NBER WORKING PAPER SERIES

BOUNDS ON THE VARIANCES OF SPECIFICATION ERRORS IN MODELS WITH EXPECTATIONS

Steven N. Durlauf

Robert E. Hall

Working Paper No. 2936

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1989

This paper is part of NBER's research program in Economic Fluctuations. Durlauf's work was supported by the Center for Economic Policy Research. Hall's work was supported in part by the National Science Foundation. We thank Andrew Bernard for superlative research assistance. Any opinions expressed are those of the authors not those of the National Bureau of Economic Research.

NBER Working Paper #2936
April 1989

BOUNDS ON THE VARIANCES OF SPECIFICATION ERRORS IN MODELS WITH EXPECTATIONS

ABSTRACT

Under rather general conditions, observed covariances place a useful lower bound on the variance of the misspecification or noise in models based on expectations. Such models are widely used for securities prices, exchange rates, consumption, and output. For a correctly specified model, the lower bound will be zero. We construct an optimal bound on model noise that captures the complete set of testable restrictions on an expectations based model. Many specification tests for asset prices are easily interpreted as estimates of this lower bound. As a result, the power of different tests may be ranked according to the information restrictions employed in constructing noise estimates. Our results show that specification tests which use the history of lagged dependent variables are usually better able to uncover model noise than based on information sets that exclude those variables.

Steven Durlauf
Department of Economics
Stanford University
Stanford, California 94305

Robert E. Hall
Hoover Institution
Stanford University
Stanford, California 94305

Introduction

Robert Shiller [1981a,b] and Stephen LeRoy and Richard Porter [1981] were the first to note that expectational models have simple testable implications derived from general properties of expectations. A large literature has developed from this idea (see the survey by Gilles and LeRoy [1988] for many references). Shiller and LeRoy-Porter's idea was that the expectation or prediction of a random variable must necessarily have a lower variance that its realization. Shiller's basic test compared the variance of the two and rejected the simple expectational model if the variance of the prediction exceeded the variance of the realization. In the stock market, the comparison is between the realized present discounted value of dividends and the current market price of the stock. Shiller concluded that the stock market has a good deal of noise or model error from his finding that stock prices are much more variable than are realized discounted dividends.

Shiller looked at only two moments of the data—the variance of the realization, P^*, and the variance of the actual price, P. Subsequent researchers, notably Louis Scott [1985], Kenneth West [1987], and John Campbell and Robert Shiller [1987] have proposed tests of the hypothesis of no noise that rely on covariances as well as variances. Our purpose in this paper is to carry out a systematic investigation of the sharpening of results that is available through the use of covariances. We depart from the almost exclusive focus of the earlier research on testing the null hypothesis of the absence of noise or model error. Our objective is to obtain lower bounds on the amount of noise. A lower bound above zero refutes the null hypothesis of no noise.

Before Shiller and LeRoy-Porter, research on noise in expectation models tended to look at one-period-ahead relations. In the stock market, for example, numerous investigators looked for evidence of excess returns over brief holding periods. The finding of a relationship between an observed variable and excess returns supported a conclusion of market inefficiency or other types of noise compared to the predictions of the model. One of the contributions of this paper is to compare the one-period-ahead or flow approach to the Shiller or stock approach. The stock approach appears far superior as a method for detecting and measuring slow-moving noise. In particular, noise components that grow at a rate near the rate of interest escape detection in the flow approach. This point has been emphasized by authors whose concern was the detection of speculative bubbles. Our work here is not directed toward bubbles in particular, but rather to slower-moving noise components in general. In applications of our ideas to investment and consumption, there is no reason to look for speculative bubbles.

Our results can be explained intuitively in a regression framework. The difference between the realization and the actual price is the sum of an expectation error and noise. The actual price contains the noise plus the true expectation of the realization. The regression of the difference on the price would have zero explanatory power if there were no noise, under rational expectations. On the other hand, if there were no variation in the true expectation, but some noise, then the fitted value of the regression would measure the noise exactly. In the general case, with some noise and some variability of the unobserved true expectation, the variance of the fitted value of the regression is a lower bound on the variance of the noise.

Section 1 considers only the covariance of a variable and its perfect-foresight counterpart. Variance of noise is inferred from the variance of the fitted value of the regression of the discrepancy between the two variables on the original variable. Section 2 extends the analysis to a regression with additional regressors. It shows that such a regression provides the strongest possible bound that could be calculated from the data. Section 3 shows that the general method should be applied in the case of constant discounting; there is no better special method for this important special case. Section 4 looks at another important special case where the perfect-foresight variable is constructed from a finite number of future values of another variable. This case applies to bonds, for example. Section 5 relates the methods of this paper to the West-Casella specification tests.

