Chapter 33

EVALUATING THE PREDICTIVE ACCURACY OF MODELS

RAY C. FAIR

Contents

Handbook of Econometrics, Volume III, Edited by Z. Griliches and M.D. Intriligator
© Elsevier Science Publishers BV, 1986

1. Introduction

Methods for evaluating the predictive accuracy of econometric models are discussed in this chapter. Since most models used in practice are nonlinear, the nonlinear case will be considered from the beginning. The model is written as:

$$f_i(y_t, x_t, \alpha_i) = u_{it}, \qquad (i=1,\ldots,n), \qquad (t=1,\ldots,T), \tag{1}$$

where y_t is an n-dimensional vector of endogenous variables, x_t is a vector of predetermined variables (including lagged endogenous variables), α_i is a vector of unknown coefficients, and u_{it} is the error term for equation i for period t. The first m equations are assumed to be stochastic, with the remaining $u_{it}(i = m+1,\ldots,n)$ identically zero for all t.

The emphasis in this chapter is on methods rather than results. No attempt is made to review the results of comparing alternative models. This review would be an enormous undertaking and is beyond the scope of this *Handbook*. Also, as will be argued, most of the methods that have been used in the past to compare models are flawed, and so it is not clear that an extensive review of results based on these methods is worth anyone's effort. The numerical solution of nonlinear models is reviewed in Section 2, including stochastic simulation procedures. This is background material for the rest of the chapter. The standard methods that have been used to evaluate ex ante and ex post predictive accuracy are discussed in Sections 3 and 4, respectively. The main problems with these methods, as will be discussed, are that they (1) do not account for exogenous variable uncertainty, (2) do not account for the fact that forecast-error variances vary across time, and (3) do not treat the possible existence of misspecification in a systematic way. Section 5 discusses a method that I have recently developed that attempts to handle these problems, a method based on successive reestimation and stochastic simulation of the model. Section 6 contains a brief conclusion.

It is important to note that this chapter is not a chapter on forecasting techniques. It is concerned only with methods for *evaluating* and *comparing* econometric models with respect to their predictive accuracy. The use of these methods should allow one (in the long run) to decide which model best approximates the true structure of the economy and how much confidence to place on the predictions from a given model. The hope is that one will end up with a model that for a wide range of loss functions produces better forecasts than do other techniques. At some point along the way one will have to evaluate and compare other methods of forecasting, but it is probably too early to do this. At any rate, this issue is beyond the scope of this chapter.[1]

[1] For a good recent text on forecasting techniques for time series, see Granger and Newbold (1977).

2. Numerical solution of nonlinear models

The Gauss–Seidel technique is generally used to solve nonlinear models. [See Chapter 14 (Quandt) for a discussion of this technique.] Given a set of estimates of the coefficients, given values for the predetermined variables, and given values for the error terms, the technique can be used to solve for the endogenous variables. Although in general there is no guarantee that the technique will converge, in practice it has worked quite well.

A "static" simulation is one in which the actual values of the predetermined variables are used for the solution each period. A "dynamic" simulation is one in which the predicted values of the endogenous variables from the solutions for previous periods are used for the values of the lagged endogenous variables for the solution for the current period. An "ex post" simulation or forecast is one in which the actual values of the exogenous variables are used. An "ex ante" simulation or forecast is one in which guessed values of the exogenous variables are used. A simulation is "outside-sample" if the simulation period is not included within the estimation period; otherwise the simulation is "within-sample." In forecasting situations in which the future is truly unknown, the simulations must be ex ante, outside-sample, and (if the simulation is for more than one period) dynamic.

If one set of values of the error terms is used, the simulation is said to be "deterministic." The expected values of most error terms in most models are zero, and so in most cases the errors terms are set to zero for the solution. Although it is well known [see Howrey and Kelejian (1971)] that for nonlinear models the solution values of the endogenous variables from deterministic simulations are not equal to the expected values of the variables, in practice most simulations are deterministic. It is possible, however, to solve for the expected values of the endogenous variables by means of "stochastic" simulation, and this procedure will now be described. As will be seen later in this chapter, stochastic simulation is useful for purposes other than merely solving for the expected values.

Stochastic simulation requires that an assumption be made about the distributions of the error terms and the coefficient estimates. In practice these distributions are almost always assumed to be normal, although in principle other assumptions can be made. For purposes of the present discussion the normality assumption will be made. In particular, it is assumed that $u_t = (u_{1t}, \ldots, u_{mt})'$ is independently and identically distributed as multivariate $N(0, \Sigma)$. Given the estimation technique, the coefficient estimates, and the data, one can estimate the covariance matrix of the error terms and the covariance matrix of the coefficient estimates. Denote these two matrices as $\hat{\Sigma}$ and $\hat{V}$, respectively. The dimension of $\hat{\Sigma}$ is $m \times m$, and the dimension of $\hat{V}$ is $K \times K$, where K is the total number of coefficients in the model. $\hat{\Sigma}$ can be computed as $(1/T)\hat{U}\hat{U}'$, where $\hat{U}$ is the $m \times T$ matrix of values of the estimated error terms. The computation of $\hat{V}$ depends on

the estimation technique used. Given $\hat{V}$ and given the normality assumption, an estimate of the distribution of the coefficient estimates is $N(\hat{\alpha}, \hat{V})$, where $\hat{\alpha}$ is the $K \times 1$ vector of the coefficient estimates.

