8 Evaluating Predictive Accuracy

8.1 Introduction

This chapter deals with one of the most important issues in macroeconomics: the evaluation and testing of models. The central question in this area is how to decide which model out of a number best approximates the structure of the economy. Although an obvious answer is to choose the model that fits the data best, the problem comes in deciding what criterion to use to judge which model fits the data best. In the next two sections the standard ways in which this problem has been treated are discussed: Section 8.2 considers the evaluation of ex ante forecasts, and Section 8.3 considers the evaluation of ex post forecasts. My method for dealing with this problem is explained in Section 8.4. Results for various models are presented in Sections 8.5 and 8.6.

The three most common measures of predictive accuracy that have been used to evaluate ex ante and ex post forecasts are root mean squared error (RMSE), mean absolute error (MAE), and Theil's (1966, p. 28) inequality coefficient (U). Let $\hat{y}_{it}$ be the forecast of variable i for period t, and let y_{it} be the actual value. $\hat{y}_{it}$ can be a prediction for more than one period ahead. Assuming that observations on $\hat{y}_{it}$ and y_{it} are available for $t = 1, \ldots, T$, the three measures are

$$\text{(8.1)} \qquad \text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{y}_{it})^2},$$

$$\text{(8.2)} \qquad \text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|y_{it} - \hat{y}_{it}|,$$

$$\text{(8.3)} \qquad \text{U} = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{it} - \Delta\hat{y}_{it})^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{it})^2}},$$

where Δ in (8.3) 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.

8.2 Evaluation of Ex Ante Forecasts

The procedure followed to evaluate ex ante forecasts is simply to collect the forecast data for a certain period and to compute one or more of the three measures just mentioned. Forecasts from different models are evaluated by comparing the error measures across models. 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 values 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 adding the forecasted change ($\Delta\hat{y}_{it}$), which is based on the old data, to the new lagged value (y_{it-1}). The adjusted forecast value is then compared 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, no adjustments are needed.

A number of studies have examined ex ante forecasting accuracy using one or more of the above measures; some of the more recent ones are McNees (1973, 1974, 1975, 1976) and Zarnowitz (1979). It is usually the case that forecasts from both model builders and non-model 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 as 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 predicted 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 8.4.) The second, and more important, reason is that almost all the forecasts examined in these studies are generated from subjectively adjusted models. (The use of add factors is discussed in Section 7.4.) It is thus the accuracy of the forecasting performance of the model builders rather than that of the models that is being examined.

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.

In regard to nonsubjectively-adjusted ex ante forecasts, there is some evidence that their accuracy 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 1971b). No add factors were used for these forecasts. The accuracy of these forecasts is examined in Fair (1974b), and the results indicate that the accuracy 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 (nonsubjectively-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 (that is, the forecasts based on the actual values of the exogenous variables) were on average of about the same degree of accuracy as the subjectively adjusted forecasts (Fair 1974b).

8.3 Evaluation of Ex Post Forecasts

The RMSE, MAE, and U measures have also been widely used to evaluate the accuracy of ex post forecasts. One of the better-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 little more needs to be said about it. Sometimes the accuracy of a given model is compared to the accuracy of a "naive" model, which 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; 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 a similar measure is widespread, there are two serious problems associated with this 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 of comparing RMSEs or similar measures across models 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), 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, 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.

Another problem associated with within-sample calculations of the error measures 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 structure. Within-sample error calculations are not likely to discover this, and thus they may give a very misleading impression of the true accuracy of the model. Outside-sample error calculations should pick up this problem, however, 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. If one ignores 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, although of some interest, 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.

8.4 A Method for Evaluating Predictive Accuracy

My method for evaluating predictive accuracy, in contrast to previous procedures, 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 relies heavily on the use of stochastic simulation.

8.4.1 Uncertainty from the Error Terms and Coefficient Estimates

Estimating the uncertainty from the error terms and coefficient estimates is simply a matter of computing $\tilde{\sigma}^2_{itk}$ in (7.8). $\tilde{\sigma}^2_{itk}$ is a stochastic-simulation estimate of σ^2_{itk}, the variance of the forecast error for a k-period-ahead forecast of variable i from a simulation beginning in period t. It is based on draws from both the distribution of the error terms and the distribution of the coefficient estimates. If an estimate of the uncertainty from the error terms only is desired, the draws should be only from the distribution of the error terms, with the coefficient estimates fixed at some set of values.

8.4.2 Uncertainty from the Exogenous Variables

There are two polar assumptions that can be made about the uncertainty of the exogenous variables: one is that there is no 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-simulation 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 macroeconometric 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 basic assumption that I have used in my work so far is in between the two polar assumptions. The procedure that I have followed is to estimate an eighth-order autoregressive equation for each exogenous variable (with a constant term and time trend included 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 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.

A procedure similar to the second polar assumption was used in an earlier stochastic simulation study of Haitovsky and Wallace (1972), where third-order autoregressive equations were estimated for the exogenous variables and then these equations were added to the model. This procedure is consistent with the second polar assumption *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 the investigator, 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 (7.8) in that the trials for $\tilde{\tilde{\sigma}}^2_{itk}$ include draws of exogenous-variable errors.

The procedure that I have used to estimate exogenous-variable uncertainty is implemented as follows. Let $\hat{s}_i$ denote the estimated standard error from the eighth-order autoregressive equation for exogenous variable i. Let v_{it} be a normally distributed random variable with mean zero and variance $\hat{s}^2_i$: $v_{it} \sim N(0, \hat{s}^2_i)$ for all t. Let $\hat{x}_{it}$ be the "base" value of exogenous variable i for period t. The base values can either be the actual values, if the period in question is within the period for which data exist, or guessed values otherwise. If the values are guessed, they need *not* be the predictions from the autoregressive equations; the latter are used merely to get the values for $\hat{s}_i$. Let x^*_{it} be the value of variable i used on a given trial. Then for a given trial x^*_{it} is taken to be $\hat{x}_{it} + v_{it}$, where v_{it} is drawn from the above distribution. If, say, the simulation period were 8 quarters in length and there were 100 exogenous variables, 800 draws would be taken, one for each of the 100 i's and one for each of the 8 t's. There would be 100 autoregressive equations estimated.

For some of my work I have taken the estimated standard error from the autoregressive equation for each variable to be an estimate of the degree of uncertainty attached to forecasting the *change* in the variable for each period. Given the way that many exogenous variables are forecast, by extrapolating past trends or taking variables to be unchanged from their last observed values, it may be that any error in forecasting the level of a variable in, say, the first period will persist throughout the forecast period. If this is true, the assumption that the errors pertain to the changes in the variables may be better than the assumption that they pertain to the levels. This procedure is implemented as follows. Let quarter 1 be the first quarter of the prediction period, and assume that the prediction period is of length T. The values of x_{it}^* $(t = 1, \ldots, T)$ for a given trial are taken to be

$$
\begin{aligned}
x_{i1}^* &= \hat{x}_{i1} + v_{i1}, \\
x_{i2}^* &= \hat{x}_{i2} + v_{i1} + v_{i2}, \\
&\;\;\vdots \\
x_{iT}^* &= \hat{x}_{iT} + v_{i1} + v_{i2} + \ldots + v_{iT},
\end{aligned}
\tag{8.4}
$$

where each v_{it} $(t = 1, \ldots, T)$ is drawn from the $N(0, \hat{s}_i^2)$ distribution. Because of the assumption that the errors pertain to changes, the error term v_{i1} is carried along from quarter 1 on. Similarly, v_{i2} is carried along from quarter 2 on, and so forth.

8.4.3 Uncertainty from the Possible Misspecification of the Model

The most difficult and costly part of the method is estimating the uncertainty from the possible misspecification of the model, which 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. As will be seen, the expected value of the difference between the two estimated variances for a given variable and period is zero for a correctly specified model. The expected value is not in general zero for a misspecified model, and this fact can be used to try to account for misspecification effects.

All of the stochastic simulations that are referred to in this section are with respect to error terms and coefficients only. In other words, there is assumed to be no exogenous-variable uncertainty. Section 8.4.4 discusses the way in which the estimates of exogenous-variable uncertainty that were discussed in Section 8.4.2 are combined with the estimates of misspecification effects.

Assume that the prediction period begins one period after the end of the estimation period, and call this period t. From stochastic simulation one obtains an estimate of the variance of the forecast error, $\tilde{\sigma}^2_{itk}$ in (7.8). One also obtains an estimate of the expected value of the k-period-ahead forecast of variable i, $\tilde{\bar{y}}_{itk}$ in (7.7). The difference between this estimate and the actual value, y_{it+k-1}, is the mean forecast error:

$$(8.5) \qquad \hat{\epsilon}_{itk} = y_{it+k-1} - \tilde{\bar{y}}_{itk}.$$

If it is assumed that $\tilde{\bar{y}}_{itk}$ exactly equals the true expected value, $\bar{y}_{itk}$, then $\hat{\epsilon}_{itk}$ in (8.5) is a sample draw from a distribution with a known mean of zero and variance σ^2_{itk}. The square of this error, $\hat{\epsilon}^2_{itk}$, is thus under this assumption an unbiased estimate of σ^2_{itk}. One therefore 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:

$$(8.6) \qquad d_{itk} = \hat{\epsilon}^2_{itk} - \tilde{\sigma}^2_{itk}.$$

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, that is, $\tilde{\bar{y}}_{itk} = \bar{y}_{itk}$ and $\tilde{\sigma}^2_{itk} = \sigma^2_{itk}$, is obviously only approximately correct at best. As noted in Section 7.3.1, 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}. For this reason, nothing more will be said about stochastic-simulation error. The emphasis instead will be on possible assumptions about the mean of the distribution of d_{itk}, given the assumption of no stochastic-simulation error.

If the model is misspecified, it is not in general true that the expected value of d_{itk} is zero. Misspecification has two effects on d_{itk}. 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.

Because of the common practice in macroeconometric work of searching for equations that fit the data well, it seems likely that the estimated means of d_{itk} will be positive in practice for a misspecified model. If the model fits the data well within sample, the stochastic-simulation estimates of the forecast error variances will be small. This is because they are based on draws from estimated distributions of the error terms and coefficient estimates that have small (in a matrix sense) covariance matrices. If the model, although fitting the data well, is in fact misspecified, this should result in large outside-sample forecast errors. The estimated mean of d_{itk} is thus likely to be positive: $\tilde{\sigma}^2_{itk}$ is small because of small estimated covariance matrices, and $\hat{\epsilon}^2_{itk}$ is large because of large outside-sample forecast errors.

The procedure described so far uses one estimation period and one prediction period. It results in one value of d_{itk} for each variable i and length ahead k. Since one observation is obviously not adequate for estimating the mean of d_{itk}, more observations must be generated. This can be done by using successively new estimation periods and new prediction periods. Assume, for example, that one has data from period 1 through period 100. The model can be estimated through, say, period 70, with the prediction period beginning with period 71. Stochastic simulation for the prediction period will yield for each i and k a value of d_{i71k} in (8.6). The model can then be reestimated through period 71, with the prediction period now beginning with period 72. Stochastic simulation for this prediction period will yield for each i and k a value of d_{i72k} in (8.6). 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} ($t = 71, \ldots, 99$); and so on. If the assumption of no stochastic-simulation error holds for all t, then the expected value of d_{itk} is zero for all t for a correctly specified model.

The final step in the process is to make an assumption about the mean of

d_{itk} that allows the computed values of d_{itk} to be used to estimate the mean. A variety of assumptions are possible. One is simply that the mean is constant across time. In other words, misspecification is assumed to affect the mean in the same way for all t. If this assumption is made, the mean can be estimated by merely averaging the computed values of d_{itk} for each i and k. Another possible assumption is that the mean is a function of other variables, where the other variables are specified. (A simple example of this is the assumption that the mean follows a linear time trend.) Given this assumption, the mean can be estimated from a regression of d_{itk} on the specified variables. (In the linear trend case, the explanatory variables would be a constant and a time trend.) The predicted value from this regression for period t, denoted $\hat{d}_{itk}$, is the estimated mean for period t. In this case the estimated mean obviously varies over time if the explanatory variables vary. This second assumption would be used if it were felt that the degree of misspecification of the model varies in a systematic way with other variables.

A version of the first assumption is that the mean of d_{itk} is proportional to $\tilde{\bar{y}}^2_{itk}$, which implies that the mean of $d_{itk}/\tilde{\bar{y}}^2_{itk}$ is constant across time. d_{itk} is in units of the variable squared, and this assumption is equivalent to the constant mean assumption in percentage terms. For variables with trends it may be more reasonable to couch the assumption in percentage terms, since the mean may vary as a function of the size of the variable.

