

Chapter 14

Volatility

Stephen F. LeRoy

*Curtis L. Carlson School of Management, University of Minnesota, Minneapolis, MN 55455,
U.S.A.*

Douglas G. Steigerwald

Department of Economics, University of California, Santa Barbara, CA 93106, U.S.A.

1. Introduction

Until the mid-1970s, capital market efficiency was tested using what are now known as returns tests. In these tests the analyst sought to determine whether information currently available to investors is correlated with future asset returns. For example, past returns, detailed firm data and such macroeconomic variables as GNP were all considered as possible explanatory variables in a linear model of equity returns. If nonzero correlations were found, one inferred that the current price could be improved upon as a predictor of future prices, suggesting that capital markets are not informationally efficient.

If capital markets are not informationally efficient in this sense then (under additional assumptions) investors who trade actively to exploit these correlations can hope to do better on average than investors who buy and hold. If, however, markets are informationally efficient, then prices ‘fully reflect’ available information and there are no exploitable correlations between future returns and current information. In the absence of exploitable correlations, active trading rules cannot succeed on average.

Most of the evidence accumulated prior to the 1980s implied that asset markets are informationally efficient, at least as a first approximation. Fama’s influential [1970] paper surveyed the empirical evidence available up to that date and established the conclusion in favor of market efficiency to most academic readers’ satisfaction (see Fama 1991 for an update).

LeRoy & Porter [1981] and Shiller [1981], working independently, proposed an alternative test of market efficiency. This test exploited Samuelson’s [1965] demonstration that, in the context of the stock market, the present-value relation is equivalent (subject to a convergence assumption) to the null hypothesis tested in the returns tests. The present-value relation states that the current actual stock price is the best predictor of the discounted value of future actual dividends, referred to as the ex-post rational price. More precisely, the present-value

relation states that actual current price equals the mathematical expectation of ex-post rational price conditional on whatever information investors have available:

$$p_t = E(p_t^* | I_t), \quad (1.1)$$

where

$$p_t^* \equiv \sum_{i=1}^{\infty} \beta^i d_{t+i}. \quad (1.2)$$

This version of the market efficiency hypothesis suggested to LeRoy & Porter and Shiller a strategy for empirical testing which was different from the returns tests surveyed by Fama. If price equals discounted expected dividends, then price changes occur only when expected dividends change. Expected dividends, unfortunately, are not easily measured. However, under rational expectations the volatility of changes in expected dividends is related systematically to the volatility of changes in actual dividends, which can be estimated. LeRoy & Porter and Shiller exploited this line of reasoning to derive the bounds on stock price volatility implied by measured dividend volatility. The simplest variance bounds relation is

$$V(p_t) \leq V(p_t^*), \quad (1.3)$$

which follows from (1.1) plus the result from probability theory that the variance of the conditional expectation of a random variable is less than or equal to the variance of the random variable itself. LeRoy & Porter and Shiller found that these bounds were violated, suggesting that the common journalistic observation that stock price changes are in some sense excessive relative to fundamentals may have substance.

Returns tests appeared to accept market efficiency, volatility tests to reject it. Both LeRoy & Porter and Shiller suggested a possible explanation for this discrepancy: volatility tests have greater power than returns tests. To be sure, Shiller provided little, and LeRoy & Porter nothing, by way of specific support for this claim. The subsequent variance-bounds literature has returned periodically to the question of power, particularly in comparing volatility tests with conventional returns tests of market efficiency. The results achieved so far have not been conclusive. We believe that this concern with statistical power is not misplaced: power is the central concept in analyzing econometric aspects of the variance-bounds tests. The related topic of bias in parameter estimation, which has been the main focus in much discussion of econometric aspects of the variance-bounds literature, is in fact subsidiary: bias causes problems only insofar as it reduces the power of a test (for given size), and there is generally no presumption that this occurs.

In Sections 2–4 of this paper we survey the variance-bounds literature and the parallel evolution of the literature on returns tests of market efficiency, taking power as the organizing principle. In Section 5 we present some new results on

the comparative power of volatility and returns tests. Section 6 is the conclusion.

We recognize that some readers — particularly those who want to get the flavor of the variance-bounds literature without delving deeply into its econometric side — may prefer a more general treatment. Such readers should consult LeRoy [1984] and West [1988]. LeRoy [1989] discussed the variance-bounds literature in its relation to the general question of asset market efficiency. Shiller's [1989] book gave introductory treatment, as did Cochrane's [1991] review. Finally, Gilles & LeRoy's [1991] survey provided an elementary econometric survey of the variance-bounds literature, as here, but focused on bias in parameter estimators rather than power.

2. Variance-bounds tests of capital market efficiency

2.1. Statistical hypothesis testing

It became clear at an early stage of the development of the variance-bounds literature that there is a trivial sense in which variance-bounds tests — some, at least — are in fact more powerful than returns tests. To explain this, it is necessary first to review the basics of statistical hypothesis testing.

Whenever a statistical test is constructed, inferences made about the null or alternative models are probabilistic rather than deterministic. There is virtually always some probability that even if the true data-generating process is the null model, the particular data realization sampled by the researcher leads to rejection of the null model. Correspondingly, there is generally some chance that even though the null model is incorrect, the data under observation conform to it so closely that the model cannot be rejected. These ideas lead directly to the concepts of size and power.

A test of the null hypothesis is based upon the estimated values of certain parameters in the model. The *critical region* denotes the range of estimated parameter values for which the null model is rejected. The *size* of the test is the probability that the parameter estimate lies in the critical region when the null hypothesis is true, so that the null model is incorrectly rejected. The *power* of the test is the probability of the same event when the alternative model is true, so that the null model is correctly rejected.

One wants a test that is powerful against all relevant alternative models. This can be achieved simply by specifying a large critical region. Unfortunately, specifying a large critical region also increases the size of the test. Because of this relation between size and power, there is a trivial sense in which any test can be made to have high power: enlarge the critical region. Such a test will produce the correct answer with high probability when the null hypothesis is false, but only because the test has a high probability of rejecting the null whether it is true or false.

It follows that in comparing the power of two econometric tests, size must be held constant. Otherwise the comparison is pointless.

2.2. Size and power in variance-bounds tests

Let us see how these considerations bear on the variance-bounds tests. The simplest and best-known of the variance-bounds theorems, introduced as (1.3) above, says that the variance of actual stock prices p_t is bounded above by the variance of the ex-post rational price p_t^* . One natural way to test (1.3) is to examine the volatility statistic, defined to be the difference between the estimated variance of p_t^* and that of p_t . There are several problems involved in constructing the statistic, of which the most serious involves trend correction. If p_t and p_t^* have not been corrected for trend the population means and variances are time-dependent, implying that population variances cannot be consistently estimated by sample variances. The first-generation variance-bounds papers of LeRoy & Porter and Shiller ran into difficulty with trend-correction. Subsequent papers, however, have remedied this problem [Campbell & Shiller, 1988; LeRoy & Parke, 1992]. This material is reviewed in Gilles & LeRoy [1991]; repetition here would divert attention from the topic of power, so we simply assume that a satisfactory trend-correction algorithm has been implemented, implying that p_t and p_t^* are stationary.