1. Information contained in the covariance of the actual and perfect-foresight variables

Let P_t^* be the observed realization of a random variable. In the application to the stock market, for example, P_t^* is the actual discounted value of dividends. Let P_t^e be the mathematical expectation of P_t^* conditional on the unobserved vector of information, Φ_t,

$$P_t^e = E(P_t^* | \Phi_t) \tag{1.1}$$

Under rational expectations, the two P's should differ by an unpredictable random variable, ν_t, which measures the impact on P_t^* of information not available at the time that the expectation, P_t^e, is formed:

$$P_t^* \;=\; P_t^e + \nu_t \tag{1.2}$$

The unobserved random variable P_t^e is the value assigned by a particular model of the determination of an observed random variable, P_t (the actual stock price, in the stock market application). The model presumably omits some features of the actual process by which P_t is determined (in the stock market, the model might assume a constant discount rate, when the actual discount rate varies randomly over time, for example). Hence the value mandated by the model, P_t^e, differs from the actual value, P_t, by a specification error or noise variable, S_t:

$$S_t \;=\; P_t - P_t^e \tag{1.3}$$

In this unobserved components model, we seek to make inferences concerning P_t^e, ν_t, and S_t given data on P_t and P_t^*. Our particular interest is in finding the smallest admissable value of σ_s^2, the variance of S_t. The null hypothesis of no specification error is $P_t = P_t^e$ or $\sigma_s^2 = 0$.

Louis Scott [1985] developed a test of the null hypothesis of no noise based on the regression of P_t^* on P_t. If there is no noise, P_t should be an efficient prediction of P_t^* and thus receive a coefficient of one in the regression. He found coefficients much less than one and consequently rejected the null hypothesis.

Our approach pursues a similar strategy, but our interest is in inferring the magnitude of the specification error or noise, rather than simply demonstrating its existence. We consider the regression of the discrepancy, $P_t - P_t^*$, on the price, P_t. The coefficient in this regression is one minus Scott's coefficient; the coefficient is zero absent noise. Our basic result is that the variance of the fitted value for this regression is the sharpest available lower bound on the variance of the noise.

The left-hand variable in the regression is

$$P_t - P_t^* = S_t - \nu_t \tag{1.4}$$

and the right-hand variable is

$$P_t = P_t^e + S_t \tag{1.5}$$

The regression coefficient is

$$\frac{Cov(S_t - \nu_t, P_t^e + S_t)}{Var(P_t^e + S_t)} \tag{1.6}$$

and the variance of the fitted value is

$$\frac{[Cov(S_t - \nu_t, P_t^e + S_t)]^2}{Var(P_t^e + S_t)} \tag{1.7}$$

Let σ_S^2 be the variance of the unobserved noise, $\sigma_{P^e}^2$ be the variance of the unobserved expectation, and σ_{SP^e} be their covariance. By hypothesis, the covariances of the expectation error, ν_t, with S_t and P_t^e are both zero. Hence the variance of the fitted value is

$$\frac{(\sigma_S^2 + \sigma_{SP^e})^2}{\sigma_S^2 + 2\sigma_{SP^e} + \sigma_{P^e}^2} \tag{1.8}$$

Before showing that this expression is never greater than the noise variance σ_S^2, we consider some special cases. First, suppose that the model noise is uncorrelated with the expectation, P_t^e: $\sigma_{SP^e} = 0$. Then the variance of the fitted value is

$$\sigma_S^2 \; \frac{\sigma_S^2}{\sigma_S^2 + \sigma_{P^e}^2} \tag{1.9}$$

which is plainly less than σ_S^2. If the variability of the expectation is a fairly small part of the total variability of the observed variable, then the variance of the fitted value is a good indicator of the variance of noise. Otherwise, the variance of noise is quite a bit greater than the variance of the fitted value.

As a second case, suppose that noise is a multiple of the unobserved expectation, P_t^e:

$$S_t = \lambda P_t^e \; . \tag{1.10}$$

A stock market that reacted in the right direction, but systematically more than justified by fundamentals, would fit this case. For any value of λ other than -1, it is easy to show that the variance of the fitted value is exactly the variance of S_t. The bound works extremely well whenever the nature of the specification error is for the model to under- or over-predict movements, provided that it does not under-predict by just the right amount to cancel all movements in P_t.

As a third example, suppose that

$$S_t = -\frac{1}{2}P^e_t + u_t \ , \tag{1.11}$$

with $\sigma^2_u = \frac{1}{4}\sigma^2_{P^e}$. Substitution shows that the variance of the fitted value is then exactly zero: The positive covariance between $S_t - \nu_t$ and $P^e_t + S_t$ caused by the presence of S_t in both is exactly offset by the negative covariance of S_t and P^e_t. In this case, the bound is uninformative even though noise could be very large. In general, there is a one-dimensional space of cases described by

$$S_t = -\theta P^e_t + u_t \ , \tag{1.12}$$

and

$$\sigma^2_u = \theta(1-\theta)\sigma^2_{P^e} \tag{1.13}$$

for $0 < \theta < 1$ in which the bound is completely uninformative even though noise is present. We conclude that the bound is more useful when noise is positively correlated with the model's fitted value than when it is negatively correlated.