8.4.4 Total Uncertainty

Given $\hat{d}_{itk}$, the estimate of the mean of d_{itk} for period t, it is possible to estimate the total variance of the forecast error, denoted $\hat{\sigma}^2_{itk}$. This 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 $\hat{d}_{itk}$:

$$\hat{\sigma}^2_{itk} = \tilde{\tilde{\sigma}}^2_{itk} + \hat{d}_{itk}. \tag{8.7}$$

The use of $\tilde{\tilde{\sigma}}^2_{itk}$ instead of $\tilde{\sigma}^2_{itk}$ in (8.7) is where the estimate of exogenous variable uncertainty is brought into the analysis.

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 $\hat{\sigma}^2_{itk}$ than another, this would be fairly strong evidence for concluding that it is a more accurate model, that is, a better approximation to the true structure.

It may be useful at this stage to review the steps that are involved in arriving at $\hat{\sigma}^2_{itk}$ in (8.7). Consider the example used in Section 8.4.3, where data are available for periods 1 through 100. Assume that one is interested in estimating the uncertainty of an eight-period-ahead forecast that begins in period 90. In other words, one is interested in computing $\hat{\sigma}^2_{itk}$ for $t = 90$ and $k = 1, \ldots, 8$. Assume that the main set of coefficient estimates of the model is based on an estimation period through period 100. Given (1) these estimates and the associated estimates of the distributions of the error terms and coefficient estimates, (2) the actual values of the exogenous variables for periods 90–97, and (3) some assumption about exogenous-variable uncertainty, $\tilde{\tilde{\sigma}}^2_{itk}$ can be computed using stochastic simulation for $t = 90$ and $k = 1, \ldots, 8$. Each trial consists of one eight-period dynamic simulation beginning in period 90. It requires draws of the error terms, coefficients, and (possibly) exogenous-variable errors. If, say, 250 trials are taken, the model must be solved 250 times for the eight quarters.

Since computing $\tilde{\tilde{\sigma}}^2_{itk}$ requires only one stochastic simulation, this is the relatively inexpensive part of the method. The expensive part consists of the successive reestimation and stochastic simulation that are needed in computing the d_{itk} values. In the example in Section 8.4.3, the model would be estimated 30 times and stochastically simulated 30 times in computing the d_{itk} values. If 250 trials for each stochastic simulation were used, the model would be solved $250 \times 30 = 7{,}500$ times, where each solution is a dynamic eight-period simulation. After the d_{itk} values are computed for, say, periods 70 through 99, $\hat{d}_{itk}$ can be computed for $t = 90$ and $k = 1, \ldots, 8$ using whatever assumption has been made about the distribution of d_{itk}. This procedure then allows $\hat{\sigma}^2_{itk}$ in (8.7) to be computed for $t = 90$ and $k = 1, \ldots, 8$.

8.4.5 General Remarks about the Method

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.

Any assumption that one makes about the mean of d_{itk} is at best likely to be only a rough approximation to the truth. It is unlikely that the effects of misspecification on the two estimated variances are so systematic as to lead to any assumption that one might make about the mean of the difference between the two being exactly right. One useful thing that can be done is simply to plot the d_{itk} values over time for a given i and k and see if there are systematic tendencies. One might observe trend or cyclical movements in these plots, which could be useful either in deciding what to assume about the mean of d_{itk} or in deciding how to change the model to try to eliminate the misspecification. If the latter is done, one is using the d_{itk} values to reveal weaknesses in the model that might be corrected rather than to adjust the stochastic-simulation estimates of the variances for misspecification. The individual d_{itk} values may thus be of interest in their own right aside from their possible use in estimating total predictive uncertainty. If the values are used solely to reveal weaknesses of the model, no assumption about the mean of d_{itk} is needed.

Although I have been interpreting the d_{itk} values as measuring the misspecification of the model, this is not exactly right. Since misspecification affects both $\tilde{\sigma}^2_{itk}$ and $\hat{\epsilon}^2_{itk}$ in (8.6), it may be for a particular model that both are affected about the same. In this case the expected value of d_{itk} would be close to zero and yet the model could be seriously misspecified. In other words, misspecification can make both $\tilde{\sigma}^2_{itk}$ and $\hat{\epsilon}^2_{itk}$ larger and leave the difference between the two about the same. The more common case, as discussed in Section 8.4.3, seems likely to be one in which extensive searching for equations that fit the data well has resulted in an estimate of $\tilde{\sigma}^2_{itk}$ that is too small. In this case the d_{itk} values are likely to be on average large. Whatever the case, one should be aware that interpreting the d_{itk} values as measures of misspecification is using the word "misspecification" in a very special way. A better but more awkward way of stating what the d_{itk} values are is that they are a measure of the misspecification of the model that is not already reflected in the stochastic-simulation estimate of the forecast error variance.

It is important to note that the interpretation of the d_{itk} values does not affect the interpretation of $\hat{\sigma}^2_{itk}$ in (8.7) as an estimate of the total variance of the forecast error. If misspecification affects the stochastic-simulation estimate of the variance about as much as it affects the estimate based on the outside-sample forecast error (so that $\hat{d}_{itk}$ is close to zero), misspecification effects will be reflected in $\tilde{\tilde{\sigma}}^2_{itk}$ in (8.7) rather than in $\hat{d}_{itk}$. The term $\hat{d}_{itk}$ is merely the adjustment for the misspecification effects that are not captured by $\tilde{\tilde{\sigma}}^2_{itk}$.

The estimates of the mean of d_{itk} that have been proposed in Section 8.4.3 are 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 d_{itk} is not constant across time. It is true, however, that

$\hat{\epsilon}_{itk}/\sqrt{\tilde{\sigma}^2_{itk} + \hat{d}_{itk}}$ has unit variance for all t under the null hypothesis, and therefore it is reasonable to assume that $\hat{\epsilon}^2_{itk}/(\tilde{\sigma}^2_{itk} + \hat{d}_{itk})$ has a constant variance for all t. This then suggests the following iterative procedure. (1) For each i and k, calculate $\hat{d}_{itk}$ from the d_{itk} regression, as discussed earlier; (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 (that is, for the case in which it is assumed that the mean of the d_{itk} distribution is constant across time).

If one is willing to assume that $\hat{\epsilon}_{itk}$ is normally distributed, which may or may not be a good approximation, Litterman (1979) has shown that the iterative procedure just described 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 d_{itk} 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 d_{itk} may not be a bad approximation in many cases. The results for the US model, which are reported in Section 8.5, also suggest that the assumption may be a reasonable approximation.

Another interpretation of the mean of d_{itk} is that it is a measure of the average unexplained forecast error variance (that is, that part not explained by $\tilde{\sigma}^2_{itk}$). Using this interpretation, Litterman (1980) has examined the question of whether the use of the estimated mean of d_{itk} leads to more accurate estimates of the forecast error variance. The results of his tests, which are based on the normality assumption, show that substantially more accurate estimates are obtained using the estimated means.

It should finally be noted that although the method is designed to catch a model that fits the data well within sample but is in fact poorly specified, there is a subtle form of data mining that the method does not account for. If, say, a model is specified in period 100, estimated through period 90, and tested with respect to its outside-sample forecasting accuracy for periods 91 – 100, it is clear that this is not a strict outside-sample test. Information on what happened between periods 91 and 100 may have been used in the specification of the model, and thus one cannot be sure that the model's "outside-sample" accuracy that is estimated for periods 91 – 100 will hold for, say, periods 101 – 110. Within the context of the present method, this means that the computed values of d_{itk} for periods 91 – 100 are too low, which will result in values of $\hat{d}_{itk}$ that are too low and thus values of $\hat{\sigma}^2_{itk}$ that are too low.

8.5 A Comparison of the US, ARUS, VAR1US, VAR2US, and LINUS Models

In this section five econometric models of the United States are compared using the method in Section 8.4. The main concern is to see how the US model compares to the autoregressive model (ARUS), the two vector autoregressive models (VAR1US and VAR2US), and a simple linear model (LINUS). The US model is discussed in Chapter 4, and the other models are discussed in Chapter 5.

8.5.1 Computing the d_{itk} Values

The primary cost of the method is computing the d_{itk} values. In computing these values, each of the five models was estimated 51 times. The first estimation period ended in 1969III, the second estimation period ended in 1969IV, and so on through 1982I. A stochastic simulation was then run for each of the 51 sets of estimates, where the prediction period began two quarters after the end of the estimation period. The reason for beginning the prediction period two quarters rather than only one quarter after the end of the estimation period is that in practice most of the data for the most recent quarter are preliminary. In my work I use the preliminary data as initial conditions for a forecast but not as observations for estimation. This means that there is always a two-quarter gap between the end of the estimation period and the beginning of the prediction period, and the present procedure is consistent with this practice.

The computations for the US model were as follows. The first of the 51 estimation periods was 1954I–1969III (63 observations). The coefficients were estimated by 2SLS, and the covariance matrix of the coefficient estimates was computed. Let $\hat{\alpha}_2$ denote the coefficient estimates, and let $\hat{V}_2$ denote the estimated covariance matrix. The correct formula for the covariance matrix is (6.20) in Chapter 6, where the off-diagonal blocks of the matrix are not zero. Computing this matrix is fairly expensive in that it requires more time than is required to compute the coefficient estimates. (The times reported in Section 6.5.1 for the IBM 4341 are 3.0 minutes for the coefficient estimates and 5.5 minutes for the covariance matrix.) If the off-diagonal blocks are taken to be zero, there is no extra cost in computing the covariance matrix because the diagonal blocks are available from the estimates of the individual equations. For the work here, the off-diagonal blocks were taken to be zero for all 51 sets of estimates.

Given the coefficient estimates, the covariance matrix of the error terms ($\hat{S}$)

was estimated as $(1/63)\hat{U}\hat{U}'$, where $\hat{U}$ is the 30×63 matrix of values of the estimated error terms. Using $N(0, \hat{S})$ as the distribution of the error terms and $N(\hat{\alpha}_2, \hat{V}_2)$ as the distribution of the coefficient estimates, a stochastic simulation was then run for the 1970I–1971IV period, where both error terms and coefficients were drawn. The number of trials was 50. The results from this simulation allowed values of d_{itk} to be computed for all i, for $k = 1, \ldots, 8$, and for t equal to 1970I. The simulation produces values of $\tilde{\sigma}^2_{itk}$ and $\tilde{\tilde{y}}_{itk}$. Given $\tilde{\tilde{y}}_{itk}$ and given the actual data on the endogenous variables, $\hat{\epsilon}_{itk}$ can be computed. d_{itk} is then merely $\hat{\epsilon}^2_{itk} - \tilde{\sigma}^2_{itk}$.

The results for one variable in the model (real GNP) from this simulation are presented in the first row of Table 8-1. The first eight values, $100(\tilde{\sigma}_{itk}/\tilde{\tilde{y}}_{itk})$, are the stochastic simulation estimates of the standard errors of the forecast, expressed as a percentage of the forecast mean. The second eight values, $100(|\hat{\epsilon}_{itk}|/\tilde{\tilde{y}}_{itk})$, are the estimates of the standard errors of the forecast based on the actual outside-sample forecast errors, again expressed as a percentage of the forecast mean.

There are a few dummy variables in the model that are not relevant for the early estimation periods, which means that there are slightly fewer than 169 coefficients to estimate for the early periods. For the first period, for example, there are 165 coefficients to estimate.

The second estimation period was 1954I–1969IV (64 observations), which differs from the first period by the addition of one quarter at the end. The first quarter of the period was left unchanged. The coefficients were estimated by 2SLS for this period, and new estimates of $\hat{V}_2$ and $\hat{S}$ were obtained. Stochastic simulation was then performed for the 1970II–1972I period, which allowed values of d_{itk} to be computed for all i, for $k = 1, \ldots, 8$, and for t equal to 1970II. The results for real GNP from this simulation are presented in the second row of Table 8-1. A total of 50 trials were also used for this simulation.

This process was repeated for the remaining 49 estimation periods. Since only data through 1982III exist, the length of the prediction periods for the last seven sets of estimates was less than eight, as can be seen in Table 8-1. The last estimation period was 1954I–1982I (113 observations), and for this set of estimates the prediction period was merely one quarter, 1982III.

