Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?

By ROBERT J. SHILLER*

A simple model that is commonly used to interpret movements in corporate common stock price indexes asserts that real stock prices equal the present value of rationally expected or optimally forecasted future real dividends discounted by a constant real discount rate. This valuation model (or variations on it in which the real discount rate is not constant but fairly stable) is often used by economists and market analysts alike as a plausible model to describe the behavior of aggregate market indexes and is viewed as providing a reasonable story to tell when people ask what accounts for a sudden movement in stock price indexes. Such movements are then attributed to "new information" about future dividends. I will refer to this model as the "efficient markets model" although it should be recognized that this name has also been applied to other models.

It has often been claimed in popular discussions that stock price indexes seem too "volatile," that is, that the movements in stock price indexes could not realistically be attributed to any objective new information, since movements in the price indexes seem to be "too big" relative to actual subsequent events. Recently, the notion that financial asset prices are too volatile to accord with efficient markets has received some econometric support in papers by Stephen LeRoy and Richard Porter on the stock market, and by myself on the bond market.

To illustrate graphically why it seems that stock prices are too volatile, I have plotted in Figure 1 a stock price index p_t with its *ex post* rational counterpart p_t^* (data set 1).[1] The stock price index p_t is the real Standard and Poor's Composite Stock Price Index (detrended by dividing by a factor proportional to the long-run exponential growth path) and p_t^* is the present discounted value of the actual subsequent real dividends (also as a proportion of the same long-run growth factor).[2] The analogous series for a modified Dow Jones Industrial Average appear in Figure 2 (data set 2). One is struck by the smoothness and stability of the *ex post* rational price series p_t^* when compared with the actual price series. This behavior of p^* is due to the fact that the present value relation relates p^* to a long-weighted moving average of dividends (with weights corresponding to discount factors) and moving averages tend to smooth the series averaged. Moreover, while real dividends did vary over this sample period, they did not vary long enough or far enough to cause major movements in p^*. For example, while one normally thinks of the Great Depression as a time when business was bad, real dividends were substantially below their long-run exponential growth path (i.e., 10–25 percent below the

*Associate professor, University of Pennsylvania, and research associate, National Bureau of Economic Research. I am grateful to Christine Amsler for research assistance, and to her as well as Benjamin Friedman, Irwin Friend, Sanford Grossman, Stephen LeRoy, Stephen Ross, and Jeremy Siegel for helpful comments. This research was supported by the National Bureau of Economic Research as part of the Research Project on the Changing Roles of Debt and Equity in Financing U.S. Capital Formation sponsored by the American Council of Life Insurance and by the National Science Foundation under grant SOC-7907561. The views expressed here are solely my own and do not necessarily represent the views of the supporting agencies.

[1] The stock price index may look unfamiliar because it is deflated by a price index, expressed as a proportion of the long-run growth path and only January figures are shown. One might note, for example, that the stock market decline of 1929–32 looks smaller than the recent decline. In real terms, it was. The January figures also miss both the 1929 peak and 1932 trough.

[2] The price and dividend series as a proportion of the long-run growth path are defined below at the beginning of Section I. Assumptions about public knowledge or lack of knowledge of the long-run growth path are important, as shall be discussed below. The series p^* is computed subject to an assumption about dividends after 1978. See text and Figure 3 below.

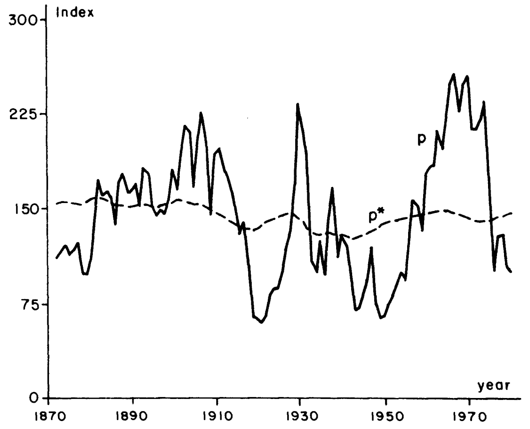

FIGURE 1

Note: Real Standard and Poor's Composite Stock Price Index (solid line p) and *ex post* rational price (dotted line p^*), 1871–1979, both detrended by dividing a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 1, Appendix.

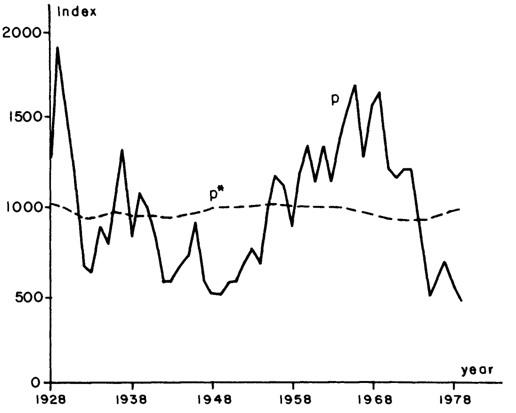

FIGURE 2

Note: Real modified Dow Jones Industrial Average (solid line p) and *ex post* rational price (dotted line p^*), 1928-1979, both detrended by dividing by a long-run exponential growth factor. The variable p^* is the present value of actual subsequent real detrended dividends, subject to an assumption about the present value in 1979 of dividends thereafter. Data are from Data Set 2, Appendix.

growth path for the Standard and Poor's series, 16–38 percent below the growth path for the Dow Series) only for a few depression years: 1933, 1934, 1935, and 1938. The moving average which determines p^* will smooth out such short-run fluctuations. Clearly the stock market decline beginning in 1929 and ending in 1932 could not be rationalized in terms of subsequent dividends! Nor could it be rationalized in terms of subsequent earnings, since earnings are relevant in this model only as indicators of later dividends. Of course, the efficient markets model does not say $p=p^*$. Might one still suppose that this kind of stock market crash was a rational mistake, a forecast error that rational people might make? This paper will explore here the notion that the very volatility of p (i.e., the tendency of big movements in p to occur again and again) implies that the answer is no.

To give an idea of the kind of volatility comparisons that will be made here, let us consider at this point the simplest inequality which puts limits on one measure of volatility: the standard deviation of p. The efficient markets model can be described as asserting that $p_t = E_t(p_t^*)$, i.e., p_t is the mathematical expectation conditional on all information available at time t of p_t^*. In other words, p_t is the optimal forecast of p_t^*. One can define the forecast error as $u_t = p_t^* - p_t$. A fundamental principle of optimal forecasts is that the forecast error u_t must be uncorrelated with the forecast; that is, the covariance between p_t and u_t must be zero. If a forecast error showed a consistent correlation with the forecast itself, then that would in itself imply that the forecast could be improved. Mathematically, it can be shown from the theory of conditional expectations that u_t must be uncorrelated with p_t.

If one uses the principle from elementary statistics that the variance of the sum of two uncorrelated variables is the sum of their variances, one then has $var(p^*) = var(u) + var(p)$. Since variances cannot be negative, this means $var(p) \leqslant var(p^*)$ or, converting to more easily interpreted standard deviations,

$$\sigma(p) \leqslant \sigma(p^*) \tag{1}$$

This inequality (employed before in the

papers by LeRoy and Porter and myself) is violated dramatically by the data in Figures 1 and 2 as is immediately obvious in looking at the figures.[3]

This paper will develop the efficient markets model in Section I to clarify some theoretical questions that may arise in connection with the inequality (1) and some similar inequalities will be derived that put limits on the standard deviation of the innovation in price and the standard deviation of the change in price. The model is restated in innovation form which allows better understanding of the limits on stock price volatility imposed by the model. In particular, this will enable us to see (Section II) that the standard deviation of Δp is highest when information about dividends is revealed smoothly and that if information is revealed in big lumps occasionally the price series may have higher kurtosis (fatter tails) but will have *lower* variance. The notion expressed by some that earnings rather than dividend data should be used is discussed in Section III, and a way of assessing the importance of time variation in real discount rates is shown in Section IV. The inequalities are compared with the data in Section V.

