The Use of Volatility Measures in Assessing Market Efficiency
Author(s): Robert J. Shiller
Source: *The Journal of Finance*, Vol. 36, No. 2, Papers and Proceedings of the Thirty Ninth Annual Meeting American Finance Association, Denver, September 5-7, 1980 (May, 1981), pp. 291-304
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2327010
Accessed: 27/11/2009 08:24

THE JOURNAL OF FINANCE • VOL. XXXVI, NO. 2 • MAY 1981

The Use of Volatility Measures in Assessing Market Efficiency*

ROBERT J. SHILLER**

I. Introduction

RECENTLY a number of studies have used measures of the variance or "volatility" of speculative asset prices to provide evidence against simple models of market efficiency. These measures were interpreted as implying that prices show too much variation to be explained in terms of the random arrival of new information about the fundamental determinants of price. The first such use of the volatility measures was made independently by LeRoy and Porter [1981] in connection with stock price and earnings data, and myself [1979], in connection with long-term and short-term bond yields. Subsequently, further use of these measures was made to study efficient markets models involving stock prices and dividends (Shiller [1981b]), yields on intermediate and short-term bonds (Shiller [1981a], Singleton [1980]), preferred stock dividend price ratios and short-term interest rates (Amsler [1980]) and foreign exchange rates and money stock differentials (Huang [1981], Meese and Singleton [1980]). My intent here is to interpret the use of volatility measures in these papers, to describe some alternative models which might allow more variation in prices, and to contrast the volatility tests with more conventional methods of evaluating market efficiency.

My initial motivation for considering volatility measures in the efficient markets models was to clarify the basic smoothing properties of the models to allow an understanding of the assumptions which are implicit in the notion of market efficiency. The efficient markets models, which are described in section II below, relate a price today to the expected present value of a path of future variables. Since present values are long weighted moving averages, it would seem that price data should be very stable and smooth. These impressions can be formalized in terms of inequalities describing certain variances (section III). The results ought to be of interest whether or not the data satisfy these inequalities, and the procedures ought not to be regarded as just "another test" of market efficiency. Our confidence of our understanding of empirical phenomena is enhanced when we learn how such an obvious property of data as its "smoothness" relates to the model, and to alternative models (section IV below).

On further examination of the volatility inequalities, it became clear that the

* This research was supported by the National Science Foundation under Grant #SOC 79-07561. The views expressed here should not be attributed to the National Science Foundation or the National Bureau of Economic Research. I am indebted to Fischer Black, Olivier Blanchard, Sanford Grossman, Franco Modigliani and Jeremy Siegel for helpful discussions.

** Associate Professor, University of Pennsylvania and Research Associate, National Bureau of Economic Research.

inequalities may also suggest formal tests of market efficiency that have distinct advantages over conventional tests. These advantages take the form of greater power in certain circumstances of robustness to data errors such as misalignment and of simplicity and understandability. An interpretation of volatility tests versus regression tests in terms of the likelihood principle is offered in section V.

II. The Simple Efficient Markets Hypothesis

The structure common to the various models of market efficiency noted in the introduction can be written:

$$p_t = \sum_{k=0}^{\infty} \gamma^{k+1} E_t d_{t+k} = E_t p_t^* \tag{1}$$

where p_t is a price or yield and $p_t^* \equiv \sum_{k=0}^{\infty} \gamma^{k+1} d_{t+k}$ is a perfect foresight or ex-post rational price or yield not known at time t. E_t denotes mathematical expectation conditional on information at time t, and γ is a discount factor $\gamma \equiv 1/(1 + r)$ where r is the (constant) discount rate. Information includes current p_t, d_t, and their lagged values, as well as other variables. In the application of the basic structure (1) which shall be emphasized here, (from Shiller [1981b]) p_t for 1871–1979 is the Standard and Poor Composite Stock Price Index for January (deflated by the wholesale price index scaled to base year $T = 1979$) times a scale factor λ^{T-t} (see figure 1). The parameter λ equals one plus the long-term growth rate of