The proof of the bound is straightforward. From the Cauchy-

Schwartz inequality,

$$(\sigma_{SP^e})^2 \leq \sigma_S^2 \sigma_{P^e}^2 \ . \tag{1.14}$$

Adding $\sigma_S^4 + 2\sigma_S^2 \sigma_{SP^e}^2$ to both sides, we have

$$(\sigma_S^2 + \sigma_{SP^e})^2 \leq \sigma_S^2(\sigma_S^2 + 2\sigma_{SP^e} + \sigma_{P^e}^2) \ . \tag{1.15}$$

The expression in parentheses on the right-hand side is the variance of the observed variable, P_t. Except in the borderline case where the variance is zero, we can divide both sides of (1.15) by the variance to get the desired result,

$$\frac{(\sigma_S^2 + \sigma_{SP^e})^2}{\sigma_S^2 + 2\sigma_{SP^e} + \sigma_{P^e}^2} \leq \sigma_S^2 \ . \tag{1.16}$$

The borderline case of zero variance in the observed right-hand variable corresponds to the situation in the second example with $\lambda = -1$. If we adopt the convention that the coefficient resulting from regressing a variable on another variable with zero variance is any arbitrary finite value, then the left-hand side of (1.16) will have the value zero and the bound will be correct but vacuous.

Recall that the left-hand variable in the regression is $P_t - P_t^*$ and the right-hand variable is P_t. Let σ_P^2 and σ_{PP*} be their observed moments. Stating the result in terms of the observed variables, we have

Theorem 1.1. The bound on the variance of the noise is

$$\sigma_S^2 \geq \frac{(\sigma_{PP*} - \sigma_P^2)^2}{\sigma_P^2} \tag{1.17}$$

This bound is attainable and is therefore the tightest possible bound.□

The second example shows that the bound is attained by noise that is proportional to the expectation. Absent additional information that rules this form of noise out, the bound cannot be improved.

The contribution of Theorem 1.1 is to show that the simple regression procedure gives the tightest possible bound on the variance of the noise if only the contemporaneous moments are available. Our framework shows how Scott's [1985] simple regression approach is optimal conditional on a particular information set. We also note that the idea that noise can be detected by regressing the difference between the perfect-foresight variable and the actual variable on variables known at the time the actual variable is formed is implicit in the work of other authors such as West [1987] and Campbell-Shiller [1987] and is the basis for the more general results in the next section of this paper.

Inference without the covariance

If the covariance σ_{PP*} is unobservable, then the best available bound on σ_S^2 can be found by finding the least restrictive bound over all choices for this covariance that satisfy the Cauchy-Schwartz inequality, $(\sigma_{PP*})^2 \leq \sigma_P^2 \sigma_{P*}^2$. The Cauchy-Schwartz constraint means that if $\sigma_P^2 \leq \sigma_{P*}^2$, then $\sigma_{PP*} = \sigma_P^2$ is admissable, and the data are consistent with the hypothesis that the model specification error S_t is zero. Alternatively, if $\sigma_P^2 \geq \sigma_{P*}^2$ then σ_S^2 must be positive. The bound in Theorem 1.1 is minimized subject to the Cauchy-Schwartz inequality when $\sigma_{PP*} = \sigma_P \sigma_{P*}$. This establishes

Theorem 1.2. Absent information on the covariance, σ_{PP*}, if $\sigma_P^2 \leq \sigma_{P*}^2$, there is no informative bound on the noise variance, σ_S^2. If $\sigma_P^2 \geq \sigma_{P*}^2$, then

$$\sigma_S^2 \geq (\sigma_P - \sigma_{P*})^2 . \tag{1.18}$$

□

This bound is by necessity less informative than the bound in Theorem 1.1. This derivation formally justifies the excess volatility test derived by Shiller [1981a], in the sense that his test is optimal when the variances are the only observable moments. This derivation also illustrates how the Shiller test has zero power against alternative hypotheses which represent large deviations from the null. His test is incapable of detecting noise that makes the variance of P less than the variance of P^*. Only the covariance can tell whether this situation arises because of expectation errors or because of noise.

2. *Bounds on the noise variance based on observed predictors—the general case*

The results of Section 1 illustrate the powerful information restrictions implicit in both regression and excess volatility tests of model noise. We generalize our results in this section to the restrictions on model noise which are generated by data on some of the information available to the market. In this section, we maintain the assumption that P_t^* has a general form and consequently ν_t has arbitrary time-series properties. In the next two sections, we consider the case where P_t^* is a geometric weighted average of a future variable. In that case, ν_t follows a prescribed AR(1) process.

Throughout our discussion, we shall assume that there exists a vector time series x_t observable to the econometrician which is a strict subset of the information set Φ_t employed by market participants to construct forecasts. Let $L_x(t)$ be the linear space of time series running up to time t, each of which is a linear combination of x_t, x_{t-1}, x_{t-2},... with square-summable weights. We let $M_x(t)$ denote the projection operator for $L_x(t)$. We assume that the vector x_t has a subvector D_t (dividends in the stock market example) and that there is a known function $P_t^*(D)$ of the entire set of past and future values, D, of D_t. As in the previous section, the difference between noise and the expectation error is observed after the fact:

$$P_t - P_t^*(D) = S_t - \nu_t \tag{2.1}$$

Because ν_t is orthogonal to any element in $L_x(t)$, the projection of the observed $P_t - P_t^*$ onto $L_x(t)$ is the same as the projection of the unobserved S_t onto $L_x(t)$. We define $S_{t|t}$ as the fitted values $M_x(t)S_t$. Then we have

$$S_t = S_{t|t} + [1 - M_x(t)]S_t \tag{2.2}$$

Because the residual term $[1 - M_x(t)]S_t$ is orthogonal to the projection, $S_{t|t}$, the variance of the projection is less than the variance of S_t. Hence $\sigma^2_{S_{t|t}}$ is a lower bound for the noise variance, σ^2_S. Further, because the model does not rule out the possibility that the unobserved noise is a linear combination of the observed x_t's (i.e., $S_t \in L_x(t)$), the bound cannot be improved with the available information; it is attained in that case.