It is known that $V(p_t)$ is accurately estimated by its sample variance for reasonable sample sizes and specifications of investors' information sets, but $V(p_t^*)$ is more problematic. The fact that p_t^* depends on dividends beyond any finite horizon implies that p_t^* is unobservable in any finite sample. Grossman & Shiller [1981] resolved this problem by defining an observable proxy $p_{t|T}^*$ for the unobservable p_t^* by setting the terminal value $p_{T|T}^*$ equal to p_T and deriving earlier values from the recursion

$$p_{t|T}^* = \beta(p_{t+1|T}^* + d_{t+1}), \quad (2.1)$$

an implication of (1.2). They estimated $V(p_t^*)$ by taking the sample variance of $p_{t|T}^*$.

It happens that the estimator of $V(p_t^*)$ just described is severely downward-biased in small samples [Flavin, 1983; Kleidon, 1986; see Gilles & LeRoy, 1991, for an exposition]. This implies that the mathematical expectation of the volatility statistic is smaller than the corresponding population value, and in fact may be negative even if the present-value model is true. If the critical region associated with a test of the null hypothesis that the present-value model is correct consists of all negative values of the volatility statistic, then the downward bias just described ensures that the size of the test will be large. Hence even if (1.3) holds in the population our sample estimators will typically indicate excess volatility (the inequality in the sample counterpart of (1.3) is reversed).

It has become common to conclude from such reasoning that variance-bounds tests are biased toward rejection, and therefore are not credible. This reasoning is incorrect, as the choice of zero as a critical value for the volatility statistic is entirely arbitrary. To the extent that $V(p_t^*)$ is estimated with downward bias, the probability of incorrectly rejecting the present-value model is increased, but since the size of the test was unspecified in the first place, it is difficult to understand in

what sense we have a biased test. It would be more accurate to conclude that in the absence of some means of controlling the size, we have no real test at all. This is true whether or not point estimates of parameters are biased.

It is now easy to see the trivial sense in which variance-bounds tests may have higher power than conventional returns tests. The downward bias in the estimate of $V(p_t^*)$ just described means that if the rejection region is specified to be the negative values of the volatility statistic, the variance-bounds tests may be prone to reject the null of market efficiency whether it is false or not. This is a point against variance-bounds tests if market efficiency is in fact true, but is a point in their favor if it is false. Thus by adopting a biased estimator of $V(p_t^*)$, we have increased the power of the test at the cost of increasing its size. This does not provide a useful benchmark for comparing variance-bounds tests with returns tests since we could increase both the size and power of a returns test by enlarging its critical region. A claim for the superiority of variance-bounds tests can be based only on a showing that they are more powerful than returns tests for given size, and a claim for their inferiority must demonstrate the opposite. Nothing in the discussion so far shows either point.

To construct a true econometric test based on the volatility statistic and to compare this test with conventional market efficiency tests, it is necessary to determine how to choose the critical value of the test statistic so that the size is held to some preassigned level. That the estimator of $V(p_t^*)$ constructed as described above is biased is not necessarily a problem, even if the bias is difficult to evaluate analytically. Assuming that Monte Carlo methods are used to construct critical values, then as long as the estimate of $V(p_t^*)$ is constructed in the same way on the artificially generated data as on real-world data, the bias is automatically allowed for in the choice of the critical value.¹

2.3. *The nuisance parameter problem in testing variance bounds*

The fact that variance-bounds tests are tests of an inequality causes a major problem. The efficient markets model makes no restriction on investors' information sets; equivalently, inequality (1.3) holds no matter how much or little information investors have. Thus there are many versions of the null hypothesis, indexed by a measure of investors' information. This causes a problem in setting the critical value of the volatility statistic: different specifications of investors' information lead to different critical values. The better the information investors have about future dividends, the greater will be the volatility of stock prices under the efficient markets model, and therefore the lower will be the value of $V(p_t^*) - V(p_t)$ that is implied by market efficiency. The fact that the population value of $V(p_t^*) - V(p_t)$ associated with the present-value model depends on the

¹ This point lay behind Shiller's [1988] reply to Kleidon's [1986] criticism: Shiller granted that his estimator of $V(p_t^*)$ is biased under the conditions assumed by Kleidon, but used simulation results to argue that even so the real-world value of the volatility statistic was too far below zero to be consistent with market efficiency.

extent of investors' information suggests that the same will be true of the critical value of the associated sample statistic. (This is not necessarily the case, however, because the sample variability of the volatility statistic also depends on the extent of investors' information, which complicates the picture). The Monte Carlo simulations reported in LeRoy & Parke [1992] imply that for reasonable parameter values the critical level of the volatility statistic does indeed depend strongly on the extent of investors' information.

The question becomes how one sets the critical value of the volatility statistic, given the dependence of its distribution on a nuisance parameter. One potential solution is to vary the amount of information agents are assumed to have, and then choose the critical value so that the maximum size of the test is held to some preassigned level like 5%. The problem is that in the present setting the nuisance parameter problem is so severe that for most versions of the null hypothesis rejection will occur with probability much lower than 1% when the nominal probability of Type I error is set at 5%. With the acceptance region set so large, the test will have very low power. There is no way to avoid the fact that when a severe nuisance parameter problem occurs, the size and power of any statistical test are essentially impossible to evaluate no matter how one chooses the rejection region.

This argument, presented in LeRoy & Parke [1992], implies that hypothesis testing of the variance-bounds inequality is essentially impossible. The point was not noticed earlier, and is still not widely appreciated, because of the practice prevailing in the variance-bounds literature (discussed above) of implicitly identifying the critical value of the volatility statistic with zero — so that the present-value model is accepted if $\hat{V}(p_t) \leq \hat{V}(p_t^*)$ and rejected in the opposite case — despite the absence of any justification for this identification. We already saw that if the critical value of the volatility statistic is set equal to zero and if a downward-biased estimate of $V(p_t^*)$ is used, rejection of market efficiency is to be expected, but these tests have a large size. More recent papers, such as Cochrane [1992], use better estimates of $V(p_t^*)$ and fail to reject the efficient market model based on the sample counterpart of the variance-bounds inequality. However, these papers do not take cognizance of the nuisance parameter problem.

2.4. Summary

Thinking about statistical power in the context of variance-bounds tests leads to serious questions about their interpretation: it is pointless to compare the power of variance-bounds and returns tests without holding size constant. But because the probability distribution of the volatility statistic depends on the extent of investors' information, which is unrestricted under market efficiency, there is no way to hold constant the size of a variance-bounds test.

The practice in the variance-bounds literature has been to reject market efficiency when the sample counterpart of the variance inequality is not satisfied. This arbitrary choice of rejection region has the obvious implication that if a downward-biased estimator of $V(p_t^*)$ is used, and if investors have considerable information

about future dividends (so that $V(p_t)$ is high under the null hypothesis) then the test is likely to have a large size and, correspondingly, high power. If an unbiased estimator of $V(p_t^*)$ is used, the size is smaller, but the power is lower as well. It appears that little is learned by actually conducting such tests: there is no point in verifying that some particular variance-bounds test rejects or fails to reject market efficiency if one has no way to evaluate the probabilities of these outcomes under the null hypothesis of market efficiency or under some alternative.