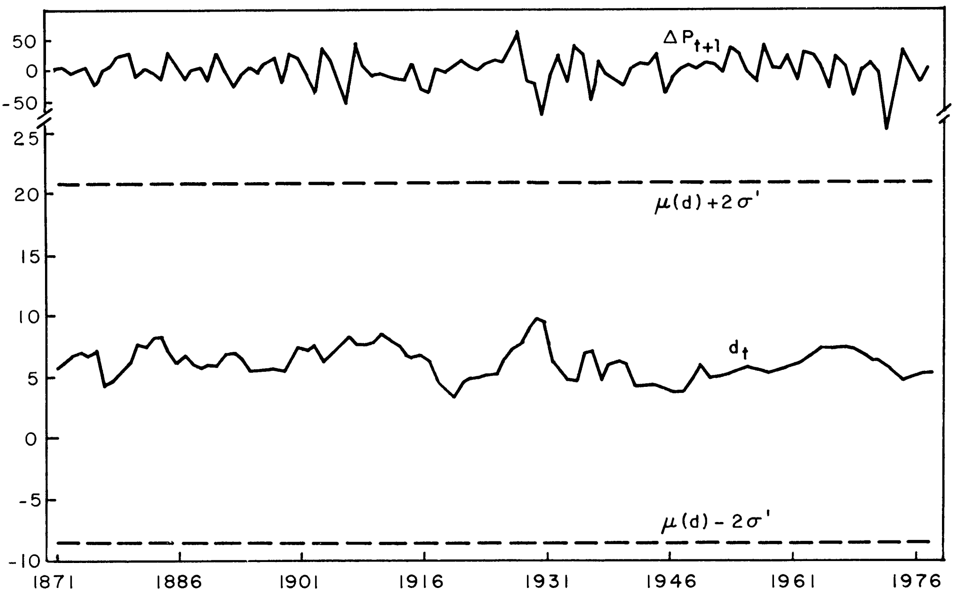

Figure 1. Upper plot: Δp_{t+1} where p_t is the real Standard and Poor Composite Stock Price Index for January scaled (to eliminate heteroscedasticity) by multiplying by $\exp(.0145(1979 - t))$. Lower plot d_t: real dividends for the p_t series times $\exp(.0145(1978 - t))$. Here, $\hat{\sigma}(d) = 1.28$. The dividend series here ranges from 3.31 to 9.62, a range of nearly 5 standard deviations. The dashed lines mark off a ± 2 standard deviation range from the mean $\mu(d)$ where the standard deviation is $\sigma' = 7.30$, or the lower bound allowed for $\sigma(d)$ by inequality (I-2), given $\sigma(\Delta p) = 24.3$. In order to justify this magnitude of $\sigma(\Delta p)$, the true $\sigma(d)$ would have to be at least this high according to (I-2).

the real value of the portfolio (1.45% per annum) and the scale factor has the effect of eliminating heteroscedasticity due to the gradually increasing size of the market. The variable d_t for 1871–1978 is the total real dividends paid over the year times λ^{T-t-1} (figure 1). The discount rate r was estimated as average d divided by average p, or 4.52% (while the average dividend price ratio was 4.92%). In LeRoy and Porter [1981] the approach is similar except that real earnings rather than real dividends are used. As is well known, one cannot claim price as the present value of expected earnings without committing a sort of double counting unless earnings are corrected for investment. Usually, such a correction means subtracting retained earnings from earnings, which would convert the earnings series into the dividend series (Miller and Modigliani [1961]). Instead, LeRoy and Porter correct both the price and earnings series by dividing by the capital stock, which is estimated as the accumulated undepreciated real retained earnings. Since the resulting series appear stationary over their shorter (postwar) sample period, they make no further scale correction.

In Shiller [1979], p_t is the yield to maturity on a long-term bond such as a consol times $\gamma/(1-\gamma)$, and corrected for a constant liquidity premium, while d_t is the one-period yield. This model is derived as an approximation to a number of different versions of the expectations model of the term structure. In Amsler [1980] the same model is applied using instead of the yield on long-term bonds her dividend price ratio series on high quality preferred stock which she argues is a proxy for a perpetuity yield. In Shiller [1981a] and Singleton [1980] p_t is related to the yield to maturity on intermediate term bonds and the model is modified slightly by truncating the summation in (1) at the maturity date.[1] In Huang [1981] p_t is the spot exchange rate and d_t is the nominal money stock differential adjusted for the real income differential between domestic and foreign countries. His model is implied by the well-known theory of the monetary approach to the balance of payments.

Although the models considered here are so simple they may be considered excessively restrictive, the econometric issues raised here are relevant to the study of alternative hypotheses as well, such as the hypotheses considered in section III below.

An implication of the structure (1) is that the innovation in price $\tilde{\Delta}_{t+1}p_{t+1}$, which is unforecastable based on information available at time t,[2] equals $\Delta p_{t+1} + d_t - rp_t$ so that:

$$E_t(\Delta p_{t+1} + d_t - rp_t) = 0. \tag{2}$$

In the context of the stock price and dividend series application, this innovation is the change in price corrected for price movements in response to short-run dividend variations (e.g., corrected for the predictable ex-dividend day movement in stock prices). That the innovation in stock prices is not very forecastable (i.e., the correlation coefficient between $\tilde{\Delta}_{t+1}p_{t+1}$ and information available is small) is certainly well known, as a result of the well-known tests of the random-walk

[1] The former paper (Shiller [1981a]) did not find any significant evidence of excess volatility with bonds maturing in less than two years.

[2] The innovation operator is defined as $\tilde{\Delta}_t = E_t - E_{t-1}$ where E_t is the expectation operator at time t.

hypothesis. For example, with our data, $\tilde{\Delta}_{t+1}p_{t+1}$ is significantly negatively correlated over 1871–1978 with p_t, but the R^2 is only .08.[3] With relatively short time series, the correlation coefficient is not statistically significantly different from zero. To detect a slight correlation statistically, one must use a great deal of data. Such results with the century of data are generally not viewed with much interest, however. It is commonly felt that such small correlations detected in long historical data are of questionable relevance to modern conditions and of minor interest given possible errors in data from remote times in history.

If we could summarize the results of the tests of the random walk hypothesis with recent data by saying that (2) held *exactly*, then, using an additional assumption that prices are not explosive, we could conclude that (1) was proven. Rearranging (2) one gets the first order rational expectations model $p_t = \gamma E_t(p_{t+1} + d_t)$. One can solve this model by recursive substitution. Substituting the same expression led for one period gives us $p_t = \gamma E_t(\gamma E_{t+1}(p_{t+2} + d_{t+1}) + d_t) = \gamma^2 E_t(p_{t+2}) + \gamma^2 E_t d_{t+1} + \gamma d_t$. One may then substitute for p_{t+2} in this expression and so on. The result is expression (1) as long as the additional requirement that $\gamma^k E_t(p_{t+k})$ goes to zero as k goes to infinity is satisfied. That the terminal condition holds in the stock price application seems quite reasonable since the dividend price ratio shows no particular trend, while on an explosive path p and d should diverge.