The total time needed to estimate the model 51 times was about 2.1 hours on the IBM 4341. The total time for the 51 stochastic simulations, which consisted of 50 trials each, was about 2.2 hours. The stochastic-simulation work consisted of $50 \times 51 = 2{,}550$ solutions of the model. For none of the draws did the Gauss-Seidel technique fail to solve the model. For the earlier work on the VAX, the model was estimated and stochastically simulated 44 times. The total time for the estimation was about 4.8 hours, and the total time for the stochastic simulation (50 trials each) was about 10.7 hours.

The same calculations were performed for the other models, the only difference being that 100 rather than 50 trials were used for each stochastic simulation for the ARUS, VAR1US, and VAR2US models. (50 trials were used for the LINUS model.) The first quarter of the estimation period was 1954I for all the models except ARUS, where it was 1954II. The estimation times for ARUS, VAR1US, VAR2US, and LINUS were, respectively, 3, 9, 3, and 36 minutes on the IBM 4341 and 5, 16, 5, and 19 minutes on the VAX. The stochastic-simulation times were 15, 28, 13, and 14 minutes on the IBM 4341 and 38, 71, 31, and 35 minutes on the VAX.

8.5.2 Discussion of the d_{itk} Values for the US Model

Since the individual d_{itk} values may be of interest in their own right, they will be examined before proceeding to the estimates of the total variance of the forecast error. Consider the results for real GNP in Table 8-1. If one looks down one of the first eight columns, it can be seen that the standard errors vary considerably across prediction periods (except for perhaps the one-quarter-ahead results in the first column). For the eight-quarter-ahead results, for example, the estimated standard errors vary from 1.43 percent in row 35 to 3.41 percent in row 17. Experimenting with more trials indicated that sampling error contributes very little to this variability. It thus appears that there is considerable variability of forecast-error variances across time (for a fixed k), at least for the US model. This variability is due to different estimated covariance matrices, different initial conditions (that is, different lagged values of the endogenous and exogenous variables), and different values of the exogenous variables. It is interesting to note that some of the largest standard errors occur in the mid-1970s, which was characterized at times by extreme initial conditions and exogenous variable values. In particular, the price of imports *(PIM)*, which is an exogenous variable, took on extreme values during much of this period. It may be that these extreme values help contribute to the larger stochastic-simulation estimates of the standard errors for the mid-1970s.

The values in the last eight columns in Table 8-1 are the absolute values of the outside-sample forecast errors in percentage terms. These values, unlike the values in the first eight columns, use the actual values of the endogenous variables for the prediction period in their calculation, which is the reason they are more erratic. In some cases the forecasts are nearly perfect, and in

TABLE 8-1. Estimated standard errors for 51 estimation periods for real GNP for the US model (each prediction period begins two quarters after the end of the estimation period)

Estimation period ending in		k :	$100(\tilde{\sigma}_{itk}/\tilde{\tilde{y}}_{itk})$								$100(\lvert\hat{\varepsilon}_{itk}\rvert/\tilde{\tilde{y}}_{itk})$							
			1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	1969	III	.55	.82	1.17	1.44	1.82	2.16	2.37	2.69	.31	.55	.17	1.32	.16	.01	.43	.05
2		IV	.39	.75	1.01	1.16	1.34	1.52	1.70	1.81	.03	.56	.43	1.17	1.06	.75	1.15	1.56
3	1970	I	.49	.86	1.20	1.52	1.94	2.34	2.80	3.07	1.03	.41	2.30	2.39	2.11	2.60	2.96	3.97
4		II	.43	.79	1.12	1.40	1.77	2.16	2.53	2.87	1.01	.81	1.01	.95	1.57	2.06	3.22	3.33
5		III	.40	.73	1.09	1.48	1.78	2.11	2.36	2.58	1.23	.26	.91	1.20	1.30	.65	.86	1.01
6		IV	.44	.66	.97	1.27	1.54	1.79	2.16	2.44	1.00	1.98	2.25	2.49	1.91	2.13	2.26	1.90
7	1971	I	.52	.82	1.14	1.44	1.74	2.13	2.50	2.69	1.50	2.07	2.53	2.03	2.50	2.88	2.72	4.64
8		II	.52	.75	1.02	1.46	1.96	2.28	2.71	3.17	.46	.62	1.86	2.08	2.23	2.61	.83	.23
9		III	.35	.67	.92	1.24	1.43	1.77	2.02	2.25	.05	1.10	1.18	1.36	2.04	.45	.10	.34
10		IV	.43	.69	.93	1.19	1.41	1.78	2.04	2.26	.85	.72	.79	1.42	.22	.60	1.11	2.25
11	1972	I	.47	.74	1.04	1.24	1.50	1.83	2.24	2.70	.00	.01	.51	1.42	2.00	2.66	4.16	4.69
12		II	.41	.72	.95	1.27	1.54	2.00	2.32	2.79	.29	.74	1.37	2.18	3.30	5.08	5.88	6.63
13		III	.42	.75	.96	1.25	1.57	2.08	2.53	2.94	.37	1.54	2.32	3.26	4.89	5.75	6.42	7.43
14		IV	.47	.73	1.00	1.31	1.55	1.93	2.45	2.86	1.53	1.92	2.49	3.99	4.60	5.08	5.74	7.73
15	1973	I	.50	.70	.92	1.12	1.32	1.63	1.96	2.37	.45	1.24	2.62	3.24	3.70	4.36	6.33	5.57
16		II	.52	.85	1.15	1.58	1.93	2.35	2.58	2.78	.47	1.50	1.63	1.46	1.73	3.47	2.24	1.67
17		III	.55	1.02	1.40	1.95	2.41	2.85	3.28	3.41	.86	1.17	1.34	2.03	4.19	3.59	3.12	3.80
18		IV	.49	.91	1.38	1.87	2.20	2.61	2.73	2.84	.29	.66	1.24	3.39	2.81	2.45	2.94	1.85
19	1974	I	.57	1.00	1.30	1.76	2.11	2.28	2.49	2.57	.27	.11	1.99	1.47	1.15	1.91	.93	.93
20		II	.43	.77	1.31	1.67	2.01	2.36	2.51	2.65	.41	.77	.45	1.41	1.10	2.30	2.27	1.60
21		III	.54	.97	1.22	1.49	1.79	2.09	2.18	2.29	1.01	.21	1.03	.58	1.73	1.82	1.40	1.33
22		IV	.56	.82	1.12	1.49	1.87	2.19	2.62	2.98	.22	.39	.29	.78	.94	.82	.97	2.27
23	1975	I	.56	.91	1.27	1.73	2.12	2.51	2.80	3.09	.67	1.91	1.29	1.44	1.87	2.01	1.02	.72
24		II	.48	.92	1.27	1.62	1.89	2.27	2.71	2.91	.77	.11	.12	.88	1.42	.93	1.27	1.65
25		III	.38	.68	.97	1.37	1.63	1.85	2.01	2.13	.78	.77	.20	.11	.82	.83	1.03	.18
26		IV	.48	.77	1.10	1.47	1.64	1.77	1.92	2.04	.08	.19	.16	.98	1.11	1.41	.60	.21
27	1976	I	.42	.76	.95	1.16	1.36	1.59	1.67	1.87	.44	.85	.02	.05	.11	.87	1.56	.56
28		II	.44	.67	.92	1.06	1.38	1.66	1.83	1.96	.14	.88	.95	1.02	.10	.68	.11	.59
29		III	.47	.85	1.17	1.34	1.46	1.58	1.72	1.83	.61	.49	.55	.39	.72	.40	.04	.01
30		IV	.45	.64	.91	1.07	1.21	1.26	1.39	1.65	.06	.12	.87	1.34	.37	1.00	1.01	1.60

31	1977	I	.42	.71	.91	1.09	1.13	1.19	1.40	1.62	.29	.67	1.21	.37	.97	.96	1.67	2.59
32		II	.42	.76	1.08	1.34	1.55	1.76	1.92	2.17	.84	1.21	.08	.66	.64	1.27	2.26	2.21
33		III	.40	.57	.88	1.09	1.29	1.51	1.63	1.65	.66	.03	.69	1.01	1.82	2.73	2.69	3.65
34		IV	.39	.56	.75	.92	1.05	1.28	1.37	1.44	1.03	.21	.12	1.14	2.28	2.38	3.31	3.94
35	1978	I	.35	.63	.88	1.10	1.29	1.37	1.39	1.43	.29	.09	.73	1.73	1.87	2.91	3.62	6.26
36		II	.51	.92	1.13	1.26	1.42	1.58	1.59	1.69	.10	.46	1.46	1.46	2.53	3.13	5.91	5.87
37		III	.44	.76	1.03	1.33	1.54	1.78	1.91	2.02	.29	.89	.71	1.58	2.08	4.68	4.64	4.30
38		IV	.45	.70	1.05	1.28	1.34	1.43	1.45	1.48	1.07	1.04	1.79	2.23	4.71	4.54	3.89	2.87
39	1979	I	.39	.57	.77	.94	1.20	1.39	1.51	1.70	.30	1.43	2.05	4.80	4.77	4.34	3.48	4.69
40		II	.39	.63	.89	1.08	1.21	1.30	1.49	1.57	1.30	1.99	4.54	4.40	3.83	2.78	4.02	4.19
41		III	.54	.67	1.05	1.20	1.47	1.70	1.84	1.88	.97	3.79	3.74	3.26	2.29	3.65	3.93	6.23
42		IV	.50	.83	1.05	1.37	1.64	1.77	1.91	1.98	2.59	2.30	1.69	.71	2.11	2.64	4.96	6.25
43	1980	I	.50	.81	1.00	1.14	1.26	1.39	1.45	1.54	.19	.39	.11	1.61	2.21	4.87	6.17	6.39
44		II	.58	.86	.98	1.28	1.50	1.61	1.72	1.73	.57	.24	1.80	2.22	4.78	5.91	5.91	6.15
45		III	.59	.81	1.08	1.35	1.62	1.72	1.79		.85	.42	.99	3.75	5.23	5.57	6.10	
46		IV	.56	.94	1.26	1.48	1.67	1.82			.97	.96	3.34	4.47	4.70	4.97		
47	1981	I	.64	.79	.99	1.33	1.43				.39	1.69	2.79	3.18	3.61			
48		II	.53	.92	1.18	1.42					1.09	1.14	.63	.95				
49		III	.60	.94	1.21						.23	.06	.19					
50		IV	.54	.82							.16	.04						
51	1982	I	.56								.41							

Notes: • Estimation technique was 2SLS. Each of the estimated covariance matrices of the coefficient estimates was taken to be block diagonal.
• 50 trials were used for each stochastic simulation.

others the errors are quite large. The largest error is for the eight-quarter-ahead forecast in row 14, which is 7.73 percent. The results in row 14 are for the prediction period beginning in 1973II, and therefore the eight-quarter-ahead forecast is for 1975I.

The square of an element in the right half of Table 8-1 minus the square of the corresponding element in the left half is equal to $d_{itk}/\tilde{\bar{y}}^2_{itk}$, which is simply d_{itk} in percentage terms. The key question is whether these values have any systematic tendencies. To examine this question, $d_{itk}/\tilde{\bar{y}}^2_{itk}$ is plotted in Figure 8-1 for i equal to real GNP and k equal to 1. The main conclusion from Figure 8-1 is that no systematic tendencies are apparent. The value for 1980II is very large relative to others, but aside from this, the values are not obviously larger for one subperiod than for another, and there is no obvious trend. Plots for many other variables were examined, and the same conclusion was reached.

The only systematic tendency that was apparent was that some of the plots showed evidence of serial correlation for values of k greater than about four or five. This can be explained as follows. If, say, quarter 85 is a difficult quarter to predict, perhaps because of a large unexplained shock in the quarter, then a dynamic simulation that runs through this quarter may also do poorly in predicting quarters 86 and beyond. In other words, the simulation may get thrown off by the bad prediction in quarter 85. This means, for example, that five-quarter-ahead forecasts for quarters 85, 86, 87, 88, and 89 may all be on average poor, thus implying large values for $\hat{\epsilon}^2_{itk}$ ($k = 5$ and $t = 81, \ldots, 85$). The shock in quarter 85 will have no effect on the stochastic-simulation estimates of the variances, since these are not based on the actual data for the endogenous variables for this quarter, and therefore the large values of the outside-sample errors imply large values of d_{itk}. In this way, serial correlation may be introduced into the d_{itk} series for values of k greater than one.

The general impression one gets from examining the plots is thus that the misspecification of the model does not appear to have changed over time or to have been different in any subperiods. One could attempt to examine this question in a less casual way by, say, regressing the d_{itk} values (for a given i and k) on variables that one thinks may be related to the misspecification of the model. Although this might be worth doing in future work, it seems unlikely to me, from having examined the plots, that much would come of it.