3. Returns tests

Defenders of efficient capital markets rejected the contention that the variance-bounds tests demonstrated the excess volatility of asset prices, and the discussion of the preceding section suggests that they had some justification for doing so. The main problem, as we saw, was that nothing can be concluded from the fact that point estimates of $V(p_t)$ and $V(p_t^*)$ reverse the variance-bounds inequality in the absence of a showing that such an outcome would be unlikely if the present-value model were correct.² Further, the critical value of the volatility statistic depends upon a nuisance parameter that reflects how much information agents have. Because the present value model does not specify the amount of information agents have, it appears that there is no way to remedy this problem and draw valid inferences from variance-bounds tests.

However, just as researchers were dismissing the conclusion of variance bounds tests in favor of those of returns tests, and thereby reaffirming the earlier conclusion in favor of the present-value model, the latter tests were undergoing a major reappraisal. Fama & French [1988] and Poterba & Summers [1988] reexamined the autocorrelation of rates of return on stock. Unlike the earlier papers discussed in Fama [1970], in which daily and weekly returns were used, Fama & French and Poterba & Summers looked at autocorrelation of returns averaged over months and years. They found that over these long horizons returns are significantly autocorrelated. The average return from $t - T$ to T predicts about 35% of the variation of the average return from t to $t + T$, for T on the order of three to five years. The correlation is negative, so that low returns in the past predict high returns in the future.

The finding that stock returns are autocorrelated at long horizons parallels the excess volatility findings: many of the alternatives to the present-value model that would produce negatively autocorrelated returns would also generate excess volatility. For example, investor overreaction to relevant information would produce both negative return autocorrelation and excess price volatility. Alternatively,

² LeRoy & Porter [1981] presented evidence that the rejections implied by the point estimates of volatility parameters were of borderline statistical significance at conventional levels, based on asymptotic distributions. However, because of bias induced by faulty trend correction this evidence was of questionable validity. Also, a showing of small-sample bias raised questions about the applicability of asymptotic distribution theory.

suppose that stock prices are modeled as the sum of the value implied by the present-value relation and a noise term which is independent of fundamentals. Again, both negatively autocorrelated returns and excess volatility would result.

The finding that stock prices contain a mean-reverting component, like the finding of excess volatility, has been subjected to criticism. Kim, Nelson & Startz [1991] contended that much or all of the evidence of mean-reversion disappears if data from the 1930s are dropped from the sample (however, this finding has been questioned by Cogley [1991]). Similarly, Richardson [1993] took issue with Fama & French's analysis on econometric grounds.

The contrasting outcomes of variance-bounds and returns tests reported in the introduction appears now to be reversed. The original variance-bounds tests rejected the present-value model, whereas the earlier returns tests accepted it; the upshot of the discussion so far, however, is that the best evidence rejecting the present-value model comes from returns tests, not variance-bounds tests.

4. Orthogonality tests

4.1. Introduction

In criticizing volatility tests, we followed precedent in identifying these with the bounds test

$$V(p_t) \leq V(p_t^*), \quad (4.1)$$

the simplest and best-known volatility implication of the present-value relation. We saw that inequality (4.1) is essentially untestable because the present-value relation leaves investors' information unrestricted, rendering it impossible to set critical values for the rejection region. However, bounds tests are not the only kind of volatility test; we also have orthogonality tests. These, although less familiar, are better-behaved econometrically than bounds tests: as we will see, they are less subject than bounds tests to the nuisance parameter problem discussed in Section 2.

The simplest orthogonality test is derived by iterating the definition of returns (as the sum of dividends plus capital gain) to obtain

$$p_t^* = p_t + \sum_{i=1}^{\infty} \beta^i (p_{t+i} + d_{t+i} - \beta^{-1} p_{t+i-1}). \quad (4.2)$$

Taking variances in (4.2) and invoking the orthogonality of the two terms on the right-hand side of (4.2) yields

$$V(p_t^*) = V(p_t) + \frac{\beta^2 V(p_{t+1} + d_{t+1} - \beta^{-1} p_t)}{1 - \beta^2}, \quad (4.3)$$

assuming that these variances are constant over time [see LeRoy, 1989, and Gilles & LeRoy, 1991, for more discussion].

The orthogonality test (4.3) has a somewhat different interpretation from the bounds test (4.1). The present-value model is seen in (4.3) to provide a joint restriction on the volatility of price and that of excess payoffs, not just on the volatility of price as with the bounds test. Because (4.3) holds as an equality for any specification of investors' information, the orthogonality test provides a way of partly circumventing the nuisance parameter problem that plagues the bounds test (4.1): variations in m affect $V(p_t)$, but the effect is precisely offset by an equal and opposite effect on the right-most term of (4.3).³ Both effects are allowed for in the test. It was for this reason that LeRoy & Porter placed primary reliance on their version of the orthogonality test rather than the bounds test in concluding that volatility is excessive (although the reader will not find a clear discussion of the distinction between bounds and orthogonality tests in that paper).

Orthogonality tests provide statistically significant evidence of excess volatility [LeRoy & Parke, 1992]: no matter how one specifies investors' information, the volatility statistic based on (4.3) (i.e., its left-hand side less its right-hand side, with population parameters replaced by their sample counterparts) is significantly negative.

4.2. Relative power of volatility and returns tests

The empirical evidence to date indicates that both returns tests and volatility tests reject the present-value model. Rejection appears to be of borderline significance in the case of returns tests (and is subject to question on econometric grounds), but is clearly significant in the case of volatility tests. Since both types of test give essentially the same answer, determining which of these tests is more powerful — the focus of this paper — is a less urgent topic than it would be if their outcomes differed. Nonetheless, it is well worth investigating, and we now do so.

It seems unlikely that any conclusion about the general superiority of one test over the other will be forthcoming. Instead, volatility tests are likely to be more powerful than returns tests under some alternatives and less powerful under others. To see that the relative power of two tests of some null hypothesis generally depends on the alternative hypothesis, we turn to a more familiar example. Consider testing the irrelevance of x_1 and x_2 in the linear regression model

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t. \quad (4.4)$$

One straightforward approach is based upon testing

$$H_0 : \quad \beta_1 = \beta_2 = 0 \quad (4.5)$$

against the alternative

$$H_1 : \quad \beta_1 \text{ or } \beta_2 \neq 0. \quad (4.6)$$

³ However, even for the orthogonality test the sample characteristics of the statistic corresponding to this parameter depend on investors' information, as will be seen below. Therefore even under the orthogonality test the nuisance parameter problem, although much mitigated, is still with us.