This paper takes as its starting point the approach I used earlier (1979) which showed evidence suggesting that long-term bond yields are too volatile to accord with simple expectations models of the term structure of interest rates.[4] In that paper, it was shown how restrictions implied by efficient markets on the cross-covariance function of short-term and long-term interest rates imply inequality restrictions on the spectra of the long-term interest rate series which characterize the smoothness that the long rate should display. In this paper, analogous implications are derived for the volatility of stock prices, although here a simpler and more intuitively appealing discussion of the model in terms of its innovation representation is used. This paper also has benefited from the earlier discussion by LeRoy and Porter which independently derived some restrictions on security price volatility implied by the efficient markets model and concluded that common stock prices are too volatile to accord with the model. They applied a methodology in some ways similar to that used here to study a stock price index and individual stocks in a sample period starting after World War II.

It is somewhat inaccurate to say that this paper attempts to contradict the extensive literature of efficient markets (as, for example, Paul Cootner's volume on the random character of stock prices, or Eugene Fama's survey).[5] Most of this literature really examines different properties of security prices. Very little of the efficient markets literature bears directly on the characteristic feature of the model considered here: that expected *real* returns for the aggregate stock market are constant through time (or approximately so). Much of the literature on efficient markets concerns the investigation of nominal "profit opportunities" (variously defined) and whether transactions costs prohibit their exploitation. Of course, if real stock prices are "too volatile" as it is defined here, then there may well be a sort of real profit opportunity. Time variation in expected real interest rates does not itself imply that any

[3]Some people will object to this derivation of (1) and say that one might as well have said that $E_t(p_t)=p_t^*$, i.e., that forecasts are correct "on average," which would lead to a reversal of the inequality (1). This objection stems, however, from a misinterpretation of conditional expectations. The subscript t on the expectations operator E means "taking as given (i.e., nonrandom) all variables known at time t." Clearly, p_t is known at time t and p_t^* is not. In practical terms, if a forecaster gives as his forecast anything other than $E_t(p_t^*)$, then high forecast is not optimal in the sense of expected squared forecast error. If he gives a forecast which equals $E_t(p_t^*)$ only on average, then he is adding random noise to the optimal forecast. The amount of noise apparent in Figures 1 or 2 is extraordinary. Imagine what we would think of our local weather forecaster if, say, actual local temperatures followed the dotted line and his forecasts followed the solid line!

[4]This analysis was extended to yields on preferred stocks by Christine Amsler.

[5]It should not be inferred that the literature on efficient markets uniformly supports the notion of efficiency put forth there, for example, that no assets are dominated or that no trading rule dominates a buy and hold strategy, (for recent papers see S. Basu; Franco Modigliani and Richard Cohn; William Brainard, John Shoven and Lawrence Weiss; and the papers in the symposium on market efficiency edited by Michael Jensen).

trading rule dominates a buy and hold strategy, but really large variations in expected returns might seem to suggest that such a trading rule exists. This paper does not investigate this, or whether transactions costs prohibit its exploitation. This paper is concerned, however, instead with a more interesting (from an economic standpoint) question: what accounts for movements in real stock prices and can they be explained by new information about subsequent real dividends? If the model fails due to excessive volatility, then we will have seen a new characterization of how the simple model fails. The characterization is not equivalent to other characterizations of its failure, such as that one-period holding returns are forecastable, or that stocks have not been good inflation hedges recently.

The volatility comparisons that will be made here have the advantage that they are insensitive to misalignment of price and dividend series, as may happen with earlier data when collection procedures were not ideal. The tests are also not affected by the practice, in the construction of stock price and dividend indexes, of dropping certain stocks from the sample occasionally and replacing them with other stocks, so long as the volatility of the series is not misstated. These comparisons are thus well suited to existing long-term data in stock price averages. The robustness that the volatility comparisons have, coupled with their simplicity, may account for their popularity in casual discourse.

I. The Simple Efficient Markets Model

According to the simple efficient markets model, the real price P_t of a share at the beginning of the time period t is given by

$$P_t = \sum_{k=0}^{\infty} \gamma^{k+1} E_t D_{t+k} \qquad 0<\gamma<1 \tag{2}$$

where D_t is the real dividend paid at (let us say, the end of) time t, E_t denotes mathematical expectation conditional on information available at time t, and γ is the constant real discount factor. I define the constant real interest rate r so that $\gamma = 1/(1+r)$. Information at time t includes P_t and D_t and their lagged values, and will generally include other variables as well.

The one-period holding return $H_t \equiv (\Delta P_{t+1} + D_t)/P_t$ is the return from buying the stock at time t and selling it at time $t+1$. The first term in the numerator is the capital gain, the second term is the dividend received at the end of time t. They are divided by P_t to provide a rate of return. The model (2) has the property that $E_t(H_t) = r$.

The model (2) can be restated in terms of series as a proportion of the long-run growth factor: $p_t = P_t/\lambda^{t-T}$, $d_t = D_t/\lambda^{t+1-T}$ where the growth factor is $\lambda^{t-T} = (1+g)^{t-T}$, g is the rate of growth, and T is the base year. Dividing (2) by λ^{t-T} and substituting one finds[6]

$$\begin{aligned} p_t &= \sum_{k=0}^{\infty} (\lambda\gamma)^{k+1} E_t d_{t+k} \\ &= \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} E_t d_{t+k} \end{aligned} \tag{3}$$

The growth rate g must be less than the discount rate r if (2) is to give a finite price, and hence $\bar{\gamma} \equiv \lambda\gamma < 1$, and defining $\bar{r}$ by $\bar{\gamma} \equiv 1/(1+\bar{r})$, the discount rate appropriate for the p_t and d_t series is $\bar{r} > 0$. This discount rate $\bar{r}$ is, it turns out, just the mean dividend divided by the mean price, i.e, $\bar{r} = E(d)/E(p)$.[7]

[6]No assumptions are introduced in going from (2) to (3), since (3) is just an algebraic transformation of (2). I shall, however, introduce the assumption that d_t is jointly stationary with information, which means that the (unconditional) covariance between d_t and z_{t-k}, where z_t is any information variable (which might be d_t itself or p_t), depends only on k, not t. It follows that we can write expressions like $var(p)$ without a time subscript. In contrast, a realization of the random variable the *conditional* expectation $E_t(d_{t+k})$ is a function of time since it depends on information at time t. Some stationarity assumption is necessary if we are to proceed with any statistical analysis.

[7]Taking unconditional expectations of both sides of (3) we find

$$E(p) = \frac{\bar{\gamma}}{1-\bar{\gamma}} E(d)$$

using $\bar{\gamma} = 1/1+\bar{r}$ and solving we find $\bar{r} = E(d)/E(p)$.

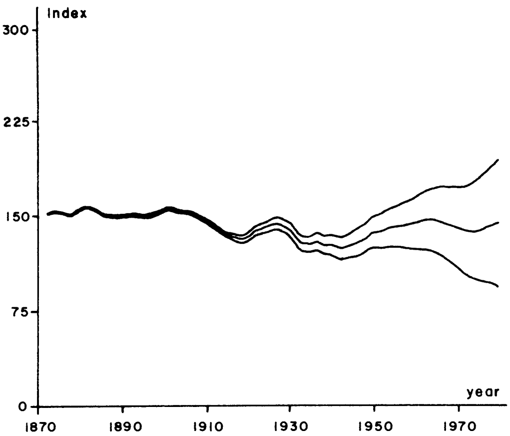

FIGURE 3

Note: Alternative measures of the *ex post* rational price p^*, obtained by alternative assumptions about the present value in 1979 of dividends thereafter. The middle curve is the p^* series plotted in Figure 1. The series are computed recursively from terminal conditions using dividend series d of Data Set 1.