The fact that the misspecification of the model does not appear to have changed over time is not in itself encouraging regarding the accuracy of the model. The misspecification may in fact be quite large, even though unchanging, and may have a large effect on total forecasting uncertainty. What is encouraging about the results is that the assumption of a constant mean for d_{itk} or $d_{itk}/\tilde{\bar{y}}^2_{itk}$ (for a given i and k) seems to be a reasonable approximation.

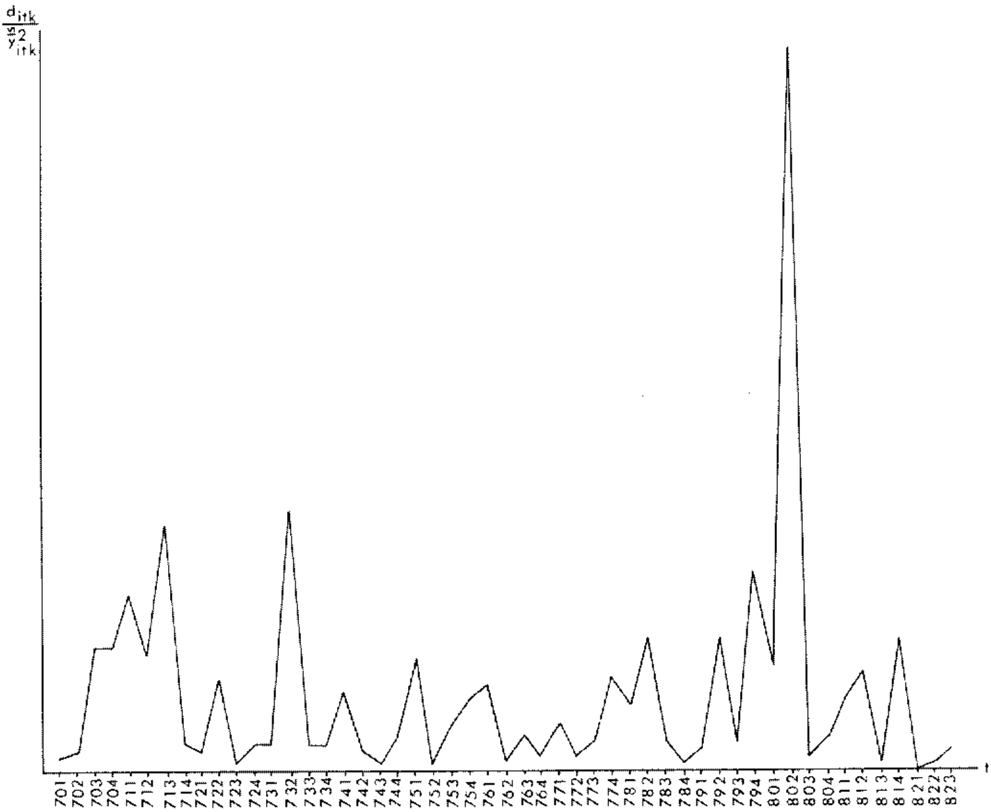

Figure 8-1 Plot of $d_{itk}/\bar{\bar{y}}^2_{itk}$ for the US model for i = real GNP, $k = 1$, t = 1970I – 1982III

8.5.3 Computing the Total Variance of the Forecast Error

The total variance of the forecast error is $\hat{\sigma}^2_{itk}$ in (8.7). The computation of $\hat{\sigma}^2_{itk}$ for the five models is discussed in this section. It is easiest to describe these computations by referring to the results in Table 8-2. The prediction period is 1978I – 1979IV. Consider first the results for real GNP for the US model. The values in the a and b rows are from the same two stochastic simulations that were used for the results in Table 7-1. For the a-row results only draws of the error terms were made, whereas for the b-row results draws of both the error terms and coefficients were made. The number of trials for each simulation was 250. The coefficient estimates that were used for these results are the 2SLS estimates for the 1954I – 1982III period (115 observations). These are the

TABLE 8-2. Estimated standard errors of forecasts for 1978 I - 1979 IV for five models

		1978				1979			
		I	II	III	IV	I	II	III	IV
GNPR: Real GNP									
US:	a	.49	.66	.81	.98	1.10	1.14	1.22	1.32
	b	.51	.69	.89	1.03	1.09	1.22	1.30	1.35
	c	.61	.71	.89	1.08	1.24	1.39	1.52	1.60
	d	.87	1.10	1.54	2.00	2.43	2.79	3.09	3.43
ARUS:	a	.78	1.20	1.56	1.81	1.97	2.08	2.14	2.17
	b	.77	1.22	1.57	1.80	2.00	2.12	2.24	2.38
	d	1.25	1.97	2.57	2.94	3.14	3.37	3.68	4.05
VAR1US:	a	.72	1.05	1.24	1.43	1.58	1.71	1.76	1.87
	b	.77	1.20	1.44	1.64	1.91	2.15	2.38	2.59
	d	1.67	2.96	3.71	4.45	5.00	5.63	6.36	7.15
VAR2US:	a	.80	1.14	1.34	1.52	1.68	1.78	1.86	1.96
	b	.85	1.26	1.49	1.80	2.01	2.17	2.44	2.60
	d	1.34	2.35	3.21	3.95	4.30	4.44	4.70	4.93
LINUS:	a	.59	.87	1.07	1.25	1.37	1.47	1.56	1.62
	b	.61	.87	1.03	1.21	1.39	1.55	1.75	1.87
	c	.67	.93	1.12	1.33	1.46	1.65	1.72	1.89
	d	.97	1.53	2.14	2.76	3.38	4.00	4.58	5.13
GNPD: GNP deflator									
US:	a	.35	.51	.64	.72	.79	.87	.90	.93
	b	.37	.52	.63	.77	.84	.93	1.09	1.15
	c	.50	.61	.68	.80	.87	.99	1.10	1.13
	d	.64	1.03	1.46	1.98	2.49	3.03	3.65	4.31
ARUS:	a	.27	.49	.72	.97	1.23	1.43	1.63	1.79
	b	.26	.51	.76	1.08	1.40	1.71	1.99	2.25
	d	.49	.98	1.50	2.09	2.78	3.48	4.16	4.90
VAR1US:	a	.27	.41	.54	.69	.82	.95	1.05	1.13
	b	.29	.45	.59	.78	.96	1.16	1.32	1.46
	d	.56	1.00	1.39	1.82	2.30	2.70	3.03	3.38
VAR2US:	a	.31	.48	.65	.82	.94	1.07	1.17	1.26
	b	.33	.52	.73	.92	1.08	1.24	1.39	1.57
	d	.54	.98	1.41	1.88	2.27	2.57	2.86	3.21
100·UR: Unemployment rate (percentage points)									
US:	a	.24	.38	.48	.54	.61	.65	.65	.68
	b	.26	.42	.52	.61	.70	.77	.78	.80
	c	.27	.41	.52	.61	.66	.76	.81	.82
	d	.45	.69	.92	1.14	1.33	1.52	1.68	1.83
ARUS:	a	.32	.58	.79	.98	1.07	1.10	1.11	1.14
	b	.30	.58	.86	1.09	1.24	1.32	1.38	1.43
	d	.37	.73	1.07	1.39	1.57	1.69	1.78	1.84
VAR1US:	a	.25	.42	.54	.61	.69	.73	.75	.77
	b	.26	.46	.61	.70	.76	.83	.88	.96
	d	.56	1.19	1.71	2.10	2.33	2.47	2.60	2.85
VAR2US:	a	.28	.48	.62	.70	.78	.82	.85	.88
	b	.29	.52	.67	.78	.88	.94	.99	1.02
	d	.44	.90	1.38	1.84	2.17	2.34	2.44	2.50

(continued)

TABLE 8-2 (continued)

		1978				1979			
		I	II	III	IV	I	II	III	IV
RS: Bill rate (percentage points)									
US:	a	.71	1.00	1.07	1.13	1.17	1.21	1.17	1.19
	b	.73	.94	1.04	1.03	1.15	1.25	1.31	1.45
	c	.72	.96	1.07	1.14	1.13	1.27	1.42	1.40
	d	1.20	1.79	1.99	2.15	2.29	2.42	2.50	2.51
ARUS:	a	.72	1.11	1.20	1.35	1.53	1.62	1.70	1.75
	b	.76	1.21	1.37	1.54	1.75	1.96	2.06	2.04
	d	1.36	2.24	2.48	2.85	3.21	3.61	3.83	3.85
VAR1US:	a	.56	.84	.93	1.03	1.07	1.19	1.26	1.22
	b	.59	.96	1.07	1.24	1.39	1.53	1.62	1.70
	d	1.31	2.26	2.59	2.80	3.13	3.64	3.72	3.73
VAR2US:	a	.63	.99	1.10	1.18	1.27	1.39	1.45	1.45
	b	.65	1.08	1.29	1.44	1.61	1.72	1.79	1.81
	d	1.24	2.19	2.71	3.18	3.58	4.01	4.31	4.40
LINUS:	a	.77	.97	1.07	1.19	1.25	1.26	1.33	1.37
	b	.80	1.06	1.18	1.29	1.37	1.42	1.47	1.60
	c	.81	1.05	1.21	1.32	1.40	1.52	1.57	1.70
	d	1.32	1.78	1.96	2.16	2.39	2.64	2.82	2.96
M1: Money supply									
US:	a	.98	1.35	1.49	1.66	1.82	2.00	2.03	1.98
	b	.95	1.37	1.57	1.77	2.11	2.32	2.38	2.54
	c	1.07	1.45	1.64	1.69	2.02	2.00	2.18	2.28
	d	1.41	1.91	2.22	2.68	3.35	3.81	4.62	5.33
ARUS:	a	.66	.77	.97	1.18	1.35	1.50	1.70	1.99
	b	.69	.85	1.09	1.32	1.55	1.74	2.04	2.33
	d	1.22	1.41	1.63	2.08	2.39	2.64	3.07	3.47
VAR1US:	a	.65	.75	.89	.96	1.03	1.11	1.21	1.38
	b	.67	.87	1.04	1.18	1.30	1.48	1.66	1.83
	d	1.25	1.35	1.25	1.10	1.02	.34	1.25	2.02
VAR2US:	a	.77	.93	1.16	1.28	1.40	1.50	1.59	1.76
	b	.82	1.00	1.28	1.53	1.76	1.95	2.09	2.37
	d	1.35	1.71	1.99	2.25	2.33	2.30	2.46	2.71
W_f: Wage rate									
US:	a	.56	.86	1.00	1.13	1.20	1.30	1.32	1.38
	b	.59	.84	1.01	1.15	1.35	1.48	1.70	1.85
	c	.60	.82	1.02	1.17	1.37	1.56	1.80	1.94
	d	.54	.86	1.30	1.81	2.43	3.12	3.96	4.83
ARUS:	a	.38	.55	.69	.81	.95	1.06	1.18	1.29
	b	.40	.58	.73	.92	1.05	1.19	1.33	1.52
	d	.67	1.03	1.35	1.71	2.03	2.36	2.75	3.14
Π_f: Profits									
US:	a	4.98	6.26	7.61	8.75	8.90	10.07	10.91	11.80
	b	5.02	6.55	8.35	9.26	10.27	11.55	13.46	15.11
	c	6.86	7.63	9.21	9.77	10.63	13.22	14.16	15.00
	d	8.49	10.30	12.83	14.59	13.27	13.68	13.29	12.17
ARUS:	a	4.84	6.87	8.21	8.72	9.55	10.33	11.50	13.24
	b	4.82	6.57	8.27	8.67	9.87	10.91	12.57	15.64
	d	9.70	14.79	18.47	20.94	23.92	26.92	32.12	37.88

(continued)

TABLE 8-2 (continued)

		1978				1979			
		I	II	III	IV	I	II	III	IV
SR: Savings rate of the household sector									
US:	a	6.88	10.05	11.98	14.89	18.41	21.88	22.40	25.99
	b	7.65	10.74	12.33	15.54	20.22	25.08	25.82	30.49
	c	10.73	13.26	14.56	17.34	21.47	26.45	27.28	32.05
	d	11.41	14.58	16.99	21.34	27.08	34.63	40.12	49.01
ARUS:	a	9.46	11.30	12.84	13.71	14.29	15.38	14.94	14.46
	b	9.86	11.88	13.77	13.92	15.18	16.00	16.59	16.58
	d	12.29	14.70	17.21	17.51	19.34	20.65	21.40	21.38
S_g: Savings of the federal government (billions of current dollars)									
US:	a	.72	1.12	1.43	1.69	1.90	2.14	2.38	2.50
	b	.76	1.11	1.45	1.80	1.98	2.33	2.79	2.84
	c	2.29	2.55	2.60	3.07	2.99	3.27	3.52	3.94
	d	2.42	2.82	3.18	3.95	4.37	4.99	5.58	6.31
ARUS:	a	2.84	3.90	4.96	5.65	6.03	6.59	6.77	7.08
	b	3.39	4.54	5.55	6.32	7.12	7.57	8.21	8.66
	d	6.07	8.42	10.26	11.44	12.45	12.89	12.96	12.81
S_r: Savings of the foreign sector (billions of current dollars)									
US:	a	.95	1.27	1.38	1.49	1.61	1.66	1.72	2.09
	b	1.11	1.29	1.45	1.58	1.66	2.05	2.31	2.51
	c	1.64	1.86	1.89	2.18	2.38	2.76	2.85	3.06
	d	2.39	3.39	4.01	4.59	5.22	6.32	7.31	8.35
ARUS:	a	1.07	1.27	1.45	1.46	1.51	1.55	1.60	1.62
	b	1.14	1.41	1.65	1.80	1.77	1.86	1.85	1.89
	d	2.38	2.83	3.44	3.53	3.61	3.61	3.52	3.66

Notes: a = Uncertainty due to error terms.
b = Uncertainty due to error terms and coefficient estimates.
c = Uncertainty due to error terms, coefficient estimates, and exogenous-variable forecasts.
d = Uncertainty due to error terms, coefficient estimates, exogenous-variable forecasts, and the possible misspecification of the model.
• For the unemployment rate, the bill rate, the savings of the federal government, and the savings of the foreign sector, the errors are in the natural units of the variables. These units are indicated in the table. For all other variables the errors are expressed as a percent of the forecast mean (in percentage points).

estimates presented in Chapter 4; they are the basic 2SLS estimates of the model.