Let u_t^* denote a particular draw of the m error terms for period t from the $N(0, \hat{\Sigma})$ distribution, and let α^* denote a particular draw of the K coefficients from the $N(\hat{\alpha}, \hat{V})$ distribution. Given u_t^* for each period t of the simulation and given α^*, one can solve the model. This is merely a deterministic simulation for the given values of the error terms and coefficients. Call this simulation a "trial". Another trial can be made by drawing a new set of values of u_t^* for each period t and a new set of values of α^*. This can be done as many times as desired. From each trial one obtains a prediction of each endogenous variable for each period. Let $\tilde{y}_{itk}^j$ denote the value on the jth trial of the k-period-ahead prediction of variable i from a simulation beginning in period t.[2] For J trials, the estimate of the expected value of the variable, denoted $\tilde{\tilde{y}}_{itk}$, is:

$$\tilde{\tilde{y}}_{itk} = \frac{1}{J} \sum_{j=1}^{J} \tilde{y}_{itk}^j. \tag{2}$$

In a number of studies stochastic simulation with respect to the error terms only has been performed, which means drawing only from the distribution of the error terms for a given trial. These studies include Nagar (1969); Evans, Klein, and Saito (1972); Fromm, Klein, and Schink (1972); Green, Liebenberg, and Hirsch (1972); Sowey (1973); Cooper and Fischer (1972); Cooper (1974); Garbade (1975); Bianchi, Calzolari, and Corsi (1976); and Calzolari and Corsi (1977). Studies in which stochastic simulation with respect to both the error terms and coefficient estimates has been performed include Cooper and Fischer (1974); Schink (1971), (1974); Haitovsky and Wallace (1972); Muench, Rolnick, Wallace, and Weiler (1974); and Fair (1980).

One important empirical conclusion that can be drawn from stochastic simulation studies to date is that the values computed from deterministic simulations are quite close to the mean predicted values computed from stochastic simulations. In other words, the bias that results from using deterministic simulation to solve nonlinear models appears to be small. This conclusion has been reached by Nagar (1969), Sowey (1973), Cooper (1974), Bianchi, Calzolani, and Corsi (1976), and Calzolani and Corsi (1977) for stochastic simulation with respect to the error terms only and by Fair (1980) for stochastic simulation with respect to both error terms and coefficients.

A standard way of drawing values of α^* from the $N(\hat{\alpha}, \hat{V})$ distribution is to (1) factor numerically (using a subroutine package) $\hat{V}$ into PP', (2) draw (again using

[2]Note that t denotes the first period of the simulation, so that $\tilde{y}_{itk}^j$ is the prediction for period $t+k-1$.

a subroutine package) K values of a standard normal random variable with mean 0 and variance 1, and (3) compute α^* as $\hat{\alpha} + Pe$, where e is the $K \times 1$ vector of the standard normal draws. Since $Eee' = I$, then $E(\alpha^* - \hat{\alpha})(\alpha^* - \hat{\alpha})' = EPee'P' = \hat{V}$, which is as desired for the distribution of α^*. A similar procedure can be used to draw values of u_t^* from the $N(0, \hat{\Sigma})$ distribution: $\hat{\Sigma}$ is factored into PP', and u_t^* is computed as Pe, where e is a $m \times 1$ vector of standard normal draws.

An alternative procedure for drawing values of the error terms, due to McCarthy (1972), has also been used in practice. For this procedure one begins with the $m \times T$ matrix of estimated error terms, $\hat{U}$. T standard normal random variables are then drawn, and u_t^* is computed as $T^{-1/2}\hat{U}e$, where e is a $T \times 1$ vector of the standard normal draws. It is easy to show that the covariance matrix of u_t^* is $\hat{\Sigma}$, where, as above, $\hat{\Sigma}$ is $(1/T)\hat{U}\hat{U}'$.

An alternative procedure is also available for drawing values of the coefficients. Given the estimation period (say, 1 through T) and given $\hat{\Sigma}$, one can draw T values of $u_t^*(t = 1, \ldots, T)$. One can then add these errors to the model and solve the model over the estimation period (static simulation, using the original values of the coefficient estimates). The predicted values of the endogenous variables from this solution can be taken to be a new data base, from which a new set of coefficients can be estimated. This set can then be taken to be one draw of the coefficients. This procedure is more expensive than drawing from the $N(\hat{\alpha}, \hat{V})$ distribution, since reestimation is required for each draw, but it has the advantage of not being based on a fixed estimate of the distribution of the coefficient estimates. It is, of course, based on a fixed value of $\hat{\Sigma}$ and a fixed set of original coefficient estimates.

It should finally be noted with respect to the solution of models that in actual forecasting situations most models are subjectively adjusted before the forecasts are computed. The adjustments take the form of either using values other than zero for the future error terms or using values other than the estimated values for the coefficients. Different values of the same coefficient are sometimes used for different periods. Adjusting the values of constant terms is equivalent to adjusting values of the error terms, given that a different value of the constant term can be used each period.[3] Adjustments of this type are sometimes called "add factors". With enough add factors it is possible, of course, to have the forecasts from a model be whatever the user wants, subject to the restriction that the identities must be satisfied. Most add factors are subjective in that the procedure by which they were chosen cannot be replicated by others. A few add factors are objective. For example, the procedure of setting the future values of the error terms equal to the average of the past two estimated values is an objective one. This procedure,

[3]Although much of the discussion in the literature is couched in terms of constant-term adjustments, Intriligator (1978, p. 516) prefers to interpret the adjustments as the user's estimates of the future values of the error terms.

along with another type of mechanical adjustment procedure, is used for some of the results in Haitovsky, Treyz, and Su (1974). See also Green, Liebenberg, and Hirsch (1972) for other examples.