In summary, we have

Theorem 2.1 The variance of the model error satisfies

$$\sigma^2_S \geq \sigma^2_{S_{t|t}} \tag{2.3}$$

and this bound is attained when $S_t \in L_x(t)$, a possibility not ruled out by our assumptions.□

The approach of Theorem 2.1 is optimal in an extremely general way. We will show that there is no other way to process the underlying data that will give more information about the variance of the noise term. First, we show that whenever there is a variable, P_t^e, with the property that the discrepancy, $P_t^e - P_t^*(D)$, satisfies an orthogonality condition, $[P_t^e - P_t^*(D)]M_x(t)=0$, then there is an information set Φ_t such that

$$P_t^e \;=\; E[P_t^*(D)|\Phi_t] \;. \tag{2.4}$$

The information set is simply $L_x(t)$ itself and the result follows from the observation that P_t^e is an element of $L_x(t)$ so it can be brought outside the projection. Thus the orthogonality restriction is a complete statement of all the restrictions put on the data by the model.

Now we can show why there is no way to get additional restrictions on the noise variance. We will show that there is a case satisfying all of our assumptions in which the noise vector, element by element, is exactly the fitted value of Theorem 2.1, $S_{t|t}$. Now the P_t^e for that case is $P_t - S_{t|t}$, which satisfies the orthogonality restriction by construction. Because the P_t^e has all the properties required by the model, no manipulation of the data can show that it is inconsistent with the model. Manipulations involving looking partway into the future, or taking second moments of the data, or anything else, cannot rule this case out. Hence there is no prospect of getting a tighter bound on the noise, without making additional assumptions or using additional x-variables. We summarize in

Theorem 2.2 The bound of Theorem 2.1 cannot be improved by any other processing of the data; there is a model with noise equal to $S_{t|t}$ capable of generating all of the observed data and satisfying all the model restrictions.□

The idea of exploiting the unpredictability of a random variable which is a forecast error under some null hypothesis as a specification test appears in a large number of studies in many different areas of research. For example, numerous authors have explored the hypothesis that forward exchange rates are forecasts of future spot exchange rates. Fama [1984], in particular, constructs a regression in an attempt to estimate a time varying risk premium (corresponding to model noise in our framework) which is equivalent to the projection of $P_t - P_t^*$ onto $P_t - P_{t-1}^*$. Tests of foreign exchange market efficiency such as Hansen and Hodrick [1980], Bilson [1981], and Hodrick and Srivastava [1984], to name a few, can be interpreted in our framework as well. Our results demonstrate how the orthogonality of forecast errors represents the total set of testable restrictions in these models. The power properties of all such proposed tests may therefore in principle be ranked on the basis of which information restrictions are imposed.

Theorem 2.1 underscores the results of Frankel and Stock [1987] comparing regression and excess volatility tests. Frankel and Stock demonstrated that regression tests of model misspecification were optimal relative to any excess volatility test which could be constructed with a given information set. Our results extend this optimality by deriving the circumstances under which the regression formulation exhausts all testable implications of the model.

3. *Inference with geometric discounting*

When $P_t^*(D)$ has the particular form of the discounted value of a future stream, with a constant discount rate, further comments on noise variance bounds are in order. Models with this expectational structure include the stock price-dividend relation in Shiller's original work, as well as models of hyperinflations (Cagan [1956], Sargent and Wallace [1973], Sargent [1977]), exchange rates (Meese [1986]), consumption (Flavin [1981]), and output and investment (Hall [1988]). The methods developed in this section could apply to any of these models.

The constant discount model hypothesizes that P_t^e reflects the discounted value of a future flow, D_{t+i}, with a constant discount factor β:

$$P_t^e \;=\; \sum_{i=0}^{\infty} \beta^i E(D_{t+i}|\Phi_t) \; . \tag{3.1}$$

The perfect-foresight variable, P_t^*, has the same form without the expectations. As before, we assume that the observed price, P_t, is the sum, $P_t^e + S_t$, of a term mandated by the model and a noise variable or specification error.

Our primary conclusion about the constant discount model is that the general method of the previous section remains the optimal approach to placing a bound on the amount of specification error or noise.

Although the special form of the model seems to invite exploitation by quasi-differencing, we show that there is nothing to gain by applying any special technique. We also show that the appropriate choice of x-variables means that there is nothing to gain from using lagged residuals in a second-stage regression; the first stage will achieve the theoretical serial correlation of β.

