NBER WORKING PAPER SERIES

DO STOCK PRICES MOVE TOO MUCH TO BE JUSTIFIED BY SUBSEQUENT CHANGES IN DIVIDENDS?

Robert J. Shiller

Working Paper No. 456

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

February 1980

The research reported here is part of the NBER's research project on the changing role of Debt and Equity Finance in the U.S. Capital Formation in the United States, which is being financed by a grant from the American Council of Life Insurance, and the National Science Foundation under grant #SOC 7907561. Christine Amsler provided research assistance. The views expressed here are solely those of the author and do not necessarily represent the views of the supporting Agencies.

NBER Working Paper 456
February, 1980

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

ABSTRACT

An ex-post rational real common stock price series, formed as the present value of subsequent detrended real dividends, is found to be a very stable and smooth series when compared with the actual detrended real stock price series. An efficient markets model which makes price the optimal forecast of the ex-post rational price is inconsistent with this data if the long-run trend of real dividends is assumed given. To reconcile the data with the efficient markets model, one must assume that the market expected real dividends deviate from their long-run trend much more than they did historically.

Professor Robert J. Shiller
Department of Economics
University of Pennsylvania
3718 Locust Walk / CR
Philadelphia, PA 19104

(215) 243-8483

I. Introduction and Basic Concepts

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 than 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 objected in popular discussions that stock price indexes are too "volatile", i.e., that the movements in stock price indexes could not realistically be attributed to any objective new information, since movements in the indexes are "too big" relative to actual subsequent movements in dividends.

To illustrate graphically why it seems that this is so, 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 detrended real Standard & Poor's Composite Stock Price Index, and p_t^* is the present discounted value of the actual subsequent detrended real dividends. 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, and moving averages tend to smooth the series averaged. While real dividends did vary over this sample period, they did not vary long enough or

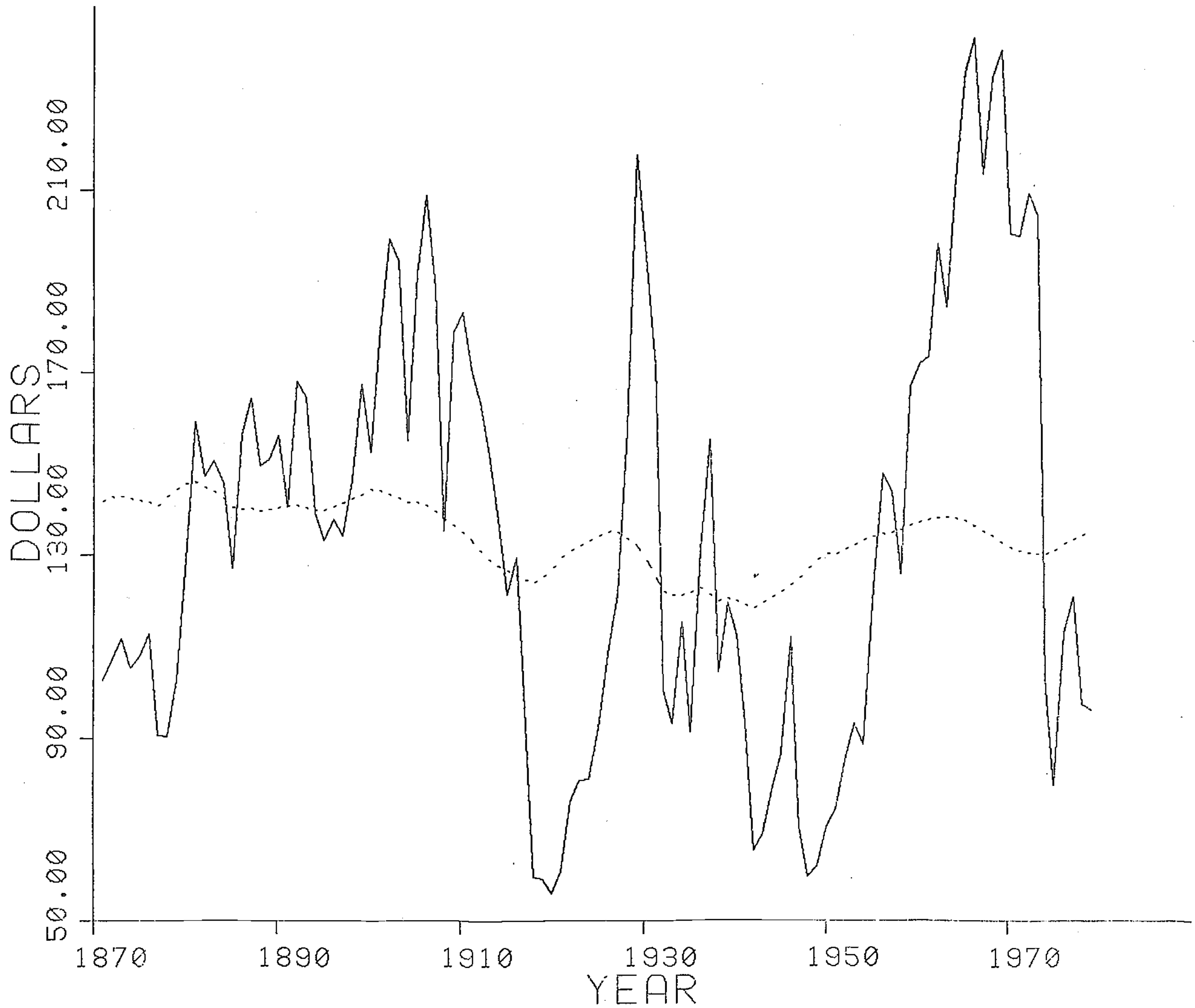

Figure 1 Detrended real Standard & Poor Composite Stock Price Index (solid line, p) and ex-post rational price (dotted line, p*), first of the year, 1871-1979. The variable p* is the present value of actual subsequent detrended real dividends, subject to an assumption about dividends after 1978. The variable p_t is from data set 1, described in appendix, and p^*_t is defined for this data set using $p^*_t = \bar{\gamma}(p^*_{t+1} + d_t)$ with p^*_{1979} set at the average value of p_t over the sample.