3. Evaluation of ex ante forecasts

The three most common measures of predictive accuracy are root mean squared error (RMSE), mean absolute error (MAE), and Theil's inequality coefficient[4] (U). Let $\hat{y}_{it}$ be the forecast of variable i for period t, and let y_{it} be the actual value. Assume that observations on $\hat{y}_{it}$ and y_{it} are available for $t = 1, \ldots, T$. Then the measures for this variable are:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it})^2}, \tag{3}$$

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y_{it} - \hat{y}_{it}|, \tag{4}$$

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{it} - \Delta \hat{y}_{it})^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{it})^2}}, \tag{5}$$

where Δ in (5) denotes either absolute or percentage change. All three measures are zero if the forecasts are perfect. The MAE measure penalizes large errors less than does the RMSE measure. The value of U is one for a no-change forecast ($\Delta \hat{y}_{it} = 0$). A value of U greater than one means that the forecast is less accurate than the simple forecast of no change.

An important practical problem that arises in evaluating ex ante forecasting accuracy is the problem of data revisions. Given that the data for many variables are revised a number of times before becoming "final", it is not clear whether the forecast values should be compared to the first-released values, to the final values, or to some set in between. There is no obvious answer to this problem. If the revision for a particular variable is a benchmark revision, where the level of the variable is revised beginning at least a few periods before the start of the prediction period, then a common procedure is to adjust the forecast value by

[4] See Theil (1966, p. 28).

adding the forecasted change ($\Delta \hat{y}_{it}$), which is based on the old data, to the new lagged value (y_{it-1}) and then comparing the adjusted forecast value to the new data. If, say, the revision took the form of adding a constant amount $\bar{y}_i$ to each of the old values of y_{it}, then this procedure merely adds the same $\bar{y}_i$ to each of the forecasted values of y_{it}. This procedure is often followed even if the revisions are not all benchmark revisions, on the implicit assumption that they are more like benchmark revisions than other kinds. Following this procedure also means that if forecast changes are being evaluated, as in the U measure, then no adjustments are needed.

There are a number of studies that have examined ex ante forecasting accuracy using one or more of the above measures. Some of the more recent studies are McNees (1973, 1974, 1975, 1976) and Zarnowitz (1979). It is usually the case that forecasts from both model builders and nonmodel builders are examined and compared. A common "base" set of forecasts to use for comparison purposes is the set from the ASA/NBER Business Outlook Survey. A general conclusion from these studies is that there is no obvious "winner" among the various forecasters [see, for example, Zarnowitz (1979, pp. 23, 30)]. The relative performance of the forecasters varies considerably across variables and length ahead of the forecast, and the differences among the forecasters for a given variable and length ahead are generally small. This means that there is yet little evidence that the forecasts from model builders are more accurate than, say, the forecasts from the ASA/NBER Survey.

Ex ante forecasting comparisons are unfortunately of little interest from the point of view of examining the predictive accuracy of models. There are two reasons for this. The first is that the ex ante forecasts are based on guessed rather than actual values of the exogenous variables. Given only the actual and forecast values of the endogenous variables, there is no way of separating a given error into that part due to bad guesses and that part due to other factors. A model should not necessarily be penalized for bad exogenous-variable guesses from its users. More will be said about this in Section 5. The second, and more important, reason is that almost all the forecasts examined in these studies are generated from subjectively adjusted models, (i.e. subjective add factors are used). It is thus the accuracy of the forecasting performance of the model builders rather than of the models that is being examined.

Before concluding this section it is of interest to consider two further points regarding the subjective adjustment of models. First, there is some indirect evidence that the use of add factors is quite important in practice. The studies of Evans, Haitovsky, and Treyz (1972) and Haitovsky and Treyz (1972) analyzing the Wharton and OBE models found that the ex ante forecasts from the model builders were more accurate than the ex post forecasts from the models, even when the same add factors that were used for the ex ante forecasts were used for the ex post forecasts. In other words, the use of actual rather than guessed values

of the exogenous variables decreased the accuracy of the forecasts. This general conclusion can also be drawn from the results for the BEA model in Table 3 in Hirsch, Grimm, and Narasimham (1974). This conclusion is consistent with the view that the add factors are (in a loose sense) more important than the model in determining the ex ante forecasts: what one would otherwise consider to be an improvement for the model, namely the use of more accurate exogenous-variable values, worsens the forecasting accuracy.

Second, there is some evidence that the accuracy of non-subjectively adjusted ex ante forecasts is improved by the use of actual rather than guessed values of the exogenous variables. During the period 1970III–1973II, I made ex ante forecasts using a short-run forecasting model [Fair (1971)]. No add factors were used for these forecasts. The accuracy of these forecasts is examined in Fair (1974), and the results indicate that the accuracy of the forecasts is generally improved when actual rather than guessed values of the exogenous variables are used.

It is finally of interest to note, although nothing really follows from this, that the (non-subjectively adjusted) ex ante forecasts from my forecasting model were on average less accurate than the subjectively adjusted forecasts [McNees (1973)], whereas the ex post forecasts, (i.e. the forecasts based on the actual values of the exogenous variables) were on average about the same degree of accuracy as the subjectively adjusted forecasts [Fair (1974)].