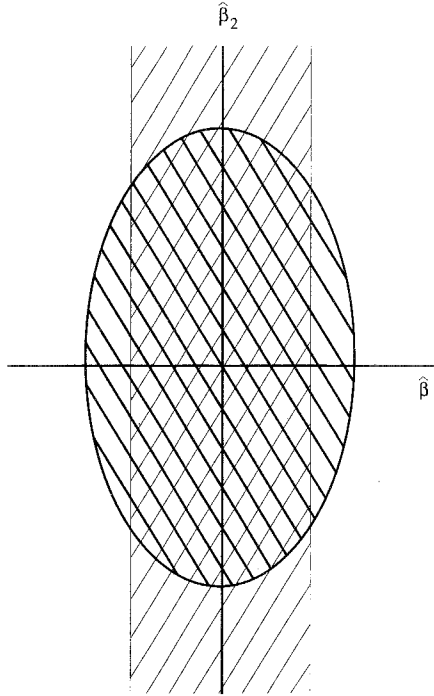


Fig. 1. Critical regions for testing $\beta_1 = \beta_2 = 0$.

Since the test statistics are the sample regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$, the critical region is given by the ellipsoid in Figure 1.

One could also form a test of $\beta_1 = \beta_2 = 0$ based only upon β_1 , leading to $H'_0 : \beta_1 = 0$ and $H'_1 : \beta_1 \neq 0$. The critical region for this test is given by the area between the two vertical lines in the figure. If the size of each test is 5%, for example, then the probability of obtaining values in the ellipsoid given H_0 is equal to the probability of obtaining values between the two vertical lines given H'_0 . Thus by construction the two tests work equally well when H_0 is true.

There is no presumption that under an alternative hypothesis the probabilities of the regions in which the tests give different outcomes is the same. For example, under an alternative hypothesis with $\beta_1 = 0$ and $\beta_2 \neq 0$ the test that ignores the implication of the model for β_2 is likely incorrectly to fail to reject the null. However, if the alternative hypothesis has $\beta_1 \neq 0$ and $\beta_2 = 0$, ignoring the implications of the model for β_2 is a virtue since the test involving only the statistic for β_1 is more likely to reject the null hypothesis when it is false.⁴

⁴ Cochrane [1991] argued that because volatility tests and return tests use the same instruments, they are likely to have the same power. This argument ignores the fact, just discussed, that the power of different tests generally depends on the alternative hypothesis, even if the tests use the same instruments. It follows that return and volatility tests do not necessarily have equal power

LeRoy & Porter and Shiller supposed that something similar was occurring with the volatility tests: by ignoring some implications of the present-value relation in order to focus on its implications for price and return variance, the volatility tests were in effect searching for violations in the right direction, whereas a test that searches in all directions for violations necessarily searches less thoroughly. Although LeRoy & Porter and Shiller suggested this possibility, they provided no concrete evidence supporting the suggestion.

4.3. Two orthogonality tests for the geometric random walk

The orthogonality test discussed above says that the variance of actual price plus that of returns (multiplied by a constant) equals the variance of the ex-post rational price (4.3). This version of the orthogonality test has the virtue of simplicity, but also has the major drawback that it requires the assumption that dividends are stationary, necessitating a trend correction. To avoid trend-correction questions, we replaced (4.3) with

$$V(p_t^*/d_t) = V(p_t/d_t) + \frac{\beta^2 \{V(p_t/d_t) + [E(p_t/d_t)]^2\}}{1 - \beta^2(\sigma^2 + \mu^2)} V(r_t) \quad (4.7)$$

[see LeRoy & Parke, 1992, for a derivation]. Here μ and σ are the mean and standard deviation, respectively, of the dividend growth rate. Equation (4.7) is analogous to (4.3) except that in (4.7) the extensive variables p_t^* , p_t and $p_{t+1} + d_{t+1} - \beta^{-1}p_t$ are replaced by the intensive variables p_t^*/d_t , p_t/d_t and r_t , (the rate of return: $r_{t+1} \equiv [p_{t+1} + d_{t+1} - p_t]/p_t$). Qualitatively, the interpretation of (4.7) is the same as that of (4.3): in (4.7) the variance of the price-dividend ratio plus (a function of) the variance of the rate of return equals the variance of the ex-post rational price-dividend ratio. The more information investors have, the higher is $V(p_t/d_t)$ and the lower is $V(r_t)$.

Just as (4.3) required that the extensive variables p_t^* , p_t and $(p_{t+1} + d_{t+1} - \beta^{-1}p_t)$ have constant variances, (4.7) requires that the intensive variables p_t^*/d_t , p_t/d_t and r_t have constant variances. If dividends are generated by a log-linear process (and assuming that information revelation is regular), this property will be satisfied for the intensive variables even though dividends, and therefore stock prices, have upward trends. Thus no further trend-correction is necessary.

The volatility statistic associated with (4.7) is computed by calculating the sample moments $\hat{V}(p_t/d_t)$, $\hat{V}(r_t)$, $\hat{E}(d_{t+1}/d_t)$ and $\hat{V}(d_{t+1}/d_t)$, where the sample variances are the average squared deviations around sample means, and substituting these for $V(p_t/d_t)$, $V(r_t)$, μ and σ^2 in (4.7). We estimated β from

$$\hat{E}(p_t/d_t) = \frac{\hat{\beta}\hat{\mu}}{1 - \hat{\beta}\hat{\mu}} \quad (4.8)$$

against any particular alternative. Also, Cochrane's argument makes no allowance for the fact that because return tests and volatility tests use different auxiliary hypotheses, they are in effect tests of different versions of market efficiency.

It remains to select an estimator of $V(p_t^*/d_t)$. There are two types of estimators of $V(p_t^*/d_t)$: model-free and model-based [Gilles & LeRoy, 1991]. A model-free estimator directly constructs an observable counterpart to the unobservable p_t^* series and takes the sample variance of the implied p_t^*/d_t series. A model-based estimator is derived by postulating a dividends model and deriving an expression for $V(p_t^*/d_t)$ in terms of β and the parameters of the dividends model. Then $V(p_t^*/d_t)$ is estimated by substituting parameter estimates in the derived expression. Model-free estimators have the advantage that they do not add a dividends model to the null model being tested, but model-based tests have the advantage that they are much less subject to sample variation [LeRoy & Parke, 1992]. Our model-free estimator of $V(p_t^*/d_t)$ was the sample variance of $p_{t|T}^*/d_t$, where $p_{t|T}^*$ is defined in (2.1).

Our model-based estimator of $V(p_t^*/d_t)$ was constructed from the geometric random walk:

$$d_{t+1} = d_t \epsilon_{t+1}, \tag{4.9}$$

where $E(\epsilon_t) = \mu$, $V(\epsilon_{t+1}) = \sigma^2$, and the ϵ_t are independently and identically distributed. Under the geometric random walk

$$V(p_t^*/d_t) = \frac{\beta^2 \sigma^2}{[1 - \beta^2(\sigma^2 + \mu^2)](1 - \beta\mu)^2}, \tag{4.10}$$

implying that the model-based estimate of $V(p_t^*/d_t)$ is:

$$\hat{V}(p_t^*/d_t) = \frac{\hat{\beta}^2 \hat{\sigma}^2}{[1 - \hat{\beta}^2(\hat{\sigma}^2 + \hat{\mu}^2)](1 - \hat{\beta}\hat{\mu})^2}. \tag{4.11}$$

In each case the sample statistic \hat{S} is just the left-hand side of (4.7) less its right-hand side, with sample statistics replacing population parameters as just indicated. If the model being tested is that which actually generated the data, \hat{S} should be near zero; if \hat{S} is significantly different from zero, the model is rejected. Specifically, excess volatility would cause $\hat{V}(p_t/d_t)$ and $\hat{V}(r_t)$ to be high, leading to a significantly negative value of \hat{S} .