TABLE 1—DEFINITIONS OF PRINCIPAL SYMBOLS

γ = real discount factor for series before detrending; $\gamma = 1/(1+r)$

$\bar{\gamma}$ = real discount factor for detrended series; $\bar{\gamma} \equiv \lambda\gamma$

D_t = real dividend accruing to stock index (before detrending)

d_t = real detrended dividend; $d_t \equiv D_t/\lambda^{t+1-T}$

Δ = first difference operator $\Delta x_t \equiv x_t - x_{t-1}$

δ_t = innovation operator; $\delta_t x_{t+k} \equiv E_t x_{t+k} - E_{t-1} x_{t+k}$; $\delta x \equiv \delta_t x_t$

E = unconditional mathematical expectations operator. $E(x)$ is the true (population) mean of x.

E_t = mathematical expectations operator conditional on information at time t; $E_t x_t \equiv E(x_t | I_t)$ where I_t is the vector of information variables known at time t.

λ = trend factor for price and dividend series; $\lambda \equiv 1+g$ where g is the long-run growth rate of price and dividends.

P_t = real stock price index (before detrending)

p_t = real detrended stock price index; $p_t = P_t/\lambda^{t-T}$

p_t^* = *ex post* rational stock price index (expression 4)

r = one-period real discount rate for series before detrending

$\bar{r}$ = real discount rate for detrended series; $\bar{r} = (1-\bar{\gamma})/\bar{\gamma}$

$\bar{r}_2$ = two-period real discount rate for detrended series; $\bar{r}_2 = (1+\bar{r})^2 - 1$

t = time (year)

T = base year for detrending and for wholesale price index; $p_T = P_T$ = nominal stock price index at time T

We may also write the model as noted above in terms of the *ex post* rational price series p_t^* (analogous to the *ex post* rational interest rate series that Jeremy Siegel and I used to study the Fisher effect, or that I used to study the expectations theory of the term structure). That is, p_t^* is the present value of actual subsequent dividends:

$$(4) \qquad p_t = E_t(p_t^*)$$

$$\text{where} \qquad p_t^* = \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} d_{t+k}$$

Since the summation extends to infinity, we never observe p_t^* without some error. However, with a long enough dividend series we may observe an approximate p_t^*. If we choose an arbitrary value for the terminal value of p_t^* (in Figures 1 and 2, p^* for 1979 was set at the average detrended real price over the sample) then we may determine p_t^* recursively by $p_t^* = \bar{\gamma}(p_{t+1}^* + d_t)$ working backward from the terminal date. As we move back from the terminal date, the importance of the terminal value chosen declines. In data set (1) as shown in Figure 1, $\bar{\gamma}$ is .954 and $\bar{\gamma}^{108} = .0063$ so that at the beginning of the sample the terminal value chosen has a negligible weight in the determination of p_t^*. If we had chosen a different terminal condition, the result would be to add or subtract an exponential trend from the p^* shown in Figure 1. This is shown graphically in Figure 3, in which p^* is shown computed from alternative terminal values. Since the only thing we need know to compute p^* about dividends after 1978 is p^* for 1979, it does not matter whether dividends are "smooth" or not after 1978. Thus, Figure 3 represents our uncertainty about p^*.

There is yet another way to write the model, which will be useful in the analysis which follows. For this purpose, it is convenient to adopt notation for the innovation in a variable. Let us define the innovation operator $\delta_t \equiv E_t - E_{t-1}$ where E_t is the conditional expectations operator. Then for any variable X_t the term $\delta_t X_{t+k}$ equals $E_t X_{t+k} - E_{t-1} X_{t+k}$ which is the change in the conditional expectation of X_{t+k} that is made in response to new information arriving between $t-1$ and t. The time subscript t may be dropped so that δX_k denotes $\delta_t X_{t+k}$ and

δX denotes δX_0 or $\delta_t X_t$. Since conditional expectations operators satisfy $E_j E_k = E_{min(j,k)}$ it follows that $E_{t-m}\delta_t X_{t+k} = E_{t-m}(E_t X_{t+k} - E_{t-1} X_{t+k}) = E_{t-m} X_{t+k} - E_{t-m} X_{t+k} = 0$, $m \geqslant 0$. This means that $\delta_t X_{t+k}$ must be uncorrelated for all k with all information known at time $t-1$ and must, since lagged innovations are information at time t, be uncorrelated with $\delta_{t'} X_{t+j}$, $t' < t$, all j, i.e., innovations in variables are serially uncorrelated.

The model implies that the innovation in price $\delta_t p_t$ is observable. Since (3) can be written $p_t = \bar{\gamma}(d_t + E_t p_{t+1})$, we know, solving, that $E_t p_{t+1} = p_t/\bar{\gamma} - d_t$. Hence $\delta_t p_t \equiv E_t p_t - E_{t-1} p_t = p_t + d_{t-1} - p_{t-1}/\bar{\gamma} = \Delta p_t + d_{t-1} - \bar{r} p_{t-1}$. The variable which we call $\delta_t p_t$ (or just δp) is the variable which Clive Granger and Paul Samuelson emphasized should, in contrast to $\Delta p_t \equiv p_t - p_{t-1}$, by efficient markets, be unforecastable. In practice, with our data, $\delta_t p_t$ so measured will approximately equal Δp_t.

The model also implies that the innovation in price is related to the innovations in dividends by

$$\delta_t p_t = \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \delta_t d_{t+k} \tag{5}$$

This expression is identical to (3) except that δ_t replaces E_t. Unfortunately, while $\delta_t p_t$ is observable in this model, the $\delta_t d_{t+k}$ terms are not directly observable, that is, we do not know when the public gets information about a particular dividend. Thus, in deriving inequalities below, one is obliged to assume the "worst possible" pattern of information accrual.

Expressions (2)–(5) constitute four different representations of the same efficient markets model. Expressions (4) and (5) are particularly useful for deriving our inequalities on measures of volatility. We have already used (4) to derive the limit (1) on the standard deviation of p given the standard deviation of p^*, and we will use (5) to derive a limit on the standard deviation of δp given the standard deviation of d.

One issue that relates to the derivation of (1) can now be clarified. The inequality (1) was derived using the assumption that the forecast error $u_t = p_t^* - p_t$ is uncorrelated with p_t. However, the forecast error u_t is not serially uncorrelated. It is uncorrelated with all information known at time t, but the lagged forecast error u_{t-1} is not known at time t since p_{t-1}^* is not discovered at time t. In fact, $u_t = \sum_{k=1}^{\infty} \bar{\gamma}^k \delta_{t+k} p_{t+k}$, as can be seen by substituting the expressions for p_t and p_t^* from (3) and (4) into $u_t = p_t^* - p_t$, and rearranging. Since the series $\delta_t p_t$ is serially uncorrelated, u_t has first-order autoregressive serial correlation.[8] For this reason, it is inappropriate to test the model by regressing $p_t^* - p_t$ on variables known at time t and using the ordinary t-statistics of the coefficients of these variables. However, a generalized least squares transformation of the variables would yield an appropriate regression test. We might thus regress the transformed variable $u_t - \bar{\gamma} u_{t+1}$ on variables known at time t. Since $u_t - \bar{\gamma} u_{t+1} = \bar{\gamma}\delta_{t+1} p_{t+1}$, this amounts to testing whether the innovation in price can be forecasted. I will perform and discuss such regression tests in Section V below.