4. Evaluation of ex post forecasts

The measures in (3)–(5) have also been widely used to evaluate the accuracy of ex post forecasts. One of the more well known comparisons of ex post forecasting accuracy is described in Fromm and Klein (1976), where eleven models are analyzed. The standard procedure for ex post comparisons is to compute ex post forecasts over a common simulation period, calculate for each model and variable an error measure, and compare the values of the error measure across models. If the forecasts are outside-sample, there is usually some attempt to have the ends of the estimation periods for the models be approximately the same. It is generally the case that forecasting accuracy deteriorates the further away the forecast period is from the estimation period, and this is the reason for wanting to make the estimation periods as similar as possible for different models.

The use of the RMSE measure, or one of the other measures, to evaluate ex post forecasts is straightforward, and there is little more to be said about this. Sometimes the accuracy of a given model is compared to the accuracy of a "naive" model, where the naive model can range from the simple assumption of no change in each variable to an autoregressive moving average (ARIMA) process for each variable. (The comparison with the no-change model is, of course,

already implicit in the U measure.) It is sometimes the case that turning-point observations are examined separately, where by "turning point" is meant a point at which the change in a variable switches sign. There is nothing inherent in the statistical specification of models that would lead one to examine turning points separately, but there is a strand of the literature in which turning-point accuracy has been emphasized.

Although the use of the RMSE or similar measure is widespread, there are two serious problems associated with the general procedure. The first concerns the exogenous variables. Models differ both in the number and types of variables that are taken to be exogenous and in the sensitivity of the predicted values of the endogenous variables to the exogenous-variable values. The procedure does not take these differences into account. If one model is less "endogenous" than another (say that prices are taken to be exogenous in one model but not in another), then it has an unfair advantage in the calculation of the error measures. The other problem concerns the fact that forecast error variances vary across time. Forecast error variances vary across time both because of nonlinearities in the model and because of variation in the exogenous variables. Although RMSEs are in some loose sense estimates of the averages of the variances across time, no rigorous statistical interpretation can be placed on them: they are not estimates of any parameters of the model.

There is another problem associated with within-sample calculations of the error measures, which is the possible existence of data mining. If in the process of constructing a model one has, by running many regressions, searched diligently for the best fitting equation for each variable, there is a danger that the equations chosen, while providing good fits within the estimation period, are poor approximations to the true structure. Within-sample error calculations are not likely to discover this, and so they may give a very misleading impression of the true accuracy of the model. Outside-sample error calculations should, of course, pick this up, and this is the reason that more weight is generally placed on outside-sample results.

Nelson (1972) used an alternative procedure in addition to the RMSE procedure in his ex post evaluation of the FRB-MIT-PENN (FMP) model. For each of a number of endogenous variables he obtained a series of static predictions using both the FMP model and an ARIMA model. He then regressed the actual value of each variable on the two predicted values over the period for which the predictions were made. Ignoring the fact that the FMP model is nonlinear, the predictions from the model are conditional expectations based on a given information set. If the FMP model makes efficient use of this information, then no further information should be contained in the ARIMA predictions. The ARIMA model for each variable uses only a subset of the information, namely, that contained in the past history of the variable. Therefore, if the FMP model has made efficient use of the information, the coefficient for the ARIMA

predicted values should be zero. Nelson found that in general the estimates of this coefficient were significantly different from zero. This test, while interesting, cannot be used to compare models that differ in the number and types of variables that are taken to be exogenous. In order to test the hypothesis of efficient information use, the information set used by one model must be contained in the set used by the other model, and this is in general not true for models that differ in their exogenous variables.

5. An alternative method for evaluating predictive accuracy

The method discussed in this section takes account of exogenous-variable uncertainty and of the fact that forecast error variances vary across time. It also deals in a systematic way with the question of the possible misspecification of the model. It accounts for the four main sources of uncertainty of a forecast: uncertainty due to (1) the error terms, (2) the coefficient estimates, (3) the exogenous-variable forecasts, and (4) the possible misspecification of the model. The method is discussed in detail in Fair (1980). The following is an outline of its main features.

Estimating the uncertainty from the error terms and coefficients can be done by means of stochastic simulation. Let σ_{itk}^2 denote the variance of the forecast error for a k-period-ahead forecast of variable i from a simulation beginning in period t. Given the J trials discussed in Section 2, a stochastic-simulation estimate of σ_{itk}^2 (denoted $\tilde{\sigma}_{itk}^2$) is:

$$\tilde{\sigma}_{itk}^2 = \frac{1}{J} \sum_{j=1}^{J} \left(\tilde{y}_{itk}^j - \tilde{\bar{y}}_{itk} \right)^2, \tag{6}$$

where $\tilde{\bar{y}}_{itk}$ is determined by (2). If an estimate of the uncertainty from the error terms only is desired, then the trials consist only of draws from the distribution of the error terms.[5]

There are two polar assumptions that can be made about the uncertainty of the exogenous variables. One is, of course, that there is no exogenous-variable uncertainty. The other is that the exogenous-variable forecasts are in some way as uncertain as the endogenous-variable forecasts. Under this second assumption one could, for example, estimate an autoregressive equation for each exogenous variable and add these equations to the model. This expanded model, which would have no exogenous variables, could then be used for the stochastic-simula-

[5]Note that it is implicitly assumed here that the variances of the forecast errors exist. For some estimation techniques this is not always the case. If in a given application the variances do not exist, then one should estimate other measures of dispersion of the distribution, such as the interquartile range or mean absolute deviation.

tion estimates of the variances. While the first assumption is clearly likely to underestimate exogenous-variable uncertainty in most applications, the second assumption is likely to overestimate it. This is particularly true for fiscal-policy variables in macroeconomic models, where government-budget data are usually quite useful for purposes of forecasting up to at least about eight quarters ahead. The best approximation is thus likely to lie somewhere in between these two assumptions.