The flow approach

Our work in this paper uses the stock approach—we look for evidence on the level of the discrepancy between the level of the observed variable and the level mandated by the model. Until Shiller, most research took the flow approach—it looked for a discrepancy between the actual expected one-period return and the expected return mandated by the model, namely zero. The earlier empirical work on the constant discount rate model studied the excess return,

$$
\begin{aligned}
r_t &= D_t - P_t + \beta P_{t+1} \\
&= (1 - \beta F)(P_t - P_t^*) \\
&= \eta_t + S_t - \beta S_{t+1} \, . \qquad (3.2)
\end{aligned}
$$

Here F is the forward shift operator and

$$\eta_t = \sum_{i=0}^{\infty} \beta^i \left[E(D_{t+i} | \Phi_{t+1}) - E(D_{t+i} | \Phi_t) \right] , \qquad (3.3)$$

the innovation in the valuation. (D_t is assumed to be observable at $t+1$.) The *intertemporal arbitrage condition* is an implication of the present value model stating that the expectation of the excess holding return conditional on information available in the market is zero. Let

$$N_t \;=\; M_{\mathcal{I}}(t) r_t \;=\; M_{\mathcal{I}}(t)(S_t \;-\; \beta S_{t+1}) \;. \tag{3.4}$$

N_t is a measure of the extent of the failure of the arbitrage condition. Finding a non-zero N_t shows the existence of noise.

Within the framework of this paper, N_t is a measure of the backward quasi-difference, $S_t - \beta S_{t+1}$ of the noise, S_t. As such, it does not directly serve our purpose of measuring the level of noise. In one important case, the noise vanishes from the arbitrage condition: This happens when S_t grows or is expected to grow in proportion to β^{-t}, a speculative bubble. Even when the noise does not take the exact form of a bubble, any slow-moving component will be essentially erased by the backward quasi-difference operation. The problem is particularly acute if the time interval is short, say weekly or monthly. Specification tests will have low power against slow-moving noise if the tests use quasi-differenced data.

The case of a rational bubble is described by,

$$S_t = \beta^{-1} S_{t-1} + \iota_t, \tag{3.5}$$

where ι_t is an unpredictable random variable with zero mean. The reason that the flow method breaks down completely in this case is that S_t cannot be recovered from the future values of ι_t; there is a term $\lim_{T\to\infty} \beta^T S_T$ in the expression for S_t that does not vanish. In other words, S_t is not an element of $L_\iota(\infty)-L_\iota(t-1)$. Note that the bubble is unbounded in variance whereas the linear space of values of S_t reconstructed after filtering contains only bounded elements. A procedure that used projections of future r_t to construct estimates of model noise cannot improve the bound (according to Theorem 2.2), and if nonequivalence of the type just described occurs, such a procedure would weaken the bound.

An example of an inferior approach is the following: Given a time series for N_t, calculate a corresponding estimated noise series, $\hat{S}_t$, from the recursion,

$$\hat{S}_t = \beta \hat{S}_{t+1} + N_t \ . \tag{3.6}$$

Although $\hat{S}_t$ is potentially useful as an indication of the amount of noise in the level of P_t, its variance is not necessarily less than the variance of the actual noise, S_t. Because of its dependence on future N's, it does not satisfy the basic orthogonality condition needed to prove that its variance places a lower bound on the noise variance. One could compute the projection of $\hat{S}_t$ on variables known at time t, resulting in a measure $\hat{S}_{t|t} = M_x(t)\hat{S}_t$ that has a variance less than the noise variance. However, this measure is inferior to the direct projection developed in the previous section.

West [1988] proposed a test of the specification of the constant geometric stock price model which essentially consists in comparing the variance of η_t defined in (3.2) to the variance of

$$\hat{\eta}_t \quad = \quad (1 - \beta F)E[(P_t - P_t^* | L)] \ . \tag{3.7}$$

The information set, L, can in principle be any subset of agents' information which includes current and lagged dividends. In the absence of noise, the variance of $\hat{\eta}_t$ must at least as great as the variance of η_t, because the two random variables are white noise transformations of forecast errors based upon nested information sets. The failure of this inequality supports a conclusion of specification error or noise in the flow sense. Our analysis shows that this test only captures a subset of the many implications for innovations of the constant geometric discount model. The complete set of implications can be expressed as the orthogonality of $P_t - P_t^*$ to all of the variables known to agents at the time P_t is determined. Durlauf and Hall [1988a] further demonstrate that in practice, the dividend innovations test is an inefficient way of recovering model noise.

Choice of regressors for the noise bound with geometric discounting

Recall that our basic approach to finding lower bounds on the noise variance is to find variables in the information set of market participants that have predictive power for $P_t - P_t^*$. The explanatory power of these regressors raises and thus sharpens the bound on the variance of the noise. Our discussion in Section 1 stressed the importance of including P_t. We will now show the value of including P and D lagged 1, 2, ... periods as additional regressors for the constant-discount case. For the rest of this section, we assume that these lagged variables are included in the x-variables. The explanatory power arises from the presence of lagged noise, S_{t-i}, in the lagged variables, which helps predict the noise, S_t, in $P_t - P_t^*$. Theorem 2.2 then establishes that the regression on this set of variables gives the best possible bound.