To find a limit on the standard deviation of δp for a given standard deviation of d_t, first note that d_t equals its unconditional expectation plus the sum of its innovations:

$$d_t = E(d) + \sum_{k=0}^{\infty} \delta_{t-k} d_t \tag{6}$$

If we regard $E(d)$ as $E_{-\infty}(d_t)$, then this expression is just a tautology. It tells us, though, that d_t $t = 0, 1, 2, \ldots$ are just different linear combinations of the same innovations in dividends that enter into the linear combination in (5) which determine $\delta_t p_t$ $t = 0, 1, 2, \ldots$. We can thus ask how large $var(\delta p)$ might be for given $var(d)$. Since innovations are serially uncorrelated, we know from (6) that the variance of the sum is

[8]It follows that $var(u) = var(\delta p)/(1-\bar{\gamma}^2)$ as LeRoy and Porter noted. They base their volatility tests on our inequality (1) (which they call theorem 2) and an equality restriction $\sigma^2(p) + \sigma^2(\delta p)/(1-\bar{\gamma}^2) = \sigma^2(p^*)$ (their theorem 3). They found that, with postwar Standard and Poor earnings data, both relations were violated by sample statistics.

the sum of the variances:

$$(7) \quad var(d) = \sum_{k=0}^{\infty} var(\delta d_k) = \sum_{k=0}^{\infty} \sigma_k^2$$

Our assumption of stationarity for d_t implies that $var(\delta_{t-k} d_t) \equiv var(\delta d_k) \equiv \sigma_k^2$ is independent of t.

In expression (5) we have no information that the variance of the sum is the sum of the variances since all the innovations are time t innovations, which may be correlated. In fact, for given $\sigma_0^2, \sigma_1^2, \ldots$, the maximum variance of the sum in (5) occurs when the elements in the sum are perfectly positively correlated. This means then that so long as $var(\delta d) \neq 0$, $\delta_t d_{t+k} = a_k \delta_t d_t$, where $a_k = \sigma_k / \sigma_0$. Substituting this into (6) implies

$$(8) \qquad \hat{d}_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$$

where a hat denotes a variable minus its mean: $\hat{d}_t \equiv d_t - E(d)$ and $\varepsilon_t \equiv \delta_t d_t$. Thus, if $var(\delta p)$ is to be maximized for given $\sigma_0^2, \sigma_1^2, \ldots$, the dividend process must be a moving average process in terms of its own innovations.[9] I have thus shown, rather than assumed, that if the variance of δp is to be maximized, the forecast of d_{t+k} will have the usual ARIMA form as in the forecast popularized by Box and Jenkins.

We can now find the maximum possible variance for δp for given variance of d. Since the innovations in (5) are perfectly positively correlated, $var(\delta p) = (\Sigma_{k=0}^{\infty} \bar{\gamma}^{k+1} \sigma_k)^2$. To maximize this subject to the constraint $var(d) = \Sigma_{k=0}^{\infty} \sigma_k^2$ with respect to $\sigma_0, \sigma_1, \ldots$, one may set up the Lagrangean:

$$(9) \quad L = \left(\sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \sigma_k \right)^2 + \nu \left(var(d) - \sum_{k=0}^{\infty} \sigma_k^2 \right)$$

[9] Of course, all indeterministic stationary processes can be given linear moving average representations, as Hermann Wold showed. However, it does not follow that the process can be given a moving average representation in terms of its own innovations. The true process may be generated nonlinearly or other information besides its own lagged values may be used in forecasting. These will generally result in a less than perfect correlation of the terms in (5).

where ν is the Lagrangean multiplier. The first-order conditions for $\sigma_j, j = 0, \ldots \infty$ are

$$(10) \quad \frac{\partial L}{\partial \sigma_j} = 2 \left(\sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \sigma_k \right) \bar{\gamma}^{j+1} - 2\nu \sigma_j = 0$$

which in turn means that σ_j is proportional to $\bar{\gamma}^j$. The second-order conditions for a maximum are satisfied, and the maximum can be viewed as a tangency of an isoquant for $var(\delta p)$, which is a hyperplane in $\sigma_0, \sigma_1, \sigma_2, \ldots$ space, with the hypersphere represented by the constraint. At the maximum $\sigma_k^2 = (1 - \bar{\gamma}^2) var(d) \bar{\gamma}^{2k}$ and $var(\delta p) = \bar{\gamma}^2 var(d)/(1 - \bar{\gamma}^2)$ and so, converting to standard deviations for ease of interpretation, we have

$$(11) \qquad \sigma(\delta p) \leq \sigma(d) / \sqrt{\bar{r}_2}$$

$$\text{where} \qquad \bar{r}_2 = (1 + \bar{r})^2 - 1$$

Here, $\bar{r}_2$ is the two-period interest rate, which is roughly twice the one-period rate. The maximum occurs, then, when $\hat{d}_t$ is a first-order autoregressive process, $\hat{d}_t = \bar{\gamma} \hat{d}_{t-1} + \varepsilon_t$, and $E_t \hat{d}_{t+k} = \bar{\gamma}^k \hat{d}_t$, where $\hat{d} \equiv d - E(d)$ as before.

The variance of the innovation in price is thus maximized when information about dividends is revealed in a smooth fashion so that the standard deviation of the new information at time t about a future dividend d_{t+k} is proportional to its weight in the present value formula in the model (5). In contrast, suppose all dividends somehow became known years before they were paid. Then the innovations in dividends would be so heavily discounted in (5) that they would contribute little to the standard deviation of the innovation in price. Alternatively, suppose nothing were known about dividends until the year they are paid. Here, although the innovation would not be heavily discounted in (5), the impact of the innovation would be confined to only one term in (5), and the standard deviation in the innovation in price would be limited to the standard deviation in the single dividend.

Other inequalities analogous to (11) can also be derived in the same way. For example, we can put an upper bound to the standard deviation of the change in price (rather than the innovation in price) for given standard deviation in dividend. The only difference induced in the above procedure is that Δp_t is a different linear combination of innovations in dividends. Using the fact that $\Delta p_t = \delta_t p_t + \bar{r} p_{t-1} - d_{t-1}$ we find

$$(12) \qquad \Delta p_t = \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \delta_t d_{t+k}$$

$$+\bar{r} \sum_{j=1}^{\infty} \delta_{t-j} \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} d_{t+k-1} - \sum_{j=1}^{\infty} \delta_{t-j} d_{t-1}$$

As above, the maximization of the variance of δp for given variance of d requires that the time t innovations in d be perfectly correlated (innovations at different times are necessarily uncorrelated) so that again the dividend process must be forecasted as an ARIMA process. However, the parameters of the ARIMA process for d which maximize the variance of Δp will be different. One finds, after maximizing the Lagrangean expression (analogous to (9)) an inequality slightly different from (11),

$$(13) \qquad \sigma(\Delta p) \leqslant \sigma(d)/\sqrt{2\bar{r}}$$

The upper bound is attained if the optimal dividend forecast is first-order autoregressive, but with an autoregressive coefficient slightly different from that which induced the upper bound to (11). The upper bound to (13) is attained if $\hat{d}_t = (1-\bar{r})\hat{d}_{t-1} + \varepsilon_t$ and $E_t d_{t+k} = (1-\bar{r})^k \hat{d}_t$, where, as before, $\hat{d}_t \equiv d_t - E(d)$.

II. High Kurtosis and Infrequent Important Breaks in Information

It has been repeatedly noted that stock price change distributions show high kurtosis or "fat tails." This means that, if one looks at a time-series of observations on δp or Δp, one sees long stretches of time when their (absolute) values are all rather small and then an occasional extremely large (absolute) value. This phenomenon is commonly attributed to a tendency for new information to come in big lumps infrequently. There seems to be a common presumption that this information lumping might cause stock price changes to have high or infinite variance, which would seem to contradict the conclusion in the preceding section that the variance of price is limited and is maximized if forecasts have a simple autoregressive structure.