5. Monte Carlo tests

5.1. Monte Carlo evaluations of volatility tests

We constructed Monte Carlo evaluations of the size and power of the model-based bounds test, and both the model-free and model-based orthogonality tests, by assuming that dividends were generated by a geometric random walk with normal innovations.⁵ We set the mean and standard deviation of the dividend

⁵ We conducted tests of the model-free bounds test, but do not report them because they are subject to the same critical shortcomings as the model-based bounds test, reported below.

growth rate equal to those reported by LeRoy & Parke [1992] for annual US aggregate dividends ($\mu = 1.0216$, $\sigma^2 = 0.0153$). Simulated stock prices were calculated using an annual discount factor of 0.9350, the value estimated by LeRoy & Parke. For each iteration, we simulated a time-series for both prices and dividends of length 117 to accord with our annual data set. We performed 10,000 iterations, each time saving the values of the test statistics, which we used to construct a simulated distribution. The critical values for our test statistics were then based on the 5%-tail of the simulated distribution in the direction associated with excess volatility.

As noted in Section 2 the present-value model being tested is a compound null hypothesis, with different versions implied by different specifications of investors' information. The stock price series associated with a given dividend sample path depends on how much information investors are assumed to have, implying that for bounds tests the population parameter corresponding to the test statistic, $V(p_t^*/d_t) - V(p_t/d_t)$, takes on different values under different versions of the null hypothesis. A major advantage of the orthogonality test is that this problem does not occur: the parameter being tested equals zero under all versions of the null hypothesis.

To parametrize the dependence of stock price on investors' information, we assumed that investors have information variables which enable them to see ahead exactly m periods, for various values of m . This device was used by Gilles & LeRoy [1991] and LeRoy & Parke [1992]. For $m = 0$ investors extrapolate future dividends from current dividends using the constant growth rate implied by the geometric random walk, implying that the price-dividend ratio is constant over time. Investors at t know actual dividends up to $t + m$ and extrapolate dividends beyond $t + m$ by applying a constant growth rate to d_{t+m} . The higher the value of m , the higher is $V(p_t/d_t)$ and the lower is $V(r_t)$; as m approaches infinity, $V(p_t^m/d_t)$ approaches $V(p_t^*/d_t)$ and $V(r_t)$ approaches zero.

We limited the Monte Carlo runs to $m \leq 5$. The reason is that with $m = 5$ the predicted volatility of the price-dividend ratio under the null hypothesis approximately equals the volatility of the real-world price-dividend ratio [LeRoy & Parke, 1992]. Thus for $m > 5$ the volatility of the price-dividend ratio predicted by the model is greater than that observed in the real-world data under both the null hypothesis and the alternative, so these cases can be discarded at the outset as empirically irrelevant.

To evaluate the power of the test just described, we repeated the calculations of the mean and standard deviation of the test statistic assuming that stock prices equal those implied under various values of m by the null hypothesis plus a white noise term:

$$p_t/d_t = E(p_t^*/d_t | I_t) + \eta_t, \quad (5.1)$$

where η_t has unit variance, as compared to a variance of p_t^*/d_t of 89.3. We then calculated power by estimating the mean and standard deviation of the test statistic over 10,000 simulated time series generated under this alternative hypothesis. Again using the simulated finite-sample distribution, we calculated the

power of each test as the probability that the test statistic generated under the alternative lies in the rejection region as computed under the null.

5.2. Regression tests as a benchmark

As a benchmark to evaluate the power of the bounds and volatility tests, we performed similar power calculations for two regression tests:

$$r_{t+1} = \alpha + \beta r_t + \xi_{1t} \quad (5.2)$$

and

$$r_{t+1} = \gamma + \delta \left(\frac{p_t}{d_t} \right) + \xi_{2t}. \quad (5.3)$$

The null hypothesis implies that $\beta = 0$ and $\delta = 0$. Under the alternative hypothesis, β and δ are negative. This is so because a positive realization of the noise term at t is correlated with high values of r_t and p_t/d_t , and also with low values of r_{t+1} . Therefore a one-tail test of the present-value model can be constructed by setting the rejection region as the lower 5% tail of $\hat{\beta}$ and $\hat{\delta}$.

The reason we concentrated on a white noise alternative — despite the fact that fads are usually modeled as highly autocorrelated noise processes — is that under a white noise alternative it is easy to determine the most powerful regression test of the present-value model: regress r_{t+1} on r_t . Including lagged returns as regressors will only dissipate degrees of freedom since lagged returns have population coefficients of zero under both the null and the alternative. However, we present evidence below that our results on power carry over when noise is autocorrelated.

The regression tests just outlined have the major advantage that, under the null hypothesis, they are free of nuisance parameter problems. For the regression tests, like the orthogonality test but unlike the bounds test, the value of the population parameter is zero under all versions of the null hypothesis (i.e., under any specification of investors' information). However, it is also true of regression tests, unlike both the volatility tests, that the probability distribution of the corresponding test statistic does not depend on investors' information. Put differently, in the regressions the t -test for $\beta = 0$ and $\delta = 0$ allows construction of a 5% rejection region under any specification of investors' information. Under the orthogonality tests, in contrast, it is necessary to make allowance for the fact that for any critical value of the test statistic the size of the test may depend on investors' information. This dependence complicates the interpretation of the results.

5.3. Monte Carlo results

Tables 1 and 2 report the mean and standard deviation of the test statistic for each of the five tests outlined above — three volatility tests and two regression tests — under the null hypothesis and the alternative hypothesis, respectively.

Table 1

Mean / standard deviation of test statistics; null hypothesis: present-value model

m	Volatility tests			Regression tests	
	(1) Bounds	(2) Model-free orthogonality	(3) Model-based orthogonality	(4) r_{t+1} on r_t	(5) r_{t+1} on p_t/d_t
1	83.2 / 13.1	-43.2 / 37.5	-0.0 / 1.0	0.0 / 0.09	0.0 / 0.004
2	77.2 / 12.7	-42.4 / 38.4	0.3 / 2.2	0.0 / 0.09	0.0 / 0.004
3	71.7 / 12.9	-41.3 / 39.8	0.9 / 3.6	0.0 / 0.09	0.0 / 0.002
4	66.8 / 13.6	-40.3 / 39.5	1.6 / 5.3	0.0 / 0.09	0.0 / 0.002
5	62.5 / 14.6	-39.2 / 38.2	2.5 / 7.2	0.0 / 0.09	0.0 / 0.002

Table 2

Mean / standard deviation of test statistics; alternative hypothesis: present-value model plus noise

m	Volatility tests			Regression tests	
	(1) Bounds	(2) Model-free orthogonality	(3) Model-based orthogonality	(4) r_{t+1} on r_t	(5) r_{t+1} on p_t/d_t
1	82.2 / 13.1	-86.7 / 42.8	-29.1 / 11.1	-0.13 / 0.09	-0.0004 / 0.005
2	76.2 / 12.7	-88.7 / 41.4	-30.4 / 11.2	-0.14 / 0.09	-0.0004 / 0.003
3	70.7 / 13.0	-90.0 / 43.0	-31.4 / 11.9	-0.16 / 0.09	-0.0005 / 0.003
4	65.8 / 13.7	-89.9 / 43.2	-31.8 / 12.6	-0.17 / 0.09	-0.0005 / 0.002
5	61.5 / 14.7	-91.2 / 41.9	-32.3 / 14.1	-0.18 / 0.09	-0.0006 / 0.002

Column 1 of Table 2 shows that, as expected, the mean value of the test statistic under the bounds tests decreases with m . Comparison of column 1 of Tables 1 and 2 shows that, again as expected, the noise term lowers the mean of the test statistic. However, the effect is minor relative to both the standard deviation of the test statistic and the effect of m on the test statistic. We see already that the bounds test is likely not to be very good at detecting the presence of noise.