The assumption that was made for the results in Fair (1980) was in between the two polar assumptions. The procedure that was followed was to estimate an eighth-order autoregressive equation for each exogenous variable (including a constant and time in the equation) and then to take the estimated standard error from this regression as the estimate of the degree of uncertainty attached to forecasting the change in this variable for each period. This procedure ignores the uncertainty of the coefficient estimates in the autoregressive equations, which is one of the reasons it is not as extreme as the second polar assumption. In an earlier stochastic-simulation study of Haitovsky and Wallace (1972), third-order autoregressive equations were estimated for the exogenous variables, and these equations were then added to the model. This procedure is consistent with the second polar assumption above *except* that for purposes of the stochastic simulations Haitovsky and Wallace took the variances of the error terms to be one-half of the estimated variances. They defend this procedure (pp. 267–268) on the grounds that the uncertainty from the exogenous-variable forecasts is likely to be less than is reflected in the autoregressive equations.

Another possible procedure that could be used for the exogenous variables would be to gather from various forecasting services data on their ex ante forecasting errors of the exogenous variables (exogenous to you, not necessarily to the forecasting service). From these errors for various periods one could estimate a standard error for each exogenous variable and then use these errors for the stochastic-simulation draws.

For purposes of describing the present method, all that needs to be assumed is that *some* procedure is available for estimating exogenous-variable uncertainty. If equations for the exogenous variables are not added to the model, but instead some in between procedure is followed, then each stochastic-simulation trial consists of draws of error terms, coefficients, and exogenous-variable errors. If equations are added, then each trial consists of draws of error terms and coefficients from both the structural equations and the exogenous-variable equations. In either case, let $\tilde{\tilde{\sigma}}^2_{itk}$ denote the stochastic-simulation estimate of the variance of the forecast error that takes into account exogenous-variable uncertainty. $\tilde{\tilde{\sigma}}^2_{itk}$ differs from $\tilde{\sigma}^2_{itk}$ in (6) in that the trials for $\tilde{\tilde{\sigma}}^2_{itk}$ include draws of exogenous-variable errors.

Estimating the uncertainty from the possible misspecification of the model is the most difficult and costly part of the method. It requires successive reestimation and stochastic simulation of the model. It is based on a comparison of

estimated variances computed by means of stochastic simulation with estimated variances computed from outside-sample forecast errors.

Consider for now stochastic simulation with respect to the structural error terms and coefficients only (no exogenous-variable uncertainty). Assume that the forecast period begins one period after the end of the estimation period, and call this period t. As noted above, from this stochastic simulation one obtains an estimate of the variance of the forecast error, $\tilde{\sigma}^2_{itk}$. One also obtains from this simulation an estimate of the *expected value* of the k-period-ahead forecast of variable i: $\tilde{\bar{y}}_{itk}$ in equation (2). The difference between this estimate and the actual value, y_{it+k-1}, is the mean forecast error:

$$\hat{\varepsilon}_{itk} = y_{it+k-1} - \tilde{\bar{y}}_{itk}. \tag{7}$$

If it is assumed that $\tilde{\bar{y}}_{itk}$ exactly equals the true expected value, $\bar{y}_{itk}$, then $\hat{\varepsilon}_{itk}$ in (7) is a sample draw from a distribution with a known mean of zero and variance σ^2_{itk}. The square of this error, $\hat{\varepsilon}^2_{itk}$, is thus under this assumption an unbiased estimate of σ^2_{itk}. One thus has two estimates of σ^2_{itk}, one computed from the mean forecast error and one computed by stochastic simulation. Let d_{itk} denote the difference between these two estimates:

$$d_{itk} = \hat{\varepsilon}^2_{itk} - \tilde{\sigma}^2_{itk}. \tag{8}$$

If it is further assumed that $\tilde{\sigma}^2_{itk}$ exactly equals the true value, then d_{itk} is the difference between the estimated variance based on the mean forecast error and the true variance. Therefore, under the two assumptions of no error in the stochastic-simulation estimates, the expected value of d_{itk} is zero.

The assumption of no stochastic-simulation error, i.e. $\tilde{\bar{y}}_{itk} = \bar{y}_{itk}$ and $\tilde{\sigma}^2_{itk} = \sigma^2_{itk}$, is obviously only approximately correct at best. Even with an infinite number of draws the assumption would not be correct because the draws are from estimated rather than known distributions. It does seem, however, that the error introduced by this assumption is likely to be small relative to the error introduced by the fact that some assumption must be made about the mean of the distribution of d_{itk}. Because of this, nothing more will be said about stochastic-simulation error. The emphasis instead is on the possible assumptions about the mean of the distribution of d_{itk}, given the assumptions of no stochastic-simulation error.