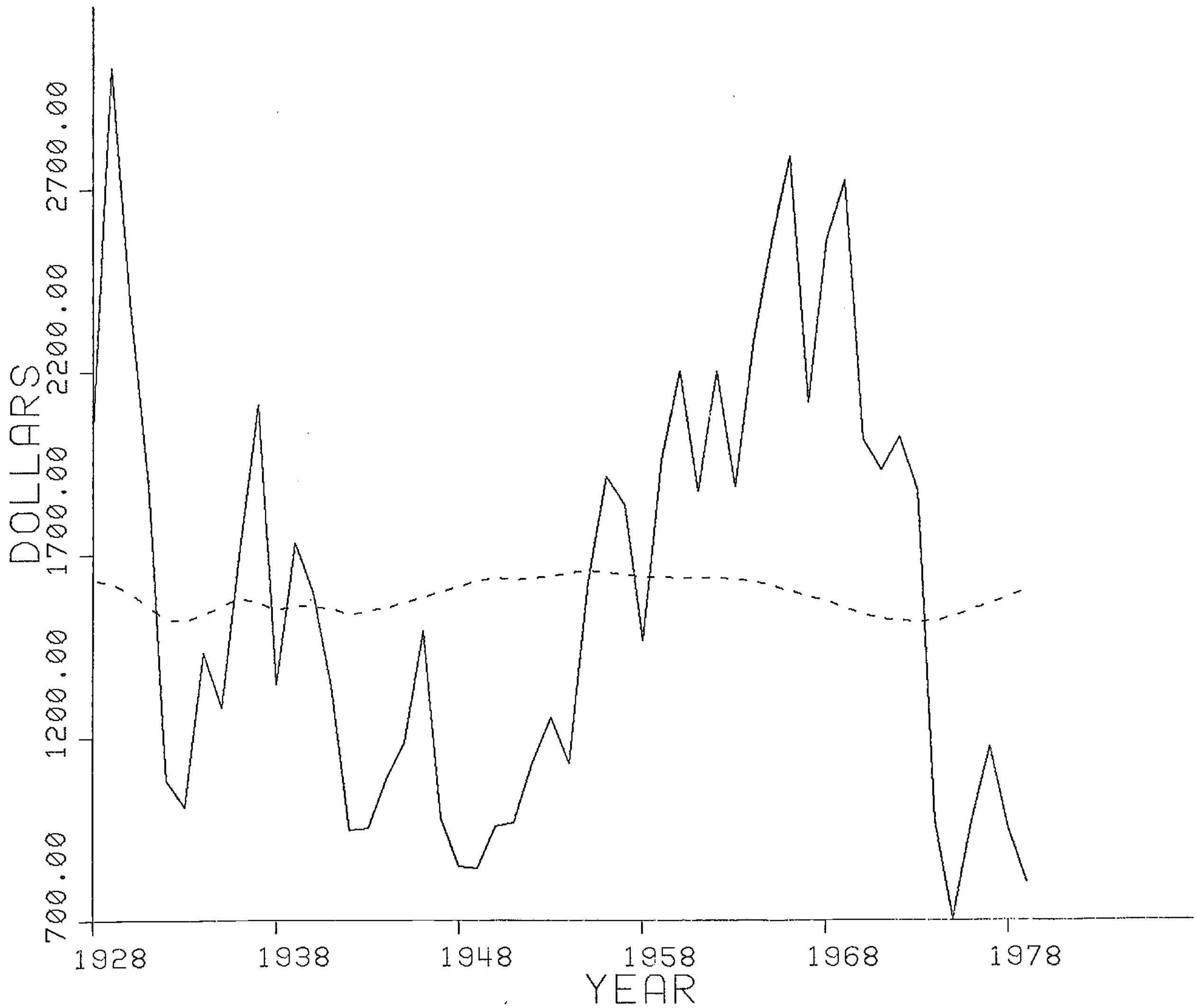

Figure 2 Detrended real modified Dow Jones Industrial Average (solid line, p) and ex-post rational price (dotted line, p*), 1928 - 1979. The variable p* is the present value of actual subsequent detrended real dividends, subject to an assumption about dividends after 1978.

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 trend (i.e., 10-25% below trend for the Standard & Poor series, 16-38% below trend for the Dow Series) only for a few years: 1933, 1934 1935 and 1938. 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 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 $p_t = E_t(p_t^*)$, that is, 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 $\text{var}(p_t^*) = \text{var}(u_t) + \text{var}(p_t)$. Since variances cannot be negative, this means $\text{var}(p_t) \leq \text{var}(p_t^*)$ or, converting to more easily interpreted standard deviations:

$$\sigma(p_t) \leq \sigma(p_t^*) \tag{1}$$

This inequality (noted before by LeRoy and Porter [1979] and Shiller [1979]) 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 II below 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 III) that the standard deviation of Δp_t 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 <u>lower</u> variance. The notion expressed by some that earnings rather than dividend data should be used is discussed in section IV, and a way of assessing the importance of time variation in real discount rates is shown in section V. The inequalities are compared with the data in section VI.

This paper takes as its starting point an approach developed earlier (Shiller [1979]) which showed 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.

LeRoy and Porter [1979] have 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. The approach in this paper, however, while it has benefited from their work is actually quite different from theirs. Some different characterizations of volatility are examined here, dividends rather than earnings are used, and long time series rather than post-war data are employed. I do not attempt Box-Jenkins [1970] modelling of the series. On the other hand, some indication is given here of how much expected real rates would have to move to explain the volatility.

It may appear that this paper is attempting to contradict the extensive literature of efficient markets (for example, as in Cootner [1964], or surveyed in Fama [1970]). 5/ This appearance is somewhat deceptive, since most of this literature really examines different properties of security prices, and in fact there is no agreement on an operational definition of the term "efficient markets". Very little of the efficient markets literature bears directly on the characteristic feature of our model: that expected <u>real</u> returns for the aggregate stock market are constant through time (or approximately so). 6/ 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 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 exploration. 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 we will make 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.