There is a big difference, however, between the statement that (2) holds approximately and (2) holds exactly. Expression (2) can hold approximately if movements Δp_{t+1} are so large as to swamp movements in $(rp_t - d_t)$, even if the movements in p do not reflect information about future d at all. For example, suppose stock prices are heavily influenced by fads or waves of optimistic or pessimistic "market psychology." Suppose this means that demeaned $p_t = \rho p_{t-1} + \epsilon_t$ where $0 < \rho < 1$ and ϵ_t is unforecastable noise and d is constant. If ρ is, say, .95, changes in p will not be highly autocorrelated, and hence $\tilde{\Delta}_{t+1}p_{t+1}$ as defined in (2) would not be very correlated with information (the theoretical R^2 of $\tilde{\Delta}_{t+1}p_{t+1}$ on p_t and a constant would, for $r = .045$, be .08). The fact that in short samples one would have trouble proving the existence of such correlation is a reflection of the fact that with such an alternative it takes a very long time to take effective advantage of the profit opportunity implied by the fad. It may take, say, 10 or 20 years for the fad to end, and hence one does not have the opportunity to observe this to happen very many times.

Such a fads alternative may be detectable using the volatility tests discussed in the next section. The reasons why such tests may be more powerful than a regression test will be indicated formally in section V. Before turning to these matters, however, it is important to reflect on the importance of the terminal condition discussed above. The terminal condition in (1) is essential to consider if we want to examine the question raised in the introduction to this paper: are movements in speculative prices too large, or are they justified by subsequent movements in dividends? If p_t were generated by the autoregressive process p_{t+1}

[3] This regression was added to the final draft of the paper in response to John Lintner's comments. The return $(\Delta p_{t+1} + d_t)/p_t$ is significantly positively correlated with d_t/p_t, but the R^2 is only .05.

$= (1 + r)p_t + d_t + \epsilon_{t+1}$ where ϵ_{t+1} is an unforecastable white noise then (2) would be satisfied regardless of the size of $\sigma(\epsilon)$. Thus, (2) itself is consistent with any degree of variability of prices. This autoregressive process is, however, necessarily explosive unless ϵ_t represents information about (i.e., is offset by) future d.

If we wish to derive formal tests of the restrictions, then we must represent the model in terms which allow us to derive the likelihood function of our observations and to represent the restrictions on the likelihood function implied by the model. A first step in this direction is to assume that the variables (or some transformations of the variables) are jointly covariance stationary. A random vector x_t is considered covariance stationary if the mean of x_t and the covariance between x_t and x_{t+k} are finite and do not depend on t. Our assumption that $[p_t, d_t]$ is covariance stationary thus rules out an explosive path for p_t, and incorporates the assumption of the terminal condition. The terminal condition could also be imposed by assuming $[\Delta p_t, \Delta d_t]$ is covariance stationary, since on an explosive path Δp would diverge. However, the terminal condition would not be represented by the assumption only that the proportional changes $[\Delta p/p, \Delta d/d]$ are stationary, since these could have constant variances along an explosive path for p. On the other hand, the assumption that $\Delta p/p$, $\Delta d/d$ and p/d are stationary would rule out an explosive solution to (2). Such an assumption may be more attractive than that made here, and may lead to similar results, but is less convenient analytically because it is a nonlinear transformation of data in a linear model and variance inequalities like those derived in the section below would depend on the distribution assumed for the variables.

Stationarity, and hence the terminal condition, is part of the maintained hypothesis and the motivation for the volatility tests is not that these tests will reject when p (or Δp) is explosive in the sample. Quite to the contrary, the stationarity assumption suggests tests which may reject far more decisively than a regression test even when p (or Δp) is not divergent in the sample at all, as will be seen in section V below.

III. Derivation of Variance Inequalities

Since p^* is a weighted moving average of d, one can use simple principles of spectral analysis to put restrictions on the spectrum of p^*: its spectrum must be relatively much more concentrated in the lower frequencies than the spectrum of d, and must be everywhere below the spectrum of d/r except at the zero frequency. Thus, informally, p^* series ought to be "smoother" than d series. It was shown in Shiller [1979] that the same formal characterization does not apply to the spectrum of p, yet there is a sense in which p ought to be "smooth." The inequality restrictions on the spectra of p and d implied by the model in fact concern weighted integrals of the spectra, and these inequality restrictions can be stated more simply in terms of variance inequalities.

The variance inequalities implied by the model (1) can be derived from: 1. the equality restrictions between variances and covariances imposed by the model, 2. the inequality restrictions regarding variances and covariances implied by positive semidefiniteness of variance matrices and 3. the stationarity assumption for the random processes. Several simple inequalities will be considered here. The first (used by LeRoy and Porter [1981], and Shiller [1979], [1981b], as well as Amsler [1980], Singleton [1980] and Meese and Singleton [1980]) puts a limit on the standard deviation of p in terms of the standard deviation of p^*:

$$\sigma(p) \leq \sigma(p^*). \tag{I-1}$$

The second (Shiller [1981b]) which is analogous to inequalities used in Shiller [1979], Amsler [1980], Shiller [1981a] and Singleton [1980] puts a limit on the standard deviation of Δp in terms of the standard deviation of d:

$$\sigma(\Delta p) \leq \sigma(d)/\sqrt{2r}. \tag{I-2}$$

The third, which is analogous to inequalities used in Shiller [1979] and Huang [1981] puts a limit on the standard deviation of Δp in terms of the standard deviation of Δd:

$$\sigma(\Delta p) \leq \sigma(\Delta d)/\sqrt{2r^3/(1 + 2r)} \tag{I-3}$$

the third inequality has the advantage that it applies even if p and d are integrated processes, i.e. the variance of the levels does not exist. An example of an integrated process is a random walk, though if d_t followed a random walk the model would imply a tighter bound on $\sigma(\Delta p)$.