The procedure just described uses a given estimation period and a given forecast period. Assume for sake of an example that one has data from period 1 through 100. The model can then be estimated through, say, period 70, with the forecast period beginning with period 71. Stochastic simulation for the forecast period will yield for each i and k a value of d_{i71k} in (8). The model can then be reestimated through period 71, with the forecast period now beginning with period 72. Stochastic simulation for this forecast period will yield for each i and k a value of d_{i72k} in (8). This process can be repeated through the estimation period

ending with period 99. For the one-period-ahead forecast ($k=1$) the procedure will yield for each variable i 30 values of d_{it1} ($t=71,\ldots,100$); for the two-period-ahead forecast ($k=2$) it will yield 29 values of d_{it2} ($t=72,\ldots,100$); and so on. If the assumption of no simulation error holds for all t, then the expected value of d_{itk} is zero for all t.

The discussion so far is based on the assumption that the model is correctly specified. Misspecification has two effects on d_{itk} in (8). First, if the model is misspecified, the estimated covariance matrices that are used for the stochastic simulation will not in general be unbiased estimates of the true covariance matrices. The estimated variances computed by means of stochastic simulation will thus in general be biased. Second, the estimated variances computed from the forecast errors will in general be biased estimates of the true variances. Since misspecification affects both estimates, the effect on d_{itk} is ambiguous. It is possible for misspecification to affect the two estimates in the same way and thus leave the expected value of the difference between them equal to zero. In general, however, this does not seem likely, and so in general one would not expect the expected value of d_{itk} to be zero for a misspecified model. The expected value may be negative rather than positive for a misspecified model, although in general it seems more likely that it will be positive. Because of the possibility of data mining, misspecification seems more likely to have a larger positive effect on the outside sample forecast errors than on the (within-sample) estimated covariance matrices.

An examination of how the d_{itk} values change over time (for a given i and k) may reveal information about the strengths and weaknesses of the model that one would otherwise not have. This information may then be useful in future work on the model. The individual values may thus be of interest in their own right aside from their possible use in estimating total predictive uncertainty.

For the total uncertainty estimates some assumption has to be made about how misspecification affects the expected value of d_{itk}. For the results in Fair (1980a) it was assumed that the expected value of d_{itk} is constant across time: for a given i and k, misspecification was assumed to affect the mean of the distribution of d_{itk} in the same way for all t. Other possible assumptions are, of course, possible. One could, for example, assume that the mean of the distribution is a function of other variables. (A simple assumption in this respect is that the mean follows a linear time trend.) Given this assumption, the mean can be then estimated from a regression of d_{itk} on the variables. For the assumption of a constant mean, this regression is merely a regression on a constant (i.e. the estimated constant term is merely the mean of the d_{itk} values).[6] The predicted value from this regression for period t, denoted $\hat{d}_{itk}$, is the estimated mean for period t.

[6] For the results in Fair (1980) a slightly different assumption than that of a constant mean was made for variables with trends. For these variables it was assumed that the mean of d_{itk} is proportional to $\bar{\bar{y}}^2_{itk}$, i.e. that the mean of $d_{itk}/\bar{\bar{y}}^2_{itk}$ is constant across time.

An estimate of the total variance of the forecast error, denoted $\hat{\sigma}^2_{itk}$, is the sum of $\tilde{\tilde{\sigma}}^2_{itk}$ – the stochastic-simulation estimate of the variance due to the error terms, coefficient estimates, and exogenous variables – and $\hat{d}_{itk}$:

$$\hat{\sigma}^2_{itk} = \tilde{\tilde{\sigma}}^2_{itk} + \hat{d}_{itk}. \tag{9}$$

Since the procedure in arriving at $\hat{\sigma}^2_{itk}$ takes into account the four main sources of uncertainty of a forecast, the values of $\hat{\sigma}^2_{itk}$ can be compared across models for a given i, k, and t. If, for example, one model has consistently smaller values of $\hat{\sigma}^2_{itk}$ then another, this would be fairly strong evidence for concluding that it is a more accurate model, i.e. a better approximation to the true structure.

This completes the outline of the method. It may be useful to review the main steps involved in computing $\hat{\sigma}^2_{itk}$ in (9). Assume that data are available for periods 1 through T and that one is interested in estimating the uncertainty of an eight-period-ahead forecast that began in period $T+1$, (i.e. in computing $\hat{\sigma}^2_{itk}$ for $t = T+1$ and $k = 1,\ldots,8$). Given a base set of values for the exogenous variables for periods $T+1$ through $T+8$, one can compute $\tilde{\tilde{\sigma}}^2_{itk}$ for $t = T+1$ and $k = 1,\ldots,8$ by means of stochastic simulation. Each trial consists of one eight-period dynamic simulation and requires draws of the error terms, coefficients, and exogenous-variable errors. These draws are based on the estimate of the model through period T. This is the relative inexpensive part of the method. The expensive part consists of the successive reestimation and stochastic simulation of the model that are needed in computing the d_{itk} values. In the above example, the model would be estimated 30 times and stochastically simulated 30 times in computing the d_{itk} values. After these values are computed for, say, periods $T-r$ through T, then $\hat{d}_{itk}$ can be computed for $t = T+1$ and $k = 1,\ldots,8$ using whatever assumption has been made about the distribution of d_{itk}. This allows $\hat{\sigma}^2_{itk}$ in (9) to be computed for $t = T+1$ and $k = 1,\ldots,8$.

In the successive reestimation of the model, the first period of the estimation period may or may not be increased by one each time. The criterion that one should use in deciding this is to pick the procedure that seems likely to correspond to the chosen assumption about the distribution of d_{itk} being the best approximation to the truth. It is also possible to take the distance between the last period of the estimation period and the first period of the forecast period to be other than one, as was done above.