The values in Table 8-2 are either estimated standard errors in units of the variable or estimated standard errors in percentage points. For real GNP the errors are in percentage points. The numbers in the b row, for example, are $\tilde{\sigma}_{itk}/\tilde{\bar{y}}_{itk}$, where $\tilde{\bar{y}}_{itk}$ is the stochastic-simulation estimate of the forecast mean (Eq. 7.7) and $\tilde{\sigma}_{itk}$ is the square root of the stochastic-simulation estimate of the variance of the forecast error (Eq. 7.8). The numbers in the a row are the same except that the estimates are based on draws of the error terms only.

The results in the a and b rows are not needed for the computations of the total variance of the forecast error; they are presented merely to show how much of the total variance can be attributed to the uncertainty from the error terms and coefficient estimates. The results that are needed are those from a stochastic simulation with respect to the error terms, coefficients, *and* exogenous variables. These results are presented in the c rows in Table 8-2. The procedure that was used for this stochastic simulation for the US model is as follows.

An eighth-order autoregressive equation (with a constant and time trend included) was estimated for each exogenous variable in Table A-4 (Appendix A) except for the dummy variables, the time trend, and variables whose value never changes or changes only once during the sample period. (These variables are $D593$ through $DD793$, H_m, t, δ_D, δ_H, δ_K, γ_g, and γ_s.) The sample period for each regression was 1954II – 1982III. A total of 88 equations were estimated. The estimated standard error from each of these regressions was taken to be the error associated with forecasts of the variable. The procedure discussed in Section 8.4.2 was used for the draws of the exogenous-variable values for the stochastic simulation. The base values of the exogenous variables were taken to be the actual values. Each trial of the stochastic simulation for the c rows consisted of eight draws of 30 values each from the distribution of the error terms, one draw of 169 values from the distribution of the coefficient estimates, and eight draws from each of the 88 distributions of the exogenous-variable errors. A total of 250 trials were taken. For none of the draws did the Gauss-Seidel technique fail to find a solution. The total time taken for this simulation was about the same as the time taken for the a-row and b-row simulations, namely about 6.7 minutes on the IBM 4341 and about 49 minutes on the VAX. (See the note to Table 7-1.)

A stochastic simulation of 250 trials was also performed under the assumption that the exogenous-variable errors pertain to changes in the variables rather than to levels. This procedure is also discussed in Section 8.4.2. The estimated standard errors from this simulation were in general larger than those from the first simulation, but the results were fairly close. These results are not reported in Table 8-2.

The c-row values in Table 8-2 are either $\tilde{\tilde{\sigma}}_{itk}$ or $\tilde{\tilde{\sigma}}_{itk}/\tilde{\tilde{y}}_{itk}$, where $\tilde{\tilde{\sigma}}_{itk}$ is the square root of $\tilde{\tilde{\sigma}}^2_{itk}$. The final step is to add to $\tilde{\tilde{\sigma}}^2_{itk}$ the estimated mean of d_{itk}. The discussion in Section 8.5.2 indicates that the assumption that the mean of d_{itk} is constant across time may be a reasonable approximation. This assumption was used for variables without trends. For variables with trends it was assumed that the mean of $d_{itk}/\tilde{\bar{y}}_{itk}^2$ is constant across time. Given the first assumption, the estimated mean of d_{itk} is the average of the d_{itk} values (for a fixed i and k); and given the second assumption, the estimated mean of $d_{itk}/\tilde{\bar{y}}_{itk}^2$ is the average of the $d_{itk}/\tilde{\bar{y}}_{itk}^2$ values. There are 51 observations for the one-quarter-ahead forecasts ($k = 1$), 50 observations for the two-quarter-ahead forecasts ($k = 2$), and so on. Let $\hat{d}_{ik}$ denote the estimated mean of d_{itk}, and let $\hat{d}'_{ik}$ denote the estimated mean of $d_{itk}/\tilde{\bar{y}}_{itk}^2$. The t subscript has been dropped from $\hat{d}_{ik}$ and $\hat{d}'_{ik}$ because the estimated means are assumed to be constant across time.

For variables without trends the estimate of the total variance of the forecast error, $\hat{\sigma}_{itk}^2$, is $\tilde{\tilde{\sigma}}_{itk}^2 + \hat{d}_{ik}$. For variables with trends the estimate is $\tilde{\tilde{\sigma}}_{itk}^2 + \hat{d}'_{ik} \cdot \tilde{\bar{y}}_{itk}^2$. For variables without trends the values in the d rows in Table 8-2 are the square roots of $\hat{\sigma}_{itk}^2$, and for variables with trends the values are the square roots of $\hat{\sigma}_{itk}^2/\tilde{\bar{y}}_{itk}^2$. The differences between the d-row and c-row values in the table are measures of the effects of misspecification on predictive accuracy, although this is subject to the qualification discussed in Section 8.4.5 about the interpretation of the word "misspecification."

The same procedure was followed for the other models. There are no exogenous variables in the ARUS, VAR1US, and VAR2US models, and thus there are no c-row values. For the LINUS model there are three exogenous variables for which autoregressive equations were estimated: Q_1, Q_2, and $\dot{M}1$.

8.5.4 Comparison of the Results for the Five Models

The US Model versus the Others

The models can be compared according to the size of the d-row values. In examining the d-row values I usually give more weight to the results the further out the forecast is. In other words, I usually give more weight to the four-quarter-ahead results than to the one-quarter-ahead results, more to the eight-quarter-ahead results than to the four-quarter-ahead results, and so on. The further out a forecast is, the more this is a test of the accuracy of the dynamic properties of the model.

For real GNP it is clear that the US model is substantially better than the other four models. The eight-quarter-ahead standard error is 3.43 percent, which compares to values of 4.05, 7.15, 4.93, and 5.13 percent for the other four models. The US model is also best for the unemployment rate and the bill rate. It is not as good as VAR1US and VAR2US for the GNP deflator. It is substantially worse for the money supply, where the eight-quarter-ahead standard error is 5.33 percent, which compares to values of 3.47, 2.02, and 2.71 for ARUS, VAR1US, and VAR2US respectively.

The poorer results for the money supply mean that the demand-for-money equations in the US model are not as accurate as autoregressive specifications. This is something that I have known for a long time, but it is not easy to remedy. I have so far been unable to find demand-for-money equations that lead to more accurate predictions within the context of the overall model. Fortunately, errors in predicting the money supply have fairly minor consequences for the other variables. Given the use of the interest rate reaction function, the only important way in which errors in predicting the money supply affect the other variables in the model is through their effect on the bill rate predictions. The lagged growth of the money supply is one of the explanatory variables in the bill rate equation, and therefore errors in predicting the money supply affect the bill rate predictions. Although errors in predicting the bill rate have important effects on many other variables in the model, the effect of the money supply on the bill rate is only moderate. The indirect effect of money supply errors on the other variables in the model (through the direct effect of the money supply on the bill rate) is thus fairly minor.

Given that the US model is more accurate for three of the key variables (real GNP, the unemployment rate, and the bill rate), the results seem encouraging for the model. More tests are needed, of course, especially against other structural models, before any strong conclusions can be drawn.

For the remaining five variables in Table 8-2, the comparisons are only between the US and ARUS models. Four of these variables—the level of profits, the savings rate, the savings of the federal government, and the savings of the foreign sector—are "residual" variables. These types of variables are generally hard to predict in structural models, and it is interesting to see how the US model does relative to an autoregressive equation for each variable. The results for the first variable, the wage rate, are about the same for the two models for the first four quarters; after that the ARUS model does somewhat better. For the savings rate of the household sector, the two models are almost the same for the first three quarters, and the ARUS model is substantially better thereafter. The US model is substantially better for profits and the savings of the federal government, and the ARUS model is substantially better for the savings of the foreign sector. The overall results for these five variables are thus mixed. It is encouraging that the US model is better with respect to profits and the savings of the federal government, but it is clear that the model could stand some improvement with respect to the savings rate of the household sector and the savings of the foreign sector.

Comparison of the Other Four Models

Consider first the LINUS model. The main variable that it is designed to explain is real GNP. For this variable it is less accurate than the US model and more accurate than the VAR1US and VAR2US models. It is more accurate than the ARUS model for the first four quarters ahead and less accurate after that. The results are thus mixed, although the fact that the model is not nearly as accurate as the US model is not encouraging in regard to the ability to collapse a large model into a relatively small one without a substantial loss of predictive accuracy.

In the comparison of VAR1US versus VAR2US, VAR2US seems somewhat better: it is more accurate for real GNP, the GNP deflator, and the unemployment rate. It is less accurate for the bill rate and the money supply. In the comparison of ARUS versus VAR2US, ARUS is more accurate for real GNP and the unemployment rate but less accurate for the GNP deflator and the bill rate. The results are mixed for the money supply. There is thus no obvious winner between ARUS and VAR2US.

There is one feature of the money supply results for VAR1US that should be noted. For the four- through seven-quarter-ahead predictions, the d-row values are less than the corresponding b-row values, which means that the estimated means of the $d_{itk}/\tilde{y}^2_{itk}$ values were negative. For the six-quarter-ahead prediction, the estimated mean was almost negative enough to make the d-row value zero. These results are due to the fact that the stochastic-simulation estimates of the variances are large relative to the estimates based on the outside-sample forecast errors. For models like VAR1US, which have a large number of coefficients to estimate relative to the number of observations and thus in general have very imprecise estimates, it sometimes happens that the stochastic-simulation estimates of the variances are very large. It is not clear in these cases whether much confidence should be put in the results; there are just too few observations for much to be said.