II. 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 (let us say) at the end of time t, E_t is the expectations operator conditional on information available at time t and γ is the constant real discount factor. We 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 detrended series $p_t = \lambda^{-t}P_t$, $d_t = \lambda^{-(t+1)}D_t$ where $\lambda = (1+g)$, and g is the rate of growth term. Multiplying (2) by λ^{-t} and substituting one finds: 7/

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

The growth rate of the firm 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 detrended series is $\bar{r} > 0$. This discount rate $\bar{r}$ is, it turns out, just the mean detrended real dividend divided by the mean detrended real price, i.e., $\bar{r} = E(d_t)/E(p_t)$. 8/

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 in Shiller and Siegel [1977]) which is the present value of actual subsequent dividends. That is,

$$p_t = E_t(p^*_t)$$

$$\text{where} \qquad p^*_t \equiv \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} d_{t+k} \tag{4}$$

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 fig. 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 .957 and $\bar{\gamma}^{108} = .0084$ 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 summarizes 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. We will 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. Since conditional expectations operators satisfy $E_j E_k = E_{\min(j,k)}$ it follows that $E_{t-1}\Delta_{t+k} = 0$, $k \geq 0$. This means that $\Delta_t X_{t+k}$

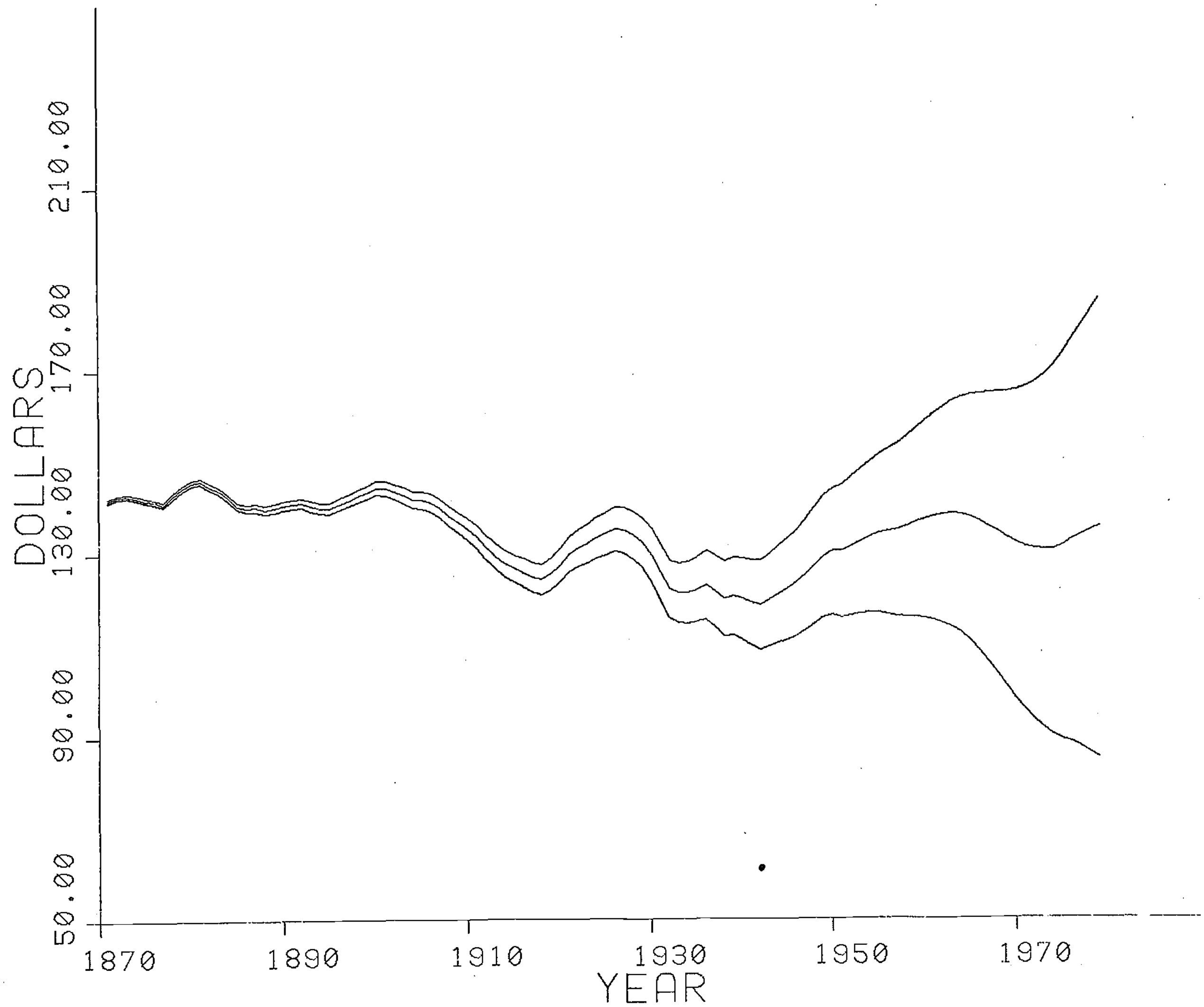

Figure 3 Alternative measures of the ex-post rational price p*, obtained by alternative assumptions about the behavior of dividends after 1978. The middle curve is the p* series plotted in figure 1. The series are computed recursively from terminal conditions using dividend series d of data set 1.

must be uncorrelated 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 (2) 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 $\Delta_t p_t$ is the variable which Granger [1975] and Samuelson [1977] 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 observable innovation in price is related to the unobservable 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.

Expressions (2) - (5) constitute four different representations of the same valuation 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 $\Delta_t p_t$ given the standard deviation of d.

One issue that relates to our 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.[9/] 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. We will perform and discuss such regression tests in section VI below.

To find a limit on the standard deviation of $\Delta_t p_t$ for a given standard deviation of d_t we first note that d_t equals its unconditional expectation plus the sum of its innovations:

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

If we regard $E(d_t)$ 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 $var(\Delta_t p_t)$ might be for given (d_t). Since innovations are serially uncorrelated, we know from (6) that the variance of the sum is the sum of the variances:

$$var(d_t) = \sum_{k=0}^{\infty} var(\Delta_{t-k} d_t) = \sum_{k=0}^{\infty} \sigma_k^2 \tag{7}$$

Our assumption of stationarity for d_t implies that $\mathrm{var}(\Delta_{t-k}d_t) \equiv \sigma_k^2$ 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 σ_0^2, σ_1^2, ..., 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 $\mathrm{var}(\Delta_t d_t) \neq 0$, $\Delta_t d_{t+k} = a_k \Delta_t d_t$, where $a_k = \sigma_k/\sigma_0$. Substituting this into (6) implies

$$\tilde{d}_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \tag{8}$$

where ~ denotes demeaned variable: $\tilde{d}_t = d_t - E(d_t)$ and $\varepsilon_t \equiv \Delta_t d_t$. Thus, if $\mathrm{var}(\Delta_t p_t)$ is to be maximized for given σ_0^2, σ_1^2, ..., the dividend process must be a moving average process in terms of its own innovations. [10] We have thus shown, rather than assumed, that if the variance of $\Delta_t p_t$ is to be maximized, the forecast of d_{t+k} will have the usual ARIMA form as in Box and Jenkins [1970].