High sample kurtosis does not indicate infinite variance if we do not assume, as did Fama (1965) and others, that price changes are drawn from the stable Paretian class of distributions.[10] The model does not suggest that price changes have a distribution in this class. The model instead suggests that the existence of moments for the price series is implied by the existence of moments for the dividends series.

As long as d is jointly stationary with information and has a finite variance, then p, p^*, δp, and Δp will be stationary and have a finite variance.[11] If d is normally distributed, however, it does not follow that the price variables will be normally distributed. In fact, they may yet show high kurtosis.

To see this possibility, suppose the dividends are serially independent and identically normally distributed. The kurtosis of the price series is defined by $K = E(\hat{p})^4/(E(\hat{p})^2)^2$, where $p \equiv \hat{p} - E(p)$. Suppose, as an example, that with a probability of $1/n$

[10] The empirical fact about the unconditional distribution of stock price changes in not that they have infinite variance (which can never be demonstrated with any finite sample), but that they have high kurtosis in the sample.

[11] With any stationary process X_t, the existence of a finite $var(X_t)$ implies, by Schwartz's inequality, a finite value of $cov(X_t, X_{t+k})$ for any k, and hence the entire autocovariance function of X_t, and the spectrum, exists. Moreover, the variance of $E_t(X_t)$ must also be finite, since the variance of X equals the variance of $E_t(X_t)$ plus the variance of the forecast error. While we may regard real dividends as having finite variance, innovations in dividends may show high kurtosis. The residuals in a second-order autoregression for d_t have a studentized range of 6.29 for the Standard and Poor series and 5.37 for the Dow series. According to the David-Hartley-Pearson test, normality can be rejected at the 5 percent level (but not at the 1 percent level) with a one-tailed test for both data sets.

the public is told d_t at the beginning of time t, but with probability $(n-1)/n$ has no information about current or future dividends.[12] In time periods when they are told d_t, $\hat{p}_t$ equals $\bar{\gamma}\hat{d}_t$, otherwise $\hat{p}_t=0$. Then $E(\hat{p}_t^4)=E((\bar{\gamma}\hat{d}_t)^4)/n$ and $E(\hat{p}_t^2)=E((\bar{\gamma}\hat{d}_t)^2)/n$ so that kurtosis equals $nE(\bar{\gamma}\hat{d}_t)^4)/E((\bar{\gamma}\hat{d}_t)^2)$ which equals n times the kurtosis of the normal distribution. Hence, by choosing n high enough one can achieve an arbitrarily high kurtosis, and yet the variance of price will always exist. Moreover, the distribution of $\hat{p}_t$ conditional on the information that the dividend has been revealed is also normal, in spite of high kurtosis of the unconditional distribution.

If information is revealed in big lumps occasionally (so as to induce high kurtosis as suggested in the above example) $var(\delta p)$ or $var(\Delta p)$ are not especially large. The variance loses more from the long interval of time when information is not revealed than it gains from the infrequent events when it is. The highest possible variance for given variance of d indeed comes when information is revealed smoothly as noted in the previous section. In the above example, where information about dividends is revealed one time in n, $\sigma(\delta p)=\bar{\gamma}n^{1/2}\sigma(d)$ and $\sigma(\Delta p)=\bar{\gamma}(2/n)^{1/2}\sigma(d)$. The values of $\sigma(\delta p)$ and $\sigma(\Delta p)$ implied by this example are for all n strictly below the upper bounds of the inequalities (11) and (13).[13]

III. Dividends or Earnings?

It has been argued that the model (2) does not capture what is generally meant by efficient markets, and that the model should be replaced by a model which makes price the present value of expected earnings rather than dividends. In the model (2) earnings may be relevant to the pricing of shares but only insofar as earnings are indicators of future dividends. Earnings are thus no different from any other economic variable which may indicate future dividends. The model (2) is consistent with the usual notion in finance that individuals are concerned with returns, that is, capital gains plus dividends. The model implies that expected total returns are constant and that the capital gains component of returns is just a reflection of information about future dividends. Earnings, in contrast, are statistics conceived by accountants which are supposed to provide an indicator of how well a company is doing, and there is a great deal of latitude for the definition of earnings, as the recent literature on inflation accounting will attest.

There is no reason why price per share ought to be the present value of expected earnings per share if some earnings are retained. In fact, as Merton Miller and Franco Modigliani argued, such a present value formula would entail a fundamental sort of double counting. It is incorrect to include in the present value formula both earnings at time t and the later earnings that accrue when time t earnings are reinvested.[14] Miller and Modigliani showed a formula by which price might be regarded as the present value of earnings corrected for investments, but that formula can be shown, using an accounting identity to be identical to (2).

Some people seem to feel that one cannot claim price as present value of expected dividends since firms routinely pay out only a fraction of earnings and also attempt somewhat to stabilize dividends. They are right in the case where firms paid out *no* dividends, for then the price p_t would have to grow at the discount rate $\bar{r}$, and the model (2) would not be the solution to the difference equation implied by the condition $E_t(H_t)=r$. On the other hand, if firms pay out a fraction of dividends or smooth short-run fluctuations in dividends, then the price of the firm will grow at a rate less than the

[12] For simplicity, in this example, the assumption elsewhere in this article that d_t is always known at time t has been dropped. It follows that in this example $\delta_t p_t \neq \Delta p_t + d_{t-1} - rp_{t-1}$ but instead $\delta_t p_t = p_t$.

[13] For another illustrative example, consider $\hat{d}_t = \bar{\gamma}\hat{d}_{t-1}+\varepsilon_t$ as with the upper bound for the inequality (11) but where the dividends are announced for the next n years every $1/n$ years. Here, even though $\hat{d}_t$ has the autoregressive structure, ε_t is not the innovation in d_t. As n goes to infinity, $\sigma(\delta p)$ approaches zero.

[14] LeRoy and Porter do assume price as present value of earnings but employ a correction to the price and earnings series which is, under additional theoretical assumptions not employed by Miller and Modigliani, a correction for the double counting.

discount rate and (2) is the solution to the difference equation.[15] With our Standard and Poor data, the growth rate of real price is only about 1.5 percent, while the discount rate is about 4.8%+1.5%=6.3%. At these rates, the value of the firm a few decades hence is so heavily discounted relative to its size that it contributes very little to the value of the stock today; by far the most of the value comes from the intervening dividends. Hence (2) and the implied p^* ought to be useful characterizations of the value of the firm.

The crucial thing to recognize in this context is that once we know the terminal price and intervening dividends, we have specified all that investors care about. It would not make sense to define an *ex post* rational price from a terminal condition on price, using the same formula with earnings in place of dividends.

IV. Time-Varying Real Discount Rates

If we modify the model (2) to allow real discount rates to vary without restriction through time, then the model becomes untestable. We do not observe real discount rates directly. Regardless of the behavior of P_t and D_t, there will always be a discount rate series which makes (2) hold identically. We might ask, though, whether the movements in the real discount rate that would be required aren't larger than we might have expected. Or is it possible that small movements in the current one-period discount rate coupled with new information about such movements in future discount rates could account for high stock price volatility?[16]

[15]To understand this point, it helps to consider a traditional continuous time growth model, so instead of (2) we have $P_0=\int_0^\infty D_t e^{-rt}dt$. In such a model, a firm has a constant earnings stream I. If it pays out all earnings, then $D=I$ and $P_0=\int_0^\infty Ie^{-rt}dt=I/r$. If it pays out only s of its earnings, then the firm grows at rate $(1-s)r$, $D_t=sIe^{(1-s)rt}$ which is less than I at $t=0$, but higher than I later on. Then $P_0=\int_0^\infty sIe^{(1-s)rt}e^{-rt}dt=\int_0^\infty sIe^{-srt}dt=sI/(rs)$. If $s\neq 0$ (so that we're not dividing by zero) $P_0=I/r$.