Comparison Using Root Mean Squared Errors

Root mean squared errors (RMSEs) for the five models for the 1970I–1982III period are presented in Table 8-3. These errors were computed as follows. Outside-sample forecast errors are available from the 51 stochastic simulations that were involved in computing the d_{itk} values. These errors are the

TABLE 8-3. Root mean squared errors of outside-sample forecasts for 1970 I - 1982 III for five models

	Number of quarters ahead							
	1	2	3	4	5	6	7	8
GNPR: Real GNP								
US	.79	1.15	1.65	2.16	2.65	3.07	3.43	3.83
ARUS	1.21	1.92	2.58	3.04	3.33	3.63	3.97	4.34
VAR1US	1.77	3.06	3.91	4.80	5.49	6.25	7.21	8.29
VAR2US	1.36	2.38	3.29	4.08	4.58	4.98	5.47	6.05
LINUS	.87	1.46	2.16	2.88	3.65	4.42	5.21	5.96
GNPD: GNP deflator								
US	.57	1.05	1.56	2.12	2.70	3.29	3.97	4.71
ARUS	.48	.97	1.49	2.09	2.80	3.55	4.32	5.16
VAR1US	.61	1.08	1.52	2.03	2.60	3.11	3.58	4.09
VAR2US	.56	1.01	1.44	1.94	2.38	2.73	3.09	3.50
100·UR: Unemployment rate (percentage points)								
US	.47	.76	1.01	1.26	1.49	1.69	1.89	2.10
ARUS	.37	.74	1.06	1.36	1.52	1.63	1.73	1.79
VAR1US	.58	1.21	1.72	2.11	2.35	2.49	2.65	2.93
VAR2US	.44	.90	1.39	1.85	2.16	2.34	2.46	2.56
RS: Bill rate (percentage points)								
US	1.10	1.69	1.90	2.08	2.28	2.40	2.45	2.56
ARUS	1.28	2.11	2.34	2.67	2.97	3.30	3.48	3.50
VAR1US	1.30	2.27	2.64	2.87	3.26	3.79	3.89	3.91
VAR2US	1.20	2.13	2.64	3.13	3.53	3.97	4.27	4.39
LINUS	1.19	1.64	1.82	2.03	2.27	2.52	2.72	2.85
M1: Money supply								
US	1.33	1.91	2.42	3.16	3.92	4.68	5.64	6.60
ARUS	1.18	1.40	1.65	2.08	2.37	2.64	3.01	3.38
VAR1US	1.39	1.62	1.82	2.07	2.42	2.64	3.38	4.21
VAR2US	1.36	1.78	2.16	2.52	2.79	3.07	3.57	4.10

Notes:
- The results are based on 51 sets of coefficient estimates for each model.
- Each prediction period began two quarters after the end of the estimation period.
- The predicted values used were the mean values from the 51 stochastic simulations to get the d_{itk} values for each model.
- There are 51 observations for the one-quarter-ahead forecasts, 50 for the two-quarter-ahead forecasts, and so on.
- For the unemployment rate and the bill rate the errors are in the natural units of the variables. For the other variables the errors are expressed as a percent of the forecast mean (in percentage points).

differences between the mean values (the $\tilde{\bar{y}}_{itk}$) and the actual values. They are based on 51 sets of estimates of each model, where each prediction period begins two quarters after the end of the estimation period. From these errors one can compute RMSEs by merely adding the squared errors, dividing by the number of observations, and taking the squared root. For the one-quarter-ahead predictions there are 51 observations, for the two-quarter-ahead predictions there are 50 observations, and so on.

It is of interest to compare the RMSEs in Table 8-3 with the d-row values in

Table 8-2. In some loose sense the RMSEs handle the effects of misspecification because they are based on outside-sample errors only, and thus the main differences between the RMSEs and the d-row values are that the RMSEs do not handle exogenous variable uncertainty and do not account for the fact that forecast error variances vary across time. The RMSEs and the d-row values differ, in some cases by substantial amounts, but the rankings of the models are roughly (but not exactly) the same. One would probably draw similar conclusions as those given above if one looked only at the RMSE results.

The main reason for the similar rankings is that exogenous-variable uncertainty is not much of a problem in any model. For three of the models there are no exogenous variables, and for the US and LINUS models, which have exogenous variables, the differences between the c-row and b-row values in Table 8-2 are not in general very large. The US model in particular does not appear to be heavily tied to hard-to-forecast exogenous variables. For models that are heavily tied and that differ considerably in the number and types of variables that are taken to be exogenous, the difference between the rankings using the RMSEs and those using the d-row values could be substantial.

With respect to the cost of the calculations, the RMSE results are essentially as costly as the d-row results because both are based on 51 sets of estimates and 51 stochastic simulations. The RMSE results could, however, be made less costly by using deterministic simulations to compute the predicted values. As discussed in Section 7.3, predicted values from deterministic simulations are generally close to expected values from stochastic simulations, so little is likely to be lost by using deterministic simulations. In the present case this would save about half the cost, since about half the time was spent computing the estimates and about half in performing the stochastic simulations.

8.5.5 Other Results for the US Model

Comparison across Rows

It should be clear from examining the a and b rows in Table 8-2 that more of the forecasting uncertainty is due to the error terms than to the coefficient estimates: the differences between the b and a rows are small relative to the size of the a-row values. It should also be clear, as noted earlier, that exogenous-variable uncertainty does not contribute very much to total uncertainty: the differences between the c and b rows are small. The variable most affected by exogenous-variable uncertainty in Table 8-2 is the savings of the federal government. This is, of course, as expected, since many of the key exogenous variables in the model are federal government variables.

It should be noted that there is no requirement that each c-row value be greater than its corresponding b-row value. Although this is rare, an increase in the variability of one endogenous variable may be associated with a decrease in the variability of another. In the results for the US model in Table 8-2, one of the c-row values is less than the corresponding b-row value for the GNP deflator, three are less for the unemployment rate, three are less for the bill rate, five are less for the money supply, and one is less for the wage rate.

The d-row values are sometimes more than twice as large as the corresponding c-row values, which means that misspecification contributes substantially to overall uncertainty. For real GNP the d-row value for the eight-quarter-ahead prediction is 3.43 percent, which compares to the c-row value of 1.60 percent. For the GNP deflator the numbers are 4.31 and 1.13 percent. Only one d-row value is less than the corresponding c-row value for the US model, which is for the one-quarter-ahead prediction of the wage rate. When this happens, as noted earlier, it means that the estimated mean of d_{itk} or $d_{itk}/\tilde{\bar{y}}^2_{itk}$ is negative. It is argued in Section 8.4.3 that the estimated means are in general likely to be positive, and the results in Table 8-2 certainly confirm this.

An Alternative Measure of Dispersion

In order to see whether the possible nonexistence of moments is a problem, an alternative measure of dispersion from the variance was computed for some of the variables. This measure, $\tilde{\delta}_{itk}$, is discussed in Section 7.3.2. It is equal to $(\tilde{y}^b_{itk} - \tilde{y}^a_{itk})/2$, where $\tilde{y}^a_{itk}$ it the value for which 34.135 percent of the trial values lie above it and below the median and $\tilde{y}^b_{itk}$ is the value for which 34.135 percent of the trial values lie below it and above the median. If the nonexistence of moments is a problem, one might expect $\tilde{\sigma}_{itk}$ to be much larger than $\tilde{\delta}_{itk}$.

The results for one stochastic simulation for the US model are presented in Table 8-4. This is the same simulation that was used for the b-row results in Table 8-2. The draws are with respect to the error terms and coefficients. The number of trials was 250. None of the draws resulted in a failure of the Gauss-Seidel technique to find a solution, and therefore no "extreme" draws had to be discarded. The values in the a rows in Table 8-4 are either estimated standard errors, $\tilde{\sigma}_{itk}$, or estimated standard errors as a percentage of the

TABLE 8-4. Stochastic simulation with respect to error terms and coefficient estimates:* two measures of dispersion

	1978				1979			
	I	II	III	IV	I	II	III	IV
GNPR: Real GNP								
a	.51	.69	.89	1.03	1.09	1.22	1.30	1.35
b	.53	.67	.89	1.03	1.11	1.29	1.42	1.33
GNPD: GNP deflator								
a	.37	.52	.63	.77	.84	.93	1.09	1.15
b	.34	.49	.61	.77	.82	.97	1.14	1.24
100·UR: Unemployment rate (percentage points)								
a	.26	.42	.52	.61	.70	.77	.78	.80
b	.26	.43	.55	.58	.69	.76	.73	.74
RS: Bill rate (percentage points)								
a	.73	.94	1.04	1.03	1.15	1.25	1.31	1.45
b	.72	.96	1.03	1.03	1.17	1.20	1.22	1.39
M1: Money supply								
a	.95	1.37	1.57	1.77	2.11	2.32	2.38	2.54
b	.94	1.45	1.55	1.74	2.06	2.41	2.33	2.45

Notes: a = $\tilde{\sigma}_{itk}$ for UR and RS, $\tilde{\sigma}_{itk}/\tilde{\tilde{y}}_{itk}$ for the others.
b = $\tilde{\delta}_{itk}$ for UR and RS, $\tilde{\delta}_{itk}/\tilde{y}^m_{itk}$ for the others, where $\tilde{y}^m_{itk}$ is the stochastic-simulation estimate of the median.
* This stochastic simulation is the same as the one used for the b-row results for the US model in Table 8-2.

forecast mean, $\tilde{\sigma}_{itk}/\tilde{\tilde{y}}_{itk}$. The values in the b rows are either $\tilde{\delta}_{itk}$ or $\tilde{\delta}_{itk}$ as a percentage of the forecast median, $\tilde{\delta}_{itk}/\tilde{y}^m_{itk}$.

It is clear from Table 8-4 that the results are very close. The measures are almost indistinguishable, and any conclusions drawn from using one measure would also be drawn from using the other. It thus does not appear that the possible nonexistence of moments is a practical problem for models like the US model, and therefore the common practice of ignoring this problem may be justified. It is true, however, that the cost of computing alternative measures is fairly low, and as a check on the results these measures should probably be computed from time to time.

Comparison of the Predictive Accuracy of Eight Sets of Estimates

In Section 6.6 the eight sets of estimates of the US model were compared in various ways. Another way to do this is to see how they compare in terms of predictive accuracy of the overall model. One procedure that could be used would be to compute d-row values like those in Table 8-2 for each estimator, which would require estimating the model 51 times for each estimator and performing 51 stochastic simulations for each estimator. This procedure is too expensive for present purposes, especially given the cost of estimating the model just one time by FIML and 3SLS. One also runs into the problem that the numbers of observations for the early estimation periods are not sufficient to estimate all 107 coefficients that were estimated for the basic period by FIML.

An easier procedure is simply to compute root mean squared errors for each set of estimates for some prediction period, and this is what was done. The prediction period that was used is 1970I – 1982III, which is within the estimation period that was used for each set of estimates, 1954I – 1982III. Although this procedure is a poor one for comparing alternative models because of possible differences in exogenous variables and the possible misspecification of the models, it is not as bad for comparing alternative estimates of the same model. The exogenous variables are the same for each set of estimates, and the misspecification of the model may not vary too much across the different sets. In future work, however, it would be better to try to use the more expensive procedure to compare the estimates.

The results are presented in Table 8-5. Remember that the main conclusion from the comparisons in Section 6.6 is that all the estimates are fairly close to each other except for the FIML estimates. One of the key questions here, therefore, is how the FIML estimates compare to the others in terms of predictive accuracy.

The main conclusion that one can draw from the results in Table 8-5 is that they are not conclusive. The ranking of the estimates varies across variables and across the length of the prediction period. The biggest difference in the results concerns the one- through four-quarter-ahead results for FIML for real GNP. The one- and two-quarter-ahead FIML errors are much larger than the others, and the three- and four-quarter-ahead FIML errors are smaller. Part of this difference is probably due to the fact, as discussed in Section 6.6, that the FIML estimates of the coefficients of the lagged dependent variables are generally smaller than the other estimates. (See, for example, the results in Table 6.5.) In other words, the FIML results are less dependent on the values of the lagged endogenous variables, which may hurt for the first few quarters ahead and help thereafter.