We can now find the maximum possible variance for $\Delta_t p_t$ for given variance of d_t. Since the innovations in (5) are perfectly positively correlated, $\mathrm{var}(\Delta_t P_t) = (\sum_{k=0}^{\infty} \gamma^{-k+1} \sigma_k)^2$. To maximize this subject to the constraint $\mathrm{var}(d_t) = \sum_{k=0}^{\infty} \sigma_k^2$ with respect to σ_0, σ_1, ..., we set up the Lagrangean:

$$L = (\sum_{k=0}^{\infty} \gamma^{-k+1} \sigma_k)^2 + \nu(\mathrm{var}(d_t) - \sum_{k=0}^{\infty} \sigma_k^2) \tag{9}$$

where ν is the Lagrangean multiplier. The first order conditions for σ_j, j=0, ... ∞ are:

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

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 $\mathrm{var}(\Delta_t p_t)$, which is a hyperplane in $\sigma_0, \sigma_1, \sigma_2, \ldots$ space, with the hypersphere represented by the constraint. At the maximum $\sigma_k^2 = (1-\bar{\gamma}^{2}) \cdot \mathrm{var}(d_t)\bar{\gamma}^{2k}$ and $\mathrm{var}(\Delta_t p_t) = \bar{\gamma}^{2}\mathrm{var}(d_t)/(1-\bar{\gamma}^{2})$ and so, converting to standard deviations for ease of interpretation, we have:

$$\sigma(\Delta_t p_t) \leq \sigma(d_t)/\sqrt{\bar{r}_2} \qquad (11)$$
$$\text{where } \bar{r}_2 = (1+\bar{r})^2 - 1$$

Here, $\bar{r}_2$ is the two-period interest rate, which is roughly twice the one-period rate. The maximum occurs, then, when d_t is a first order autoregressive process, $\tilde{d}_t = \bar{\gamma}\tilde{d}_{t-1} + \varepsilon_t$, and $E_t\tilde{d}_{t+k} = \bar{\gamma}^{k}\tilde{d}_t$. In contrast, if d_t were revealed, let us say, at t-50 ($\sigma_{50}^2 = \mathrm{var}(d_t)$) then the innovation in dividend would be so heavily discounted in (5) that it would contribute little to $\mathrm{var}(\Delta_t p_t)$. Alternatively, if nothing were revealed about d_t until time t ($\sigma_0^2 = \mathrm{var}(d_t)$) then still the innovation in only one dividend would contribute to $\mathrm{var}(\Delta_t p_t)$.

The same maximum variance for $\Delta_t p_t$ can also be derived in another way. We will illustrate this by a maximum for the variance of the price change Δp_t, though the procedure we use could also be employed to derive expression (11) as well. [11/] Under the stationarity assumption, the variance of Δp_t may be written $\mathrm{var}(\Delta p_t) = 2\mathrm{var}(p_t) - 2\mathrm{cov}(p_t, p_{t+1})$. Since $\Delta_{t+1}p_{t+1}$ cannot be forecasted using p_t, we know that $\mathrm{cov}(\Delta_t p_{t+1}, p_t) = \mathrm{cov}(p_t, p_{t+1}) + \mathrm{cov}(d_t, p_t) - \mathrm{var}(p_t)/\bar{\gamma} = 0$ or, rearranging: $\mathrm{cov}(p_t, p_{t+1}) = \mathrm{var}(p_t)/\bar{\gamma} - \mathrm{cov}(d_t, p_t)$. Substituting this into the expression for $\mathrm{var}(\Delta p_t)$ gives:

$$\begin{aligned}\mathrm{var}(\Delta p_t) &= 2(1 - \tfrac{1}{\bar{\gamma}})\,\mathrm{var}(p_t) + 2\mathrm{cov}(d_t, p_t) \\ &= 2(1 - \tfrac{1}{\bar{\gamma}})\,\mathrm{var}(p_t) + 2\rho_{dp}\sqrt{\mathrm{var}(d_t)}\sqrt{\mathrm{var}(p_t)} \qquad (12)\end{aligned}$$

where ρ_{dp} is the correlation coefficient between d_t and p_t. Maximizing with respect to $\mathrm{var}(p_t)$ for given ρ_{dp} and $\mathrm{var}(d_t)$ we set the first derivative to zero:

$$\frac{\partial \mathrm{var}(\Delta p_t)}{\partial \mathrm{var}(p_t)} = 2(1 - \frac{1}{\bar{\gamma}}) + \rho_{dp} \frac{\sqrt{\mathrm{var}(d_t)}}{\sqrt{\mathrm{var}(p_t)}} = 0 \tag{13}$$

and since $\rho_{dp} > 0$, the second order condition for a maximum is satisfied. Hence at the maximum:

$$\mathrm{var}(\Delta p_t) = \frac{\bar{\gamma}}{2(1-\bar{\gamma})} \rho_{dp}^2 \mathrm{var}(d_t) \tag{14}$$

since $\rho_{dp}^2 \leq 1$ and since $(1-\bar{\gamma})/\bar{\gamma} = \bar{r}$, we thus have the inequality

$$\sigma(\Delta p_t) \leq \frac{1}{\sqrt{2\bar{r}}} \sigma(d_t) \tag{15}$$

The maximum standard deviation of Δp_t for given standard deviation of d_t is attained if dividends are paid every period and the demeaned dividend series, $\tilde{d}_t$ follows the first order autoregression $\Delta\tilde{d}_t = -\bar{r}\tilde{d}_{t-1} + \varepsilon_t$ where ε_t is unforecastable white noise, and $E_t(\tilde{d}_{t+k}) = (1-\bar{r})^k\tilde{d}_t$. In this case, there is a perfect correlation between prices and dividends, and $\tilde{p}_t = \tilde{d}_t/(2\bar{r})$ (while at the same time $E(p_t) = E(d_t)/\bar{r}$).

Intuitively, one can see why $\sigma(\Delta p_t)$ is maximized with such a d_t process. Such a d_t process moves fairly smoothly, but not too smoothly, through time. If d_t were much less smooth (i.e, a choppy irregular series) then the long average in the present value formula would in effect average these movements out leaving little variation in p_t^* and hence little room for movements in Δp_t. If the d_t process were much smoother than in this autoregression, then the moving average in the definition of p_t^* would not effectively average out the movements in d so that p_t^* would vary a lot. However, p_t^* would then be very smooth itself, again leaving little room for variation in Δp_t.