Information contained in regression residuals

It might appear that a second way to add explanatory variables to the regression to measure noise could be based on the observation that, absent noise, the disturbance in the regression is AR(1) with parameter β. To the extent that the actual serial correlation of the residual is different, added predictive power would be available. However, it turns out that adding lagged P and lagged D, or, equivalently, the lagged excess returns, as regressors already exploits all of that source of predictive power.

To demonstrate this proposition, we will show that the residuals are already serially correlated with parameter β once lagged P and lagged D are included in the regressors. The residuals are

$$v_t \;=\; [1 - M_x(t)](P_t - P_t^*) \;. \tag{3.8}$$

Let z_t be the backward autoregressive transformation of the residuals:

$$z_t \;=\; (1-\beta F)v_t \;=\; P_t - \beta P_{t+1} - D_t - (1-\beta F)M_x(t)(P_t - P_t^*) \tag{3.9}$$

We seek to show that z_t is serially uncorrelated. If lagged values of z are in the space of the regressors, $L_X(t)$, then the residual, v_t, will be orthogonal to the lagged zs. Moreover, v_{t+1} will also be orthogonal. Hence z_t will be orthogonal to its own lagged values and thus will be serially uncorrelated. Our assumption that lagged Ps and Ds are among the regressors guarantees that the first half of z_t in (3.9) is in $L_X(t)$. The second half involves a projection onto $L_X(t)$ and so must lie in $L_X(t)$. Thus we have

Theorem 3.1 With geometric discounting, if the set of regressors for the noise-detecting regression includes the lagged values of P and D, then the residuals from the regression are AR(1) with parameter β; there is no remaining predictive power from lagged transformed residuals.□

Note that the particular way that the lagged P and D enter the lagged transformed residuals is through the excess return, r_t. Another way to express the requirement that there be no remaining predictive power in the lagged residuals is that the lagged excess returns be included in the regressors. As a practical matter, our procedure boils down to running a regression of the form,

$$P_t - P_t^* = \pi(L)x_t + \gamma(L)r_{t-1} + v_t \quad . \tag{3.10}$$

Then the fitted value $\pi(L)x_t + \gamma(L)r_{t-1}$ is the most informative measure of the noise in the model and its variance is a good lower bound on the true noise variance.

Theorem 3.1 provides two ways in which to think about the efficiency of specification tests. First, a noise estimate is *informationally efficient* if all variables observable to the econometrician are employed in forming the noise bound. Second, conditional on a given $P_t - P_t^*$ and $L_x(t)$, a noise estimate is an *optimal smoothing estimator* if the history of excess returns is contained in $L_x(t)$. The latter notion of efficiency is consistent with the interpretation of our results in a Kalman smoothing framework, an idea which is explored in Durlauf and Hall [1988b]. We emphasize that the serial correlation of the residuals from the regressions recommended in this paper is not a nuisance issue but is intrinsic to specification testing. The noise-detecting regression is not complete if it does not satisfy the theoretical serial correlation requirement, both in the case of geometric discounting and in other cases. In Section 4 we consider one other case.

Related work

Campbell and Shiller [1987] studied the behavior of excess holding returns in the dividend stock price model. These authors treated the present discounted value of expected excess holding returns as one measure of violations of the model. Absent bubbles, it is apparent from equation (3.2) that this present discounted value is just our noise variable S_t. Our analysis provides a precise metric which justifies the use of the measure in assessing model deviations and also demonstrates how the measure may serve as a comprehensive model specification test.

Fama and French [1988] and Lo and MacKinlay [1987] have argued in a related context that the stock approach also possesses superior power properties in finite samples in uncovering model noise to the flow approach. Specifically, they demonstrate that for a time series exhibiting long run mean reversion, the first differences of the series might appear to be uncorrelated at short lags. These authors conclude that the behavior of long changes in the series is a superior way of uncovering noise, as opposed to analyzing first differences in isolation. The first differences correspond to our measure of excess holding returns. Of course, our results deal exclusively with asymptotic noise measurement, and therefore say nothing about finite sample properties. For actual empirical work, the critique based on finite sample considerations may very well be more important in justifying the use of the stock approach.

The case of an unidentified discount rate

The previous discussion has assumed that the discount rate, β, is known. In effect, we are assuming that it is econometrically identified—all of our previous results would apply if β were not known from prior considerations but could be estimated by the use of an instrumental variable known to be orthogonal to the noise in the equation at hand. If the noise is known to have zero mean, then the constant can serve as an instrument; in that case, the instrumental variable estimator of β is just the value that sets the sample mean of $P_t - D_t - \beta P_{t+1}$ to zero. Absent the assumption that noise has zero mean, a truly exogenous instrument is needed. An example of such a variable in a macroeconomic setting might be military spending, which is unlikely to be a response to noise.

A more challenging task is putting bounds on noise when the discount rate is not identified. When noise is present, the standard approach to estimation of β will fail. That approach is to use the set of variables known in the market at time t as instruments in the excess return equation,

$$P_t - D_t \;=\; \beta P_{t+1} + \eta_t + S_t - \beta S_{t+1} \,. \tag{3.11}$$

The approach will work only if $S_t - \beta S_{t+1}$ is zero or uncorrelated with the instruments.