It is important to note that the above estimate of the mean of the d_{itk} distribution is not in general efficient because the error term in the d_{itk} regression is in general heteroscedastic. Even under the null hypothesis of no misspecification, the variance of the d_{itk} distribution is not constant across time. It is true, however, that $\hat{\varepsilon}_{itk}/(\tilde{\sigma}^2_{itk} + \hat{d}_{itk})^{1/2}$ has unit variance under the null hypothesis, and so it may not be a bad approximation to assume that $\hat{\varepsilon}^2_{itk}/(\tilde{\sigma}^2_{itk} + \hat{d}_{itk})$ has a constant variance across time. This then suggests the following iterative proce-

dure. 1) For each i and k, calculate $\hat{d}_{itk}$ from the d_{itk} regression, as discussed above; 2) divide each observation in the d_{itk} regression by $\tilde{\sigma}^2_{itk} + \hat{d}_{itk}$, run another regression, and calculate $\hat{d}_{itk}$ from this regression; 3) repeat step 2) until the successive estimates of $\hat{d}_{itk}$ are within some prescribed tolerance level. Litterman (1980) has carried out this procedure for a number of models for the case in which the only explanatory variable in the d_{itk} regression is the constant term (i.e. for the case in which the null hypothesis is that the mean of the d_{itk} distribution is constant across time).

If one is willing to assume that $\hat{\varepsilon}_{itk}$ is normally distributed, which is at best only an approximation, then Litterman (1979) has shown that the above iterative procedure produces maximum likelihood estimates. He has used this assumption in Litterman (1980) to test the hypothesis (using a likelihood ratio test) that the mean of the d_{itk} distribution is the same in the first and second halves of the sample period. The hypothesis was rejected at the 5 percent level in only 3 of 24 tests. These results thus suggest that the assumption of a constant mean of the d_{itk} distribution may not be a bad approximation in many cases. This conclusion was also reached for the results in Fair (1982), where plots of d_{itk} values were examined across time (for a given i and k). There was little evidence from these plots that the mean was changing over time.

The mean of the d_{itk} distribution can be interpreted as a measure of the average unexplained forecast error variance, (i.e. that part not explained by $\tilde{\sigma}^2_{itk}$) rather than as a measure of misspecification. Using this interpretation, Litterman (1980) has examined whether the use of the estimated means of the d_{itk} distributions lead to more accurate estimates of the forecast error variances. The results of his tests, which are based on the normality assumption, show that substantially more accurate estimates are obtained using the estimated means. Litterman's overall results are thus quite encouraging regarding the potential usefulness of the method discussed in this section.

Aside from Litterman's use of the method to compare various versions of Sims' (1980) model, I have used the method to compare my model [Fair (1976)], Sargent's (1976) model, Sims' model, and an eighth-order autoregressive model. The results of this comparison are presented in Fair (1979).

6. Conclusion

It should be clear from this chapter that the comparison of the predictive accuracy of alternative models is not a straightforward exercise. The difficulty of evaluating alternative models is undoubtedly one of the main reasons there is currently so little agreement about which model best approximates the true structure of the economy. If it were easy to decide whether one model is more accurate than another, there would probably be by now a generally agreed upon

model of, for example, the U.S. economy. With further work on methods like the one described in Section 5, however, it may be possible in the not-too-distant future to begin a more systematic comparison of models. Perhaps in ten or twenty years time the use of these methods will have considerably narrowed the current range of disagreements.

References

Bianchi, C., G. Calzolari and P. Corsi (1976) "Divergences in the Results of Stochastic and Deterministic Simulation of an Italian Non Linear Econometric Model", in: L. Dekker, ed., *Simulation of Systems*. Amsterdam: North-Holland Publishing Co.

Calzolari, G. and P. Corsi (1977) "Stochastic Simulation as a Validation Tool for Econometric Models". Paper presented at IIASA Seminar, Laxenburg, Vienna, September 13–15.

Cooper, J. P. (1974) *Development of the Monetary Sector, Prediction and Policy Analysis in the FRB-MIT-Penn Model*. Lexington: D. C. Heath & Co.

Cooper, J. P. and S. Fischer (1972) "Stochastic Simulation of Monetary Rules in Two Macroeconometric Models", *Journal of the American Statistical Association*, 67, 750–760.

Cooper, J. P. and S. Fischer (1974) "Monetary and Fiscal Policy in the Fully Stochastic St. Louis Econometric Model", *Journal of Money, Credit and Banking*, 6, 1–22.

Evans, Michael K., Yoel Haitovsky and George I. Treyz, assisted by Vincent Su (1972) "An Analysis of the Forecasting Properties of U.S. Econometric Models", in: B. G. Hickman, ed., *Econometric Models of Cyclical Behavior*. New York: Columbia University Press, 949–1139.

Evans, M. K., L. R. Klein and M. Saito (1972) "Short-Run Prediction and Long-Run Simulation of the Wharton Model", in: B. G. Hickman, ed., *Econometric Models of Cyclical Behavior*. New York: Columbia University Press, 139–185.

Fair, Ray C. (1971) *A Short-Run Forecasting Model of the United States Economy*. Lexington: D. C. Heath & Co.

Fair, Ray C. (1974) "An Evaluation of a Short-Run Forecasting Model", *International Economic Review*, 15, 285–303.

Fair, Ray C. (1976) *A Model of Macroeconomic Activity. Volume II: The Empirical Model*. Cambridge: Ballinger Publishing Co.

Fair, Ray C. (1979) "An Analysis of the Accuracy of Four Macroeconometric Models", *Journal of Political Economy*, 87, 701–718.