At the maximum, the R^2 between Δp_t and p_{t-1} is $\bar{r}/2$ which is a very small number. Hence, the case where $\sigma(\Delta p_t)$ is maximized is also a case where p_t resembles a random walk. Even with fairly sizeable samples this correlation will generally not be significant. However, the above maximum is not also the minimum for the R^2 of a regression of Δp_t on p_{t-1}. If $\tilde{d}_t = \mu\tilde{d}_{t-1} + \varepsilon_t$ and $E_t(\tilde{d}_{t+k}) = \mu^k\tilde{d}_t$ then $R^2 = (1 - \mu)/2$ and so the R^2 approaches zero as μ approaches 1.

III. 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 $\Delta_t p_t$ or Δp_t, one sees long stretches of time when their (absolute) values are small and then an infrequent 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 our 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. [12/] Our model does not suggest that price changes have a distribution in this class. Our model instead suggests that the real issue is the existence of moments for the dividends series.

As long as d_t is jointly stationary with information and has a finite variance, then p_t, p_t^*, $\Delta_t p_t$ and Δp_t will be stationary and have a finite variance. [13/] If d_t 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 kurtosis = $E(\tilde{p}_t^4)/(E(\tilde{p}_t^2))^2$. Suppose that with a probability of 1/n we are told d_t at the beginning of time t but with probability (n-1)/n have no information. In time periods when we are told d_t, $\tilde{p}_t = \bar{\gamma}\tilde{d}_t$, otherwise $\tilde{p}_t = 0$. Then $E(\tilde{p}_t^4) = \frac{1}{n} E((\bar{\gamma}\tilde{d}_t)^4)$ and $E(\tilde{p}_t^2) = \frac{1}{n} E((\bar{\gamma}\tilde{d}_t)^2)$ so that kurtosis equals $nE(\bar{\gamma}\tilde{d}_t)^4)/E((\bar{\gamma}d_t)^2))$ which equals n times the kurtosis of the normal distribution. Hence, by choosing n high enough we can acheive an arbitrarily high kurtosis, and yet the variance will always exist. Moreover, the distribution of $\tilde{p}_t$ conditional on the information that the dividend has been revealed is also normal, so that, as Rosenberg [1972] suggested, conditional distributions may always be normal.

If information is revealed in big lumps occasionally (so as to induce high kurtosis as suggested in the above example) $\mathrm{var}(\Delta p_t)$ or $\mathrm{var}(\Delta_t p_t)$ are not especially large. The variance of Δp_t 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 of Δp_t for given variance of d_t indeed comes when $\tilde{d}_t$ has a simple autoregressive forecast as noted in the previous section. In the above example, where information about dividends is revealed one time in n, $\sigma(\Delta p_t) = (2\bar{\gamma}^2/n)^{1/2}\sigma(d_t)$. The value of $\sigma(\Delta p_t)$ implied by this example is for all n strictly below the upper bound of the inequality (15). 14/

IV. Dividends or Earnings?

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. Earnings are statistics conceived by accountants which are supposed to provide an indicator of how well a company is doing. Unfortunately, there is a great deal of latitude for the definition of earnings, as the recent literature on inflation accounting will attest. Historically, earnings appear inaccurate and overstated in that retained earnings appear to earn far less than the discount rate and are a poor indicator of future dividends, as noted by Cowles [1938], Little [1962] and Baumol, Heim, Malkiel and Quandt [1970].

There is no reason why price per share ought to be the present value of expected earnings per share if earnings are retained. In fact, as Miller and Modigliani [1961] 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. 15/ That is, however, what the present value of earnings per share would include. 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, 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. That feeling apparently stems from a careless extrapolation of the case where firms paid out <u>no</u> dividends or paid constant dividends in a growing economy. Simple growth models such as those described in Fama and Miller [1971] in fact show that as long as the payout fraction is non-zero, one may regard price as present value of dividends. 16/

In these models, one can always describe price as the present value of dividends as long as the dividend retention policy doesn't cause the firm to grow at the discount rate. With our Standard and Poor data, the growth rate of real price is only about 1.5%, while the discount rate is about 4.5% + 1.5% = 6%. At these rates, the price of the firm a few decades hence is of little concern to investors.

The crucial thing to recognize in our 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 and using our same formula with earnings in place of dividends.

V. 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? 17/

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

$$P_t = E_t\left(\sum_{k=0}^{\infty} \prod_{j=0}^{k} \frac{1}{1+r_{t+j}} D_{t+k}\right) \tag{16}$$

which has the property that $E_t(H_t) = r_t$, i.e., expected one-period holding returns equal the one-period real discount rate at time t. As before, we can rewrite the model in terms of detrended series:

$$p_t = E_t(p_t^*) \tag{17}$$

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

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

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 (15). 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 $d = E(d_t)$and $\bar{r} = E(\bar{r}_t)$:

$$\tilde{p}_t^* \cong \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \tilde{d}_{t+k} - \frac{E(d_t)}{E(\bar{r}_t)} \sum_{k=0}^{\infty} \bar{\gamma}^{k+1} \tilde{\bar{r}}_{t+k} \tag{18}$$

where $\bar{\gamma} \doteq 1/(1+E(\bar{r}_t))$. 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 p* in (4) except that d_{t+k} is replaced by $\tilde{r}_{t+k}$ and the expression is premultiplied by $E(d_t)/E(\bar{r}_t)$.

It is possible to offer a simple intuitive interpretation for this linearization. First note that the linearization of $1/(1+\bar{r}_{t+k})$, demeaned, around $E(\bar{r}_t)$is $-\bar{\gamma}^2\tilde{\bar{r}}_{t+k}$. 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 one percent, or slightly less than one percent. Note that all terms in (17) 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%. We can approximate the sum of all these terms as $\bar{\gamma}^{-k-1}E(d_t)/E(\bar{r}_t)$. $E(d_t)/E(\bar{r}_t)$ is the value at the beginning of time t+k of a constant dividend stream $E(d_t)$ discounted by $E(\bar{r}_t)$, and $\bar{\gamma}^{-k-1}$ discounts it to the present. So, we see that a 1 percentage point increase in $\bar{r}_{t+k}$, all else constant, decreases p^*_t by about $\bar{\gamma}^{-k+1}E(d_t)/E(\bar{r}_t)$, which corresponds to the k^{th} term in expression (18). There are two sources of innacuracy with this linearization. First, the present value of all future dividends starting with time t+k is not exactly $\bar{\gamma}^{-k-1}E(d_t)/E(\bar{r}_t)$. 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 one percent. To some extent, however, these errors in the effects on p^*_t of $\bar{r}_t$, $\bar{r}_{t+1}$, $\bar{r}_{t+2}$, ... should average out, and one can use (18) to get an idea of the effects of changes in discount rates.