It is trivial to prove the first inequality. The model implies that $p_t^* - p_t \equiv u_t$ is a forecast error which must be uncorrelated with p_t. Hence, $\text{var}(p_t^*) = \text{var}(p_t) + \text{var}(u_t)$. Since positive semidefiniteness requires $\text{var}(u_t) \geq 0$, it follows that $\text{var}(p_t^*) \geq \text{var}(p_t)$. The derivation of the other inequalities may proceed in much the same way, though the previous papers used a somewhat different derivation. To prove the second, note that the model implies that the innovation $\Delta p_{t+1} + d_t - rp_t$ is a forecast error which must be uncorrelated with p_t. Hence, $\text{cov}(p_{t+1}, p_t) + \text{cov}(d_t, p_t) - (1 + r)\text{var}(p_t) = 0$. Stationarity requires that $\text{var}(p_{t+1}) = \text{var}(p_t)$, and hence $\text{cov}(p_{t+1}, p_t) = \text{var}(p_t) - ½\text{var}(\Delta p_t)$. If this is substituted into the restriction we get $-r\text{var}(p_t) - ½\text{var}(\Delta p_t) + \text{cov}(d_t, p_t) = 0$, or, in terms of standard deviations and the correlation coefficient ρ between d_t and p_t, $r\sigma(p)^2 - \rho\sigma(d)\sigma(p) + ½\sigma^2(\Delta p) = 0$. Positive semidefiniteness can be interpreted as requiring all standard deviations to be positive and $|\rho| \leq 1$. One easily verifies that if $\sigma(\Delta p)$ exceeds the upper bound of the inequality (I-2) then no values of $\sigma(p)$ or ρ can be consistent with the model restriction and positive semidefiniteness. Solving the quadratic expression for $\sigma(p)$ one finds $\sigma(p) = (\rho\sigma(d) \pm \sqrt{\rho^2\sigma(d)^2 - 2r\sigma(\Delta p)^2})/2r$. If a real value of $\sigma(p)$ is to satisfy the model restriction, then the expression inside the square root operator (discriminant) must be nonnegative, $\rho^2\sigma(d)^2 \geq 2r\sigma(\Delta p)^2$ which in turn implies, since $\rho^2 \leq 1$, the inequality (I-2). To prove the third inequality, note first that if Δd_t is jointly stationary with information variables, then the model (1) implies that both Δp_t and $rp_t - d_t$ are also jointly stationary. The forecast error $\Delta p_{t+1} - (rp_t - d_t)$ must be uncorrelated with information at time t and hence with $(rp_t - d_t)$, and so the model implies the restriction $\text{cov}(\Delta p_{t+1}, (rp_t - d_t)) - \text{var}(rp_t - d_t) = 0$. Stationarity implies $\text{cov}(\Delta p_{t+1}, (rp_t - d_t)) = \{2\,\text{cov}(\Delta d_t, (rp_t - d_t)) - \text{var}(r\Delta p_t) + \text{var}(\Delta d_t)\}/2r$. One then finds, converting to standard deviations that the model restriction can be written $2r\sigma(rp_t - d_t)^2 - 2\rho\sigma(\Delta d_t)\sigma(rp_t - d_t) + r^2\sigma(\Delta p)^2 - \sigma(\Delta d)^2 = 0$, where ρ

is the correlation coefficient between Δd_t and $(rp_t - d_t)$. One can then verify that this restriction cannot be satisfied for $|\rho| < 1$ and positive standard deviations if (I-3) is violated, by solving the quadratic expression in $\sigma(rp_t - d_t)$ as with the proof of (I-2).

Intuitive interpretations of the inequalities are also possible. For the first, we require that at any point of time p_t is an unbiased forecast of p_t^* which means, roughly speaking, that p_t^* must be equally likely to be above or below p_t. If p_t had a higher variance than p_t^*, then in periods when p_t was especially high (or low) it would tend to lie above (below) p_t^* and hence in these periods would be an upwardly (downwardly) biased forecast of p_t^*. For the second, one must bear in mind that if $\sigma(\Delta p)$ is large, then either $\sigma(p)$ is very large or Δp is highly forecastable. Changes in p which are not highly forecastable would cumulate into big swings in p, while on the other hand, if p were confined to relatively narrow range then the successive movements in Δp would have to show strong negative correlation and hence be forecastable. Either way, if $\sigma(\Delta p)$ is above the upper bound in (I-2) then the excess returns $\Delta p_{t+1} + d_t - rp_t$ would be forecastable, either because Δp_{t+1} makes large predictable movements relative to dividend and price movements or because rp_t moves a lot relative to d_t and the predictable component of Δp_{t+1}. For the third, note that if $\sigma(\Delta p)$ is very large relative to $\sigma(\Delta d)$, then either $\sigma(rp_t - d_t)$ is very large or Δp_{t+1} is highly forecastable, either way the forecast error $\Delta p_{t+1} - (rp_t - d_t)$ would be forecastable.

The first two of these inequalities were compared with the Standard and Poor data described above over the 1871–1979 sample period in my paper on stock prices (Shiller [1981b]). The variable p^* was computed by recursive substitution from a terminal value of p^* equal to the average p over the sample. With this data the estimated standard deviation of p, $\hat{\sigma}(p)$, was 47.2 and $\hat{\sigma}(p^*) = 7.51$ so that (I-1) was dramatically violated by sample statistics. The $\hat{\sigma}(d)$ over the entire sample was 1.28, while $\hat{\sigma}(\Delta p)$ was 24.3. The upper bound on the standard deviation of Δp allowed by the inequality (I-2) was thus (disregarding sampling error in the estimation of standard deviations) 4.26, so that the upper bound was exceeded by almost six fold. Putting it another way, the standard deviation of d would have to be at least 7.30, which is greater than the sample mean of d, in order to justify $\sigma(\Delta p)$.[4] Figure 1 gives an impression as to how dramatically (I-2) is violated.