It is possible that the four LAD estimators (LAD and the three 2SLAD estimators) are hurt by the use of the root mean squared error measure rather than the mean absolute error (MAE) measure. In order to determine this, MAEs were also computed for the eight sets of estimates. The results for real GNP and the GNP deflator are presented in Table 8-6. It is clear from this

TABLE 8-5. Root mean squared errors for eight sets of coefficient estimates for 1970 I - 1982 III for the US model

	Number of quarters ahead 1	2	3	4	5	6	7	8
GNPR: Real GNP								
2SLS	.66	.81	1.08	1.25	1.43	1.61	1.73	1.81
FIML	.86	.93	1.03	1.16	1.36	1.58	1.74	1.86
3SLS	.65	.79	1.05	1.20	1.39	1.60	1.76	1.88
2SLAD q = 0.0	.65	.79	1.06	1.24	1.45	1.65	1.77	1.85
2SLAD q = 0.5	.65	.78	1.05	1.19	1.39	1.62	1.77	1.90
2SLAD q = 1.0	.68	.82	1.12	1.28	1.48	1.70	1.84	1.98
LAD	.68	.83	1.15	1.31	1.49	1.68	1.78	1.88
OLS	.67	.84	1.11	1.27	1.43	1.59	1.69	1.77
GNPD: GNP deflator								
2SLS	.44	.69	.88	1.05	1.18	1.23	1.25	1.22
FIML	.45	.70	.90	1.08	1.21	1.26	1.28	1.27
3SLS	.45	.70	.90	1.08	1.21	1.26	1.27	1.24
2SLAD q = 0.0	.45	.71	.91	1.11	1.26	1.35	1.38	1.37
2SLAD q = 0.5	.44	.69	.89	1.08	1.22	1.28	1.31	1.29
2SLAD q = 1.0	.44	.69	.88	1.07	1.21	1.29	1.33	1.33
LAD	.44	.69	.88	1.07	1.21	1.27	1.30	1.29
OLS	.44	.69	.87	1.04	1.16	1.21	1.22	1.20
100·UR: Unemployment rate (percentage points)								
2SLS	.29	.43	.55	.66	.75	.83	.90	.95
FIML	.33	.48	.58	.71	.79	.87	.93	.97
3SLS	.30	.46	.59	.71	.80	.87	.95	1.01
2SLAD q = 0.0	.29	.42	.52	.63	.71	.77	.84	.88
2SLAD q = 0.5	.29	.42	.52	.62	.70	.76	.82	.87
2SLAD q = 1.0	.32	.49	.63	.74	.83	.88	.94	.99
LAD	.30	.44	.56	.67	.76	.82	.87	.91
OLS	.30	.45	.57	.69	.78	.84	.91	.96
RS: Bill rate (percentage points)								
2SLS	.97	1.32	1.37	1.42	1.47	1.47	1.50	1.56
FIML	1.04	1.47	1.55	1.66	1.79	1.82	1.88	1.93
3SLS	.98	1.36	1.41	1.48	1.56	1.58	1.62	1.69
2SLAD q = 0.0	.98	1.35	1.40	1.44	1.49	1.48	1.48	1.54
2SLAD q = 0.5	1.01	1.35	1.39	1.44	1.50	1.50	1.50	1.54
2SLAD q = 1.0	1.03	1.42	1.46	1.56	1.66	1.71	1.77	1.83
LAD	1.03	1.42	1.46	1.57	1.69	1.74	1.81	1.88
OLS	.97	1.33	1.36	1.42	1.49	1.49	1.53	1.60
M1: Money supply								
2SLS	1.07	1.29	1.37	1.63	1.87	1.90	2.05	2.23
FIML	1.05	1.23	1.33	1.64	1.89	1.96	2.15	2.35
3SLS	1.06	1.25	1.35	1.62	1.88	1.94	2.11	2.32
2SLAD q = 0.0	1.05	1.24	1.33	1.60	1.87	1.94	2.11	2.31
2SLAD q = 0.5	1.12	1.41	1.56	1.86	2.14	2.23	2.39	2.59
2SLAD q = 1.0	1.15	1.50	1.72	2.08	2.37	2.46	2.65	2.90
LAD	1.12	1.41	1.56	1.88	2.11	2.14	2.29	2.50
OLS	1.06	1.25	1.32	1.56	1.80	1.82	1.97	2.15

Notes:
- The sample period for all estimates is 1954 I - 1982 III, and so all forecasts are within sample.
- The actual values of the exogenous variables were used for all the forecasts.
- All simulations were deterministic.
- There are 51 observations for the one-quarter-ahead forecasts, 50 for the two-quarter-ahead forecasts, and so on.
- For the unemployment rate and the bill rate the errors are in the natural units of the variables. For the other variables the errors are in percentage points.

TABLE 8-6. Mean absolute errors for eight sets of coefficient estimates for 1970 I - 1982 III for the US model

	Number of quarters ahead							
	1	2	3	4	5	6	7	8
GNPR: Real GNP								
2SLS	.52	.67	.86	1.01	1.10	1.26	1.38	1.48
FIML	.71	.73	.80	.93	1.09	1.28	1.44	1.50
3SLS	.52	.64	.84	.99	1.11	1.30	1.45	1.56
2SLAD q = 0.0	.52	.65	.84	.97	1.13	1.29	1.37	1.44
2SLAD q = 0.5	.52	.64	.84	.98	1.14	1.30	1.41	1.51
2SLAD q = 1.0	.54	.68	.90	1.05	1.22	1.35	1.39	1.54
LAD	.53	.67	.92	1.05	1.17	1.30	1.36	1.47
OLS	.53	.69	.88	1.03	1.10	1.27	1.39	1.47
GNPD: GNP deflator								
2SLS	.34	.55	.72	.85	.95	.95	.97	.95
FIML	.37	.57	.74	.88	.95	.97	.97	.97
3SLS	.36	.57	.74	.89	.97	.97	.97	.93
2SLAD q = 0.0	.35	.57	.75	.88	1.00	1.03	1.09	1.07
2SLAD q = 0.5	.35	.55	.73	.90	1.01	1.05	1.07	1.07
2SLAD q = 1.0	.35	.56	.73	.88	.99	1.03	1.05	1.02
LAD	.35	.56	.73	.88	.99	1.03	1.05	1.00
OLS	.35	.55	.71	.84	.93	.94	.96	.94

Notes:
- The sample period for all estimates is 1954 I - 1982 III, and so all forecasts are within sample.
- The actual values of the exogenous variables were used for all the forecasts.
- All simulations were deterministic.
- There are 51 observations for the one-quarter-ahead forecasts, 50 for the two-quarter-ahead forecasts, and so on.
- The errors are in percentage points.

table that the main conclusion is not changed by the use of the MAE measure: the same inconclusive results are obtained for both measures.

One way of looking at these results is the following. It is clear from the results in Table 8-2 that the US model is misspecified when estimated by 2SLS. Table 8-2 provides quantitative estimates of this misspecification, and for some variables the estimates are fairly large. One might expect that estimating the model by other techniques would change the degree of misspecification, either positively or negatively. The results in Tables 8-5 and 8-6, however, suggest that this is not the case. However the model is misspecified, the size of the misspecification is not sensitive to the use of alternative estimators. An interesting question for future research is whether this conclusion holds for other models and for later versions of the US model.

8.6 A Comparison of the MC and ARMC Models

The cost of solving the MC model is too large for it to be feasible to use the method in Section 8.4 to analyze it. As discussed in Section 7.5.2, the time taken to solve the model for one quarter varies between about 20 and 40 seconds on the IBM 4341, which compares to about .2 seconds for the US model. The MC model is thus between about 100 and 200 times more expensive to solve than the US model, which for present purposes rules out for the MC model many of the experiments that could be performed for the US model. Aside from the cost, the number of observations available for the flexible exchange rate period is also not large enough to allow the method in Section 8.4 to be used. The method requires that a model be successively reestimated over a number of periods, and in the MC case there are barely enough observations to estimate the equations that pertain to the flexible exchange rate period once.

Because the method in Section 8.4 could not be used, the present comparison of the MC and ARMC models is very crude, and not much weight should be placed on the results. What was done is the following. Three eight-quarter periods were chosen: a fixed exchange rate period, 1970II–1972I, and two flexible rate periods, 1974I–1975IV and 1976I–1977IV. For each of these periods both static and dynamic predictions were generated using deterministic simulation, where the error terms were set equal to zero. The actual values of the exogenous variables were used for the MC model; the ARMC model has no exogenous variables. The MC model was solved both for the case in which trade shares are exogenous and for the case in which they are determined by the trade share equations. This allows one to examine how much accuracy is lost by having to predict trade shares rather than knowing them exactly. Given these predictions, RMSEs were computed for each run.

The results are presented in Tables 8-7, 8-8, and 8-9. For the results in Table 8-7 a weighted average of the RMSEs across all countries except the United States was taken for each variable. The RMSEs were weighted by the ratio of the country's real GNP (in 75$) in the last (that is, eighth) quarter of the prediction period to the total real GNP of all the countries. This provides a summary measure of the overall fit of the MC model with respect to each variable. The RMSEs for the individual countries are presented in Table 8-8 for one run, the dynamic simulation for the period 1974I–1975IV. This is the period of the large increase in the price of oil by OPEC, and it is not a particularly easy period to explain. The RMSEs for the United States are presented in Table 8-9.

As mentioned in Section 5.1.2, the ARMC model does not contain estimated equations for variables that are determined by identities in the MC model. Four of the variables listed in Tables 8-7 and 8-8 are determined by identities, *Y*, *PM*, *X*75$, and *PW*$, and therefore no ARMC results are presented for these variables.

TABLE 8-7. Weighted RMSEs for all countries except the US

Equation number in Table B-3 or B-4	Variable		1970 II - 1972 I						1974 I - 1975 IV					
			STA			DYN			STA			DYN		
			MC*	MC	ARMC	MC*	MC	ARMC	MC*	MC	ARMC	MC*	MC	ARMC
4	Real GNP	Y	1.54	2.00	—	2.79	2.90	—	1.84	2.58	—	3.25	4.67	—
5	GNP deflator	PY	.88	.89	.93	2.79	2.87	2.68	1.35	1.36	1.55	3.08	3.10	5.24
7a,7b	Interest rate	RS	.58	.60	.69	.82	.94	.97	.89	.94	.97	1.63	1.71	1.73
9b	Exchange rate	e	a	a	a	a	a	a	3.87	3.96	4.16	5.03	5.30	6.22
V	Import price	PM	.98	2.90	—	3.01	5.40	—	2.81	2.89	—	6.55	6.96	—
6	Money supply	M1*	2.74	2.78	2.48	5.21	5.23	4.14	2.83	2.84	2.85	4.78	4.54	4.22
1	Imports	M	4.37	4.65	4.81	7.95	7.22	6.88	4.91	5.11	6.06	6.64	7.94	10.75
2	Consumption	C	1.54	1.59	1.32	2.96	2.83	2.17	2.24	2.34	2.07	3.40	4.08	3.97
3	Investment	I	2.84	2.84	2.72	6.25	5.96	5.35	3.47	3.47	3.62	7.17	7.23	6.84
8	Interest rate	RB	.27	.28	.30	.51	.51	.77	.46	.48	.52	1.02	1.07	1.13
11	Export price	PX	1.90	1.85	2.17	4.83	4.95	6.92	3.84	3.91	3.83	8.16	8.05	11.00
II	Exports	X75$	1.93	7.28	—	5.17	10.45	—	2.27	9.22	—	3.98	14.40	—
VI	World price	PW$	.93	.83	—	3.50	3.70	—	4.06	4.25	—	7.98	7.88	—

Equation number in Table B-3 or B-4	Variable		1976 I - 1977 IV					
			STA			DYN		
			MC*	MC	ARMC	MC*	MC	ARMC
4	Real GNP	Y	1.53	1.90	—	2.41	2.80	—
5	GNP deflator	PY	1.15	1.15	1.18	2.16	2.37	2.75
7a,7b	Interest rate	RS	.86	.85	.96	1.72	1.79	1.81
9b	Exchange rate	e	2.43	2.42	2.51	4.38	4.19	5.73
V	Import price	PM	2.21	2.22	—	4.31	4.23	—
6	Money supply	M1*	2.32	2.34	2.35	3.38	3.48	3.43
1	Imports	M	4.48	4.63	5.04	6.19	7.03	7.16
2	Consumption	C	1.62	1.67	1.48	2.82	3.01	2.33
3	Investment	I	2.82	2.82	2.83	5.01	5.02	4.23
8	Interest rate	RB	.43	.43	.45	.89	1.00	1.00
11	Export price	PX	2.33	2.31	2.42	3.79	3.74	4.98
II	Exports	X75$	1.59	6.87	—	2.83	9.38	—
VI	World price	PW$	1.60	1.56	—	2.46	2.50	—

Notes: a. Fixed exchange rate period for almost all countries.
- STA = Static simulation.
- DYN = Dynamic simulation.
- MC* = MC model with trade shares exogenous.
- All errors are in percentage points.
- Weights are GNP in 75$ in the last quarter of the period.