To give an impression as to the accuracy of the linearization (18), I computed p^*_t for data set 2 in two ways: first using (17) and then using (18), with the same terminal condition p^*_{1979}. In place of the unobserved $\bar{r}_t$ series I used the actual 4-6 month prime commercial paper rate plus a constant to give it the mean $\bar{r}$ of Table I. 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 a real rate. The commercial paper rate ranges, in this sample, from 0.53% to 9.87%. It stayed below 1% for over a decade (1935-46) and, at the end of the sample, stayed generally well above 5% for over a decade. In spite of this erratic behavior, the correlation coefficient between p* computed from (17) and p* computed from (18) was .996, and $\sigma(p^*_t)$ was 250.5 and 268.0 by (17) and (18) respectively. Thus, the linearization (18) 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, $\sigma(p_t^*)$ is still less than half of $\sigma(p_t)$. This suggests that the variability $\bar{r}_t$ that is needed to save the efficient markets model is much larger yet, as we shall indeed see in the empirical section below.

Under some assumption about the correlation between $\bar{r}_t$ and d_t we can use analogies to the above inequalities to gauge the maximum effect of variation in $\bar{r}_t$ for given variance of $\bar{r}_t$ on the variation of p_t, $\Delta_t p_t$ or Δp_t. For example, if we assume $\bar{r}_t$ and d_t are uncorrelated, then using the inequality (15) we see that the maximum possible contribution V_{max} of time variation in $\bar{r}_t$ to the variance of Δp_t is:

$$V_{max} = (E(d_t)/E(\bar{r}_t))^2 var(\bar{r}_t)/2E(\bar{r}_t) \tag{19}$$

Of course, this maximum can occur only when the unobserved real interest rate has the right stochastic structure, i.e., it is first order autoregressive: $\Delta\tilde{\bar{r}}_t = -E(\bar{r}_t)\tilde{\bar{r}}_{t-1} + \varepsilon_t$. We will use expression (19) in the next section of the paper to help us to interpret our results.

VI. Empirical Evidence

The elements of the inequalities (1), (11) and (15) are displayed for the two data sets (described in the appendix) in Table I. In both data sets, the trend was estimated by regressing $\ln(P_t)$ on a constant and time and then setting λ in (3) equal to e^b where b is the coefficient of time. The detrended real series were then multiplied by a scale factor chosen so that p for 1979 equalled the nominal value of the index for that date. The discount rate $\bar{r}$ is estimated

TABLE I

SAMPLE STATISTICS FOR PRICE AND DIVIDEND SERIES

1	2	3	4	5	6	7	8	9	10	11	12
						Elements of Inequalities					
						Inequality 1		Inequality 11		Inequality 15	
Data Set	Sample Period	$E(p)$ $E(d)$	$\bar{r}$ $\bar{r}_2$	$b = \ln\lambda$ $\hat{\sigma}(b)$	$\mathrm{cov}(p, p^*)$ $\sigma(d)$	$\sigma(p_t)$	$\sigma(p_t^*)$	$\sigma(\Delta_t p_t)$ $\min(\sigma)$	$\sigma(d)/\sqrt{\bar{r}_2}$	$\sigma(\Delta p_t)$ $\min(\sigma)$	$\sigma(d)/\sqrt{2\bar{r}}$
#1 Standard & Poor	1871–1979	135.3 6.119	.0452 .0925	.0145 (.0011)	.3962 1.281	47.17	7.505	24.59 22.13	4.212	24.29 21.86	4.260
#2 Modified Dow Industrial	1928–1979	1599. 68.56	.0429 .0876	.0176 (.0036)	.1789 14.75	595.6	44.18	401.8 346.8	49.83	397.5 342.5	50.36

NOTE: In this table, E denotes sample mean, σ denotes standard deviation and $\hat{\sigma}$ denotes standard error. Min (σ) is the lower bound on σ computed as a one-sided χ^2 95% confidence interval. The symbols p, d, $\bar{r}$, $\bar{r}_2$, b, p* and $\Delta_t p_t$ are defined in the text. Data sets are described in the appendix.

Inequality 1 in the text asserts that the standard deviation in column 7 should be less than or equal to that in column 8, inequality 11 that σ in column 9 should be less than or equal to that in column 10, and inequality 15 that σ in column 11 should be less than that in column 12.

as the average detrended real dividend divided by the average detrended real price. 18/

With data set 1 the series are the real Standard & Poor's Composite Stock Price Index and the associated real dividend series. The earlier observations for this series are due to Cowles [1938] 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" ([1938], p.2). In updating his series, Standard & 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 & 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 series are a modified real Dow Jones Industrial Average and associated real dividend series. With this data set, the advantages and disadvantages of data set 1 are reversed. Our modifications in the Dow Jones Industrial Average assure that our series reflect the performance of a single unchanging portfolio. The disadvantage is that the performance of only 30 stocks is recorded.

The table 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 5 times as great as the right hand side, and as much as 13 times as great.

We saw above that the inequality (15) could be derived assuming only that innovations in price are uncorrelated with the price level, $\text{cov}(\Delta_{t+1}p_{t+1}, p_t) = 0$ and the assumption that processes are stationary. Since the inequality (15) is violated dramatically, and processes do appear stationary we would expect that the sample covariance between $\Delta_{t+1}p_{t+1}$ and p_t is not zero. 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 $-.1576$ ($t = -3.271$, $R^2 = .0831$) for data set 1 and a coefficient of $-.2382$ ($t = -2.618$, $R^2 = .1048$) for data set 2. These results are not due to the detrending of the data. 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 3.875 ($t = 2.669$, $R^2 = .0541$) for data set 1 and a coefficient of 4.954 ($t = 1.843$, $R^2 = .0457$) 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 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 $\tilde{p}_t = \rho \tilde{p}_{t-1} + \varepsilon_t$, 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 expression (19), we can compute how big the standard deviation of real discount rates would have to be to possibly account for the discrepancy $\delta = \sigma^2(\Delta p_t) - \sigma^2(d_t)/(2\bar{r})$ between Table I results (columns 11 and 12) and the inequality (15). Setting $\delta \leq V_{max}$ in (19), assuming Table I $\bar{r}$ (column 4) equals $E(\bar{r}_t)$ and that sample variances equal population variances, we find that $\sigma^2(\tilde{r}_t) \geq 2\delta E(\bar{r}_t)^3/E(d_t)^2$. We find that the standard deviation of $\tilde{\bar{r}}_t$ would have to be at least 5.32 percentage points for data set 1 and 7.22 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 $\pm$ 2 standard deviation range around the $\bar{r}$ given in Table I, then $\bar{r}_t$ would have to range from -6.12% to 15.16% for data set 1 and -10.14% to 18.72% 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 noted above![20/]