Fair, Ray C. (1980) "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review*, 21, 355–378.

Fair, Ray C. (1982) "The Effects of Misspecification on Predictive Accuracy," in: G. C. Chow and P. Corsi, eds., *Evaluating the Reliability of Macro-economic Models*. New York: John Wiley & Sons, 193–213.

Fromm, Gary and Lawrence R. Klein (1976) "The NBER/NSF Model Comparison Seminar: An Analysis of Results", *Annals of Economic and Social Measurement*, Winter, 5, 1–28.

Fromm, Gary, L. R. Klein and G. R. Schink (1972) "Short- and Long-Term Simulations with the Brookings Model", in: B. G. Hickman, ed., *Econometric Models of Cyclical Behavior*. New York: Columbia University Press, 201–292.

Garbade, K. D. (1975) *Discretionary Control of Aggregate Economic Activity*. Lexington: D. C. Heath & Co.

Granger, C. W. J. and Paul Newbold (1977) *Forecasting Economic Time Series*. New York: Academic Press.

Green, G. R., M. Liebenberg and A. A. Hirsch (1972) "Short- and Long-Term Simulations with the OBE Econometric Model", in: B. G. Hickman, ed., *Econometric Models of Cyclical Behavior*. New York: Columbia University Press, 25–123.

Haitovsky, Yoel and George Treyz (1972) "Forecasts with Quarterly Macroeconometric Models: Equation Adjustments, and Benchmark Predictions: The U.S. Experience", *The Review of Economics*

and Statistics, 54, 317–325.
Haitovsky, Yoel, G. Treyz and V. Su (1974) *Forecasts with Quarterly Macroeconometric Models*. New York: National Bureau of Economic Research, Columbia University Press.
Haitovsky, Y. and N. Wallace (1972) "A Study of Discretionary and Non-discretionary Monetary and Fiscal Policies in the Context of Stochastic Macroeconometric Models", in: V. Zarnowitz, ed., *The Business Cycle Today*. New York: Columbia University Press.
Hirsch, Albert A., Bruce T. Grimm and Gorti V. L. Narasimham (1974) "Some Multiplier and Error Characteristics of the BEA Quarterly Model", *International Economic Review*, 15, 616–631.
Howrey, E. P. and H. H. Kelejian (1971) "Simulation versus Analytical Solutions: The Case of Econometric Models", in: T. H. Naylor, ed., *Computer Simulation Experiments with Models of Economic Systems*. New York: Wiley.
Intriligator, Michael D. (1978) *Econometric Models, Techniques, and Applications*. Amsterdam: North-Holland Publishing Co.
Litterman, Robert B. (1979) "Techniques of Forecasting Using Vector Autoregression". Working Paper No. 115, Federal Reserve Bank of Minneapolis, November.
Litterman, Robert B. (1980) "Improving the Measurement of Predictive Accuracy", mimeo.
McCarthy, Michael D. (1972) "Some Notes on the Generation of Pseudo-Structural Errors for Use in Stochastic Simulation Studies", in: B. G. Hickman, ed., *Econometric Models of Cyclical Behavior*. New York: Columbia University Press, 185–191.
McNees, Stephen K. (1973) "The Predictive Accuracy of Econometric Forecasts", *New England Economic Review*, September/October, 3–22.
McNees, Stephen K. (1974) "How Accurate Are Economic Forecasts?", *New England Economic Review*, November/December, 2–19.
McNees, Stephen K. (1975) "An Evaluation of Economic Forecasts", *New England Economic Review*, November/December, 3–39.
McNees, Stephen K. (1976) "An Evaluation of Economic Forecasts: Extension and Update", *New England Economic Review*. September/October, 30–44.
Muench, T., A. Rolnick, N. Wallace and W. Weiler (1974) "Tests for Structural Change and Prediction Intervals for the Reduced Forms of the Two Structural Models of the U.S.: The FRB-MIT and Michigan Quarterly Models", *Annals of Economic and Social Measurement*, 3, 491–519.
Nagar, A. L. (1969) "Stochastic Simulation of the Brookings Econometric Model", in: J. S. Duesenberry, G. Fromm, L. R. Klein, and E. Kuh, eds., *The Brookings Model: Some Further Results*. Chicago: Rand McNally & Co.
Nelson, Charles R. (1972) "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy", *The American Economic Review*, 62, 902–917.
Sargent, Thomas J. (1976) "A Classical Macroeconometric Model for the United States", *Journal of Political Economy*, 84, 207–237.
Schink, G. R. (1971) "Small Sample Estimates of the Variance–Covariance Matrix Forecast Error for Large Econometric Models: The Stochastic Simulation Technique", Ph.D. Dissertation, University of Pennsylvania.
Schink, G. R. (1974) "Estimation of Small Sample Forecast Error for Nonlinear Dynamic Models: A Stochastic Simulation Approach", mimeo.
Sims, Christopher A. (1980) "Macroeconomics and Reality", *Econometrica*, 48, 1–48.
Sowey, E. R. (1973) "Stochastic Simulation for Macroeconomic Models: Methodology and Interpretation", in: A. A. Powell and R. W. Williams, eds., *Econometric Studies of Macro and Monetary Relations*. Amsterdam: North-Holland Publishing Co.
Theil, Henri (1966) *Applied Economic Forecasting*. Amsterdam: North-Holland Publishing Co.
Zarnowitz, Victor (1979) "An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level", *Journal of Business*, 52, 1–33.