Although the value of β cannot be known if the identifying condition fails, it is still possible to put a bound on the noise variance. The idea we use is that the residuals from applying two-stage least squares to (3.11) provide information about S_t even though TSLS does not provide any information about β. We can exploit the exclusion restrictions in (3.11) even though β is not identified. The idea of TSLS is to form an instrument, z_t, by projecting the right-hand endogenous variable, P_{t+1}, onto the x-variables:

$$z_t \;=\; M_x(t) P_{t+1} \,. \tag{3.12}$$

It can be shown that the projection of the TSLS residuals on the x-variables is

$$\hat{N}_t = [M_x(t) - M_z(t)](S_t - \beta S_{t+1}) \ . \qquad (3.13)$$

Comparing this to the similar expression for the flow noise estimate when β is known, equation (3.4), we see that they differ by the term $M_z(t)(S_t - \beta S_{t+1})$ representing the projection of the flow noise on the instrument. Because of this term, the variance of $\hat{N}_t$ is necessarily no more than the variance of N_t and the flow noise bound is established.

As we noted earlier, there is no close connection between information about the flow or quasi-difference of noise and the level of noise. Again, the noise estimate $\hat{N}_t$ can be cumulated to get a level estimate, but that cumulation does not give rise to a bound on the variance of the noise. Projecting the cumulation on the x-variables gives a bound, but we have not been able to show that the bound is tight.

4. *Inference with overlapping forecast errors*

In numerous expectations based models, the data on forecasts and realizations are both available and observable, but the horizons for forecasts overlap. For example, every month there may be a k-month ahead forward exchange rate observation. As emphasized by Hansen and Hodrick [1980], this implies that the forecast errors ν_t will be MA(k). We will show that the application of our general method of projecting $P_t - P_t^*$ onto appropriate x-variables is optimal. As in the case of constant discounting of the infinite future, the variables should include lagged P, D, and x. With these variables, the residuals from the projection will satisfy the MA(k) constraint. Recall that the residual is

$$v_t \;=\; [1 - M_x(t)]S_t - \nu_t \;. \tag{4.1}$$

We seek to show that the residuals satisfy the MA(k) restriction,

$$Cov(v_t, v_{t-j}) \;=\; 0, \quad \forall\; j>k \;. \tag{4.2}$$

or

$$Cov\{[1 - M_x(t)]S_t - \nu_t, v_{t-j}\} \;=\; 0 \;. \tag{4.3}$$

With respect to the first part of the covariance, we note that $v_{t-j} \in L_x$ because $v_{t-j} = P_{t-j} - P^*_{t-j} - M_x(t-j)(P_{t-j} - P^*_{t-j})$ and these are included in the x-variables by assumption, since all of the ingredients in P^*_{t-j} are observed after k periods. With respect to the second part, $Cov(\nu_t, v_{t-j}) = 0$ by rational expectations.

5. *Relation to the West-Casella specification test*

The bounds we have developed bear a close relation to the test developed by West [1987] and Casella [1988], which was originally applied to stock prices and hyperinflations and has been subsequently used to analyze exchange rates (Meese [1986]). The idea is to test the model by comparing the reduced form coefficients in the regression

$$P_t = \pi(L)D_t + \mu_t \tag{5.1}$$

to with the coefficients predicted by the model

$$P_t^e = \sum_{i=0}^{\infty} \beta^i E(D_{t+i}|\Phi_t) \tag{5.2}$$

$$D_t = \gamma(L)\delta_t \tag{5.3}$$

where P_t^e denotes the fundamental stock price (which equals P_t absent noise) and δ_t denotes the Wold innovation in D_t. West and Casella test the model specification by comparing the projections of P_t and P_t^e onto $L_D(t)$. That is, if the projection of $P_t - P_t^e$ onto $L_D(t)$ differs from zero, the model is mis-specified. Now the projections of $P_t - P_t^*$ and $P_t - P_t^e$ onto $L_D(t)$ are identical, because the two time series can differ only by the expectation error which is necessarily orthogonal to the current and lagged dividend series. We conclude that the West-Casella analysis is equivalent to a procedure which constructs a lower model noise bound based upon the information set $L_D(t)$.

The West-Casella test therefore exploits only a subset of the testable implications of the stock price model. The omission of the current market price, P_t, means that the West-Casella test is probably weaker than a one-variable regression test based only on P_t for uncovering model noise. Expansion of the information set to include current and lagged prices as well as dividends is essential for a powerful noise test. The omission of prices is a particularly serious mistake if the specification error contains a rational bubble, as a projection of stock prices onto the dividend information set will generate sample error variances which diverge to infinity, regardless of whether dividends are an integrated process. Durlauf [1988] elaborates this point.

These theoretical observations are supported by the data. In Durlauf and Hall [1988a], we demonstrate that the entire history of dividends provides relatively little information on the nature of stock price noise, even when compared to the contemporaneous price in isolation. In fact, excess volatility tests generate greater estimates of noise variance than dividend based tests. Our empirical research has found the history of prices alone to be extremely effective in capturing nearly all potential model noise.

