced	(2) q	(3) $\hat{b}$	(4) $\hat{\delta}_1$	(5) $\hat{\sigma}_u^2$	(6) $\hat{\sigma}_v^2$	(7) Eqn. (16)	(8) Eqn. (17)
S and P	No	2^a	.9311 (.0186)	10.82 (3.47)	215.2 (79.0)	.1501 (.0543)	-230.66 (87.10)	92.92
	Yes	2^a	.9413 (.0170)	18.06 (6.22)	214.1 (80.2)	.1485 (.0523)	-193.22 (71.07)	79.95
	No	4	.9315 (.0158)	10.76 (3.10)	219.4 (73.4)	.1502 (.0510)	-235.51 (90.12)	93.12
	Yes	4	.9449 (.0136)	18.45 (5.64)	218.2 (81.1)	.1538 (.0511)	-192.05 (73.63)	78.58
Modified Dow-Jones	No	3^a	.9402 (.0301)	8.28 (2.85)	19653 (5836)	9.980 (3.383)	-21545 (5978)	96.92
	Yes	4^a	.9379 (.0188)	15.78 (8.13)	19342 (5871)	9.014 (2.655)	-19740 (5852)	89.79
	No	4	.9271 (.0253)	7.55 (3.12)	19228 (3912)	10.453 (2.427)	-21777 (4309)	97.34

NOTES:
(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, $P_{1979} = 468.94$, $d_{1978} = 30.91$.

Table III A

Distribution of Equation (17) in Monte Carlo Experiment

Percentile	5	10	50	66
Equation (17)	91.8	74.7	15.0	0

Table III B

Distribution of Equation (21) in Monte Carlo Experiment

Percentile	5	10	50	78	90
Equation	25.26	22.29	10.80	5.99	3.07

Table IV A

Percentage Excess Price Variability

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

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