On the other hand (I-3) (which was not reported in that paper) was not violated by the data. Since $\hat{\sigma}(\Delta d)$ was .768 for this data, the upper bound on $\sigma(\Delta p)$ allowed by (I-3) is 59.0 which is greater than the sample $\sigma(\Delta p)$. Of course, we do not expect the data to violate all inequalities even if the model is wrong. Moreover, the choice of annual data was arbitrary, and the inequality is violated when longer differencing intervals are used. The definition of p^* implies $p_t^* = \sum_{k=0}^{\infty} \bar{\gamma}^{k+1}\bar{d}_{t+nk}$ where n is any integer greater than one, $\bar{\gamma} = \gamma^n$ and $\bar{d}_t = d_t(1 + r)^{n-1} + d_{t+1}(1 + r)^{n-2} + \cdots d_{t+n-1}$, or the accumulated dividends over n periods. Defining the n-period rate $\bar{r}$ implicitly by $\bar{\gamma} = 1/(1 + \bar{r})$, we see that for data

[4] Even though dividends cannot be negative, $\sigma(d)$ could exceed $\mu(d)$ if the distribution of d were sufficiently skewed.

sampled at n-period intervals (I-3) becomes[5] $\sigma(p_t - p_{t+n}) \leq \sigma(\bar{d}_t - \bar{d}_{t+n})/\sqrt{2\bar{r}^3/(1+2\bar{r})}$. Using n of ten years, for example, annual data from 1871 to 1969 produces $\hat{\sigma}(p_t - p_{t+10}) = 69.4$ and $\hat{\sigma}(\bar{d}_t - \bar{d}_{t+10}) = 16.5$ so that the upper bound on $\hat{\sigma}(p_t - p_{t+10})$ is 40.8 which is violated by the data.

A sampling theory for these standard deviation and formal tests of the model can be derived if we further specify the model with distributional assumptions for d_t and p_t. If Δp_t is assumed normal and serially uncorrelated, the lower bound of a χ^2 one-sided confidence interval for $\sigma(\Delta p)$ is 21.9, which is still over 5 times the maximum given by (I-2) using the sample $\sigma(d)$, so that we can safely conclude that the model requires a much bigger standard deviation for d than was observed in the sample. It is possible, however, that we could adopt distributional assumptions for d_t which would imply that the sample standard deviation of d is not a good measure of the population standard deviation, as will be discussed in the next section. Other distributional assumptions and formal tests were employed by LeRoy and Porter [1981] and Singleton [1980].

IV. Some Alternative Hypotheses

The only alternative discussed above to the efficient markets model was a "fads" alternative. Such an alternative seems appealing, given the observed tendency of people to follow fads in other aspects of their lives and based on casual observation of the behavior of individual investors. Such fads do not necessarily imply any quick profit opportunity for investors not vulnerable to fads. It is worthwhile exploring, however, alternatives more in accordance with economic theory and with rational optimizing behavior.

Another alternative to the simple efficient markets model is one in which real price is the expected present value of real dividends discounted by a time varying real discount factor $\gamma_t = 1/(1 + r_t)$:

$$p_t = E_t\left(\sum_{k=0}^{\infty} \left(\prod_{j=0}^{k} \gamma_{t+j}\right) d_{t+k}\right). \tag{3}$$

One might study this model by seeking some measure of time varying real interest rates. In one paper (Grossman and Shiller [1981]) consumption data and an assumed utility function were used to provide a measure of the marginal rate of substitution between present and future consumption. A more parsimonious approach is to treat the discount rate as an unobserved variable and ask merely how big its movements will have to be if information about future real rates accounts for the stock price movements not accounted for by divided movements according to the above analysis. Since with the stock data mentioned above the standard deviation of Δp is 18% of the mean of p, if we are to attribute most of the variance of Δp to changes in one-period real interest rates, movements in r_t must be quite large. Expression (3) effectively contains a moving average which ought to average out the movements in the one-period real interest rate. Based on a linearization which makes p equal to an expected present value of dividends

[5] It is easily verified that if p_t and d are stationary indeterministic processes this inequality approaches (I-1) as n goes to infinity.

plus a term related to a "present value" of real interest rates, it was concluded in Shiller [1981b] that the expected one-year real interest r_t in percent would have a standard deviation in excess of 5 percentage points to account for the variance of Δp.[6] This number is substantially in excess of the standard deviation of nominal short-term interest rates over the sample period and in fact implies a $\pm$ two standard deviation range of over 20 percentage points for the expected real interest rate.

Because of the very large movements in real interest rates required to explain the observed variance of Δp by this model, some have reacted by claiming that the historical standard deviations of real dividends around their trend are inherently poor measures of subjective uncertainty about dividends. One can propose alternative models for dividends for which, say, the actual $\sigma(d)$ is infinite, yet of course sample standard deviations will be finite and spurious as a measure of population standard deviations. The most commonly mentioned specific alternative hypothesis is that d_t is an unforecastable random walk with normal increments. We saw in the preceding section that even such a model, where $\sigma(\Delta d)$ is finite, can be subjected to volatility tests. In the particular case of an unforecastable random walk, $p_t = d_t/r$ and $\sigma(\Delta p) = \sigma(\Delta d)/r$. With our data using one year differencing $\hat{\sigma}(\Delta p) = 24.3$ and $\hat{\sigma}(\Delta d)/r = 17.0$ while with ten year differencing (as described above) $\hat{\sigma}(p_t - p_{t+10}) = 69.4$ and $\hat{\sigma}(\bar{d}_t - \bar{d}_{t+10})/\bar{r} = 29.6$. Prices appear to be too volatile for either differencing interval, the dividend ratio is not constant but moves so as to cause p to be relatively more volatile than dividends. One should also ask whether such an alternative hypothesis is plausible. Since $\hat{\sigma}(\Delta d)$ is only 8 times the sample mean, this hypothesis would imply that aggregate dividends often stood a substantial chance of hitting zero in a matter of years (and then forecasted dividends would also be zero), and that the apparent resemblance of d_t to an exponential growth path over the century[7] was due to chance rather than the physical process of investment and technical process. It seems that a more plausible alternative hypothesis should be sought. Other integrated processes (i.e., processes stationary in Δd but not d) do not look promising, given that I-3 appears violated. There are, however, certainly hypothetical forms for the d process which imply that the volatility tests are incorrect, the question is whether such alternatives are plausible.