6. Summary and conclusions

This paper has explored a number of issues in assessing the degree of misspecification or model noise in expectations based models. We have derived a lower bound on the variance of misspecification or noise in expectational models, as opposed to merely detecting its presence. Our approach has treated the model noise and market expectations as the objects of interest in an unobserved components problem. By varying the information set available to the econometrician, we have been able to characterize different conditionally optimal lower bounds on model noise. These characterizations have shown how various specification tests have simple regression interpretations.

We have shown how for constant geometric discount models, widely used in research on asset prices, the autocovariance structure of regression residuals provides a guide to the measurement of noise or specification error. Our optimal bound results further permitted a characterization of all testable implications of the model.

Finally, we have compared a number of different asset price tests in the literature in terms of power. We have found that the original Shiller excess volatility test has good power against many alternative hypotheses, when compared to dividend-based specification tests because it makes partial use of the current market price. The reason is simple. If a stock price contains slowly moving noise, the history of prices is an effective way of detecting the noise. Dividends may not be correlated with the noise at all.

Areas for future research, which we hope to pursue, fall into two categories. First, under additional assumptions about the noise variable, instrumental variables procedures may be available to improve the noise bounds. Second, the noise bounds may be generalized to multivariate systems. The relationship between noise in different markets could then be explored. An important undertaking would attempt to relate the model noise estimates for aggregate stock prices to the model noise estimates in other asset markets as well as commodity and labor markets (as found by Hall [1988]) in order to begin to develop a more complete characterization of the limitations of current expectations based theories.

Bibliography

Bilson, J. F. O. [1981]. "The Speculative Efficiency Hypothesis." *Journal of Business*, 54, 435-451.

Cagan, P. [1956]. "The Monetary Dynamics of Hyperinflation." in *Studies in the Quantity Theory of Money*, ed. M. Friedman, Chicago: University of Chicago.

Campbell, J. Y. and R. J. Shiller. [1987]. "Cointegration and Tests of Present Value Models." *Journal of Political Economy*, 95, 1062-88.

Casella. A. [1988]. "Price Bubbles and the German Hyperinflation." mimeo. UC Berkeley.

Durlauf, S. N. [1988]. "Sample Path Properties of Speculative Bubbles." Working Paper, Stanford University.

_________ and R. E. Hall. [1988a]. "Determinants of Noise in the Dividend Based Stock Price Model." Working Paper, Stanford University.

_________. [1988b]. "A Kalman Smoothing Approach to Recovering Model Noise in Expectations Based Models." Working Paper in progress, Stanford University.

Fama, E. [1984]. "Forward and Spot Exchange Rates." *Journal of*

Monetary Economics, 14, 319-338.

_________ and K. French. [1987]. "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, 96, 246-273.

Flavin, M. [1981]. "The Adjustment of Consumption to Changing Expectations about Future Income." *Journal of Political Economy*, 89, 974-1009.

Frankel, J. A. and J. H. Stock. [1987]. "Regression Versus Volatility Tests of the Efficiency of Foreign Exchange Markets." *Journal of International Money and Finance*, 6, 49-56.

Gilles, C. and S. F. LeRoy. [1987]. "The Variance Bounds Tests: A Critical Survey." Working paper, UC Santa Barbara.

Hall, R. E. [1988]. "Spontaneous Volatility of Output in the Business Sector." Lecture 2 of the Okun Memorial Lectures presented at Yale University, October.

Hansen, L. P. and R. J. Hodrick. [1980]. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis." *Journal of Political Economy, 88, 829-853.*

Hodrick, R. J. and S. Srivastava. [1984]. "An Investigation of Risk and Return in Forward Foreign Exchange." *Journal of International Money and*

Finance, 3. 5-29.

LeRoy, S. and R. Porter. [1981]. "The Present Value Relation: Tests Based on Implied Variance Bounds." *Econometrica*, LXIX, 555-74.

Lo, A. W. and A. C. MacKinlay. [1987]. "Stock Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." Working Paper 29-87, Rodney L. White Center for Financial Research, Wharton School, University of Pennsylvania.

Meese, R. [1986]. "Testing for Bubbles in Exchange Markets: A Case of Sparkling Rates?" *Journal of Political Economy*, 94, 345-73.

Sargent, T. [1977]. "The Demand for Money During Hyperinflations Under Rational Expectations." *International Economic Review*, 18, 59-82.

________ and N. Wallace. [1973]. "Rational Expectations and the Dynamics of Hyperinflation." *International Economic Review*, 14, 328-50.

Scott, L. O. [1985]. "The Present Value Model of Stock Prices: Regression Tests and Monte Carlo Results." *Review of Economics and Statistics*, 599-605.

Shiller, R. J. [1981a]. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, LXXI, 421-36.

__________ [1981b]. "The Use of Volatility Measures in Assessing Market Efficiency." *Journal of Finance*, XXXV, 291-304.

West, K. [1987]. "A Specification Test for Speculative Bubbles." *Quarterly Journal of Economics,* CII, 553-80.

__________ [1988]. "Dividend Innovations and Stock Price Volatility." *Econometrica*, 56, 37-62.