[16]James Pesando has discussed the analogous question: how large must the variance in liquidity premia be in order to justify the volatility of long-term interest rates?

The natural extension of (2) to the case of time varying real discount rates is

$$(14)\qquad P_t=E_t\left(\sum_{k=0}^{\infty} D_{t+k}\prod_{j=0}^{k}\frac{1}{1+r_{t+j}}\right)$$

which has the property that $E_t((1+H_t)/(1+r_t))=1$. If we set $1+r_t=(\partial U/\partial C_t)/(\partial U/\partial C_{t+1})$, i.e., to the marginal rate of substitution between present and future consumption where U is the additively separable utility of consumption, then this property is the first-order condition for a maximum of expected utility subject to a stock market budget constraint, and equation (14) is consistent with such expected utility maximization at all times. Note that while r_t is a sort of *ex post* real interest rate not necessarily known until time $t+1$, only the conditional distribution at time t or earlier influences price in the formula (14).

As before, we can rewrite the model in terms of detrended series:

$$(15)\qquad p_t=E_t(p_t^*)$$

$$\text{where}\qquad p_t^*\equiv\sum_{k=0}^{\infty} d_{t+k}\prod_{j=0}^{k}\frac{1}{1+\bar{r}_{t+j}}$$

$$1+\bar{r}_{t+j}\equiv(1+r_t)/\lambda$$

This model then implies that $\sigma(p_t)\leq\sigma(p_t^*)$ as before. Since the model is nonlinear, however, it does not allow us to derive inequalities like (11) or (13). On the other hand, if movements in real interest rates are not too large, then we can use the linearization of p_t^* (i.e., Taylor expansion truncated after the linear term) around $d=E(d)$ and $\bar{r}=E(\bar{r})$; i.e.,

$$(16)\qquad \hat{p}_t^*\cong\sum_{k=0}^{\infty}\bar{\gamma}^{k+1}\hat{d}_{t+k}-\frac{E(d)}{E(\bar{r})}\sum_{k=0}^{\infty}\bar{\gamma}^{k+1}\hat{\bar{r}}_{t+k}$$

where $\bar{\gamma}=1/(1+E(\bar{r}))$, and a hat over a variable denotes the variable minus its mean. The first term in the above expression is just the expression for p_t^* in (4) (demeaned). The second term represents the effect on p_t^* of movements in real discount rates. This second term is identical to the expression for $\hat{p}^*$ in (4) except that d_{t+k} is replaced by $\hat{\bar{r}}_{t+k}$ and the expression is premultiplied by $-E(d)/E(\bar{r})$.

It is possible to offer a simple intuitive interpretation for this linearization. First note that the derivative of $1/(1+\bar{r}_{t+k})$, with respect to $\bar{r}$ evaluated at $E(\bar{r})$ is $-\bar{\gamma}^2$. Thus, a one percentage point increase in $\bar{r}_{t+k}$ causes $1/(1+\bar{r}_{t+k})$ to drop by $\bar{\gamma}^2$ times 1 percent, or slightly less than 1 percent. Note that all terms in (15) dated $t+k$ or higher are premultiplied by $1/(1+\bar{r}_{t+k})$. Thus, if $\bar{r}_{t+k}$ is increased by one percentage point, all else constant, then all of these terms will be reduced by about $\bar{\gamma}^2$ times 1 percent. We can approximate the sum of all these terms as $\bar{\gamma}^{k-1}E(d)/E(\bar{r})$, where $E(d)/E(\bar{r})$ is the value at the beginning of time $t+k$ of a constant dividend stream $E(d)$ discounted by $E(\bar{r})$, and $\bar{\gamma}^{k-1}$ discounts it to the present. So, we see that a one percentage point increase in $\bar{r}_{t+k}$, all else constant, decreases p_t^* by about $\bar{\gamma}^{k+1}E(d)/E(\bar{r})$, which corresponds to the kth term in expression (16). There are two sources of inaccuracy with this linearization. First, the present value of all future dividends starting with time $t+k$ is not exactly $\bar{\gamma}^{k-1}E(d)/E(\bar{r})$. Second, increasing $\bar{r}_{t+k}$ by one percentage point does not cause $1/(1+\bar{r}_{t+k})$ to fall by exactly $\bar{\gamma}^2$ times 1 percent. To some extent, however, these errors in the effects on p_t^* of $\bar{r}_t, \bar{r}_{t+1}, \bar{r}_{t+2}, \dots$ should average out, and one can use (16) to get an idea of the effects of changes in discount rates.

To give an impression as to the accuracy of the linearization (16), I computed p_t^* for data set 2 in two ways: first using (15) and then using (16), with the same terminal condition p_{1979}^*. In place of the unobserved $\bar{r}_t$ series, I used the actual four–six-month prime commercial paper rate plus a constant to give it the mean $\bar{r}$ of Table 2. The commercial paper rate is a *nominal* interest rate, and thus one would expect its fluctuations represent changes in inflationary expectations as well as real interest rate movements. I chose it nonetheless, rather arbitrarily, as a series which shows much more fluctuation than one would normally expect to see in an expected *real* rate. The commercial paper rate ranges, in this sample, from 0.53 to 9.87 percent. It stayed below 1 percent for over a decade (1935–46) and, at the end of the sample, stayed generally well above 5 percent for over a decade. In spite of this erratic behavior, the correlation coefficient between p^* computed from (15) and p^* computed from (16) was .996, and $\sigma(p_t^*)$ was 250.5 and 268.0 by (15) and (16), respectively. Thus the linearization (16) can be quite accurate. Note also that while these large movements in $\bar{r}_t$ cause p_t^* to move much more than was observed in Figure 2, $\sigma(p^*)$ is still less than half of $\sigma(p)$. This suggests that the variability $\bar{r}_t$ that is needed to save the efficient

TABLE 2—SAMPLE STATISTICS FOR PRICE AND DIVIDEND SERIES

	Data Set 1: Standard and Poor's	Data Set 2: Modified Dow Industrial
Sample Period:	1871–1979	1928–1979
1) $E(p)$	145.5	982.6
$E(d)$	6.989	44.76
2) $\bar{r}$	.0480	0.456
$\bar{r}_2$	.0984	.0932
3) $b=ln\lambda$	.0148	.0188
$\hat{\sigma}(b)$	(.0011)	(1.0035)
4) $cor(p, p^*)$	.3918	.1626
$\sigma(d)$	1.481	9.828
Elements of Inequalities:		
Inequality (1)		
5) $\sigma(p)$	50.12	355.9
6) $\sigma(p^*)$	8.968	26.80
Inequality (11)		
7) $\sigma(\Delta p+d_{-1}-\bar{r}p_{-1})$	25.57	242.1
$min(\sigma)$	23.01	209.0
8) $\sigma(d)/\sqrt{\bar{r}_2}$	4.721	32.20
Inequality (13)		
9) $\sigma(\Delta p)$	25.24	239.5
$min(\sigma)$	22.71	206.4
10) $\sigma(d)/\sqrt{2\bar{r}}$	4.777	32.56

Note: In this table, E denotes sample mean, σ denotes standard deviation and $\hat{\sigma}$ denotes standard error. $Min(\sigma)$ is the lower bound on σ computed as a one-sided χ^2 95 percent confidence interval. The symbols p, d, $\bar{r}$, $\bar{r}_2$, b, and p^* are defined in the text. Data sets are described in the Appendix. Inequality (1) in the text asserts that the standard deviation in row 5 should be less than or equal to that in row 6, inequality (11) that σ in row 7 should be less than or equal to that in row 8, and inequality (13) that σ in row 9 should be less than that in row 10.

markets model is much larger yet, as we shall see.