In discussions about plausible alternative hypotheses the hypothesis has been repeatedly suggested to me that the market may indeed be legitimately concerned with a major disaster with low probability each period which would, let us say, completely destroy the value of all stock. The fact that such a disaster did not occur in the last century in the U.S. and that dividends roughly followed a growth path does not prove that disaster wasn't a distinct possibility. Such disasters might include nationalization, punitive taxation of profits, or conquest by a

[6] An alternative approach to the same issues is used by Pesando [1980].

[7] A regression of $\log(d_t)$ on a constant, $\log(d_{t-1})$ and time for 1872–1978 gives $\log(d_{t-1})$ a coefficient of .807 with an estimated standard deviation of .058. This regression would always forecast that log d_t should return half way to the trend in three years. According to Table 8.5.1 in Fuller [1976], if $\log(d_t)$ were a Gaussian random walk the probability in a sample of this size of obtaining a smaller coefficient for $\log(d_{t-1})$ is .05.

foreign power. Similar events are not uncommon by world standards, and my selection of the long uninterrupted U.S. stock price series may involve a selection bias. In the context of foreign exchange market efficiency studies, the possibility of such a disaster has been called the "Peso problem," referring to the fluctuation in the peso forward rate in anticipation of a devaluation that did not occur in the sample period, Krasker [1980].

To evaluate this possibility, suppose that the probability of a disaster ("nationalization") *during* period t as evaluated at the beginning of period t is given by a stochastic process π_t, $0 \leq \pi_t \leq 1$, and suppose π_t is independent of an underlying dividend process d_t in the absence of disaster. If the disaster occurs, the stock is worthless and d_t will not be received. This means that the expected value as of time t of the dividend *received* at time $t + k$ is $E_t[(\prod_{j=0}^{k} (1 - \pi_{t+j}))d_{t+k}]$ where $E_t \prod_{j=0}^{k} (1 - \pi_{t+j})$ is the probability that no disaster occurs between time t and $t + k$. Thus, the model (1) should be modified as:

$$p_t = E_t \sum_{k=0}^{\infty} (\prod_{j=0}^{k} (\gamma(1 - \pi_{t+j})))d_{t+k}. \tag{4}$$

This model is identical to the model (3) if we set $\gamma_{t+j} = \gamma(1 - \pi_{t+j})$, and collapses to our basic model (1) if the probability of disaster is constant, except that our estimated γ ought to be interpreted as the estimate of $\gamma(1 - \pi)$. Clearly, then, the disaster model can explain the volatility of stock prices only if the probability of disaster changes substantially from period to period. Indeed, using the earlier conclusion about real interest rate variation the standard deviation of the probability that disaster will occur *within the current year* must have exceeded .05 if movements in p_t are to be attributed to new information about current and future π. One wonders whether the probability of disaster in a single year could have changed so much. At the very least the theory relies very heavily on the selection bias for its plausibility since such a standard deviation for π_t implies π_t must often be high. If the probability of disaster is, say, .1 each year then the probability that no disaster occurred over 108 years is roughly 1 in 100,000.

V. Regression Tests of the Model

The most obvious way to test the model (1) as noted above is to run a regression to see whether innovations are forecastable. Such a procedure would be to regress $y_t = (p_{t+1} + d_t)$ on a constant term and independent variables p_t and d_t. The coefficient of p_t is, by our model, $(1 + r)$ (which is greater than one) and the constant term and coefficient of d_t are zero. One could perform an F-test on these coefficient restrictions. Why, one might ask, ought one to examine the volatility measures as a way of evaluating the model?

There are potentially many ways of justifying a test other than a regression test. For example, one may justify it by claiming special interest in an alternative hypothesis (e.g., a "fads" alternative) implying excess volatility, or one may justify it on the basis of robustness or simplicity. Here, however, I will emphasize a sort of data alignment problem for the motivation. Since the distributed lead in (1) is infinitely long, movements in p may be due to information about dividends in the indefinite future. However, we have data about dividends only for the relatively short sample period. The price movements corresponding to information regarding the dividends in the sample period may have occurred before our sample began. The price movements during the sample may reflect information about dividends beyond the sample. Of course, one might doubt that dividends beyond the sample ought to be so important. This, however, is just the point of the volatility tests, a point in effect missed by the regression tests.

For the purpose of motivating the discussion below, consider the simplest example of a rational expectations model in which $x_t = E_t y_t$ and the maintained hypothesis is that $[x_t, y_t]$ is jointly normal with zero mean and successive observations are independent. Suppose the data are *completely* misaligned, and there are T observations on x_t (for 1950–60, say) and T nonoverlapping observations on y_t (for 1970–80, say). Clearly, it is not possible to test the model by regressing y_t on x_t. It does not follow that the model cannot be tested. The likelihood function is then

$$l(\sigma_x, \sigma_y \mid X, Y) = (2\pi\sigma_x\sigma_y)^{-T}\exp\left(-\frac{\sum x^2}{2\sigma_x^2} - \frac{\sum y^2}{2\sigma_y^2}\right).$$

The model imposes only one restriction on this likelihood function, namely that $\sigma_x \leq \sigma_y$. The likelihood ratio test statistic is then:

$$\lambda = \begin{cases} 1 \quad \text{if} \quad \hat{\sigma}_x \leq \hat{\sigma}_y \\ \left(\dfrac{\sqrt{\hat{\sigma}_x^2\hat{\sigma}_y^2}}{(\hat{\sigma}_x^2 + \hat{\sigma}_y^2)/2}\right)^T \quad \text{if} \quad \hat{\sigma}_x > \hat{\sigma}_y \end{cases}$$

where $\hat{\sigma}$ denotes sample standard deviation about zero. The likelihood ratio test which is the unique uniformly most powerful test with this data thus works out to be a volatility test. Since if $\sigma_x = \sigma_y$, the distribution of $-2 \log \lambda$ is asymptotically a mixture of a χ^2 with one degree of freedom and a spike at the origin, one can easily find an approximate 95% critical value for λ for a given level of significance. The test may enable us to reject the model with confidence. If $\hat{\sigma}_y = 0$ and $\hat{\sigma}_x \neq 0$, $-2 \log \lambda$ is infinite, so that the hypothesis would be rejected at any significance level. The power of the test approaches one as σ_y/σ_x approaches zero.