TABLE 8-8. RMSEs for the individual countries: dynamic simulation, 1974 I - 1975 IV

Country	Real GNP Y	GNP deflator PY		Interest rate RS		Exchange rate e		Import price PM	Money supply M1*		Imports M		Consumption C		Investment I		Interest rate RB		Export price PX		Exports X75$
	MC	MC	ARMC	MC	ARMC	MC	ARMC	MC	MC	ARMC	MC	ARMC	MC	ARMC	MC	ARMC	MC	ARMC	MC	ARMC	MC
Canada	2.4	2.0	3.0	.6	1.3	2.5	2.1	5.2	3.0	3.1	6.7	3.9	1.4	1.8	3.3	3.6	.5	.7	9.1	8.1	16.5
Japan	3.7	1.9	3.5	.7	1.7	3.6	4.2	5.9	6.3	1.7	4.3	21.5	3.1	9.0	8.4	11.0	—	—	7.1	13.1	7.5
Austria	1.7	2.5	3.3	.4	.9	4.2	4.6	7.1	3.7	3.5	9.7	7.1	1.7	1.4	—	—	—	—	8.2	8.2	6.7
Belgium	3.5	4.1	7.0	1.9	1.8	4.1	5.3	7.3	3.6	3.8	10.5	12.4	3.6	2.1	5.5	5.6	.6	.5	8.1	8.5	12.6
Denmark	4.5	3.0	4.4	2.6	4.5	6.4	4.3	5.4	5.0	6.7	8.1	12.9	6.1	5.8	13.2	15.9	.6	1.7	3.6	7.5	5.8
France	3.9	2.3	3.1	2.6	2.2	7.2	8.3	8.0	3.1	4.1	9.0	8.2	2.5	1.2	2.9	3.5	1.1	.9	7.2	9.7	13.3
Germany	4.3	1.6	2.9	3.0	1.7	3.9	5.3	8.7	1.7	3.5	5.9	4.7	1.0	1.6	10.5	6.1	.9	.9	8.5	9.0	9.6
Italy	3.8	1.6	6.8	3.4	4.2	7.0	9.1	4.2	3.0	9.7	8.9	14.1	4.2	3.7	4.0	3.3	1.4	1.5	8.6	11.7	10.4
Netherlands	5.4	1.5	3.5	3.3	2.1	4.5	6.6	5.4	3.1	4.8	8.8	7.2	6.5	2.3	9.5	7.4	.8	.9	8.9	10.5	8.0
Norway	3.9	2.2	2.5	2.3	2.1	6.0	4.6	5.2	8.0	5.2	3.2	4.8	3.1	1.6	9.5	10.8	.5	.6	8.0	9.1	15.9
Sweden	2.4	1.8	3.4	1.5	1.2	6.0	5.7	6.6	7.4	4.8	6.8	6.1	2.2	3.8	7.6	5.5	.8	.3	11.0	9.9	11.6
Switzerland	8.0	4.5	1.9	1.5	.9	10.5	6.7	2.2	8.8	7.2	17.5	15.7	4.9	6.0	18.8	17.1	.6	.7	2.3	6.5	11.6
United Kingdom	3.3	5.0	8.7	.3	.8	2.7	5.2	7.4	6.6	3.9	6.3	4.3	2.8	2.3	4.5	3.9	2.6	2.6	7.8	9.2	8.4
Finland	3.7	6.9	8.5	.2	.3	4.5	5.4	6.1	11.3	10.7	8.1	10.0	3.5	2.3	16.0	9.8	—	—	16.6	18.4	35.1
Greece	3.3	2.9	4.4	1.5	1.3	—	—	7.7	7.6	4.0	10.5	13.5	—	—	20.4	25.2	—	—	10.1	12.1	7.4
Ireland	4.7	4.7	5.0	.9	.7	4.7	5.4	9.9	5.0	5.7	8.1	14.5	3.2	5.6	6.8	8.3	2.2	2.6	5.7	5.5	6.8
Portugal	4.7	6.8	5.3	2.5	1.0	5.9	7.5	12.0	6.6	2.3	17.6	28.4	4.7	5.4	7.9	10.3	1.9	1.0	—	—	17.3
Spain	3.0	1.6	1.3	.2	.3	5.1	7.9	4.8	2.4	2.3	4.1	4.0	2.8	2.5	3.8	4.2	—	—	6.3	8.9	18.0
Turkey	2.2	3.9	5.0	—	—	—	—	7.3	4.1	2.4	8.7	13.5	2.9	9.6	—	—	—	—	9.4	8.2	34.1
Yugoslavia	2.5	4.0	4.6	—	—	—	—	8.5	—	—	6.3	8.2	3.7	3.2	—	—	—	—	7.7	10.9	9.6
Australia	3.1	7.6	6.5	.8	.6	6.9	2.3	12.2	8.5	8.5	7.5	8.8	2.4	2.4	2.3	2.7	.7	1.2	4.7	6.4	8.3
New Zealand	6.2	6.9	5.2	1.8	.6	29.2	18.7	23.4	2.8	6.3	12.9	17.5	7.7	2.0	10.4	18.8	.8	.7	19.9	17.0	24.7
South Africa	12.4	2.1	1.7	1.6	1.5	—	—	8.1	3.4	6.4	7.0	8.5	10.7	1.6	17.3	11.3	.9	.4	5.8	9.6	11.9
Libya	23.7	—	—	—	—	—	—	9.6	—	—	9.8	12.6	6.9	9.2	—	—	—	—	—	—	49.8
Nigeria	9.7	—	—	—	—	—	—	8.5	—	—	5.6	12.1	—	—	—	—	—	—	—	—	32.2
Saudi Arabia	17.3	—	—	—	—	—	—	7.0	—	—	19.9	9.6	17.0	7.3	—	—	—	—	—	—	22.5
Venezuela	8.5	—	—	—	—	—	—	8.0	—	—	20.9	13.9	13.0	11.1	—	—	—	—	—	—	18.1
Argentina	21.0	10.2	38.7	—	—	—	—	8.2	—	—	—	—	33.8	10.2	10.3	10.9	—	—	—	—	23.5
Brazil	4.4	4.9	8.9	—	—	13.6	17.8	8.1	—	—	12.7	13.6	4.5	2.5	—	—	—	—	11.2	29.4	15.7
Chile	3.3	19.2	81.0	—	—	—	—	6.3	—	—	32.9	101.0	—	—	—	—	—	—	—	—	32.0
Colombia	10.1	2.3	2.1	—	—	—	—	7.6	8.0	9.7	12.0	17.1	9.2	3.3	—	—	—	—	8.2	8.5	46.2
Mexico	6.6	—	—	—	—	—	—	7.7	—	—	10.9	8.6	6.5	2.4	—	—	—	—	—	—	74.3
Peru	14.8	—	—	—	—	—	—	7.2	17.7	7.8	—	—	16.6	5.0	—	—	—	—	—	—	24.3
Israel	12.0	7.5	10.4	—	—	—	—	8.3	—	—	10.5	8.7	9.8	5.1	18.4	15.0	—	—	3.7	6.7	17.2
Jordan	11.1	16.3	14.5	—	—	—	—	6.9	—	—	9.9	20.5	15.4	12.5	—	—	—	—	—	—	25.4
Syria	9.7	6.4	8.6	—	—	—	—	8.7	—	—	12.0	13.3	18.4	7.2	—	—	—	—	—	—	26.0
India	4.1	7.5	7.1	—	—	3.3	3.7	3.6	—	—	12.0	10.0	5.3	7.5	6.8	5.4	.1	.3	7.9	12.7	12.5
Korea	3.0	5.6	9.1	2.1	1.8	—	—	5.4	—	—	8.7	10.4	5.2	5.0	—	—	—	—	6.2	4.4	15.7
Malaysia	9.3	4.9	3.3	—	—	—	—	6.3	—	—	4.1	15.3	8.7	6.2	—	—	—	—	9.4	7.9	10.6
Pakistan	2.8	2.5	3.2	1.0	.9	—	—	5.8	—	—	7.5	5.3	—	—	—	—	—	—	14.3	5.8	29.6
Philippines	13.6	5.9	6.1	2.6	2.2	—	—	5.0	5.2	4.7	5.4	6.5	16.8	6.5	—	—	—	—	22.3	21.8	21.2
Thailand	3.7	3.8	3.2	.7	.6	—	—	5.5	3.6	4.7	9.4	18.6	4.3	2.9	—	—	—	—	12.2	9.5	22.0
Weighted	4.7	3.1	5.2	1.7	1.7	5.3	6.2	7.0	4.5	4.2	7.9	10.8	4.1	4.0	7.2	6.8	1.1	1.1	8.0	11.0	14.4

Note: • All errors are in percentage points.

TABLE 8-9. RMSEs for the US with and without the MC model

Variable		1970 II - 1972 I						1974 I - 1975 IV					
		STA			DYN			STA			DYN		
		a	b	c	a	b	c	a	b	c	a	b	c
Real GNP	R	.82	.89	.96	.53	.68	.87	.51	.60	.48	1.45	2.03	2.35
GNP deflator	PY	.49	.46	.51	1.88	1.50	1.45	.56	.63	.63	1.36	2.22	2.19
Interest rate	RS	.62	.64	.68	1.16	1.21	1.27	.42	.47	.47	.72	.81	.86
Import price	PM	—	1.18	4.10	—	4.50	5.76	—	3.23	3.27	—	6.95	7.12
Money supply	M1*	1.17	1.13	1.25	3.09	2.80	2.80	1.17	1.11	1.12	2.37	1.86	1.84
Imports	IM	2.35	2.41	2.38	6.86	7.66	7.76	4.39	4.55	4.58	5.48	8.03	8.39
Interest rate	R	.28	.28	.29	.44	.46	.47	.24	.25	.24	.27	.37	.34
Export price	PX	.99	1.00	1.06	1.20	1.26	1.32	1.78	2.12	2.15	5.45	7.38	7.31
Exports	X75$	—	1.37	9.04	—	2.13	8.57	—	1.70	2.98	—	2.79	5.01
World price	PW$	—	1.06	.90	—	3.96	4.16	—	4.63	4.86	—	8.14	8.03

Variable		1976 I - 1977 IV					
		STA			DYN		
		a	b	c	a	b	c
Real GNP	R	.60	.66	.68	1.56	1.65	1.22
GNP deflator	PY	.54	.56	.54	1.46	1.61	1.69
Interest rate	RS	.36	.36	.39	.57	.59	.73
Import price	PM	—	1.15	1.16	—	2.48	2.88
Money supply	M1*	.97	.99	.97	1.68	1.80	1.80
Imports	IM	2.14	2.16	2.12	3.60	3.91	3.69
Interest rate	R	.11	.11	.11	.16	.17	.24
Export price	PX	1.15	1.28	1.27	2.32	2.82	2.95
Exports	X75$	—	1.44	4.62	—	2.48	7.46
World price	PW$	—	1.78	1.74	—	2.41	2.48

Notes: a = US model alone.
b = MC model with trade shares exogenous (including US model).
c = MC model (including US model).
• STA = Static simulation.
• DYN = Dynamic simulation.
• All errors are in percentage points.

The following general conclusions can be drawn from Table 8-7. (1) MC is generally slightly less accurate than ARMC for consumption and investment. It is generally the same as ARMC or more accurate for other variables: the GNP deflator, the two interest rates, the exchange rate, the money supply, imports, and the price of exports. (2) The best period for the accuracy of MC relative to that of ARMC is probably 1974I – 1975IV, the period of the large OPEC price increase, although the relative results across periods are close. (3) The use of the trade share equations increases the RMSEs for the export variable, $X75\$$, by a factor of between about two and four. For the dynamic prediction for 1976I–1977IV, for example, the RMSE increased from 2.83 percent to 9.38 percent. The variable next most affected by the trade share equations is GNP, which is as expected since exports are part of GNP. (4) The largest RMSE for the exchange rate for the MC model is only 5.30 percent (dynamic simulation for the 1974I–1975IV period), which seems fairly good. The largest RMSE for the short-term interest rate is 1.79 percentage points.

The RMSEs in Table 8-8 for the individual countries are generally larger for the smaller countries. This is as expected, given the poor quality of much of the data for the smaller countries and the likelihood that the model approximates less well the structure of these economies. For the first 18 countries in the table (Canada through Spain), the real GNP RMSEs range from 1.7 percent for Austria to 8.0 percent for Switzerland. The range for the GNP deflator is from 1.5 percent for the Netherlands to 6.9 percent for Finland, and the range for the exchange rate is from 2.5 percent for Canada to 10.5 percent for Switzerland.

With respect to the results in Table 8-9 for the United States, the fit of the US model for most variables worsens when it is embedded in the MC model. In the full MC model the two variables that are exogenous in the US model alone, the price of imports (PM) and exports ($X75\$$), are endogenous and thus predicted with error. The RMSEs for PM for the MC model with trade shares endogenous (the c columns) range from 1.16 percent to 7.12 percent, and the RMSEs for $X75\$$ range from 2.98 percent to 9.04 percent. These two additional sources of error generally lead to larger errors for the other variables in the US model, although in some cases the error cancellation is such that the RMSEs are smaller in the full MC model. The largest increase in the RMSE for real GNP occurred for the dynamic simulation for the 1974I–1975IV period, which was from 1.45 percent to 2.35 percent.

As stressed at the beginning of this section, it is not possible to draw any definitive conclusions from the present comparison. In general the MC model seems to do fairly well compared to the ARMC model, and thus the results are at least encouraging. In particular, the exchange rate RMSEs seem small enough for the MC model to warrant at least a small amount of optimism that the exchange rate equations are reasonable approximations.