To put a formal lower bound on $\sigma(\bar{r})$ given the variability of Δp, note that (16) makes $\hat{p}_t^*$ the present value of $z_t, z_{t+1}, \ldots$ where $z_t \equiv \hat{d}_t - \hat{r}_t E(d)/E(\bar{r})$. We thus know from (13) that $2E(\bar{r})var(\Delta p) \leqslant var(z)$. Moreover, from the definition of z we know that $var(z) \leqslant var(d) + 2\sigma(d)\sigma(\bar{r})E(d)/E(\bar{r}) + var(\bar{r})E(d)^2/E(\bar{r})^2$ where the equality holds if d_t and $\bar{r}_t$ are perfectly negatively correlated. Combining these two inequalities and solving for $\sigma(\bar{r})$ one finds

(17)

$$\sigma(\bar{r}) \geqslant \left(\sqrt{2E(\bar{r})}\,\sigma(\Delta p) - \sigma(d)\right)E(\bar{r})/E(d)$$

This inequality puts a lower bound on $\sigma(\bar{r})$ proportional to the discrepancy between the left-hand side and right-hand side of the inequality (13).[17] It will be used to examine the data in the next section.

V. Empirical Evidence

The elements of the inequalities (1), (11), and (13) are displayed for the two data sets (described in the Appendix) in Table 2. In both data sets, the long-run exponential growth path was estimated by regressing $ln(P_t)$ on a constant and time. Then λ in (3) was set equal to e^b where b is the coefficient of time (Table 2). The discount rate $\bar{r}$ used to compute p^* from (4) is estimated as the average d divided by the average p.[18] The terminal value of p^* is taken as average p.

With data set 1, the nominal price and dividend series are the real Standard and Poor's Composite Stock Price Index and the associated dividend series. The earlier observations for this series are due to Alfred Cowles who said that the index is

> intended to represent, ignoring the elements of brokerage charges and taxes, what would have happened to an investor's funds if he had bought, at the beginning of 1871, all stocks quoted on the New York Stock Exchange, allocating his purchases among the individual stocks in proportion to their total monetary value and each month up to 1937 had by the same criterion redistributed his holdings among all quoted stocks. [p. 2]

In updating his series, Standard and Poor later restricted the sample to 500 stocks, but the series continues to be value weighted. The advantage to this series is its comprehensiveness. The disadvantage is that the dividends accruing to the portfolio at one point of time may not correspond to the dividends forecasted by holders of the Standard and Poor's portfolio at an earlier time, due to the change in weighting of the stocks. There is no way to correct this disadvantage without losing comprehensiveness. The original portfolio of 1871 is bound to become a relatively smaller and smaller sample of *U.S.* common stocks as time goes on.

With data set 2, the nominal series are a modified Dow Jones Industrial Average and associated dividend series. With this data set, the advantages and disadvantages of data set 1 are reversed. My modifications in the Dow Jones Industrial Average assure that this series reflects the performance of a single unchanging portfolio. The disadvantage is that the performance of only 30 stocks is recorded.

Table 2 reveals that all inequalities are dramatically violated by the sample statistics for both data sets. The left-hand side of the inequality is always at least five times as great as the right-hand side, and as much as thirteen times as great.

The violation of the inequalities implies that "innovations" in price as we measure them can be forecasted. In fact, if we regress $\delta_{t+1}p_{t+1}$ onto (a constant and) p_t, we get significant results: a coefficient of p_t of $-.1521$ ($t=-3.218$, $R^2=.0890$) for data set 1 and a coefficient of $-.2421$ ($t=-2.631$, $R^2=.1238$) for data set 2. These results are

[17] In deriving the inequality (13) it was assumed that d_t was known at time t, so by analogy this inequality would be based on the assumption that r_t is known at time t. However, without this assumption the same inequality could be derived anyway. The maximum contribution of $\bar{r}_t$ to the variance of ΔP occurs when $\bar{r}_t$ is known at time t.

[18] This is not equivalent to the average dividend price ratio, which was slightly higher (.0514 for data set 1, .0484 for data set 2).

not due to the representation of the data as a proportion of the long-run growth path. In fact, if the holding period return H_t is regressed on a constant and the dividend price ratio D_t/P_t, we get results that are only slightly less significant: a coefficient of 3.533 ($t=2.672$, $R^2=.0631$) for data set 1 and a coefficient of 4.491 ($t=1.795$, $R^2=.0617$) for data set 2.

These regression tests, while technically valid, may not be as generally useful for appraising the validity of the model as are the simple volatility comparisons. First, as noted above, the regression tests are not insensitive to data misalignment. Such low R^2 might be the result of dividend or commodity price index data errors. Second, although the model is rejected in these very long samples, the tests may not be powerful if we confined ourselves to shorter samples, for which the data are more accurate, as do most researchers in finance, while volatility comparisons may be much more revealing. To see this, consider a stylized world in which (for the sake of argument) the dividend series d_t is absolutely constant while the price series behaves as in our data set. Since the actual dividend series is fairly smooth, our stylized world is not too remote from our own. If dividends d_t are absolutely constant, however, it should be obvious to the most casual and unsophisticated observer by volatility arguments like those made here that the efficient markets model must be wrong. Price movements cannot reflect new information about dividends if dividends never change. Yet regressions like those run above will have limited power to reject the model. If the alternative hypothesis is, say, that $\hat{p}_t=\rho\hat{p}_{t-1}+\varepsilon_t$, where ρ is close to but less than one, then the power of the test in short samples will be very low. In this stylized world we are testing for the stationarity of the p_t series, for which, as we know, power is low in short samples.[19] For example, if postwar data from, say, 1950–65 were chosen (a period often used in recent financial markets studies) when the stock market was drifting up, then clearly the regression tests will not reject. Even in periods showing a reversal of upward drift the rejection may not be significant.

Using inequality (17), we can compute how big the standard deviation of real discount rates would have to be to possibly account for the discrepancy $\sigma(\Delta p)-\sigma(d)/(2\bar{r})^{1/2}$ between Table 2 results (rows 9 and 10) and the inequality (13). Assuming Table 2 $\bar{r}$ (row 2) equals $E(\bar{r})$ and that sample variances equal population variances, we find that the standard deviation of $\bar{r}_t$ would have to be at least 4.36 percentage points for data set 1 and 7.36 percentage points for data set 2. These are very large numbers. If we take, as a normal range for $\bar{r}_t$ implied by these figures, a ± 2 standard deviation range around the real interest rate $\bar{r}$ given in Table 2, then the real interest rate $\bar{r}_t$ would have to range from -3.91 to 13.52 percent for data set 1 and -8.16 to 17.27 percent for data set 2! And these ranges reflect lowest possible standard deviations which are consistent with the model only if the real rate has the first-order autoregressive structure and perfect negative correlation with dividends!

These estimated standard deviations of *ex ante* real interest rates are roughly consistent with the results of the simple regressions noted above. In a regression of H_t on D_t/P_t and a constant, the standard deviation of the fitted value of H_t is 4.42 and 5.71 percent for data sets 1 and 2, respectively. These large standard deviations are consistent with the low R^2 because the standard deviation of H_t is so much higher (17.60 and 23.00 percent, respectively). The regressions of $\delta_t p_t$ on p_t suggest higher standard deviations of expected real interest rates. The standard deviation of the fitted value divided by the average detrended price is 5.24 and 8.67 percent for data sets 1 and 2, respectively.