VII. Conclusion

We did not attempt to test formally whether the inequalities on volatility which follow from the efficient markets model are violated.[21/] Instead, we sought to describe the data so as to clarify what kinds of assumptions are necessary to reconcile it with the model. Formal tests could be undertaken only under some maintained hypothesis about the stochastic properties of the dividend series (e.g., that they are an ARIMA process of low order) and our model tells us nothing about the dividend series. Standard data analysis procedures: detrending, first differencing and estimating autoregressions or moving average representations introduce certain biases in the testing procedures that we do not wish to casually accept. Depending on what we assumed as a maintained hypothesis about the stochastic properties of the dividend series, we might or might not reject the model. We could not reject the model if our maintained hypothesis allowed for a small probability of really enormous movement in d_t which was not observed in the sample. Such a maintained hypothesis might also be described as allowing for a small probability each period of a major change in trend. Investors may well have been rationally adjusting their forecasts in response to new information about this possible big event which did not occur.

The efficient markets model does tell us that the price innovation series $\Delta_t p_t$ is serially uncorrelated, and Δp_t is approximately serially uncorrelated, which enables us to put a χ^2 95% lower bound on their standard deviations (Table I). With this information alone we can summarize our basic conclusion: the movements in detrended real price p_t over the last century can be justified as the rational response to new information about anticipated future movements in detrended real dividends d_t only if these anticipated future movements were many times bigger than those actually observed over the last century. Thus, the efficient markets model is at best an "academic" model about an unobservable (new information about the trend) and does not describe observed movements in data. Moreover, we have seen that if movements in the unobserved real interest rates are instead invoked to explain the high volatility of prices (taking the observed variance of d_t as the true variance) then these real interest rate movements would have to be very large.

APPENDIX

SOURCES OF DATA

Data Set 1
Standard and Poor Series

Annual 1871 - 1979. The price series P_t is Standard & 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). The Standard & Poor Monthly Composite Stock Price index, which may be found in Standard & Poor [1978] p. 119, is a continuation of the Cowles Commission Common Stock Index (Cowles [1938]), and currently is based on 500 stocks. Prior to 1918 the prices on which the index is based are simple averages of the high and low price for the month. Starting in 1918 the prices are monthly averages of Wednesday closing prices. Rosenberg [1972] suggested a correction to the sample variance of monthly changes to estimate end of month to end of month price change variance. With our annual data, this correction is not so important, and we ignore it.

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. Starting in 1926 these total dividends are the series "Dividends per share ... 12 months moving total adjusted to index" from Standard & Poor statistical serivce [1978]. For 1871 to 1925 total dividends are Cowles [1938] series Da-1 multiplied by .1264 to correct for change in base year.

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 compirsing 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 2 for 1 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 [1977] was used to trace changes in the companies from 1928 - 1979. Of the original 30 companies of the Dow Jones Industrial Average, today (1979) nine have the identical names, 12 have changed only their names, and nine were acquired, merged or consolidated. For these latter nine, 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 common shares of equal value at time of acquisition. 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 first trading day of the year. 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 prices indexes as in data set 1.

FOOTNOTES

1/ The stock price index may look unfamiliar because it is deflated by a price index, detrended, 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 undetrended series show a gradual increase in scale of about 1.5% a year, both for dividends and prices. Assumptions about public knowledge or lack of knowledge of this trend are important, as we shall discuss below. p^* is computed subject to an assumption about dividends after 1978 and uses discount rate $\bar{r}$, average dividend d over average price p from Table I. See text and figure 3 below.

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 reverse of the inequality (1). This objection stems, however, from a faulty understanding of the meaning of conditional expectation. The subscript t on the expectations operator E means "taking as given (i.e., non-random) 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 his 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 existence of this "noise" in p_t in precisely our interest here. Further discussion of the robustness of such inequalities is in Shiller [1979].

4/ This analysis was extended to yields on preferred stocks by Amsler [1979].

5/ It should not be inferred that the literature on efficient markets uniformly supports the notions of efficiency put forth there, e.g., that no assets are dominated or that no trading rule dominates a buy and hold strategy. Notable papers which claim to find evidence against efficiency so defined are Alexander [1964], Basu [1977], Jensen et. al. [1978] and Modigliani and Cohn [1979].

6/ The claim that real short-term interest rates on default-free fixed loans are roughly constant has received a great deal of attention recently. This literature is discussed critically in Shiller [1980].

7/ No assumptions are introduced in going from (2) to (3), since (3) is just an algebraic transformation of (2). We 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. We shall continue to include the time subscript in expressions such as $\text{var}(d_t)$ or $E(d_t)$ even though the expressions are not functions of time. 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. In section VII we discuss this assumption.

8/ Taking unconditional expectations of both sides of (3) we find

$$E(p_t) = \frac{\bar{\gamma}}{1-\bar{\gamma}} E(d_t)$$

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

9/ It follows that $\text{var}(u_t) = \text{var}(\Delta_t p_t)/(1-\bar{\gamma}^2)$ as LeRoy and Porter [1979] noted. They base their volatility tests on our inequality (1) (which they call theorem 2) and a stronger inequality $\sigma^2(p) + \sigma^2(\Delta_t p_t)/(1-\bar{\gamma}^2) \leq \sigma^2(p^*)$, (for which the equality should always hold, their theorem 3). They found that, with postwar Standard and Poor earnings data, both inequalities were violated by sample standard deviations.

10/ Of course, all indeterministic stationary processes can be given linear moving average representations (Wold [1948]). 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).

11/ To derive (11) as illustrated here, find the maximum for the variance of $p_{t+1} - p_t/\bar{\gamma} = (1 + 1/\bar{\gamma}^2)\ \mathrm{var}(P_t) - (2/\gamma)\mathrm{cov}(P_t, P_{t+1})$ with the same substitution as the above. Then use the fact that $\mathrm{var}(\Delta_t p_t) = \mathrm{var}(p_{t+1} - p_t/\bar{\gamma}) - \mathrm{var}(d_t)$.