Column 2 of Table 1 shows that under the null hypothesis the test statistics for the model-free orthogonality test average about one standard deviation below zero, reflecting the well-documented downward bias in the model-free estimate of $V(p_t^*/d_t)$. The corresponding column of Table 2 shows that the test statistic averages about two standard deviations below zero when white noise is present. In sharp contrast to the bounds test, the behavior of the test statistic under either the null or the alternative does not depend significantly on investors' information. The reason that the noise has a greater effect on the test statistic under the model-free orthogonality test is that in the rightmost term in (4.7), $V(r_t)$ is multiplied by a large number; the effect of the noise term on $V(p_t/d_t)$ is minor by comparison. In the model-based orthogonality test the effect of noise on the test statistic (column 3) is comparable in magnitude to its effect in the

model-free test. However, sampling variation is much lower with the model-based orthogonality test, especially under the null hypothesis. In the regression tests, the white noise term induces negative autocorrelation of borderline statistical significance in successive rates of return (Table 2, column 4). In contrast, the coefficient in the regression of r_{t+1} on p_t/d_t is essentially zero under the alternative.

Tables 3 and 4 show the size and power of the five tests under consideration. We computed size and power under a representative value of m : $m = 2$ (Table 3), and under the maximum reasonable value of m : $m = 5$ (Table 4). Column 1 of either Table 3 or Table 4 shows that the size and power of the bounds test depend

Table 3*

Size / power of tests; critical region based on $m = 2$

m	Volatility tests			Regression tests	
	(1) Bounds	(2) Model-free orthogonality	(3) Model-based orthogonality	(4) r_{t+1} on r_t	(5) r_{t+1} on p_t/d_t
1	0.01 / 0.01	0.04 / 0.63	0.00 / 0.99	0.06 / 0.40	0.06 / 0.06
2	0.03 / 0.04	0.05 / 0.66	0.04 / 0.99	0.06 / 0.47	0.06 / 0.06
3	0.11 / 0.13	0.04 / 0.65	0.11 / 0.99	0.06 / 0.52	0.07 / 0.05
4	0.23 / 0.26	0.04 / 0.67	0.16 / 0.99	0.06 / 0.58	0.07 / 0.05
5	0.37 / 0.39	0.04 / 0.65	0.19 / 0.99	0.06 / 0.65	0.08 / 0.06

Size = probability of rejecting the present-value model when it is true;
power = probability of rejecting the present-value model when it is false.

Table 4*

Size / power of tests; critical region based on $m = 5$

m	Volatility tests			Regression tests	
	(1) Bounds	(2) Model-free orthogonality	(3) Model-based orthogonality	(4) r_{t+1} on r_t	(5) r_{t+1} on p_t/d_t
1	0.00 / 0.00	0.06 / 0.66	0.00 / 0.97	0.06 / 0.40	0.06 / 0.06
2	0.00 / 0.00	0.05 / 0.68	0.00 / 0.98	0.06 / 0.47	0.06 / 0.06
3	0.00 / 0.00	0.04 / 0.67	0.01 / 0.98	0.06 / 0.52	0.07 / 0.06
4	0.01 / 0.01	0.05 / 0.69	0.03 / 0.98	0.06 / 0.58	0.07 / 0.05
5	0.03 / 0.04	0.05 / 0.67	0.06 / 0.99	0.06 / 0.65	0.08 / 0.06

Size = probability of rejecting the present-value model when it is true;
power = probability of rejecting the present-value model when it is false.

* Note that some of the values which should equal 0.05 from the construction of the tables actually equal 0.03, 0.04 or 0.06. This discrepancy primarily reflects the normal approximation used to set rejection regions. Also, sample variation is important in estimating tail probabilities, even with 10,000 draws. Finally, the reported values reflect roundoff error.

strongly on m , and also that size is virtually equal to power. Thus, as expected from the discussion of Tables 1 and 2, the bounds test is virtually useless in detecting the presence of the noise term.

Column 2 of Tables 3 and 4 shows that, when the rejection region for the model-free orthogonality test is chosen so that the null is rejected with 5% probability when it is true, it is rejected about two-thirds of the time when it is false. These figures are essentially unaffected by m . In contrast, the fact that the mean and standard deviation of the model-based orthogonality test statistic depend on m means that for fixed rejection region the size and power of the test do so as well (column 3). However, the very low sampling variation of the model-based orthogonality test statistic means that the nuisance parameter problem can be easily handled: setting a large acceptance region, as in Table 4, implies that the null hypothesis is almost sure not to be rejected if it is true and is almost sure to be rejected if it is false.

The regression of r_{t+1} on r_t has power that ranges from 0.40 to 0.65, depending on m , when size is set at 0.05 (column 4). Thus a regression of r_{t+1} on r_t has some ability to detect noise, but much less than the model-based orthogonality test. It is interesting to observe that the nuisance parameter problem reappears with the test based on regressing r_{t+1} on r_t : even though size does not depend on m , the power of the test does. (Note that, while regression theory guarantees that regression tests are free of nuisance parameter problems under the null, it provides no such assurance under the alternative). Finally, the regression of r_{t+1} on p_t/d_t (column 5), like the bounds test, is completely unable to detect the presence of noise. In fact, for some values of m ($m = 3, 4, 5$), the presence of noise actually increases the probability that we fail to reject the null hypothesis. This occurs because the correlation between the noise and p_t/d_t is so low that the noise acts essentially only to increase the variance of p_t/d_t , leading the coefficient in the regression of r_{t+1} to be biased toward zero. This effect is familiar from the errors-in-variables problem of econometrics.

It is important to understand why the model-based orthogonality test performs so much better than the other tests. Observe that if a researcher knew m , then both $V(p_t/d_t)$ and $V(r_t)$ could be expressed as functions of β , m , μ and σ^2 , just as $V(p_t^*/d_t)$ is given by (4.10) as a function of β , μ and σ^2 . Observe further that if these functions, along with expression (4.11) for $V(p_t^*/d_t)$, are inserted in (4.7), the result is an identity in β , m , μ and σ . Consequently, \hat{S} equals zero for any m even though the estimated values of β , μ and σ are substituted for their population counterparts in (4.7).