VI. Summary and Conclusions

We have seen that measures of stock price volatility over the past century appear to be far too high— five to thirteen times too

[19]If dividends are constant (let us say $d_t=0$) then a test of the model by a regression of $\delta_{t+1}p_{t+1}$ on p_t amounts to a regression of p_{t+1} on p_t with the null hypothesis that the coefficient of p_t is $(1+\bar{r})$. This appears to be an explosive model for which t-statistics are not valid yet our true model, which in effect assumes $\sigma(d)\neq 0$, is nonexplosive.

high—to be attributed to new information about future real dividends if uncertainty about future dividends is measured by the sample standard deviations of real dividends around their long-run exponential growth path. The lower bound of a 95 percent one-sided χ^2 confidence interval for the standard deviation of annual changes in real stock prices is over five times higher than the upper bound allowed by our measure of the observed variability of real dividends. The failure of the efficient markets model is thus so dramatic that it would seem impossible to attribute the failure to such things as data errors, price index problems, or changes in tax laws.

One way of saving the general notion of efficient markets would be to attribute the movements in stock prices to changes in expected real interest rates. Since expected real interest rates are not directly observed, such a theory can not be evaluated statistically unless some other indicator of real rates is found. I have shown, however, that the movements in expected real interest rates that would justify the variability in stock prices are very large—much larger than the movements in nominal interest rates over the sample period.

Another way of saving the general notion of efficient markets is to say that our measure of the uncertainty regarding future dividends—the sample standard deviation of the movements of real dividends around their long-run exponential growth path—understates the true uncertainty about future dividends. Perhaps the market was rightfully fearful of much larger movements than actually materialized. One is led to doubt this, if after a century of observations nothing happened which could remotely justify the stock price movements. The movements in real dividends the market feared must have been many times larger than those observed in the Great Depression of the 1930's, as was noted above. Since the market did not know in advance with certainty the growth path and distribution of dividends that was ultimately observed, however, one cannot be sure that they were wrong to consider possible major events which did not occur. Such an explanation of the volatility of stock prices, however, is "academic," in that it relies fundamentally on unobservables and cannot be evaluated statistically.

Appendix

A. *Data Set 1: Standard and Poor Series*

Annual 1871–1979. The price series P_t is Standard and Poor's Monthly Composite Stock Price index for January divided by the Bureau of Labor Statistics wholesale price index (January *WPI* starting in 1900, annual average *WPI* before 1900 scaled to 1.00 in the base year 1979). Standard and Poor's Monthly Composite Stock Price index is a continuation of the Cowles Commission Common Stock index developed by Alfred Cowles and Associates and currently is based on 500 stocks.

The Dividend Series D_t is total dividends for the calendar year accruing to the portfolio represented by the stocks in the index divided by the average wholesale price index for the year (annual average *WPI* scaled to 1.00 in the base year 1979). Starting in 1926 these total dividends are the series "Dividends per share... 12 months moving total adjusted to index" from Standard and Poor's statistical service. For 1871 to 1925, total dividends are Cowles series Da-1 multiplied by .1264 to correct for change in base year.

B. *Data Set 2: Modified Dow Jones Industrial Average*

Annual 1928–1979. Here P_t and D_t refer to real price and dividends of the portfolio of 30 stocks comprising the sample for the Dow Jones Industrial Average when it was created in 1928. Dow Jones averages before 1928 exist, but the 30 industrials series was begun in that year. The published Dow Jones Industrial Average, however, is not ideal in that stocks are dropped and replaced and in that the weighting given an individual stock is affected by splits. Of the original 30 stocks, only 17 were still included in the Dow Jones Industrial Average at the end of our sample. The published Dow Jones Industrial Average is the simple sum of the price per share of the 30 companies divided by a divisor which changes through time. Thus, if a stock splits two for one, then Dow Jones continues to include only one share but changes the divisor to prevent a sudden drop in the Dow Jones average.

To produce the series used in this paper, the *Capital Changes Reporter* was used to trace changes in the companies from 1928 to 1979. Of the original 30 companies of the Dow Jones Industrial Average, at the end of our sample (1979), 9 had the identical names, 12 had changed only their names, and 9 had been acquired, merged or consolidated. For these latter 9, the price and dividend series are continued as the price and dividend of the shares exchanged by the acquiring corporation. In only one case was a cash payment, along with shares of the acquiring corporation, exchanged for the shares of the acquired corporation. In this case, the price and dividend series were continued as the price and dividend of the shares exchanged by the acquiring corporation. In four cases, preferred shares of the acquiring corporation were among shares exchanged. Common shares of equal value were substituted for these in our series. The number of shares of each firm included in the total is determined by the splits, and effective splits effected by stock dividends and merger. The price series is the value of all these shares on the last trading day of the preceding year, as shown on the Wharton School's Rodney White Center Common Stock tape. The dividend series is the total for the year of dividends and the cash value of other distributions for all these shares. The price and dividend series were deflated using the same wholesale price indexes as in data set 1.

REFERENCES

C. Amsler, "An American Consol: A Reexamination of the Expectations Theory of the Term Structure of Interest Rates," unpublished manuscript, Michigan State Univ. 1980.

S. Basu, "The Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Markets Hypothesis," *J. Finance*, June 1977, *32*, 663–82.

G. E. P. Box and G. M. Jenkins, *Time Series Analysis for Forecasting and Control*, San Francisco: Holden-Day 1970.

W. C. Brainard, J. B. Shoven, and L. Weiss, "The Financial Valuation of the Return to Capital," *Brookings Papers,* Washington 1980, *2*, 453–502.

Paul H. Cootner, *The Random Character of Stock Market Prices*, Cambridge: MIT Press 1964.

Alfred Cowles and Associates, *Common Stock Indexes, 1871–1937*, Cowles Commission for Research in Economics, Monograph No. 3, Bloomington: Principia Press 1938.

E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, May 1970, *25*, 383–420.

______, "The Behavior of Stock Market Prices," *J. Bus., Univ. Chicago*, Jan. 1965, *38*, 34–105.

C. W. J. Granger, "Some Consequences of the Valuation Model when Expectations are Taken to be Optimum Forecasts," *J. Finance*, Mar. 1975, *30*, 135–45.

M. C. Jensen et al., "Symposium on Some Anomalous Evidence Regarding Market Efficiency," *J. Financ. Econ.*, June/Sept. 1978, *6*, 93–330.

S. LeRoy and R. Porter, "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, forthcoming.

M. H. Miller and F. Modigliani, "Dividend Policy, Growth and the Valuation of Shares," *J. Bus., Univ. Chicago*, Oct. 1961, *34*, 411–33.

F. Modigliani and R. Cohn, "Inflation, Rational Valuation and the Market," *Financ. Anal. J.*, Mar./Apr. 1979, *35*, 24–44.

J. Pesando, "Time Varying Term Premiums and the Volatility of Long-Term Interest Rates," unpublished paper, Univ. Toronto, July 1979.

P. A. Samuelson, "Proof that Properly Discounted Present Values of Assets Vibrate Randomly," in Hiroaki Nagatani and Kate Crowley, eds., *Collected Scientific Papers of Paul A. Samuelson*, Vol. IV, Cambridge: MIT Press 1977.

R. J. Shiller, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *J. Polit. Econ.*, Dec. 1979, *87*, 1190–219.

______ **and J. J. Siegel,** "The Gibson Paradox and Historical Movements in Real Interest Rates," *J. Polit. Econ.*, Oct. 1979, *85*, 891–907.

H. Wold, "On Prediction in Stationary Time Series," *Annals Math. Statist.* 1948, *19*, 558–67.

Commerce Clearing House, *Capital Changes Reporter*, New Jersey 1977.

Dow Jones & Co., *The Dow Jones Averages 1855–1970*, New York: Dow Jones Books 1972.

Standard and Poor's *Security Price Index Record*, New York 1978.