12/ The empirical fact about the unconditional distribution of stock price changes is not that they have infinite variance (which can never be demonstrated with any finite sample) but that they have high kurtosis in the sample. We may then assume finite variance in face of high sample kurtosis at the risk that our sample statistics may not be trustworthy measures of population variances. The risk is that some subsequent movements in stock prices not yet observed will have such magnitude as to swamp out our sample observations. This risk is analogous to that incurred when we assumed above that the dividend series is stationary. This risk that rare or unobserved events may yet force us to drastically change our conclusions is inherent in all statistical research, even when sample data show low kurtosis.

13/ With any stationary process, X_t, the existence of a finite $\mathrm{var}(X_t)$ implies, by Schwartz's inequality, a finite value of $\mathrm{cov}(X_t, X_{t+k})$ 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.

14/ For another illustrative example consider $\tilde{d}_t = \bar{\gamma}\tilde{d}_{t-1} + \varepsilon_t$ 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 $\tilde{d}_t$ has the autoregressive structure, ε_t is not the innovation in d_t. As n goes to infinity, $\sigma(\Delta_t p_t)$ approaches zero.

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 & Poor series and 5.37 for the Dow series. According to the David-Hartley-Person test, normality can be rejected at the 5% level (but not at the 1% level) with a one-tailed test for both data sets.

15/ LeRoy and Porter [1979] 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.

16/ These growth models are more easily understood in continuous time, so instead of (2) we have $P_0 = \int_0^\infty D_t e^{-rt}dt$. In a simple kind of growth 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 that 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) $P_0 = I/r$.

17/ Pesando [1979] has asked the analogous question: how large must the variance in liquidity premia be in order to justify the volatility of long-term interest rates?

18/ This is not equivalent to the average dividend price ratio, which was slightly higher (.0492 for data set 1, .0467 for data set 2).

19/ If dividends are constant (let us say $d_t = 0$) then a test of the model by a regression of $\Delta_{t+1}\,\tilde{p}_{t+1}$ on $\tilde{p}_t$ amounts to a regression of $\tilde{p}_{t+1}$ on $\tilde{p}_t$ with the null hypothesis that the coefficient of $\tilde{p}_t$ 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 non-explosive. Regression tests of our model when $t = 0$ have the form of tests of nonstationarity against the alternative hypothesis, under which ordinary standard errors are asymptotically valid under general conditions.

20/ If we further allow d_t and r_t to be correlated, then for a given variance of d_t and r_t, the variance of Δp_t is maximized if d_t and r_t are perfectly negatively correlated. From (18) and previous arguments we know $\sigma(\Delta p) \leq \sigma(\tilde{d}_t - E(p)\tilde{\bar{r}}_t)/\sqrt{2E(\bar{r})}$. Defining the discrepancy ratio $\delta' = \sigma(\Delta p) / [\sigma(d/\sqrt{2E(\bar{r})}]$ it follows that $\sigma(r) \geq (\delta' - 1)\sigma(d)E(\bar{r})/E(d)$. Using table 1 data, we find $\sigma(\bar{r})$ must be at least 4.45 percentage points for data set 1 and 6.35 percentage points for data set 2. Perfect negative correlation does not allow much lower standard deviation for $\bar{r}$, since d doesn't vary much.

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 4.25% and 5.19% 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 (18.29% and 24.18% 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 5.50% and 8.05% for data sets 1 and 2 respectively.

21/ LeRoy and Porter have attempted such tests.

REFERENCES

Alexander, Sidney, "Price Movements in Speculative Markets: Trends or Random Walks, No. 2," in Cootner, ed., The Random Character of Stock Prices, the MIT Press, Cambridge, 1964b, pp. 338-72.

Amsler, Christine, "An American Consol: A Reexamination of the Expectations Theory of the Term Structure of Interest Rates," unpublished, University of Pennsylvania, 1979.

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

Baumol, W.J., P. Heim, B.G. Malkiel and R.E. Quandt, "Earnings Retention, New Capital and the Growth of Firms," Review of Economics and Statistics, 62: 345-55, 1970.

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

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

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

Cowles, Alfred and Associates, Common Stock Indexes, 1871-1937, Cowles Commission for Research in Economics, Monograph #3, Principia Press, Bloomington, Indiana, 1938.

Dow Jones & Company, The Dow Jones Averages, 1855-1970, Dow Jones Books, New York, 1972.

Fama, Eugene F., "Efficient Capital Markets: A Review of Theory and Empirical Work", Journal of Finance, 25:383-420, May 1970. ?

______________, "The Behavior of Stock Market Prices", Journal of Business, 38: 34-105, January 1965.

Fama, Eugene and Merton Miller, The Theory of Finance, Holt, Rhinehart & Winston, New York, 1972.

Granger, Clive W.J., "Some Consequences of the Valuation Model when Expectations are Taken to be Optimum Forecasts", Journal of Finance, 30:135-45, March 1975.

Jensen, Michael C., et. al., "Symposium on Some Anomalous Evidence Regarding Market Efficiency," Journal of Financial Economics, 6:93-330, June/September 1978.

LeRoy, Stephen and Richard Porter, "The Present Value Relation: Tests Based on Implied Variance Bounds" mimeographed, Board of Governors of the Federal Reserve System, Washington, 1979.

Little, I.M.D., "Higgledy Piggledy Growth," Bulletin of the Oxford University Institute of Statistics, Vol 24, November 1962.

Miller, Merton H. and Franco Modigliani, "Dividend Policy, Growth and the Valuation of Shares," Journal of Business, 411-33, October 1961.

Modigliani, Franco and Richard Cohn, "Inflation, Rational Valuation and the Market," Financial Analysis Journal, March 1979.

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

Rosenberg, Barr, "The Behavior of Random Variables with Nonstationary Variance and the Distribution of Security Prices," unpublished paper, University of California at Berkeley, 1972.

Samuelson, Paul A., "Proof that Properly Discounted Present Values of Assets Vibrate Randomly", in Collected Scientific Papers of Paul A. Samuelson, Vol. IV, Hiroaki Nagatani and Kate Crowley, Editors, M.I.T. Press, Cambridge MA, 1977.

Shiller, Robert J., "Can the Fed Control Real Interest Rates?", in Stanley Fischer, Ed., Rational Expectations and Economic Policy, University of Chicago Press and National Bureau of Economic Research, 1980.

________________, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," Journal of Political Economy, December 1979 (date tentative).

Shiller, Robert J. and Jeremy J. Siegel, "The Gibson Paradox and Historical Movements in Real Interest Rates," Journal of Political Economy 85: 891-907, October 1977.

Standard & Poor's Statistical Service, Security Price Index Record, 1978.

Wold, Hermann, "On Prediction in Stationary Time Series", The Annals of Mathematical Statistics 19:558-67, 1948.