Of course, the researcher who does not know m must use the model-free estimates of $V(p_t/d_t)$ and $V(r_t)$ — i.e., their sample variances. It follows that if the null hypothesis is true, variations in \hat{S} are entirely attributable to the differences between the model-free estimates of $V(p_t/d_t)$ and $V(r_t)$ and their model-based counterparts. These differences are small, and are correlated in such a way that the orthogonality test statistic has low variance under the null hypothesis (Table 1). Therefore the orthogonality test is very likely to detect noise if it is present. This argument is similar to that of Durlauf & Hall [1989] to the effect that orthogonality tests are more powerful than bounds tests.

5.4. Robustness to respecified dividend processes

These results show that the model-free orthogonality test has somewhat better ability to detect noise than the most powerful regression test, particularly for low m . For the model-based orthogonality test, power is far greater. However, it will be objected that the playing field is not level. Both versions of the orthogonality test required, in addition to the present-value model, the specification that $V(p_t^*/d_t)$, $V(p_t/d_t)$ and $V(r_t)$ are constant over time. The model-based orthogonality test required further that dividends be generated by a geometric random walk. The regression tests, in contrast, do not require either assumption. The simulations used to evaluate the tests were based on the geometric random walk model, so that the additional specifications adopted for the orthogonality tests were satisfied by construction in the population used to evaluate the two classes of test. It is not surprising that the tests requiring a more restrictive specification perform better than the tests not requiring the more restrictive specification, given a setting in which the added restrictions are valid by assumption.

The suggestion is that the superiority of the orthogonality tests may disappear in settings where the moment variances are nonconstant or where the geometric random walk is a misspecification. That the orthogonality test lacks robustness to alternative dividend specifications is a very real possibility, and this question deserves more thorough study than we have given it. However, we have two preliminary results to report.

We considered an environment in which the geometric random walk is a gross misspecification: if dividends are generated by a stationary model rather than an integrated model the orthogonality relation (4.7) is invalid. We verified by Monte Carlo methods that application of (4.7) resulted in virtually 100% rejection frequencies if dividends are stationary whether the null hypothesis was true or false. Analysts who are not willing to stipulate that dividends have a unit root will regard as a major limitation of the orthogonality test the fact that it breaks down if the form (stationary or unit root) of the dividend process is misspecified.

What happens if dividends are log-linear and have a unit root, but are generated by a model less parsimonious than the geometric random walk? In that setting the model-free orthogonality test remains valid, but the model-based test does not since then expression (4.10) for $V(p_t^*/d_t)$ in terms of μ , σ and β is a misspecification. To investigate the extent to which our characterization of the size and power of the orthogonality test carry over in this more general setting, we based Monte Carlo simulations on the AR(2) model

$$\frac{d_{t+1}}{d_t} = 1.03 + 0.18 \frac{d_t}{d_{t-1}} - 0.19 \frac{d_{t-1}}{d_{t-2}} + \eta_t, \quad (5.4)$$

($V(\eta_t) = 0.0149$), which gives a more accurate characterization of US dividends than the more parsimonious geometric random walk [LeRoy & Parke, 1992, supplement]. Table 5 compares the size and power of the geometric random walk model on the left (this column coincides with Table 4, column 3) with the AR(2)

Table 5
Size / power of tests; critical region based
on $m = 5$

m	Geometric random walk	AR(2)
1	0.00 / 0.97	0.00 / 0.96
2	0.00 / 0.98	0.00 / 0.98
3	0.01 / 0.98	0.01 / 0.98
4	0.03 / 0.98	0.03 / 0.96
5	0.06 / 0.99	0.05 / 0.94

model on the right. As is evident, the size and power are not much affected by the geometric random walk misspecification if dividends are generated by the AR(2) model (5.4).

One interpretation of these results is that the orthogonality test is robust, at least within the class of log-linear unit-root models. Another interpretation, however, is that the geometric random walk is a good approximation to less parsimonious dividend models; given that the AR(2) is close to the geometric random walk it is not surprising that size and power are nearly the same in the two cases. From this vantage, it would be observed that little can be concluded about robustness without considering dividend models less similar than our estimated AR(2) to the geometric random walk.

In conclusion, the second of these results suggests that the orthogonality test is not critically sensitive to the geometric random walk assumption, contrary to what might have been expected. However, we again emphasize that we have not provided definitive evidence on this point.

5.5. Robustness to autocorrelated noise process

As noted above, we concentrated our attention on a white noise alternative in order to identify easily the regression test which is optimal against the alternative. However, the alternative hypothesis that is relevant empirically (at least on some accounts) incorporates highly autocorrelated noise rather than white noise. To evaluate the effect of noise autocorrelation on power, we recalculated power assuming that the noise is given by

$$\eta_t = \rho\eta_{t-1} + \epsilon_t, \quad (5.5)$$

with σ_ϵ^2 adjusted so that $\sigma_\eta^2 = 1$ for each value of ρ . Table 6 shows that the effect of noise autocorrelation on the power of both the bounds test and the model-based orthogonality test is negligible for low and moderate levels of noise: for ρ less than 0.9 the bounds test seldom rejects and the orthogonality test almost always rejects, as with white noise. However, when the noise process is almost a random walk, the noise increases the sample volatility of the test statistic: for $\rho = 0.99$ the test statistic lies in the rejection region with 39% probability (71%) for the bounds (orthogonality) test. Thus highly autocorrelated noise somewhat

Table 6

Power under autocorrelated alternative; critical region based on $m = 5$

ρ	Volatility tests		Regression tests	
	(1) Model-based bounds	(2) Model-based orthogonality	(3) r_{t+1} on r_t	(4) r_{t+1} on p_t/d_t
0.0	0.04	0.99	0.40	0.06
0.1	0.04	0.99	0.32	0.08
0.2	0.04	0.99	0.25	0.10
0.3	0.04	0.98	0.20	0.11
0.4	0.04	0.96	0.15	0.12
0.5	0.04	0.95	0.12	0.13
0.6	0.05	0.93	0.10	0.13
0.7	0.05	0.90	0.09	0.12
0.8	0.08	0.87	0.08	0.11
0.9	0.14	0.84	0.07	0.09
0.95	0.24	0.80	0.07	0.08
0.99	0.39	0.71	0.07	0.07

improves the poor performance of the bounds test, but somewhat degrades the excellent performance of the model-based orthogonality test.

As expected, the power of the regression of r_{t+1} on r_t falls as ρ rises. For very high values of ρ the noise component of r_{t+1} is virtually equal to that of r_t , so the negative autocorrelation of rates of return associated with low values of ρ disappears. Also as expected, the regression of r_{t+1} on p_t/d_t gains power as ρ rises from low to moderate values.⁶ However, for all values of ρ the regression tests of the present value model remain greatly inferior to the model-based orthogonality test.