In contrast, if the data were perfectly aligned, the likelihood ratio test would amount to a t test that the coefficient is 1.00 in a regression of y on x. If the data were only partially aligned, the likelihood function would be the product of a likelihood function of the aligned observations and the likelihood of the unaligned observations. Regression tests with the overlapping observations would be suggested by looking only at the first factor, volatility tests by looking only at the second, while the true likelihood ratio test would be a hybrid of the two. Reliance on the likelihood of the unaligned observations alone (or, more properly, treating all observations as if unaligned) might be justified if data alignment is considered inaccurate, or if a simple and easily applied test is desired.

The structure (1) is a little different from this simple example since the expectations variable p_t relates to an infinite weighted moving average of the expected variable partly within the sample and partly without, and successive observations are not independent. Since the purpose is exposition here, let us assume that the data consists of T *independent* normal observations of the vector

$z_t = [p_{\tau_t+1}, p_{\tau_t}, d_{\tau_t}]$. Imagine that, for some reason, we have observations on this vector so widely spaced in time (e.g., $z_1 = [p_{1801}, p_{1800}, d_{1800}]$, $z_2 = [p_{1901}, p_{1900}, d_{1900}]$, $z_3 = [p_{2001}, p_{2000}, d_{2000}]$, $\cdots$) that we regard separate observations as independent. Our analysis of that data set can be extended routinely to the case where observations are not widely spaced and p_t and d_t are parameterized as autoregressive or moving average processes, but this analysis would be messier so is omitted here. A discussion of this more general case in the context of the expectations model of the term structure of interest rates is in Shiller [1981a]. In that paper it is shown that the likelihood ratio test employed by Sargent [1979] to test the expectations model of the term structure does not impose the kind of restrictions implied by the model which are considered here.

The likelihood function for the T observations on the vector z_t can be written in the usual form: $l(\mu, \Omega \mid z) = (2\pi)^{-3T/2} \mid \Omega \mid^{-t/2} \exp(-\tfrac{1}{2} \sum_{t=1}^{T} (z_t - \mu)\Omega^{-1}(z_t - \mu)')$ where z is the $T \times 3$ matrix of observations, μ is the 1×3 vector of means and Ω the 3×3 covariance matrix. Since Ω is symmetric, the model has nine parameters: the three means, the three variances, and the three covariances. However, because of the stationarity assumption $E(p_t) = E(p_{t+1})$ and $\operatorname{var}(p_t) = \operatorname{var}(p_{t+1})$ there are only 7 independent parameters, a fact we shall consider below.

Given the vector z_t, we can by a linear transformation derive the vector $w_t = [p_{t+1} + d_t, p_t, d_t]$ which has as its first element the dependent variable in the regression described above and the same second and third elements. Since z_t is normally distributed, w_t is also. We can therefore write the likelihood function for w_t by a change of variables from the likelihood function (of z_t). This likelihood function can be written in various ways. For our purposes, it is convenient to write the likelihood function for w in a factored form which makes it easy notationally to consider regression tests of the model. Partitioning w_t into $w_t = [w_{1_t}, w_{2_t}]$ where w_{1_t} is a 1×1 and w_{2_t} is a 1×2, and letting w, w_1 and w_2 denote the matrices of observations or orders $T \times 3$, $T \times 1$ and $T \times 2$ respectively, and defining a T element vector Y, $Y = w_1$ of dependent variables and a $T \times 3$ matrix X, $X = [L \; w_2]$ of independent variables where L is a column vector of 1's, then the likelihood function is:

$$
\begin{aligned}
l(\sigma^2, \beta, \Phi, \mu \mid w) &= l(\sigma^2, \beta \mid w_1, w_2) l(\Phi, \mu \mid w_2) \\
&= (2\pi\sigma^2)^{-T/2} \exp(-(Y - X\beta)'(Y - X\beta)/2\sigma^2) \\
&\quad \cdot (2\pi)^{-T} \mid \Phi \mid^{-T/2} \exp(-\textstyle\sum_{t=1}^{T} (w_{2_t} - \nu)\Phi^{-1}(w_{2_t} - \nu)'/2) \qquad (5)
\end{aligned}
$$

where σ^2 is the variance of the residual in a theoretical regression of Y on X, β is the 3 element vector of theoretical regression coefficients of Y on X, ν is the 2 element vector of means of p and d and Φ the 2×2 variance matrix for the vector w_{2_t}. The nine parameters of the multivariate normal distribution for z_t are here transformed into the 3 elements of β, the 2 elements of μ, the two diagonal and one off-diagonal element of Φ, and σ^2. The first factor in the likelihood function is the usual expression for the likelihood of a regression of Y on X, while the second factor is the usual expression for the likelihood of a bivariate normal vector.

The factorization (5) of the likelihood function then makes it transparent why, if we disregard the requirement that $E(p_t) = E(p_{t+1})$ and $\text{var}(p_t) = \text{var}(p_{t+1})$, it is appropriate to test the model by regressing the variable $(p_{t+1} + d_t)$ on information available at time t, i.e., on the constant, p_t and d_t, and testing the restrictions on the regression coefficient β. The efficient markets model restricts only the first factor of the likelihood function. The likelihood ratio test statistic is the ratio of the maximized unconstrained likelihood to the maximized constrained likelihood. The maximum likelihood is the product of the maxima of the two factors and so the second factor will be the same in both numerator and denominator and hence will cancel out, leaving us with (a transformation of) the ordinary regression test statistic. A Bayesian posterior odds procedure will similarly produce the posterior odds of the simple regression of Y on X so long as there is no prior dependency between the parameters of the two factors.