6. Conclusion

Both the model-free and model-based orthogonality tests are better-behaved econometrically than the returns test that is optimal (among returns tests) against the alternative we assumed. The model-free orthogonality test has somewhat higher power than the benchmark returns test when size is held constant. Further, unlike the regression test the model-free orthogonality test is not subject to the nuisance parameter problem under either the null or the alternative. The model-based orthogonality test, in contrast, is subject to the nuisance parameter problem under both the null and the alternative. However, sampling variation is so low under both the null and the alternative — but particularly so under the

⁶ Campbell [1993] verified analytically that long-horizon return autoregressions are more powerful than short-horizon return autoregressions when the alternative is highly autocorrelated. Regressing r_{t+1} on p_t/d_t is similar to a long-horizon return autoregression.

null — that the nuisance parameter problem does not distort the outcome: in assessing the presence or absence of noise, the critical region for the model-based orthogonality test can be set so that the test almost always returns the correct diagnosis. However, the model-based test requires the assumption that dividends follow a geometric random walk, so this verdict in favor of the model-based orthogonality test might require revision to the extent that that specification is incorrect.

If the real-world data were generated by the alternative hypothesis assumed in our Monte Carlo runs — stock price equals the present value of expected dividends plus white noise, where dividends follow a geometric random walk — with approximately the same parameter values, we would expect to find exactly the pattern observed with real-world data: failure to reject or marginal rejection with returns tests: stronger rejection with variance-bounds tests. These results support LeRoy & Porter and Shiller's original conjecture that the differing outcomes of returns and variance-bounds tests reflect the greater power of the latter tests. There is nothing paradoxical about this: we argued above that different tests of the same null hypothesis will generally have different power against any particular alternative hypothesis, even if they use the same instruments.

The difference in power may also be due to a difference in the auxiliary hypotheses assumed in constructing the test statistic. The volatility tests which we used made hard use of the assumption that $V(p_t^*/d_t)$, $V(p_t/d_t)$ and $V(r_t)$ are constant over time. This constancy property was assumed satisfied under both the null and alternative hypotheses which we specified. If we had specified either a null or an alternative hypothesis that failed to satisfy this property, it is unlikely that our results would have been as favorable to the volatility test as those reported here.

There are other possible explanations for the fact that volatility tests show stronger rejection of the present-value model than returns tests. Model-based volatility tests can detect the present of rational speculative bubbles, whereas returns tests cannot since the latter do not impose a convergence condition. Therefore if stock prices have bubble components, rejection of the present-value model will be stronger under volatility tests than return tests, consistent with our stylized fact.

We have dealt with the procedural question of the relative ability of different types of statistical tests to detect departures from the present-value model. We have not considered the substantive question of how to interpret rejection of the present-value model. LeRoy [1989] discussed this question but, again, not conclusively: no consensus exists as to whether statistical rejection of the present-value model has implications for the broader question of capital market efficiency. A minimalist interpretation of the statistical rejection of the present-value model would emphasize that what is rejected is the assumption that discount rates are constant over time. However, nothing about capital market efficiency precludes time-varying discount rates. Also, it is known that small but highly autocorrelated departures from constant discount rates may be consistent with price volatility greatly in excess of that implied by the (constant discount rate version of the) present-value model. Therefore even highly excessive price volatility might give

rise to only marginally profitable trading rules. Along these lines the interpretation of the variance-bounds rejections might be that the present-value model has only weak implications for the unconditional variance of asset prices (except in the restrictive constant discount rate case), not that market efficiency is violated.

Others, such as Shiller, have drawn more sweeping conclusions from the excess volatility in financial markets. They point out that models with time-varying discount rates perform little better than the constant discount rate version in explaining real-world asset price changes. Also, they would situate excess volatility with the other anomalies of financial markets, which collectively establish a strong case against market efficiency even though individually these anomalies may be subject to question.

Acknowledgements

Without implicating them, we thank F. Black, S. Blough, T. Bollerslev, J. Cochrane, S. Durlauf, J. Geweke, R. Hodrick, R. Jagannathan, W. Parke and W. Ziemba for comments.

References

- Campbell, J. (1993). Why long horizons? A study of power against persistent alternatives. Reproduced, Princeton University.
- Campbell, J.Y., and R.J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Rev. Financial Stud.* 1, 195–228.
- Cochrane, J.H. (1991). Volatility tests and efficient markets: A review essay. *J. Monetary Econ.* 27, 463–485.
- Cochrane, J.H. (1992). Explaining the variance of price-dividend ratios. *Rev. Financial Stud.* 5, 243–280.
- Cogley, T. (1991). Out-of-sample predictive power of dividend yield regressions. Reproduced, University of Washington.
- Durlauf, S.N., and R.E. Hall (1989). Measuring noise in stock prices. Reproduced, Stanford University.
- Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *J. Finance* 25, 283–417.
- Fama, E.F. (1991). Efficient capital markets, II. *J. Finance* 46, 1575–1617.
- Fama, E.F. and K.R. French (1988). Permanent and transitory components of stock prices. *J. Political Econ.* 96, 246–273.
- Flavin, M. (1983). Excess volatility in the financial markets: A reassessment of the empirical evidence. *J. Political Econ.* 91, 929–956.
- Gilles, C., and S.F. LeRoy (1991). Econometric aspects of the variance-bounds tests: A survey. *Rev. Financial Stud.* 4, 753–791.
- Grossman, S.J., and R.J. Shiller (1981). The determinants of the variability of stock prices. *Am. Econ. Rev. Pap. Proc.* 71, 222–227.
- Kim, M.J., C.R. Nelson and R. Startz (1991). Mean reversion in stock prices? A reappraisal of the empirical evidence. *Rev. Econ. Stud.* 58, 515–528.
- Kleidon, A.W. (1986). Bias in small-sample tests of stock price rationality. *J. Business* 59, 237–261.

- LeRoy, S.F. (1984). Efficiency and the variability of asset prices. *Am. Econ. Rev. Pap. Proc.* 74, 183–187.
- LeRoy, S.F. (1989). Efficient capital markets and martingales. *J. Econ. Literature* 17, 1583–1621.
- LeRoy, S.F., and W.R. Parke (1992). Stock price volatility: Tests based on the geometric random walk. *Am. Econ. Rev.* 82, 981–992.
- LeRoy, S.F., and R.D. Porter (1981). Stock price volatility: Tests based on implied variance bounds. *Econometrica* 49, 555–574.
- Poterba, J.M., and L.H. Summers (1988). Mean reversion in stock prices: Evidence and implications. *J. Financial Econ.* 22, 27–59.
- Richardson, M. (1993). Temporary components of stock prices: A skeptic's view. *J. Business Econ. Statistics* 11, 199–207.
- Samuelson, P.A. (1965). Proof that properly anticipated prices fluctuate randomly. *Ind. Manage. Rev.* 6, 41–49.
- Shiller, R.J. (1981). Do stock market prices move too much to be justified by subsequent changes in dividends?. *Am. Econ. Rev.* 71, 421–36.
- Shiller, R.J. (1988). The probability of gross violations of a present value variance inequality. *J. Political Econ.* 96, 1089–1092.
- Shiller, R.J. (1989). *Market Volatility*, Cambridge, MIT Press.
- West, K. (1988). Bubbles, fads and stock price volatility: A partial evaluation. *J. Finance* 43, 636–656.