In fact, however, there is a constraint across the coefficients in the two factors due to the fact that $E(p_t) = E(p_{t+1})$ and $\text{var}(p_t) = \text{var}(p_{t+1})$. It was emphasized above that it is these constraints which represent the requirement that price movements be justified in terms of future dividends. One can rewrite these constraints in terms of the parameters of the likelihood of z_t. The mean restriction becomes $\beta_1 = (1 - \beta_2)\mu_p - \beta_3\mu_d$ and the variance restriction becomes $\sigma^2 = (1 - \beta_2^2)\text{var}(p) + 2\beta_2(1 - \beta_3)\text{cov}(p, d) - (1 - \beta_3)^2\text{var}(d)$. With these restrictions across the two factors of the likelihood function the second factor is no longer irrelevant to the likelihood ratio, and regression tests are no longer appropriate. If we make minus two times the log likelihood ratio our test statistic, then the power of the test approaches one as $\sigma(d)$ goes to zero so long as $\sigma(p)$ does not equal zero. If $\sigma(d)$ equals zero, the above variance restriction reduces to $\sigma^2 = (1 - \beta_2^2)\text{var}(p)$. Under the null hypotheses, $\beta_2 = 1 + r > 1$, so this restriction cannot be satisfied unless p does not vary, which is rejected with probability one if $\sigma(p) > 0$. In this case, as in our simple example with nonoverlapping observations, the sample will give $-2 \log \lambda = \infty$, while, whether or not p appears explosive in the sample $-2 \log \lambda$ for the regression test will not be infinite. The likelihood ratio test thus has power of one in a region of the parameter space where the regression test does not. In this region, the volatility test, using (I-2) and a test statistic $\hat{\sigma}(\Delta p)/\hat{\sigma}(d)$ also has power equal to one. Thus, if volatility tests reject decisively and regression tests do not, one may infer that in fact the parameters are likely to lie near this region.

VI. Conclusion and Summary

The various papers discussed here, which attempted to provide evidence against simple efficient markets models by showing that prices are too volatile, share a common econometric motivation. By assuming that both p_t (real stock prices in our main example) and d_t (real dividends), or some transformations of these, are stationary stochastic processes, it is possible to inquire whether the movements in p_t are "too big" to be explained by the model. With these assumptions, the conventional regression tests of the model are no longer suggested by the likelihood ratio principle and volatility tests have distinctly more power in certain regions of the parameter space. In the papers discussed here, the movements in p_t indeed appear to be too big. Figure 1 illustrates how dramatically our stock price and dividend data appear to violate one of the inequalities, (I-2). Roughly speaking, if the efficient markets model holds, and if the large movements of Δp are to be justified in terms of information about future d, then the d_t series ought to range over an area equal to the space between the dashed lines, while in fact it appears confined to a fraction of the area. Commonly used hypotheses that have been advanced to reconcile this data with efficient markets: a random walk model for d_t or a disaster theory, do not appear to be promising. Other possibilities are that ex ante real interest rates show very large movements or, alternatively, that markets are irrational and subject to fads.

REFERENCES

Amsler, Christine, "The Term-Structure of Interest Rates in an Expanded Market Model," unpublished Ph.D. dissertation, University of Pennsylvania, 1980.

Fuller, Wayne A., *Introduction to Statistical Time Series*, John Wiley & Sons, New York, 1976.

Grossman, Sanford and Robert Shiller, "The Determinants of the Variability of Stock Prices," *American Economic Review*, May 1981.

Huang, Roger D., "Efficiency and the Volatility of Floating Exchange Rates," forthcoming, *Journal of Finance*, 1981.

Krasker, William, "The Peso Problem in Testing the Efficiency of Forward Markets," *J. Mon. Econ.*, 6, 1980.

LeRoy, Stephen and Richard Porter, "The Present Value Relation: Tests Based on Implied Variance Bounds," *Econometrica*, March 1981.

Meese, Richard A. and Kenneth J. Singleton, "Rational Expectations, Risk Premia and the Market for Spot and Forward Exchange," mimeographed, Board of Governors of the Federal Reserve System, 1980.

Miller, Merton H. and Franco Modigliani, "Dividend Policy, Growth and the Valuation of Shares," *Journal of Business*, 411–433, October 1961.

Pesando, James E., "On Expectations, Term Premiums and the Volatility of Long-Term Interest Rates," unpublished paper, University of Toronto, 1980.

Sargent, Thomas J., "A Note on the Maximum Likelihood Estimation of the Term Structure," *J. Mon. Econ.*, 5: 133–143, January 1979.

Shiller, Robert J., "Alternative Tests of Rational Expectations Models: The Case of the Term Structure," forthcoming, *Journal of Econometrics*, 1981. (a)

Shiller, Robert J., "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", forthcoming, *American Economic Review*, June 1981. (b)

Shiller, Robert J., "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy*, 87: 1190–1219, December 1979.

Singleton, Kenneth, "Expectations Models of the Term Structure and Implied Variance Bounds," forthcoming, *Journal of Political Economy*, 1980.

DISCUSSION

JOHN B. LONG, JR.*: Cohn and Lessard present some international evidence on the joint behavior of price-earnings ratios, nominal interest rates, and inflation. For most of the countries examined, their evidence is qualitatively similar to evidence for the United States presented in 1979 by Modigliani and Cohn. In a broader context, the regularities reported by Cohn and Lessard are apparently

* University of Rochester.